Switching lectin activity: En route to a photocontrol of the bacterial lectin FimH

DISSERTATION

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Carina Spormann

Otto Diels-Institut für Organische Chemie

Kiel 2025

Erste Gutachterin:	Prof. Dr. Thisbe K. Lindhorst
Zweiter Gutachter:	Prof. Dr. Axel J. Scheidig
Tag der mündlichen Prüfung:	Freitag, 17. Oktober 2025

Für Brunhild

Die vorliegende Arbeit wurde unter Anleitung von

Prof. Dr. Thisbe K. Lindhorst

am Otto Diels-Institut für Organische Chemie

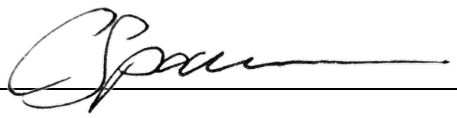
der Christian-Albrechts-Universität zu Kiel

im Zeitraum von

März 2017 bis Mai 2025 angefertigt.

Eidesstattliche Erklärung

Hiermit erkläre ich, Carina Spormann, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Inhalt und Form dieser Arbeit sind eigenständig erarbeitet und verfasst worden. Die Arbeit ist unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden. Weder die gesamte Arbeit noch Teile davon sind an anderer Stelle im Rahmen eines Prüfungsverfahrens eingereicht worden. Dies ist mein erster Promotionsversuch. Mir wurde kein akademischer Grad entzogen.

Kiel, den 

Carina Spormann

Die Bearbeitung der in Kapitel 2 beschriebenen Zielsetzung dieser Dissertation wurde bereits in der Masterarbeit begonnen. Aus dieser wurde die im Kapitel 3.2 erwähnte alternative, lineare Syntheseroute der Zielverbindungen **H1** und **H2**, sowie der Verbindungen **11**, **12**, **29**, **42**, **44** und **45** übernommen. Die Synthese des ebenfalls in Kapitel 3.2 vorgestellten Moleküle **25**, **27**, **32** und **36** wurde von Jule Bergfeld im Rahmen ihres F3-Praktikums durchgeführt und von mir teilweise wiederholt bzw. die Synthese vervollständigt. Die vorausgehende Deletion des N-terminalen Signalpeptids der *fimH*$_L$-Sequenz auf dem Plasmid pPKL241, sowie die Expression der im Kapitel 3.5 beschriebenen Expression der ΔN-Sig-FimH-Variante wurde von Sofie Todt im Rahmen ihrer Bachelorarbeit ausgeführt. Sämtliche Daten der Protein-massenspektrometrischen Untersuchungen, die in Kapitel 3.6 und 5.8.8 gezeigt sind, wurden von Dr. Philipp T. Kaulich unter der Betreuung von Prof. Dr. Andreas Tholey (Institut für Experimentelle Medizin an der Christian-Albrechts-Universität zu Kiel) gewonnen.

Im Laufe der Promotion wurden folgende Arbeiten veröffentlicht:

T. J. Paul, S. Rübel, M. Hildebrandt, A. K. Strzelczyk, C. Spormann, T. K. Lindhorst, S. Schmidt, "Thermosensitive Display of Carbohydrate Ligands on Microgels for Switchable Binding of Proteins and Bacteria", *ACS Appl. Mater. Interfaces*, **2019**, *11*, 26674-26683. (DOI: 10.1021/acsami.9b08537)

E. Fast, A. Schlimm, I. Lautenschläger, K. U. Clausen, T. Strunskus, C. Spormann, T. K. Lindhorst, F. Tuczek, "Improving the Switching Capacity of Glyco-Self-Assembled Monolayers on Au(111)", *Chemistry.* **2020**, *26*, 485-501. (DOI: 10.1002/chem.201903644)

S. Balser, M. Röhrl, C. Spormann, T. K. Lindhorst, A. Terfort, "Selective Quantification of Bacteria in Mixtures by Using Glycosylated Polypyrrole/Hydrogel Nano-Layers", *ACS Appl. Mater. Interfaces*, **2024**, *16*, 14243-14251. (DOI: 10.1021/acsami.3c14387)

G. Despras, C. Spormann, F. Klockmann, D. Schollmeyer, A. Terfort, T. K. Lindhorst, "Trivalent Heteroglycoclusters as Focal Point Pseudoenantiomers: Synthesis and Preliminary Biological Evaluation", *Eur. J. Org. Chem.*, **2024**, *24*, e20240080. (DOI: 10.1002/ejoc.202400809)

C. Spormann, T. Dorn, S. Schwaiger, S. Sdunnus, A. Koschella, H. P. Kählig, T. Heinze, T. K. Lindhorst, T. M. Wrodnigg, "Lectin-mediated adhesion: Testing of tailor-made cellulose derivatives with ConA and live *E. coli* bacteria", *Bioorg. Med. Chem.*, **2025**, *127*, 118236. (DOI: 10.1016/j.bmc.2025.118236)

Abstract

Light, particularly light in the visible wavelength range with low phototoxicity, represents an ideal stimulus to influence biological systems as it generally enables a traceless, non-invasive application with a high degree of spatiotemporal precision. Many uni- and multicellular organisms are capable of using the energy of light to induce and modulate various cellular photoresponses. For this purpose, nature has evolved a variety of photoreceptor proteins that contain chromophores as light-sensitive cofactors. Despite the rich diversity of photosensors in nature, only a few different classes of chromophores have been discovered, some of which have presumably emerged through convergent evolution. Upon photon absorption, the chromophore transmits the excitation stimulus to the *apo*-protein and possibly to other interaction partners, usually by the induction of conformational rearrangements. In an analogue fashion, artificial photosensor proteins can be generated by covalent bioconjugation with synthetic organomolecules serving as reversibly photoswitchable ligands. Ideally, the function of the corresponding proteins can thus be reversibly controlled by exploiting the photoisomerisation of the introduced photoswitch. This concept, more precisely referred to as the photoswitchable tethered ligand (PTL) approach, constitutes one aspect of the emerging field of photopharmacology. The chemical modification usually requires genetic engineering of the respective target proteins in order to introduce appropriate sites that allow highly selective and efficient bioconjugation at the desired position. While the vast majority of publications focusses on the photopharmacological targeting of transmembrane ion channels, transporters and receptors, no successful covalent modification of a lectin with photoswitchable ligands has been described in the literature.

A few years ago, the Lindhorst group initiated preliminary attempts to modify the bacterial lectin FimH with photochromic azobenzene derivatives as part of the so-called Gate-Keeper project, but the yield of modified protein was not satisfactory. This work represents one of the first applications of the PTL approach to a lectin, using the bacterial adhesin FimH as a representative example, and is pursued with the aim of achieving light-mediated functional control. The adhesin FimH is located at the distal ends of type 1 fimbriae. These are long, filamentous organelles expressed in high numbers by *Escherichia coli* and other Gram-negative bacteria that mediate adhesion to mannosylated glycans on various host cell surfaces. In the context of type 1 fimbriae-expressing pathogenic enterobacteria, FimH is regarded as an important virulence factor since adhesion usually represents the first step and prerequisite for bacterial infections.

This work, *inter alia*, encompasses genetic engineering, expression, purification and chemical modification of truncated forms of recombinant FimH. A variety of photoswitchable ligands were synthesised and evaluated with regard to their photochemical properties. Specifically functionalised ligands were either successfully tested as photochromic FimH antagonists in solution or covalently attached to the lectin according to the PTL approach. Although the bioconjugation of the lectin with the respective photoswitchable ligands eventually resulted in extensive abolishment of FimH function, many intermediate goals on the way to realise photocontrol of lectin function could be accomplished. Among other things, a straightforward assay system is presented that facilitates the assessment of lectin activity by employing only simple equipment. Since the approach presented here did not produce the desired result, a possible new realignment of the Gate-Keeper project is proposed and described in detail. The new concept could prospectively be applied to fimbriae-associated FimH, and thus may eventually allow light-mediated control over the adhesion of live *E. coli.*

Kurzzusammenfassung

Licht stellt ein geeignetes Medium dar, um biologische Systeme zu stimulieren und auf gezielte Weise zu beeinflussen. Insbesondere der Bereich des sichtbaren Lichtspektrums ermöglicht eine weitgehend nicht-invasive und wenig zellschädigende Anwendung mit überaus hoher räumlicher sowie zeitlicher Auflösung. Eine Vielzahl von ein- bis mehrzelligen Lebewesen ist ebenfalls in der Lage, die Energie des Lichts zu nutzen, um verschiedenste zelluläre Prozesse zu induzieren bzw. zu modulieren. Zu diesem Zweck hat die Natur eine Reihe von Photorezeptor-Proteinen hervorgebracht, die ein oder mehrere Chromophore als geeignete photoresponsive Cofaktoren enthalten. Interessanterweise wurden trotz der großen Vielfalt an natürlichen Photosensoren nur vier verschiedene Klassen von Chromophoren identifiziert, die teilweise durch konvergente Evolution entstanden sein müssen. Infolge der Anregung durch die Absorption eines Photons überträgt das Chromophor den Stimulus auf den *apo*-Photorezeptor. Auf diese Weise wird in der Regel eine Konformationsänderung induziert, welche im Anschluss die Wechselwirkung mit weiteren Interaktionspartnern beeinflussen und Signaltransduktionskaskaden initiieren kann. Auf analoge Weise lassen sich artifizielle Photosensor-Proteine durch die kovalente Einführung von chemisch synthetisierten Molekülen generieren, die als reversible Photoschalter fungieren. Im Optimalfall lässt sich damit die Funktion der entsprechenden Proteine durch die Isomerisierung der eingebauten Photoschalter-Einheit reversibel steuern. Dieses Vorgehen wird in der Literatur auch als Photoswitchable-Tethered-Ligand (PTL)-Methode bezeichnet und stellt einen wichtigen Teilaspekt in dem interdisziplinären Feld der Photopharmakologie dar. Die Umsetzung der PTL-Methode geht in der Regel mit der Mutagenese der entsprechenden Ziel-Proteine auf Genebene einher, um die Einführung von geeigneten Aminosäureresten für eine selektive und effiziente chemische Modifizierung zu gewährleisten. Bei einem Großteil der entsprechenden Publikationen, die sich mit kovalenten Modifikationen von Zielproteinen für photopharmakologische Zwecke befassen, werden transmembranständige Ionenkanäle, Rezeptoren, sowie Transporter adressiert.

Die erfolgreiche Anwendung der PTL-Methode mit dem Ziel, die photochemische Kontrolle von Lektinen zu erzielen, wurde dagegen noch nicht in der Literatur beschrieben. Vor einigen Jahren wurde im Arbeitskreis von T. K. Lindhorst bereits im Rahmen des sogenannten „Gate-Keeper-Projektes“ damit begonnen, das Lektin FimH mit Azobenzolderivaten chemisch zu modifizieren. Dabei konnten jedoch bisher nur geringe Ausbeuten an unselektiv markiertem Protein erhalten werden. Mit dieser Arbeit wird das Gate-Keeper-Projekt fortgesetzt, wobei sich das Vorgehen näher an der selektiveren und daher erfolgversprechenderen PTL-Methode orientiert. FimH ist das Adhäsin, welches am distalen Ende von Typ-1-Fimbrien lokalisiert ist. Letztere sind lange, filamentöse Organellen, die von *Escherichia coli* und anderen gramnegativen Bakterien in großer Zahl exprimiert werden, um die Adhäsion an mannosylierte Glykanstrukturen auf verschiedenen Wirtsoberflächen zu ermöglichen. Auch einige pathogene Spezies exprimieren Typ-1-Fimbrien, weshalb FimH als ein wichtiger Virulenzfaktor gilt, da die zelluläre Adhäsion oftmals einen ersten und entscheidenden Schritt bei Infektionsprozessen darstellt. Im Rahmen dieser Arbeit wurde u. A. die Mutagenese, Expression, Reinigung und kovalente Modifizierung von verschiedenen trunkierten FimH-Varianten vorgenommen. Außerdem wurde eine Reihe von photoschaltbaren Liganden hergestellt und hinsichtlich ihrer photochromen Eigenschaften untersucht. Die entsprechend funktionalisierten Liganden wurden sowohl als lösliche FimH-Antagonisten in bakteriellen Antiadhäsions-Assays getestet, als auch für die chemische Modifizierung gemäß der PTL-Methode verwendet.

Auch wenn letzteres in dem weitgehenden Verlust der Lektinaktivität resultierte, konnten viele Zwischenziele auf dem Weg zur Realisierung der Photokontrolle von FimH erreicht werden. Aufgrund des negativen Resultates wird am Ende des entsprechenden Kapitels ein neues Konzept für eine mögliche Fortsetzung des Projektes im Detail vorgestellt. Dieses ist voraussichtlich auch für die Umsetzung mit fimbriierten *E. coli* geeignet, mit dem weiterhin gefassten Ziel, die bakterielle Adhesion in Zukunft photochemisch kontrollieren zu können.

Abbreviations

A	Adenine
aa	Amino acid
AB	Azobenzene
Ac	Acetyl
ADP + P_i	Adenosine diphosphate + inorganic phosphate
AEC	Anion exchange chromatography
AFM	Atomic force microscopy
AIEC	Adherent-invasive *E. coli*
Amp	Ampicillin
APS	Ammonium persulfate
Ar	Aryl
ATP	Adenosine triphosphate
a. u.	Arbitrary unit
BCA	Bicinchoninic acid
bla	β-Lactamase gene
BLUF	Blue-light using FAD
BOLT	Biorthogonal ligand tethering
Boc	*tert*-Butyloxycarbonyl
bp	Base pair
BR	Bacteriorhodopsin
br.	Broad
BSA	Bovine serum albumin
c, conc.	Concentration
C	Cytosine, cysteine (Cys)
calc.	Calculated
Cam	Chloramphenicol
cat. no.	Catalogue number
CCA	*α*-Cyano-4-hydroxycinnamic acid
CD	Crohn's disease
CD_3OD	Deuterated methanol
$CDCl_3$	Deuterated chloroform
CDS	Coding DNA sequence
CFU	Colony-forming unit
ChR	Channelrhodopsin
CID	Collision-induced dissociation
ClCCA	4-Chloro-*α*-cyanocinnamic acid
cNMP	Cyclic nucleotide monophpsphate
ConA	ConcanavalinA
COSY	Correlation spectroscopy
C. reinhardtii	*Chlamydomonas reinhardtii*
CSP	Chemical shift perturbation
CTD	C-terminal domain
CuAAc	Copper(I)-catalysed azide-alkyne cycloaddition
CV	Column volume, compensation voltage

d	Doublet
D	Aspartic acid, aspartate (Asp)
Δ	Deletion
D_2O	Deuterated water
DCM	Dichloromethane
ddH_2O	Double-deionised water (ultrapure)
DEG	Diethylene glycol
DegP	Periplasmic serine endoprotease
DHB	2,5-Dihydroxybenzoic acid
dH_2O	Deionised water
Di-Boc	Di-*tert*-butyl dicarbonate
DMAP	4-Dimethylaminopyridine
DMSO	Dimethyl sulfoxide
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DSC	Donor strand complementation
dsDNA	Double-stranded DNA
DSE	Donor strand exchange
DTT	Dithiothreitol
E	Glutamic acid, glutamate (Glu)
E. coli	*Escherichia coli*
EDG	Electron-donating group
EDTA	Ethylenediaminetetraacetic acid
EI	Electron ionisation
ELISA	Enzyme-linked immunosorbent assay
eq.	Equivalent(s)
ESI	Electrospray ionisation
Et	Ethyl
EThcD	Electron-transfer/higher-energy collision dissociation
EWG	Electron withdrawing group
F	Phenylalanine (Phe)
FAD	Flavin adenine dinucleotide
FAIMS	Field asymmetric ion mobility spectrometry
$FimH_L$	Truncated FimH variant, expressed from pPKL241
FimH-LD	Lectin domain of FimH
FimH-PD	Pilin domain of FimH
$FimH_{trc}$	Truncated FimH variant, expressed from pFimH-His_6-NT-trc
FL	Full-length
Flp	Flippase
FMN	Flavin mononucleotide
for	Forward
FP	Fluorescence polarisation
FRT	Flippase recognition target
G	Guanine, glycine (Gly)
GFP	Green fluorescent protein

GPCR	G protein-coupled receptor
GSL	Glycosphingolipid
GTP	Guanosine triphosphate
HA	High-affinity
HCD	Higher-energy collisional dissociation
HEK	Human embryonic Kidney
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HF	High fidelity
H, His, His_6 tag	Histidine, hexahistidine tag
HMEC	Human microvascular endothelial cells
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HRP	Horseradish peroxidase
HSQC	Heteronuclear single-quantum correlation spectroscopy
IAA	Iodoacetamide
IC_{50}	Half-maximal inhibitory concentration
iGluR	Ionotropic glutamate receptor
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl *β*-D-1-thiogalactopyranoside
IR	Infrared
ITC	Isothermal titration calorimetry
K	Lysine (Lys)
Kan	Kanamycin
KLD	Kinase-ligase-*Dpn*I
LA	Low-affinity
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LD	Lectin domain
LED	Light emitting diode
LG	Leaving group
LOV	light, oxygen, voltage
m	moderate
m	*meta*
M	Methionine (Met)
MALDI	Matrix-assisted laser desorption/ionisation
mAU	Milli-absorption unit
MD	Molecular dynamics
Me	Methyl
MEG	Monoethylene glycol
MS	Mass spectrometry
M_w	Molecular weight
MWCO	Molecular weight cut-off
m/*z*	Mass-to-charge ratio
N	Asparagine (Asn)
NAD	Nicotinamide adenine dinucleotide
n. d.	No data, not determined
NMR	Nuclear magnetic resonance

n.s.	Not statistically significant
NTD	N-terminal domain
Nte	N-terminal extension
o	*ortho*
OD_{600}	Optical density at 600 nm
OEG	Oligoethylene glycol
OPLS	Optimised potentials for liquid simulations
orf	Open reading frame
p	*para*
p	Plasmid
P	Promoter, Proline (Pro)
PAA	Polyacrylamide
P. aeruginosa	*Pseudomonas aeruginosa*
PAGE	Polyacrylamide gel electrophoresis
Pap	pyelonephritis-associated pili
PBS	Phosphate buffered saline
PCL	Photochromic ligand
PCTL	Photoswitchable closely tethered ligand
PCR	Polymerase chain reaction
PD	Pilin domain
PDB	Protein Data Bank
PEG	Polyethylene glycol
PG	Protecting group
PMSF	Phenylmethylsulfonyl fluoride
PORTL	Photoswitchable orthogonal remotely tethered ligand
pOSIP-KL	One-step integration plasmid-kanamycin-lambda
PPh_3	Triphenylphosphine
PPy	Polypyrrole
PRP	Photoreceptor protein
PSS	Photostationary state
PVA	Polyvinyl alcohol
q	Quartet
quant.	Quantitative
quint	Quintet
R	Arginine (Arg), unspecified substituent
RBS	Ribosome binding site
rev	reverse
R_f	Retardation factor
RFP	Red fluorescent protein
RIP	Relative inhibitory potency
RMSD	Root mean square deviation
RNA	Ribonucleic acid
rt	Room temperature
s	Singlet (NMR), strong (IR)
S	Serine (Ser)
SA	Squaric acid, sinapinic acid

sat.	Saturated
S. cerevisiae	*Saccaromyces cerevisiae*
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
Sec	General secretory pathway in bacteria
SecYEG	Heterotrimeric translocon in bacteria
sept	Septet
SMARTS	SMILES arbitrary target specification
SMILES	Simplified molecular input line entry system
sp.	Species (unspecified, singular)
spp.	Species (unspecified, plural)
SPR	Surface plasmon resonance
STD	Saturation transfer difference
t	Triplet
T	Thymine, threonine (Thr)
T_A	Annealing temperature
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TB	Terrific broth
TCEP	Tris(2-carboxyethyl)phosphine
TEG	Triethylene glycol
TEMED	*N*,*N*,*N*′,*N*′-Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TLC	Thin-layer chromatography
T_m	Melting temperature
TM	Transmembrane
TOF	Time of flight
Tos, Ts	Tosyl
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer-RNA
TSS	Transformation and Storage Solution
U	Uracil
UPEC	Uropathogenic *E. coli*
UPK	Uroplakin
UTI	Urinary tract infection
UV	Ultraviolet
vis	Visible
w	Weak (IR)
W	Tryptophane (Trp)
WT	Wild type
X	Arbitrary amino acid
Y	Tyrosine (Tyr)

Contents

Abstract ..... **XIII**

Kurzzusammenfassung ..... **XV**

Abbreviations ..... **XVII**

A guide to this thesis ..... **XXV**

1. General introduction **1**

1.1 Photoreceptor proteins *1*
1.1.1 Plant photoreceptor proteins: flavoproteins and biliproteins 2
1.1.2 Opsins: Retinylidene proteins 5
1.1.3 Fusion of type I opsins and genetics: Optogenetics 12

1.2 Artificial molecular photoswitches to control biological systems *14*
1.2.1 Azobenzenes and their usefulness in biological systems 16
1.2.2 Principles and examples of photopharmacological research 19

1.3 Lectins and mannose-specific bacterial adhesion *25*
1.3.1 The role of carbohydrates and their recognition in the field of glycobiology 25
1.3.2 Type 1 fimbriae 29
1.3.3 The bacterial lectin FimH 33

2. Objectives **45**

3. Photocontrol of the bacterial lectin FimH **51**

3.1 In silico-docking simulations *51*
3.1.1 Glide docking 54
3.1.2 Covalent docking 56

3.2 Synthesis of photosensitive ligands *59*
3.2.1 Linear synthesis of iodoacetamide-functionalised 3-fluoroazobenzene *α*-D-mannosides 60
3.2.2 Semi-convergent synthesis of iodoacetamide- and *N*-acetyl-functionalised azobenzene *α*-D-mannoside derivatives 62

3.3 Photochromic properties of synthesised azobenzene mannosides *66*

3.4 Biological evaluation of photosensitive N-acetylated analogues *71*
3.4.1 Adhesion and adhesion inhibition experiments with and without UV exposure 73
3.4.2 Control experiments 76

3.5 Expression and purification of FimH variants *77*
3.5.1 Expression and purification of $FimH_L$ variants with pPKL241-based expression systems 79
3.5.2 Expression and purification of $FimH_{trc}$ variants in HM125 pFimH-His_6-NT-trc 84

3.6 Chemical modification of FimH cysteine variants *91*

3.7 Mass spectrometric analysis of alkylated and unmodified FimH variants *92*
3.7.1 Analysis of $FimH_L$ variants by LC-MS/MS 92
3.7.2 Analysis of $FimH_{trc}$ variants by SDS-PAGE and MALDI-TOF MS 96

3.8 Evaluation of FimH activity *97*
3.8.1 Conception of the FITC-*p*AP mannoside binding assay 99

3.8.2 FITC-*p*AP mannoside binding activity of unmodified $FimH_{trc}$ variants 101
3.8.3 FITC-*p*AP mannoside binding activity of $FimH_{trc}$ variants with UV exposure 104
3.8.4 Discussion 107

3.9 Summary and conclusion 111

3.10 Future prospects 115
3.10.1 Covalent inhibition of FimH 115
3.10.2 Modification of FimH using an expanded genetic code 118

4. Red-fluorescent bacteria for adhesion studies 121

4.1 Introduction 121

4.2 Objectives 127

4.3 Clonetegration of mcherry into the E. coli chromosome 129

4.4 Bacterial fluorescence and adhesion inhibition experiments 133

4.5 Conclusion and perspectives 137

5. Experimental and supplemental section 138

5.1 Chemical material and methods 138
5.1.1 General methods 138
5.1.2 Chemicals and solvents 138
5.1.3 Thin-layer chromatography 138
5.1.4 Preparative column chromatography 139
5.1.5 IR spectroscopy 139
5.1.6 Mass spectrometry 139
5.1.7 NMR spectroscopy 139
5.1.8 Polarimetry 140
5.1.9 Melting point determination 140
5.1.10 Fluorescence and UV/vis spectroscopy 140

5.2 Chemical syntheses 141

5.3 ^{1}H and ^{13}C NMR spectra of synthesised compounds 141

5.4 Investigation of photochromic properties 243
5.4.1 Photoirradiation experiments 243
5.4.2 NMR spectra of synthesised azobenzene compounds prior and after irradiation 244
5.4.3 UV/vis absorption spectra of synthesised azobenzene compounds prior and after irradiation 251
5.4.4 Thermal relaxation experiments 263

5.5 Computational methods 270
5.5.1 Preparation of X-ray crystal structures and ligands 270
5.5.2 Glide docking, *in silico*-alkylation and energy minimisation 270
5.5.3 Covalent docking 297

5.6 General biological methods 315
5.6.1 Biochemicals 315
5.6.2 Buffers and Media 315
5.6.3 Bacterial strains 317
5.6.4 Plasmids, vectors and expression constructs 319
5.6.5 Oligonucleotides 321
5.6.6 General molecular biological methods 323

5.6.7 General protein-biochemical methods 326
5.6.8 General protocol of adhesion experiments under competitive and non-competitive conditions 328

5.7 Supplemental information to Chapter 3: "Photocontrol of the bacterial lectin FimH" 331
5.7.1 Repair of pPKL241: Renewed subcloning of the *fimH* gene in pQE30 331
5.7.2 Site-directed mutagenesis of pPKL241 and pFimH-His_6-NT-trc 334
5.7.3 Bacterial adhesion experiments with *N*-acetyl-protected azobenzene mannosides 335
5.7.4 Expression of FimH-His_6 variants encoded on pPKL241- and pFimH-His_6-NT-trc-based constructs 347
5.7.5 Purification strategy for $FimH_L$-His_6 variants encoded on pPKL241-based constructs 349
5.7.6 Purification strategy for $FimH_{trc}$-His_6 variants encoded on pFimH-His_6-NT-trc-based constructs 355
5.7.7 Chemical modification of FimH with photoswitchable iodoacetamide derivatives 364
5.7.8 Mass spectrometric analysis of FimH variants 365
5.7.9 FITC-*p*AP mannopyranoside binding assay 379

5.8 Supplemental information to Chapter 4: "Red-fluorescent bacteria for adhesion studies: Chromosomal integration of mcherry in E. coli" 386
5.8.1 Chromosomal integration of pOSIP-KL-*mcherry* 386
5.8.2 Removal of the integration module from the pOSIP-KL-*mcherry* insert 389
5.8.3 Transformation of HB101-$mCherry^{Flp}$ with fimbriae-coding plasmids 391
5.8.4 Fluorescence measurements 393
5.8.5 Adhesion and adhesion inhibition experiments with HB101-$mCherry^{Flp}$ 395

6. Additional tools 398

7. References 399

Danksagung XXVIII

Appendix XXIX

DNA sequences XXIX
Sequence of pPKL241 XXIX
Sequence of pPKL241 N136C XXIX
Sequence of pPKL241 S139C XXX
Sequence of pPKL241 ΔN-sigpep N136C XXX
Sequence of pFimHs-NT-trc XXXI
Sequence of pFimH-His_6-NT-trc XXXI
Sequence of pFimH-His_6-NT-trc XXXII
Sequence of pFimH-His_6-NT-trc XXXIII
Sequence of pFimH-His_6-NT-trc S139C XXXIII
Sequence of the *E. coli* K12 chromosomal *attB* site with inserted pOSIP-KL-*mcherry* XXXIV
Sequence of the *E. coli* K12 chromosomal *attB* site with inserted *mcherry* gene XXXV
Sequence alignment of the *fimH* inserts on pPKL241 and pFimH-His_6-NT-trc XXXVI

MALDI mass spectra of FimH variants XXXVII
MALDI mass spectra from the MALDI-TOF analysis of $FimH_{trc}$ XXXVII
m/z list from the MALDI-TOF analysis of $FimH_{trc}$ XXXIX
MALDI mass spectra from the MALDI-TOF analysis of $FimH_L$ N136C ΔN-sigpep XL
m/z list from the MALDI-TOF analysis of $FimH_L$ N136C ΔN-sigpep XLII
MALDI mass spectra from the MALDI-TOF analysis of $FimH_{trc}$ R60P XLIII
m/z list from the MALDI-TOF analysis of $FimH_{trc}$ R60P XLV

A guide to this thesis

This thesis comprises two independent projects, each of which is presented in separate parts of this work. The main project is described in Chapters 1 – 3 and addresses the light-mediated control of the activity of the bacterial lectin FimH using the photoswitchable tethered ligand (PTL) approach.

Chapter 1 (General introduction) gives a broad overview of the topic, providing relevant background on photoreceptor proteins and optogenetics (Chapter 1.1), molecular photoswitches and photo-pharmacology (Chapter 1.2), type 1 fimbriae and the bacterial lectin FimH (Chapter 1.3).
Chapter 2 (Objectives) expands on the previous introduction by describing the main purpose of this thesis and outlining the experimental strategies used to pursue this goal.
Chapter 3 (Photocontrol of the bacterial lectin FimH) represents the central part of this thesis and comprises the main research project. Chapter 3.1 – 3.8 present the different experimental methods used and various results obtained. Chapter 3.9 summarises the most important findings, while Chapter 3.10 includes potential future prospects and alternative strategies for the continuation of this project.

Chapter 4 (Red-Fluorescent bacteria for adhesion studies) describes a shorter, mostly independent side project that can be read separately from the previous chapters. The Chapters 4.1 – 4.4 guide the reader through the rationale, experimental methods, and results. Chapter 4.5 concludes the chapter presenting the side project and discusses potential future applications of the fluorescent, adhesive bacteria generated in this work.

Chapter 5 (Experimental and supplemental section) encompasses all methods, procedures, as well as supplemental results for both projects. General procedures are provided in Chapters 5.1 – 5.6, while Chapters 5.7 and 5.8 include supplementary information for the main and side projects, respectively.
The appendix includes several DNA sequences of expression plasmids, computed genomic data, and MALDI mass spectra of different FimH variants.

1. General introduction

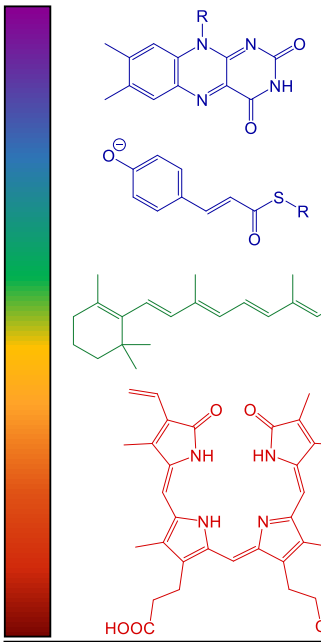
Light is the central energy source for most primary producers that generate the vast majority of global biomass and thus provides the basis for life on our planet. Evolution has arranged it in such a way that a small carbohydrate molecule is initially produced from water and carbon dioxide – glucose – which represents the key starting point for energy metabolism and serves as a precursor for further small biomolecules extending to complex biomacromolecules. The biomass obtained through photosynthesis is passed on to the subsequent trophic levels, whereby glucose also plays an important role as an energy supplier, albeit often in its polymerised forms, for instance as starch. Beyond its significance for energy metabolism and biomass production, nature utilises light also in other ways, particularly for orientation and motility of many organisms. Therefore, nature has evolved a variety of photoreceptor proteins that allow light-sensing and photoresponse to the thereby received information about the surrounding environment.

1.1 Photoreceptor proteins

Photoreceptor proteins (PRPs) are usually composed of an *apo*-protein part that possesses conserved domains for covalent binding or noncovalent attachment of light-sensitive cofactors, which are referred to as chromophores or photopigments. Depending on the photochemistry of the respective chromophores, light activation is accompanied either by photoreduction or photoisomerisation. Either way, the chromophore induces a conformational change in the *apo*-protein part upon light absorption, which in turn initiates specific signalling pathways that finally result in cellular photoresponses.[1,2] Astonishingly, nature only requires four different chromophore types to cover the entire visible and near-ultraviolet (UV) wavelength range. A notable exception constitutes the dimeric UVB resistance protein 8 (UVR8) in plants, which employs an internal system of 18 circularly arranged tryptophan (Trp) residues for light harvesting. These 18 Trp residues transmit the excitation energy to an inner core, consisting of a further 8 Trp residues. As a consequence, UVR8 requires no external chromophore for light harvesting.[3]

Some PRPs are integral transmembrane (TM) proteins that activate signal transduction pathways either through kinase domains, located on the intracellular side, or by activating receptor-associated heterotrimeric G proteins. The majority of PRPs are soluble proteins that, upon photoactivation, may serve as kinases or regulate ubiquitin ligases in the cytosol or may enter the nucleus, where they directly act on the gene expression level through interaction with transcription factors.[4,5] PRPs are categorised into six photosensor families (Table 1.1-1), of which the best-known and most extensively studied representatives are presented in more detail in the following sections.[2]

Table 1.1-1: Natural chromophores and classes of photoreceptor proteins.

Chromophors		PRP family	Photochemistry
	Flavins	Cryptochromes, Phototropin, BLUF proteins	Photoreduction, Cysteinyl adduct formation
	Coumaric acid	Xanthopsins	*trans* ↔ *cis* photoisomerisation
	Retinal	Rhodopsins	*trans* ↔ *cis* photoisomerisation
	Bilins	Phytochromes	*trans* ↔ *cis* photoisomerisation

PRP classification according to van der Horst and Hellingwerf.[2] BLUF protein: PRP sensing blue-light using FAD.

1.1.1 Plant photoreceptor proteins: flavoproteins and biliproteins

Photoreception is especially relevant in plants due to their generally photoautotrophic and often sessile mode of life. Among other things, photoreceptors enable optimisation of plant growth, development and reproduction in response to the variable light conditions at their respective habitats. Such plant PRPs include cryptochromes, photolyases and phototropins, which are, *inter alia*, responsible for light-controlled movements (phototropism) and development (photomorphogenesis) in higher plants.[6–8] They furthermore mediate repair of UV-induced DNA damage and light-oriented locomotion in unicellular species.[9] The latter process is termed phototaxis, describing the photophilic or photophobic behaviour of freely moving organisms.[10–12] The aforementioned PRPs are representatives of flavoproteins and thus feature flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as light-sensitive chromophores. As these usually mediate photoresponse by cysteinyl adduct formation and photoreduction rather than photoisomerisation, they will not be described further.

Phytochromes constitute a further type of photosensor proteins, which belong to the class of biliproteins.[13] Phytochromes are sensitive to wavelengths in the red to near infrared (IR) spectrum and are widely present in fungi and bacteria, but their occurrence and diversity are particularly pronounced in plants.[14,15] Phytochromes provide the measurement of day length as well as the intensity and spectral composition of light.[5] Similar to cryptochromes, they are largely involved in phototropism and photomorphogenesis in plants[16–18] and furthermore regulate physiological processes dependent on day length (photoperiodism) such as flowering.[19] Phytochromes are soluble, dimeric proteins, in which each monomer may exist in two interconvertible, isomeric forms with partially overlapping absorption spectra but different absorption maxima.[20,21]

The inactive ground state (Pr) absorbs red light ($\lambda_{max,\ Pr}$ = 660 nm) and subsequently undergoes a conformational change towards the active state (Pfr). The latter features kinase activity, induces phosphorylation cascades and furthermore interacts with various transcription factors, resulting in extensive transcriptional reprogramming.[21,22]

As the abbreviation (fr) indicates, Pfr absorbs in the far red range ($\lambda_{max,\ Pfr}$ = 730 nm) and immediately undergoes photoconversion to the inactive Pr isoform. The reversion process also occurs thermally in prolonged darkness. The colour change between the two isoforms is visible to the human eye and manifests itself as a slight turquoise-blue to green shift (Pr → Pfr). The phenomenon of a reversible absorption shift, induced by light absorption at a distinct wavelength, is referred to as photochromism, derived from the Greek words *phos* (light) and *chroma* (colour).[23] Due to the partial spectral overlap, the two isoforms are in a wavelength-dependent photoequilibrium that enables sensing of spectral changes and light intensity over longer periods of time. Each monomer features an open-chain bilin chromophore, which is largely responsible for the photochromic properties of the phytochrome systems.[24] The tetrapyrrole chromophore is usually phytochromobilin (PΦB) in plants, phycocyanobilin (PCB) in cyanobacteria as well as in red algae and biliverdin (BV) in bacteria, which are covalently bound in each case to a conserved cysteine residue (Cys) via a thioether bridge (Figure 1.1-1).[25,26]

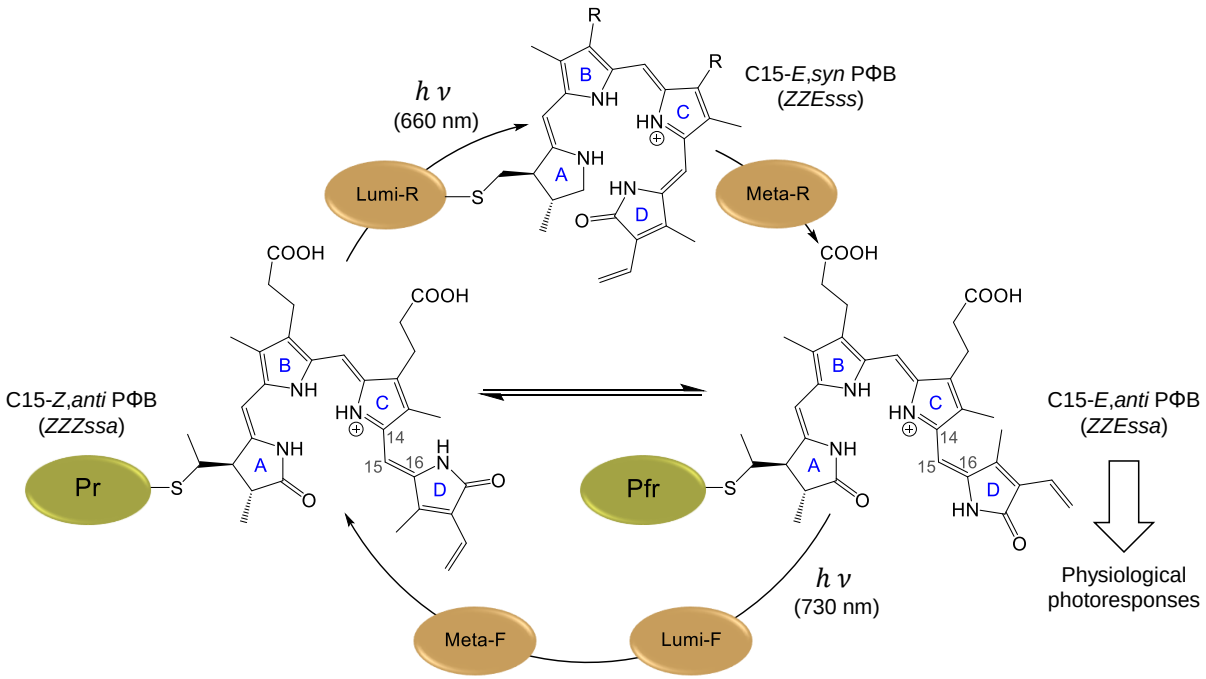


Figure 1.1-1: Proposed photocycle of phytochromobilin chomophors in plant phytochromes.
The configuration and conformation of the C4=C5, C10=C11, and C15=C16 double bonds of the phytochromobilin chromophore (PΦB) is given in brackets as: *E* = *trans*, *Z* = *cis*, *a* = *anti*, *s* = *syn*.

In many PRPs, chromophore photoisomerisation and induced conformational changes occur on the timescale from approx. several hundred femtoseconds up to a few milliseconds. The photocycle of the bilin chromophore in phytochromes has been extensively studied with several time-resolved spectroscopic techniques[27–30] and the findings were supplemented by X-ray crystallography.[31–34] Due to the developments in the field of X-ray free electron laser (XFEL) technology, conformational changes within protein crystals can now be time-resolved down to the required femtosecond scale. In recent years, time-resolved serial femtosecond crystallography (TR-SFX) has contributed significantly to the understanding of photo-induced molecular dynamics in various PRPs.[35,36] Several metastable, spectroscopically distinctive intermediates of phytochromobilin have been identified, some of which are photo-induced (Lumi) and from which others are formed by thermal decay (Meta). In the ground state, the tetrapyrrole chromophore adopts a *Z-anti* conformation of the C15=C16 double bond between the rings C and D (Figure 1.1-1). Upon photon absorption, the methin bond undergoes a *cis* → *trans* (i.e. *Z* → *E*) photoisomerisation (C15-*Z*,*anti* → C15-*E*,*syn*), followed by a thermal *syn* → *anti* isomerisation (C15-*E*,*syn* → C15-*E*,*anti*).[37] The resulting conformation of the chromophore in the active Pfr isoform can revert to the ground state either photochemically or via thermal reversion in prolonged darkness.[38]

1.1.2 Opsins: Retinylidene proteins

Probably the most well-known representatives of PRPs are opsins, which are considered to be the most abundant photosensor proteins on Earth.[39] The term was coined two years after the first description of a reddish pigment isolated from the frog retina in 1876 and derived from the ancient Greek words *rhodo* (rose) and *opsis* (sight).[40] Opsins are generally categorised into two subtypes (type I and II). Type I opsins are widely present in Archaea and Bacteria and are therefore often referred to as microbial rhodopsins although they have also been found in some unicellular algae and in fungi. Microbial rhodopsins fulfil various physiological functions and thus serve as photosensors, photo-inducible transcription factors, as well as enzymes in some species.[41–43] Certainly, the best-known type I opsins are those that act as light-sensitive translocators for protons (e.g. bacteriorhodopsin and archaerhodopsin), cations (channelrhodopsin) and chloride anions (halorhodopsin) through cytoplasmic membranes.[39,44] Type II opsins are widespread in Animalia and belong to the superfamily of membrane-bound, G-protein-coupled receptors (GPCRs). Even if the use of the same terminology may suggest the contrary, type I and II opsins are phylogenetically not related and probably evolved through convergent evolution.[45,46]

However, both PRP types commonly share the use of retinaldehyde (retinal) as the chromophore, which triggers a conformational change of the *apo*-opsin part through photoisomerisation. One exception is represented by the xanthopsin family, which was first described in the extremophilic purple bacterial species *Halorhodospira halophilia* (formerly *Ectothiorhodospira halophilia*). On the basis of similar spectral absorption and emission properties, xanthopsins were originally presumed to be rhodopsin-like proteins.[47,48] However, subsequent studies revealed that xanthopsins utilise *p*-coumaric acid instead of retinal as a chromophore, which can also undergo *trans* → *cis* photoisomerisation of the vinyl double bond after blue light exposure.[49] Possibly for this reason, xanthopsins are often better known as photoactive yellow proteins (PYPs). In the primordial organism *H. halophilia*, members of the xanthopsin family are potentially involved in negative phototaxis (avoidance of blue light) and also serve as models for a bacterial PRP by virtue of their favourable crystallisation properties and excellent diffraction quality.[50,51] The *trans* → *cis* photoisomerisation probably interrupts the stabilising hydrogen-bond network, shown in Figure 1.1-2a.[52] This presumably has a strong impact on the protein conformation and apparently induces partial unfolding.[53] The putative active state of PYP contains deprotonated *p*-coumaric acid in the *cis*-configuration. However, corresponding interaction partners that may receive PYP transduction signals are currently unknown.[54]

Animals usually possess a heterotrophic and, in many cases, vagile mode of life. They often have to actively search for their food, a habitat, a partner for reproduction, etc., whereby image-forming visual perception offers a decisive advantage. Animalia have therefore evolved specialised neuroepithelial photoreceptor cells located in the retina, which accumulate type II opsins within invaginated membrane stacks (discs) in the outer segments, in order to mediate visual perception and phototransduction.[55] In the mammalian retina, three types of photoreceptor cells have been identified – rod and cone cells, as well as photosensitive retinal ganglion cells (pRGCs), respectively. Rod cells are responsible for vision in dim light or at night (scotopic vision). The *apo*-form of the respective opsins are accordingly referred to as scotopsins and as rhodopsins, if the retinal chromophore is bound (retinylidene form).[56] Cone cells are specialised for colour perception and the accumulated PRPs are termed photopsins (*apo*-form) and iodopsin (retinylidene form), respectively.[57]

The third class of photoreceptor cells, pRGCs, are not required for image-forming vision. Nevertheless, these cells express melanopsins, which are involved in light-controlled synchronisation of the circadian rhythm and the pupillary light reflex.[58–60] Subtypes of the animal rhodopsin family have been identified in various non-photoreceptive organs (e.g. brain, muscle, skin and liver) and apparently fulfil other functions beyond visual perception.[61,62]

As previously mentioned, representatives of animal opsins are usually membrane-bound GPCRs that activate cognate heterotrimeric G proteins and thus induce signal transduction cascades. Type II opsins therefore usually possess seven transmembrane (TM) helices and one intracellular α-helix.[63] The seventh TM helix contains a highly conserved lysine residue (in bovine rhodopsin: Lys296) at which the retinal moiety is covalently bound to the ε-amine, forming a protonated Schiff base. The latter is stabilised by ionic interactions with a likewise conserved carboxylate residue (Glu113) also referred to as an "ionic lock".[64,65] By far the most common chromophore is 11-*cis*-retinal, apart from 11-*cis*-3,4-di-dehydroretinal occurring in some vertebrate species.[66] The protonation of the Schiff base has a decisive influence on spectral sensitivity as the resulting π-electron delocalisation causes a redshift from the UV range to the visible spectrum.[67] The chromophore is also stabilised by hydrophobic interactions between the polyene chain with distinct amino acid residues in the retinal binding site, which influence the spectral sensitivity and provide the possibility for specific absorption shifts in different variants of (phot-)opsins.[68,69] Humans, for instance, are trichromats, implying the expression of three variants of photopsins sensitive to green, blue, and red light, respectively.[70] However, mutations affecting the amino acid sequence can also have a negative impact on opsin function and may lead to inherited genetic defects, such as hereditary colour vision deficiency, or retinitis pigmentosa.[71,72]

The principle of visual phototransduction is here briefly explained using the well characterised bovine rhodopsin as an example (visualised in Figure 1.1-2b).[73] Upon photon absorption, the 11-*cis*-retinal undergoes a $\pi \rightarrow \pi^*$ transition, thus enabling the photoisomerisation to all-*trans*-retinal.[74,75] The overall *cis* → *trans* photoisomerisation proceeds within 200 fs with high a quantum yield, i.e. the number of converted molecules per absorbed photon (in bovine rhodopsin: 0.65).[76,77] However, the initially formed photoproduct still features a highly distorted 11-*trans* bond conformation and subsequently undergoes thermal decay over several spectroscopically distinctive intermediates, accompanied by a series of conformational changes in the *apo*-rhodopsin part.[76,78–82] During conversion from metarhodopsin-I (M I) to the active intermediate M II, the Schiff base becomes deprotonated. The ionic lock is thus disrupted, which in turn leads to the abolition of a stabilising intrahelical hydrogen network formed between the third and fourth TM helices.[83,84] As a consequence, the binding site for the α-subunit of the heterotrimeric G protein (referred to as transducin) becomes exposed.[84,85] The latter becomes activated upon interaction with M II and binding of guanosine triphosphate (GTP) and in turn activates the phosphodiesterase, which produces guanosine monophosphate (GMP) by hydrolysis of cyclic GMP (cGMP).[86] This leads to the inactivation of cGMP-dependent cation channels in the plasma membrane, an interruption of cellular Na^+ and Ca^{2+} cation uptake, and thus to hyperpolarisation of the photoreceptor cell membrane.[73] M II is thermally unstable and decomposes spontaneously by hydrolysis of the Schiff base into scotopsin and the all-*trans*-retinal or the latter becomes enzymatically released. Both processes lead to M II-inactivation and are therefore also referred to as rhodopsin photobleaching.

Retinal is furthermore enzymatically reduced to retinol, in which form it enters the retinal pigment epithelium. Here, 11-*cis*-retinal becomes restored in the visual cycle, also known as the retinoid cycle.[87] The latter includes the esterification of all-*trans*-retinol, which is then converted to 11-*cis*-retinol by an isomerohydrolase enzyme followed by reduction of the hydroxyl function.[88] The restored 11-*cis*-retinal can eventually be conjugated to scotopsin in order to form a new rhodopsin pigment that can subsequently enter another photocycle.

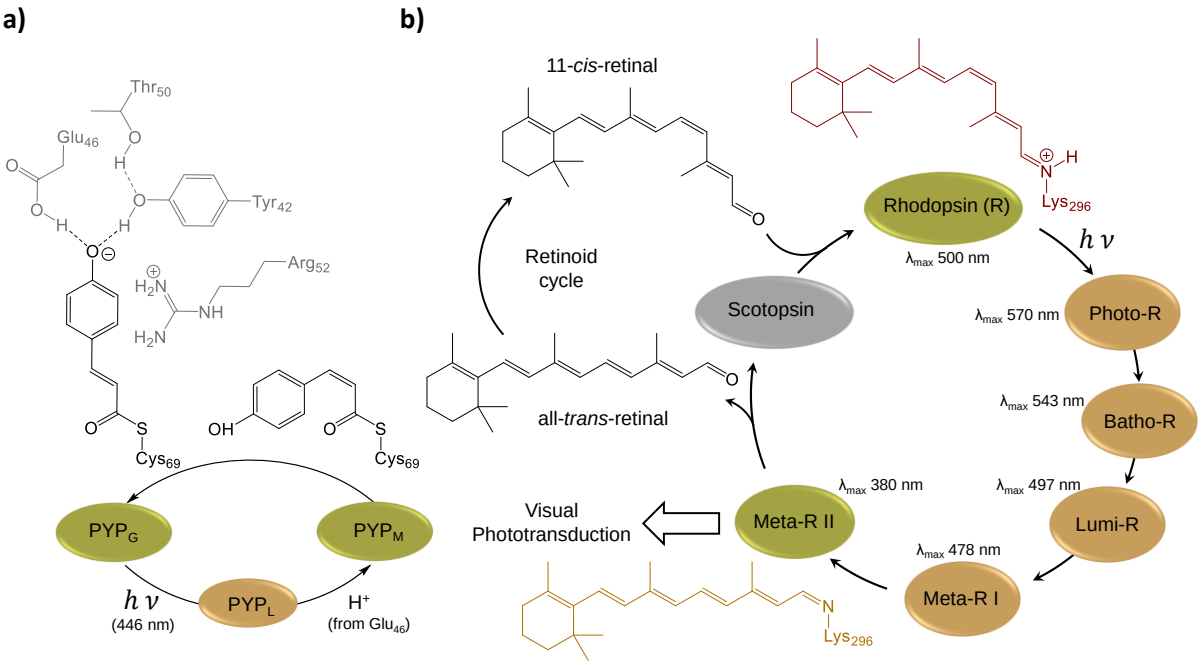


Figure 1.1-2: Proposed photocycles of **a)** *p*-couramic acid in bacterial xanthopsin, also referred to as photoactive yellow protein (PYP),[89,90] and of **b)** 11-*cis*-retinal in animal rhodopsin.[80,91]

Although they exhibit no actual sequence homology, type I opsins are strikingly similar to animal opsins in that both are integral membrane proteins composed of seven transmembrane (TM) helices. Furthermore, microbial opsins bind retinal via a protonated Schiff base with a conserved lysine (Lys) residue in the seventh helix, which is stabilised by negative counter ions.[46] However, unlike the eponymous retinylidene proteins in Animalia, microbial opsins exclusively stabilise the all-*trans* photoisomer in the ground state, which is photochemically converted to the 13-*cis*-retinal upon photon absorption, representing the thermally most stable *cis*-isomer of all natural retinal derivatives.[92,93]

Microbial rhodopsins are taxonomically and functionally diverse and highly distributed in unicellular organisms, particularly in aquatic environments.[94] Genes coding for retinylidene proteins, such as proteorhodopsin, have even been found in some viral genomes, such as in the marine giant viruses *Organic Lake Phycodnavirus* and the *Phaeocystis globosa virus*.[95] Many unicellular and simple, oligocellular organisms utilise type I opsins for phototaxis and photoorientation but, in some cases, microbial rhodopsins are required for phototrophic energy production. This applies to bacteriorhodopsin (BR), which was first described as a microbial "rhodopsin-like protein" by Stoeckenius and Oesterhelt in 1971, following its isolation from the halophilic archaeal species *Halobacterium salinarum* (formerly *H. halobium*).[96] Its function as light-sensitive outward proton pump was recognised two years later.[97]

In the following years, other BR-homologous proteins were discovered, such as archaerhodopsin (also expressed in halophilic archaea) and proteorhodopsin (in some marine bacteria, archaea and protists), which also mediate proton transport from the cytoplasmic to the extracellular side.[98,99] By this means, a proton gradient is generated along the bacterial cell membrane in order to provide the energy required for the production of ATP from ADP + P_i, which is catalysed by ATP synthase.[100] The principle of chemiosmotic coupling, i.e. the directional proton transport across membranes and its conversion into chemical energy, is ubiquitous throughout nature. Nonetheless, BR-homologous opsins are still unique as all other phototrophic organisms use chlorophyll (or bacteriochlorophyll) as photopigments. The proton gradient occurs rather indirectly, coupled to electron transport chains. The energy produced by BR homologs is estimated to be about the same level as chlorophyll-based oxygenic photosynthesis, emphasising the huge importance for energy metabolism in marine organisms.[101]

BR forms homotrimers within the membrane, which accumulate into crystalline, two-dimensional patterns, also referred to as the "purple membrane".[102] The hexagonal lattice enhances the photostability as well as thermal stability and the proton pump efficiency.[103,104] BR was also the first membrane protein that was crystallised according to the lipidic cubic phase (LCP) method, which shortly thereafter enabled the elucidation of the first high-resolution crystal structure.[105,106] The chromophore is bound to highly conserved Lys216, stabilised by three counterions (Arg82, Asp85, and Asp212). It is furthermore embedded in a hydrogen bond network formed by four other amino acid residues and seven water molecules, whith one of them (W402) directly participating in proton transport.[107] In the unexcited ground state (also referred to as the dark-adapted state), the chromophore is in equilibrium between the all-*trans*- and 13-*cis*,15-*syn*-configurations. After exposure to green-yellow light, it isomerises to the 13-*cis*,15-*anti*-retinal within 500 fs and with a high quantum yield (0.64).[108]

The BR photocycle, illustrated in Figure 1.1-3, is completed within 10 ms with recovery of the all-*trans*-retinal, thereby overcoming several spectroscopically distinguishable intermediates (K, L, M, N, and O).[109,110] The conformational changes induced in this process influence the pK_a values of pivotal amino acid residues (Asp85, Asp96, Glu194, Glu204, and Lys216), which form the proton transport pathway through the protein. The first photoproduct (K) features an energetically unfavourable conformation and rapidly relaxes to a less strained intermediate (L). In the subsequent L → M transition step, the proton is transferred from the Schiff base via W402 to the neighbouring Asp85, while another proton is simultaneously released to the extracellular medium. The Schiff base accepts a new proton from Asp96 during the M → N transition step, followed by uptake of a proton from the cytoplasmic side in the N → O step. Simultaneously, the all-*trans*-retinal becomes restored. In the final O → AT transition, a proton is transferred from Asp84 to one of the proton-releasing groups on the extracellular side, Glu194 and Glu205.

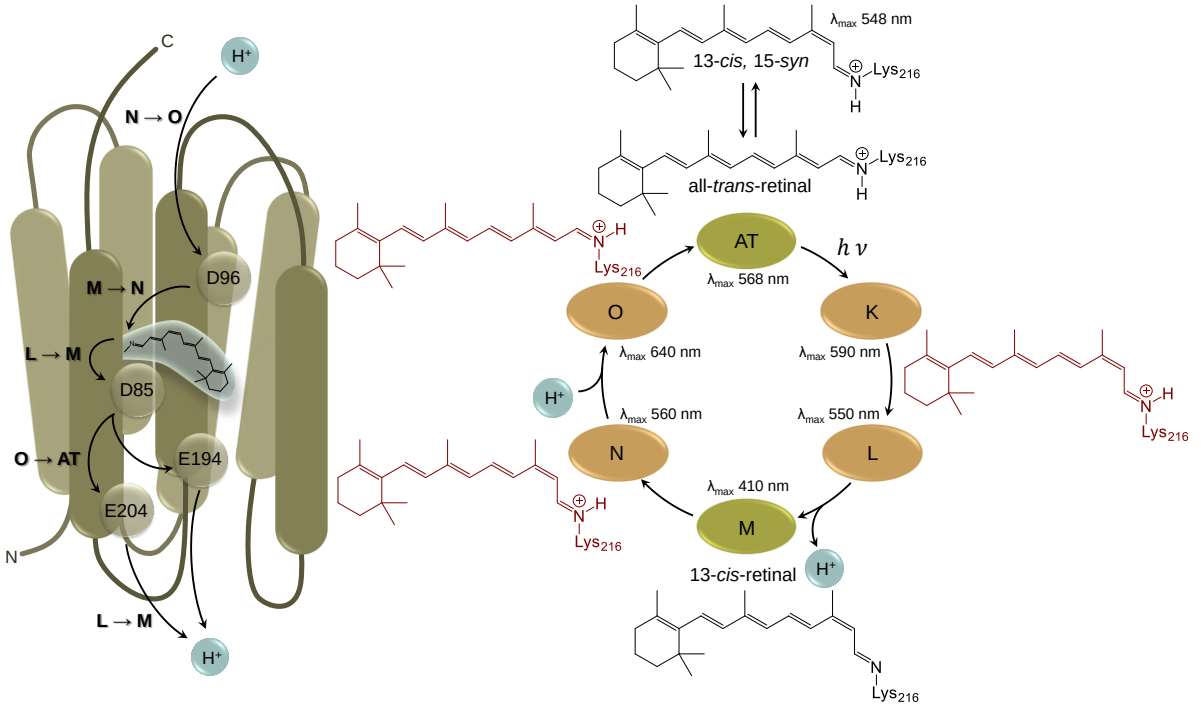


Figure 1.1-3: Proposed photocycle of all-*trans*-retinal (AT) and the suggested proton transport pathway in a bacteriorhodopsin monomer.[110]

The groundbreaking discovery of BR triggered the intensification of metagenomic research with the aim of identifying further light-sensitive microbial proteins. Thousands of potential rhodopsin genes have since been discovered in archaea, marine as well as terrestrial bacteria and in unicellular eukaryotes.[92] Besides BR, three further homologous photopigments were isolated from *H. salinarum*, namely two photosensors, (SR-I/-II), responsible for positive and negative phototaxis, respectively. The third photopigment was found to function as an inward chloride pump and was therefore termed halorhodopsin (HR; sHR, if it originates from *H. salinarium*).[111–113] HR is sensitive to green light (sHR λ_{max} = 578 nm) and occurs ubiquitously in the class of haloarchaea. Homologous chloride pumps have also been found in some flavobacteria and cyanobacteria.[114–116] According to its physiological role in osmoregulation, HR primarily mediates light-induced uptake of chloride but also of other halides and nitrates.[117] In this way, the intracellular osmotic pressure is increased and cell growth is facilitated, leading to optimal adaptation to the hypersaline environment. After bacteriorhodopsin, sHR was the second example for a membrane protein for which crystals were obtained using the LCP technique. In the solved crystal structure of a sHR homotrimer, a chloride anion was identified in direct proximity to the protonated Schiff base, attached to the conserved Lys242 and partly embedded in a solvation sphere with three water molecules.[118] The chloride ion occupies the exact position which would correspond to the proton acceptor residue Asp85 in BR. Consequently, the Schiff base in HR remains protonated and the associated M-state is missing in the HR photocycle.[92] Crystallisation in the presence of bromide and subsequent structure determination revealed a second halide binding site located near the extracellular opening.[119] Based on these findings, an ion-dragging mechanism was proposed according to which diffusion of chloride ions is controlled by a photo-induced N – H^+ dipole-flip of the protonated Schiff base. This presumably results in a rearrangement of the surrounding hydrogen bond network, mediating the release of the chloride ion into the cytosol.[120] Both the second chloride binding site and the N – H dipole-flip of the Schiff base were supported by time-resolved X-ray crystallography with an HR homolog from *Nonlabens marinus*.[121]

Another decisive breakthrough in the field of microbial opsins was marked by the discovery of cation-conducting channelrhodopsins in phototactic green algae.[122] Some unicellular, photosynthetic flagellates such as the chlorophyte species *Chlamydomonas reinhardtii* feature clearly recognisable, orange-red eyespots (stigmata). The eyespots serve as photoreceptive organelles required for light-controlled modulation of phototactic and photophobic movements, which are achieved by wave-like beating of the two flagella.[123] The involvement of rhodopsin-like photosensors in phototactic movement of microalgae has long been suspected.[124] In 2002, two type I opsin genes were identified in *C. reinhardtii*, which were proven to be functionally associated with phototaxis.[125] Shortly thereafter, their function as light-gated cation channels was recognised and the terms channelrhodopsin-1 and channelrhodopsin-2 (ChR1, ChR2) were thus coined.[122,126] ChR1 exhibits a selectivity for protons, whereas ChR2 is primarily permeable to Na^+, K^+, Ca^{2+}, and Mg^{2+} ions. Under physiological conditions, photostimulation of CHRs induces passive ion transport across the plasma membrane resulting in measurable changes of the membrane potential – the so-called photocurrents.[127] Particularly the strong influx of Ca^{2+} into the eyespot region, mediated by long or intensive light exposure of ChR2, results in activation of voltage-gated calcium channels (VGCCs), which in turn leads to massive increase of the Ca^{2+} influx and membrane depolarisation (Figure 1.1-4).[128] Probably assisted by further VGCCs located throughout the membrane, the depolarisation is propagated to the flagellar axonemes, where a distinct alteration of the flagellar waveform movements is induced, resulting in photophobic response.[129,130]

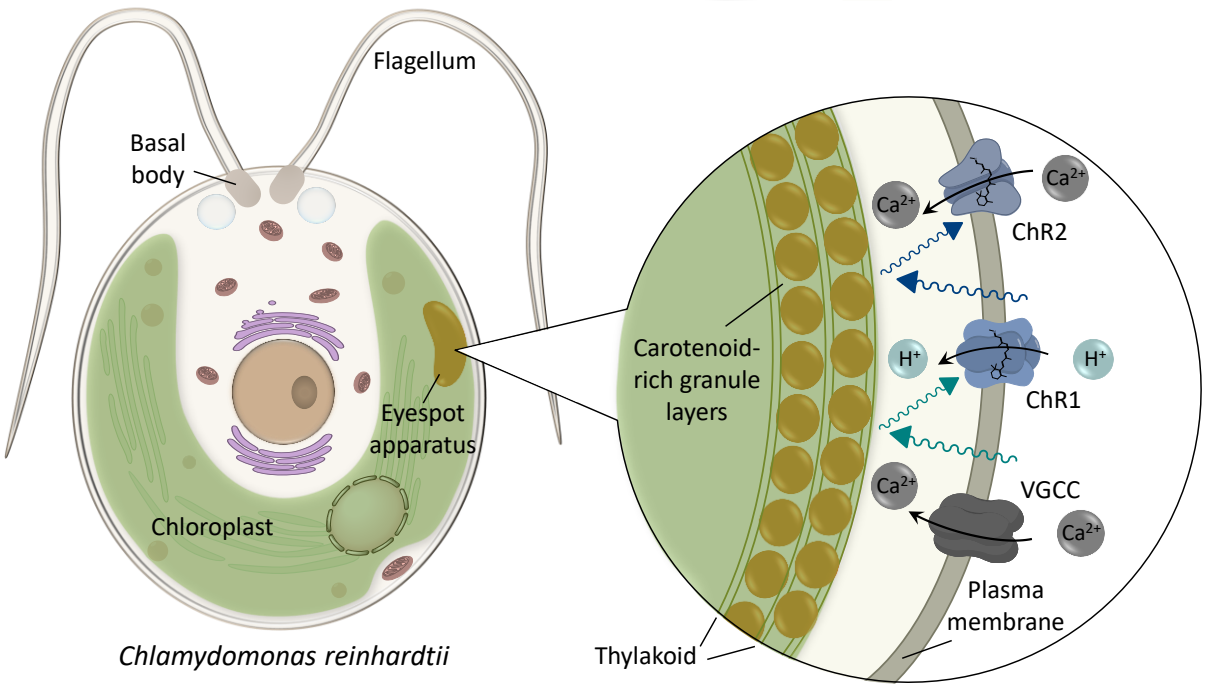


Figure 1.1-4: Schematic cell structure and magnified view of the eyespot apparatus in the biflagellate unicellular green alga *Chlamydomonas reinhardtii*.[127,131]

The inward cation pumps channelrhodopsin-1 and -2 (ChR1, ChR2) are embedded in the plasma membrane, in close proximity to the carotenoid-rich pigment granules within the thylakoid membranes, which are responsible for the characteristic orange-red colour of the eyespot apparatus. Light scattering at the chloroplast envelope acts as an amplifier of the incoming light signals that activate ChR1 and ChR2. The resulting photocurrents, in turn, activate adjacent voltage-gated calcium channels (VGCCs). Massive influx of Ca^{2+} ions results in membrane depolarisation, which is transmitted to the flagellar membrane, where photophobic response is triggered.[132]

ChR1 features a strictly pH-dependent absorption in the blue to green range (λ_{max} = 465 – 505 nm), reduced susceptibility to inactivation and higher expression levels in their native organism compared to ChR2.[133] The latter displays a rather low, constitutive expression under physiological conditions in *C. reinhardtii*, but heterologous expression (i.e. in foreign organisms) is considerably facilitated. ChR2 also absorbs blue light (λ_{max} = 460 – 480 nm) and exhibits increased selectivity for sodium ions.[92] Both variants possess a prolonged C-terminal extension, which is not necessary for channel function but potentially required for subcellular localisation, specifically, for tethering to the stacked membranes of the eyespot apparatus.[134] The first structural models of truncated ChR2 as well as a chimera of ChR1 and ChR2 (C1C2) revealed a homodimeric structure, in contrast to the trimeric BR and HR homologs.[135,136] Time-resolved spectroscopic investigation of the ChR photocycle identified six intermediates corresponding to the respective states in the BR photocycle (Figure 1.1-3), and indicated sequential deprotonation and re-protonation of the Schiff base formed with Lys296.[137,138]

TR-SFX crystal structures of the ChR chimera recently provided more profound insights into the heretofore somewhat controversial structural basis for the dynamics of the ChR photocycle.[139] In the dark-adapted state, the “inner gate”, which is formed by an intrahelical hydrogen network between TM helices 1, 3, and 7 (Glu121, Glu122, His173, and Arg307), prevents water molecules from entering the cation channel. The residues Cys167 and Asp195 (also called DC pair and conserved in many ChRs), have been proven to be essential for efficient channel function.[140] Asp195 is furthermore considered a transient proton donor for the deprotonated Schiff base. Cys167 is located on TM helix 3 in proximity to the chromophore. Photoisomerisation induces a downward conformational shift of TM helix 7, associated with a twist and movement of the 13-*cis*-retinal towards the direction of Cys167. This in turn induces an outward kink of TM helix 3 and thereby disrupts the hydrogen bonding network of the inner gate. The subsequent influx of water molecules probably results in a further opening of the channel pore, reaching a maximal radius of more than 6 Å, approx. 1 ms after photoexcitation.[141]

Eventually, the chromophore re-isomerises to the all-*trans* state, leading to rapid closing of the channel pore. As will be presented in the following chapter, ChR2 has become one of the most popular tools in the field of optogenetics. The molecular details of ion channel opening in ChR2 may differ from the chimera C1C2, which has photocycle properties similar to ChR1.[137] However, no time-resolved crystal structures of ChR2 are yet available and most of the heretofore used structural models are either based on the unexcited state, a ChR2 variant (C128T) with considerably longer open-state lifetime, cryo-electron microscopy (cryo-EM) imaging, or on molecular dynamics (MD) simulations.[141–144]

1.1.3 Fusion of type I opsins and genetics: Optogenetics

The term optogenetics describes the combination of genetic engineering and optical methods with the aim of exerting photocontrol over individual cells, tissues or even complex organisms. Fluorescent reporter proteins fused to distinct cellular targets (e.g. for monitoring purposes or as markers for gene expression) can also be regarded as an optogenetic tool, but these reporter proteins are perhaps better described as optogenetic sensors. In the narrower sense, optogenetics is often associated with the heterologous expression of PRPs (mainly type I opsins) in mammalian neurons as it was initially performed to allow imaging, control and investigation of neural activity *in vivo*.[145] Photostimulation of neurons (and other excitable cell types) genetically modified in this way can be exploited to induce (or suppress) the emergence of action potentials, whose depolarisation signals can be further transmitted to postsynaptic cells. Moreover, light as a stimulus bears great potential, as it is generally considered low or non-invasive and can be applied on a cellular level with eminently high spatiotemporal precision, which could not be achieved using heretofore available methods such as electrode stimulation. For this reason, optogenetics has become one of the most pivotal experimental tools in neuroscience, and has contributed significantly to the understanding of communication in neural circuits as well as the functioning of the nervous system and the brain.[146,147]

Zemelman and Miesenböck are often recognised as pioneers in the field of optogenetics. As part of their groundbreaking work published in 2002, the authors transfected a three-gene cascade required for rhodopsin function from *Drosophila melanogaster* into mammalian hippocampal neurons, and were thus able to monitor light-induced changes in the membrane potential.[148] However, similar approaches had already been pursued by other groups in the mid-1990s, who performed heterologous expression of BR in non-neural cells, such as baker's yeast (*Saccharomyces cerevisiae*) or in frog (*Xenopus laevis*) oocytes.[149,150] A genuine breakthrough was achieved with the discovery of channelrhodopsins.[146] Already in the first description by Nagel et al., truncated variants of ChR1/2 were expressed in *Xenopus* oocytes and in human embryonic kidney (HEK) cells and the functionality was proven by measuring the induced photocurrents.[122,126] When Boyden and colleagues expressed ChR2, C-terminally fused to yellow fluorescent protein (YFP) in rat hippocampal neurons, a pronounced neural depolarisation was detected.[151] Photoexcitation occurred on a millisecond timescale and no external addition of a retinal source was required. Thus, only by transfecting a single gene coding for a light-sensitive, ionotropic (i.e. ligand-gated) receptor, the first practical optogenetic tool for efficient photocontrol of neuronal activity was created. This constitutes a decisive advantage over animal rhodopsins, which require transfection of at least three genes and whose metabotropic phototransduction takes significantly longer.

The properties of channelrhodopsins were expanded and optimised in the following years by identifying further ChR variants in other organisms and/or genetic engineering with regard to enhanced photocurrents, kinetics, altered ion selectivity and red-shifted absorption wavelengths.[152] In addition to the depolarising (excitatory) ChR variants introduced so far, other microbial opsins further expanded the optogenetic toolkit. For instance, neuronal expression of photo-inducible inward anion pumps (e.g. halorhodopsin) results in excessive reduction of the resting membrane potential (hyperpolarisation), making these pumps suitable to be employed as optogenetic silencers (Figure 1.1-5). The same applies to light-sensitive outward pumps for the transport of protons and cations, respectively (e.g. archaeorhodopsin and synthetic LOV-based potassium channels).[153,154]

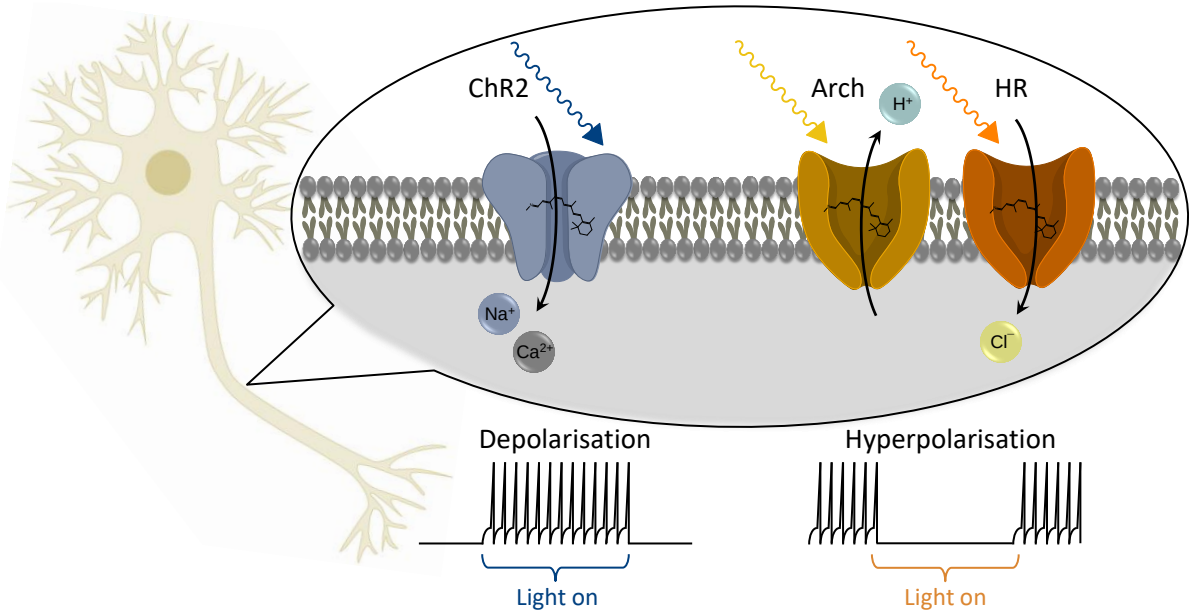


Figure 1.1-5: Light-gated ion channels and pumps used as optogenetic actuators for neuronal photostimulation.

Optogenetic activator: ChR2, channelrhodopsin-2 from *C. reinhardtii* (inward cation channel); optogenetic inactivators: Arch, archaerhodopsin (outward proton pump); HR, halorhodopsin (inward anion transporter).

In the last two decades, ChR2 variants have been expressed in various cells, tissues and organisms, including avian spinal cords and embryos, transgenic nematodes, mice, zebrafish, and primates.[155–159] The scope of application for optogenetic tools extends far beyond the field of neuroscience, and these tools are now also employed in bioscientific and clinical research. In addition to ion channels, light-sensitive transcription factors and enzymes are now applied. A well-known example is a photosensitive adenylyl cyclase, which can be employed to control cAMP levels, thereby enabling investigation and manipulation of second messenger systems.[160] A major area of medical research is the use of optogenetic tools in ganglion cells with the aim to restore vision from degenerative eye diseases and blindness.[161] The optogenetic treatment of the first human patient in 2021 yielded partial sight restoration, which represented a promising initial result for patients suffering from retinitis pigmentosa.[162]

1.2 Artificial molecular photoswitches to control biological systems

The examples of PRP-based systems presented thus far only rely on naturally occurring chromophores. An alternative approach to the introduction of light-responsiveness into biological systems that are not inherently photosensitive is the bioorthogonal implementation of synthetic chromophores. The use of reversible bistable molecular photoswitches is particularly important in this context. Photoswitches are molecules that, upon irradiation and absorption of light at a distinct wavelength, can usually switch between two isomeric states that possess distinct structural and physicochemical properties.[163,164] Since this generally also affects their absorption spectra, photoswitches can be considered the epitome of photochromic compounds. The isomerisation principle in the majority of molecular photoswitches is based on the reversible *cis* ↔ *trans* photoisomerisation of double bonds. Apart from that, photo-induced intramolecular 6π-cyclisation or [2+2] cyclisation are also common isomerisation mechanisms.[165]

The photoisomeric states of a given photoswitch can considerably differ from one another in terms of thermal stability. If both isomers are thermally stable and can be photochemically converted, this refers to a P-type photochromism.[166] If photoexcitation results in a metastable isomer, the latter can spontaneously convert back to the stable isomer by thermal relaxation in prolonged darkness (T-type photochromism). The metastable isomer can additionally be photochemically converted to the stable isomer in certain cases.[166] Some of the most popular and best-studied molecular photoswitches may include azobenzenes (ABs), thiophenefulgides, hemithioindigos, stilbenes, spiropyrans and diarylethenes (Figure 1.2-1).[167–172]

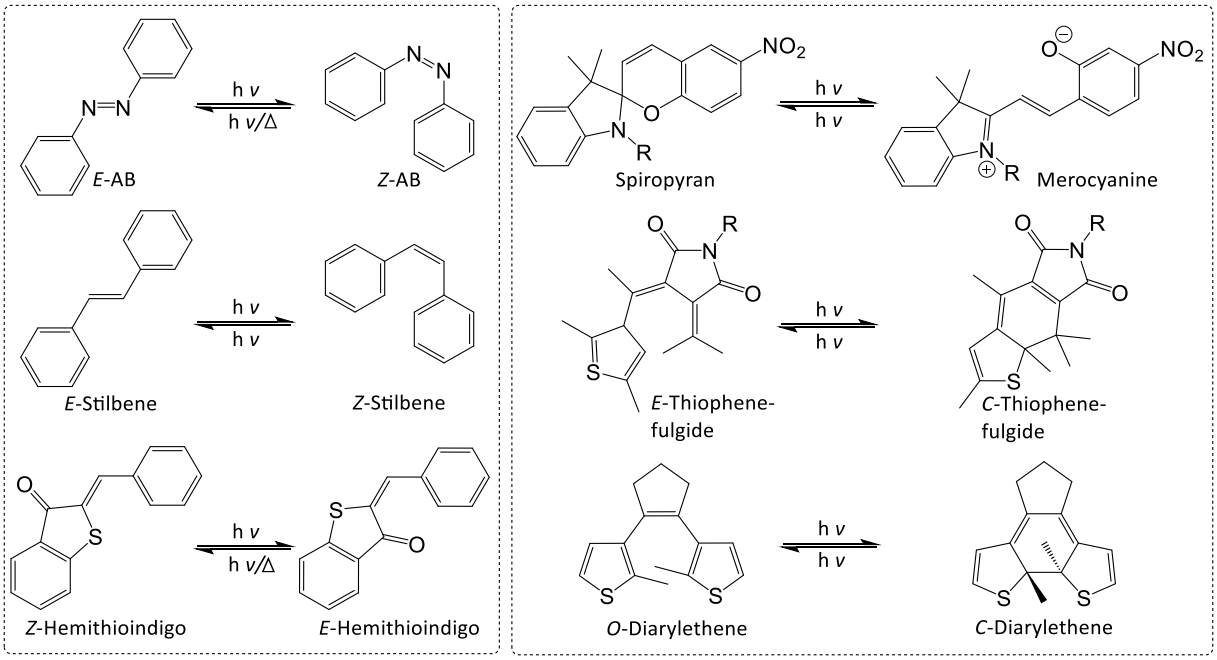


Figure 1.2-1: A selection of photoswitches that are commonly introduced into biomolecules.

Possible applications are just as diverse as the plethora of known artificial photoswitches and their respective physicochemical characteristics. Particularly suitable are fast-switchable compounds with a large extinction coefficient and high quantum yield preferably at wavelengths with low damaging potential when intended for application in biological systems.[173] Distinct properties, such as the water solubility, the photostationary state (PSS, i.e. the isomeric equilibrium at a certain wavelength) and the thermal stability (expressed as thermal half-life, $\tau_{1/2}$) can be customised to the challenges of a desired application.

Molecular photoswitches have been incorporated into a wide range of biological and bioactive molecules, including nucleosides, DNA, RNA, peptides, proteins, enzymes, carbohydrates and lipids.[200–202] The use of molecular photoswitches to control biological activity, particularly in the medical and therapeutic context, is categorised as part of the emerging field of photo-pharmacology.[176] Compared to conventional chemotherapy, photopharmaceuticals offer the decisive advantage of selective release and activation (or inactivation), which can immensely increase the spatiotemporal precision of drug targeting and minimises undesirable side effects, such as drug resistance and toxicity. In contrast to optogenetic approaches, photopharmacology does not rely on genetic engineering or gene therapy, which significantly increases its suitability with regard to clinical applications.

1.2.1 Azobenzenes and their usefulness in biological systems

Azobenzene derivatives are probably the best characterised and most commonly used photoswitches in the context of photopharmacology. This class of aryl azo compounds undergoes a characteristic *trans* → *cis* (i.e. *E* → *Z*) isomerisation of the N=N (diazene) double bond upon UV light exposure (300 – 400 nm) but also under mechanical stress or electrical stimulation.[177,167] The *trans*-isomer of the unsubstituted AB is planar, features no dipole moment and is thermally more stable than the *cis*-isomer (10 – 12 kcal mol^{-1}).[178,179] The latter possesses a bent geometry, features an explicit dipole moment (μ = 3 D) and the phenyl rings are distorted against each other by 8° (CNNC torsion) resulting from steric repulsion between the aromatic π-clouds.[177,180] The NNCC dihedral angle amounts to 53.3° in *cis*-AB, indicating that the co-planarity of the phenyl residues has been overcome.[180] Partial re-isomerisation to the *trans*-isomer occurs upon absorption of blue light (~ 450 nm), whereas thermal relaxation proceeds quantitatively to restore the *trans*-form ($\tau_{1/2}$ ~ 2 d at room temperature, rt).[181] Quantitative photoconversion is generally not possible due to the partial overlap in the UV/vis spectra of *trans*- and *cis*-AB. Both isomers feature three absorption bands in the UV and visible range, which each correspond to electronic transitions from the π → π* (pi-bonding to antibonding pi orbital) and n → π* (nitrogen lone pair to π*) orbitals. The most energetic π → π* band of the aromatic phenyl electrons (λ_{max} *trans*-AB = 225 nm, λ_{max} *cis*-AB = 250 nm), which lies in the UV-C range is usually not considered for photochemical utilisation.[182] The π → π* transition of the diazene bond electrons is associated with an excitation to the second singlet state ($S_2 \leftarrow S_0$).[183] The corresponding absorption band is very pronounced for *trans*-AB (λ_{max} ~ 320 nm, ε ~ 22 000 L mol^{-1} cm^{-1}) and the same band is hypsochromically shifted (i.e. to shorter wavelengths) and hypochromically shifted (i.e. less intense) in the *cis*-isomer (λ_{max} ~ 270 nm, ε ~ 5000 L mol^{-1} cm^{-1}).[184] The excitation to the first singlet state ($S_1 \leftarrow S_0$) corresponds to the n → π* transition, which is symmetry-forbidden in the case of *trans*-AB (λ_{max} ~ 450 nm, ε ~ 400 L mol^{-1} cm^{-1}) and therefore comparably weak. The transition becomes allowed upon isomerisation to the *cis*-isomer (λ_{max} ~ 430 nm, ε ~ 1500 L mol^{-1} cm^{-1}), therefore features a higher absorbance (hyperchromic effect) and is slightly blue-shifted compared to the *trans*-isomer.[184,185] The mechanistic details underlying the *trans* → *cis* photoisomerisation are under recurring debate. Four isomerisation pathways are most commonly proposed, namely rotation ($S_2 \leftarrow S_0$), inversion ($S_1 \leftarrow S_0$), concerted inversion ($S_2 \leftarrow S_1 \leftarrow S_0$), and inversion-assisted rotation ($S_2 \leftarrow S_1 \leftarrow S_0$).[167]

The orientation and distance between the carbons in *para* (*p*)-position (4, 4') changes considerably upon photoisomerisation, from 5.5 Å (*cis*-AB) to 9 Å (*trans*-AB).[186] These predictable, well-defined changes in geometry can be easily exploited to implement spatial changes within abiotic materials as well as in biomolecules.[187,188] The dipole moment change, which is associated with photo-isomerisation also influences the hydration sphere in aqueous environments. This may affect the affinity to receptors when AB derivatives serve as photoswitchable ligands in photopharmacological applications.[189] After photoexcitation, the isomerisation occurs on a time scale of picoseconds, thereby avoiding the emergence of triplet states. The latter could negatively affect the photostability and biocompatibility in consequence of the possible formation of singlet oxygen.[167,189] AB derivatives appear to be suitable candidates for the incorporation of biomolecules owing to the pronounced fatigue resistance, good photoisomerisation efficiency and the possibility to obtain excellent isomeric ratios at distinct $PSS_{\lambda max}$ in well-separated UV/vis spectra. The convenient availability through a variety of synthetic methods, which allow for the synthesis of even complex substitution patterns, and the low molecular weight are further advantages of azobenzene derivatives.[190]

Potential genotoxic and cytotoxic effects are sometimes considered as major disadvantage of azobenzene derivatives, which is usually attributed to the reductive diazene cleavage activity of the bacterial azoreductase, leading to a release of aniline derivatives in the gut.[191] However, this does not apply equally to all AB derivatives but appears to be a particular problem with aminoazobenzenes and in oral application.[192,193] In addition to occasional problems with solubility in water, the absorption wavelengths in the UV/blue range are also a concern for the use of AB derivatives in living systems. These wavelengths are not only harmful as they can induce DNA lesions, but may also cause tissue damage. UV and blue light also have a low penetration depth into tissue (~ 0.5 – 1.5 mm), due to pronounced absorption and scattering, which consequently precludes its use for non-invasive treatments.[194,195] A bathochromic shift (i.e. to longer wavelengths), preferably to the red or even far-red spectrum, enables deeper tissue penetration and is therefore especially desirable for *in vivo*-use.[196]

The electronic and photochemical properties of AB can be modified through targeted selection of substitution patterns in order to adapt, for instance, the absorption wavelengths and thermal stability to the intended applications. Substitution in *ortho* (*o*)- or *para* (*p*)-position with electron-donating groups (EDG, e.g. 4-amino, 4-hydroxyl or 4-alkoxy groups) results in a bathochromic shift of the $\pi \rightarrow \pi^*$ transition but this also leads to a pronounced overlap with the $n \rightarrow \pi^*$ band (Figure 1.2-2a).[197] Spectral overlap of transition bands can considerably hamper the achievement of high isomeric ratios in the photostationary equilibrium. This effect is even more pronounced in push-pull systems, which feature an asymmetric *ortho*- or *para*-substitution pattern with an EDG, as well as an electron-withdrawing group (EWG, e.g. 4'-cyano, 4'-carboxy, 4'-nitro) on either side of the AB molecule.[198] The increased delocalisation of the conjugated electrons not only reduces the energy barrier to the most red-shifted $\pi \rightarrow \pi^*$ transition, but also destabilises the *cis*-isomer as reflected by a rapid thermal relaxation (up to the millisecond range).[167] The extent of the observed effects increases with the number and strength of the electron-donating or -withdrawing nature of substituents. Tetra-*ortho*-substitution therefore appeared to be a reasonable approach to obtain redshifted absorption wavelengths. Beharry and co-workers introduced tetra-*o*-methoxy substituted AB derivatives and demonstrated a good thermal stability of the respective *cis*-isoforms.[199,200] Admittedly, the $\pi \rightarrow \pi^*$ band of the *trans*-isomers features a hypsochromic shift and therefore still lies in the UV-A range. However, the $n \rightarrow \pi^*$ transition of the *trans*-isomers displays a bathochromic effect (~ 490 nm) thus significantly reducing the spectral overlap with the $n \rightarrow \pi^*$ band of the *cis*-isomers (~ 404 nm). This allows for the possibility to induce *trans* → *cis* photoisomerisation by exposure with green light rather than with UV irradiation (Figure 1.2-2b). The band separation was rationalised with an impairment of planarity in the *trans*-configuration, due to steric repulsion of the methoxy groups as well as by an electronic interaction of the latter with the nitrogen lone pair orbitals.[199] The authors reported that tetra-*o*-methoxy-AB was prone to reduction by thiols, e.g. in the presence of glutathione. Tetra-*o*-chloro-AB derivatives likewise feature the desirable redshift of the $n \rightarrow \pi^*$ absorption band corresponding to the *trans*-isomer, but were insensitive to thiol reduction and therefore considered more suited for applications *in vivo*.[200]

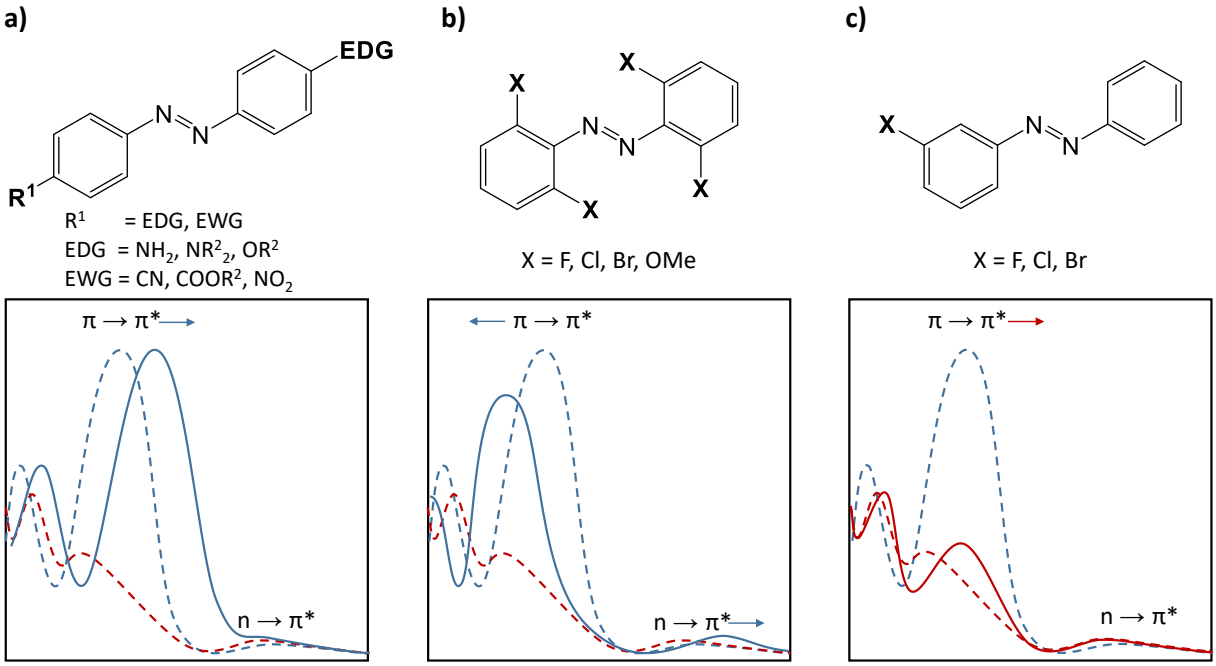


Figure 1.2-2: Substitution patterns of AB derivatives and the resulting changes in UV/vis absorbance. **a)** Electron-donating substituents or push-pull systems: bathochromic π → π* shift (*trans*-isomer); **b)** tetra-*ortho*-substituted AB: hypsochromic π → π* shift and bathochromic n → π* shift (*trans*-isomer); **c)** mono-*meta*-halogenated AB: bathochromic π → π* shift (*cis*-isomer); blue: absorbance of *trans*-AB, red: absorbance of *cis*-AB, dashed lines: absorbance of the unsubstituted parent compound.

An even more pronounced n → π* band separation was reported by Hecht et al., who implemented tetra-*o*-fluoro substitution of AB.[201] The negative inductive effect of fluorine decreases the n-orbital energy by reducing the electron density of the diazene bond, without affecting the planarity of the *trans*-isomer. This is attributed to the small atomic radius of fluorine and the hypsochromic π → π* shift is therefore not pronounced. The *cis*-isomers exhibit a remarkably high thermal stability (calc. $\tau_{1/2} \sim 2$ years). Moreover, very good isomeric ratios (> 90 %) were demonstrated, particularly for tetra-*o*-fluoro-AB derivatives that were additionally substituted with EWGs in *para*-position.[201] Other well-known strategies to obtain AB derivatives with prolonged thermal stability and responsiveness to visible light only, include heteroaryl-ABs and *o*-ethylene-bridged ABs, also known as diazocines.[202,203]

A targeted substitution of AB derivatives in *meta* (*m*)-position has been less frequently addressed in the literature. Das and co-workers synthesised a library of mono-, di- and tetra-halogenated AB derivatives with different substitution patterns (both symmetrical and asymmetrical) but did not exclusively focus on *meta*-substitution.[204] However, mono-fluorination in *meta*-position resulted in a drastic overlap of the π → π* absorption of both isomeric forms, owing to a pronounced red-shift (~ 50 nm) of the respective transition band corresponding to the *cis*-isomer (Figure 1.2-2c). By contrast, the other absorption bands remain largely unaffected. All tested fluorine-derivatives featured significantly elevated kinetic stabilities of the *cis*-isomers. This effect was more pronounced in fluoro-ABs than with analogously substituted chloro- and bromo-derivatives and increased with the number of substituents.[204] Berardino et al. investigated the London dispersion forces in differently tetra-*m*-alkyl substituted AB derivatives and reported a linear correlation between increasing substituent size and thermal stability of the *cis*-isomers.[205] Sforazzini and co-workers recently analysed the influence of asymmetric *meta*-substitution on mono-*p*-methoxy-AB derivatives with electron-directing groups.[206]

Two EWGs in *meta*-position (methyl ester) on the opposing phenyl ring induced a minor bathochromic shift of the π → π* band as well as a small hypsochromic shift of the n → π* transition corresponding to the *trans*-configuration. The π → π* band of the *cis*-isomer exhibited a red-shift by 21 nm, and the *cis*-form was also characterised by a slightly increased thermal stability. The described effects resemble those of a very weak push-pull system, since the electron-donating effect of the *p*-methoxy group is rather low. By contrast, two EDGs (methoxy) or two non-directing groups (methyl) in *meta*-positions on the opposing phenyl ring only exerted marginal effects on absorption wavelengths and half-life. Although computational studies indicated an influence of electron-directing substituents on the energy levels of frontier orbitals, the impact of *meta*-substitution on the photochromic properties seems to be rather small compared to substitution at the *para*- and *ortho*-position.

1.2.2 Principles and examples of photopharmacological research

Besides photodynamic therapy – a treatment that employs photosensitisers to induce targeted cell death via the generation of reactive oxygen species – which dates back to the 1960s and optogenetics, which was established in the early 2000s, photopharmacology is the third approach that couples visible light with therapeutic action.[207,148] In general, photopharmacology encompasses all areas, in which the photochemistry and pharmacology of bioactive molecules are intertwined. The incorporation of photosensitive molecules into pharmaceutics is intended to allow for selective release or activation in order to alter their pharmacokinetic and pharmacodynamic properties. A wide array of biological targets has been previously addressed, including soluble kinases, proteases, histone deacetylases and methyltransferases as well as acetylcholinesterases, microtubules and membrane lipids.[208,209] However, the vast majority of publications focus on transmembrane proteins, particularly ion channels, transporters and GPCRs.[189] Although there are still some challenges to be mastered in this relatively new discipline, it has the potential to overcome the major limitations of modern pharmacotherapy. In particular, this applies to the continuing problem of antibiotic and drug resistance as well as poor efficacy and selectivity of medicinal products. The latter results in the requirement of high doses, which in turn increase the risk of toxic side effects.[176]

Photopharmacological approaches are basically differentiated in terms of reversibility and interaction mode with the target, as illustrated in Figure 1.2-3. The irreversible mode of action refers to the use of photoactive prodrugs from which the bioactive agents can be released upon photoirradiation (photodecaging). Irreversible photoinactivation can also be envisaged to investigate the respective effects on biological targets.[210] The former is usually achieved by the incorporation of photolabile protecting groups at crucial positions in pharmaceuticals. Commonly used photo-caging groups include 2-nitrobenzyl and coumaryl derivatives, which are further functionalised with EDGs to obtain spectral redshift to less damaging wavelengths.[211,212] Reversibility of photopharmacological approaches, i.e. the opportunity for photochemically induced or thermally driven inactivation, is unsurprisingly achieved by the implementation of photoswitches. Corresponding derivatives can be free-diffusible ligands that interact non-covalently with a target (photochromic ligands, PCLs). Alternatively, respective photopharmaceuticals can be covalently attached to the target (so-called photoswitchable tethered ligands, PTLs) either at a natively encoded amino acid or a (genetically/chemically) engineered residue.[213] Multiple intramolecular conjugations are also possible, referred to as photoswitchable cross-linkers.

Trauner et al. further differentiate between photoswitchable closely tethered ligands (PCTLs) and photoswitchable orthogonal remotely tethered ligands (PORTLs), depending on the length of linkers employed to bridge the distance between the covalent attachment and ligand binding sites.[208] The general advantage of tethered ligands is that they are not further dispersed by diffusion. In addition, the concentration of the ligand in the vicinity of the receptor remains constant (PCTL), or at least changes only slightly (PORTL).[214]

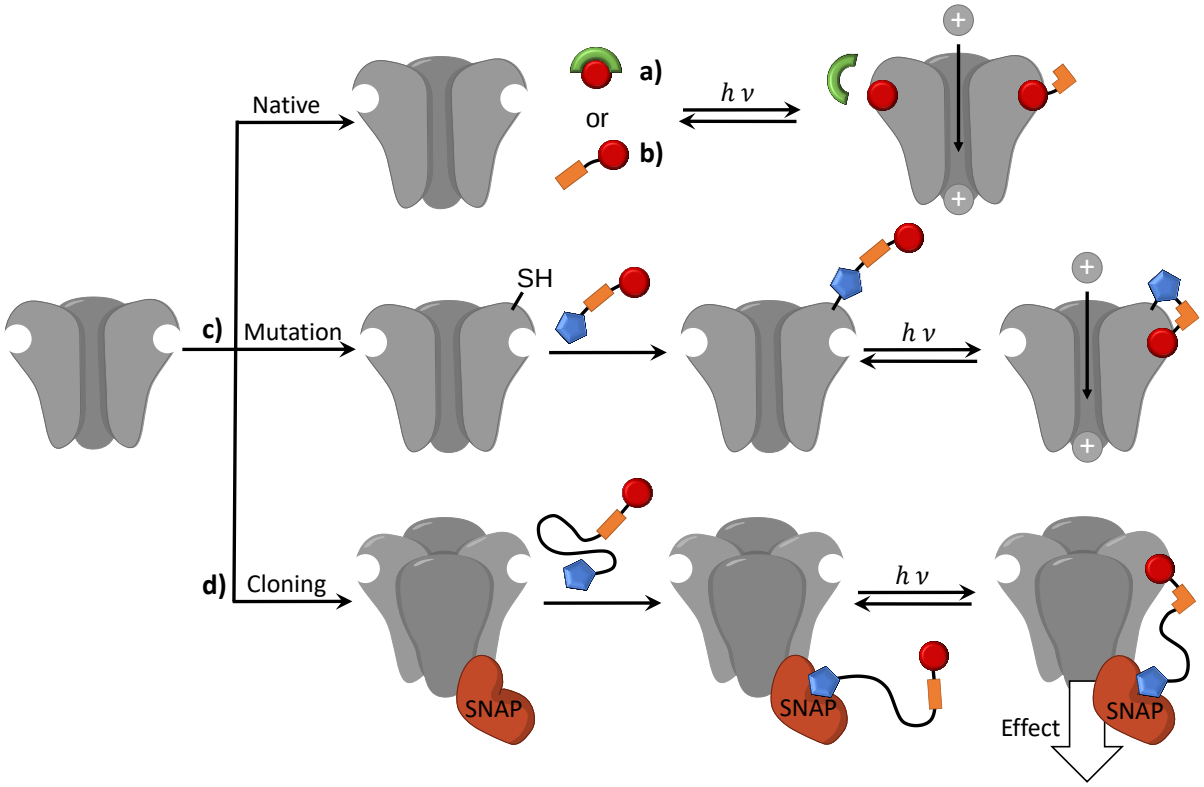


Figure 1.2-3: Common interaction modes used in photopharmacological targeting of ion channels.
a) Irreversible photoactivation (decaging); **b)** photochromic ligand (PCL); **c)** photoswitchable closely tethered ligand (PCTL); **d)** photoswitchable orthogonal remotely tethered ligand (PORTL) attached to a self-labelling protein (SNAP). Indicated in green: photocaging group; blue: bioconjugation group; red: ligand; orange: photoswitch. The illustration was adapted from a model by Trauner et al.[215]

Close tethering involves the use of short linkers or the photoswitch serves this role concomitantly. Optimally, the closesly tethered ligand can only reach the binding site in one configuration of the photoswitch but not effectively in the other. Small bioconjugation motifs (e.g. cysteine modification) are beneficial for PCTL approaches to avoid undesirable steric constraints or tension. If no suitable covalent attachment site is naturally available in the immediate proximity of the ligand binding site, and if genetic engineering is not an alternative, the PORTL approach can be considered. The latter allows the bioconjugation motif to be sterically more demanding since the (intramolecular) distance to the covalent attachment site is considerably larger. The bioconjugation site can be located on another subunit of the same protein or even be positioned on another biomolecule. Due to the very low natural abundance and the good nucleophilic properties of the thiol function, cysteine residues are in general particularly suited for chemoselective protein modification. Furthermore, the point mutation required for a single amino acid exchange is comparably straightforward and should not overly interfere with the protein structure and function.

The rapid and highly selective Michael reaction of maleimides with cysteine residues represents the conjugation method of choice for photopharmaceutical applications *in vivo*.[216,217] In addition to maleimides (Figure 1.2-4a), cysteine residues can also be selectively modified under biocompatible conditions by other means. Commonly used reagents for cysteine modification include, for instance, α-haloacids and α-haloamides – particularly iodoacetamides (IAAs, Figure 1.2-4b) – as well as sulfonates (c), active disulfides (d), haloarenes (e), and Michael acceptors other than maleimides (f).[218,219] Bioconjugation methods are also available for a variety of other natural amino acids (e.g. Lys, Tyr, Trp), which are frequently addressed in chemical biology.[220] However, these residues typically occur at higher frequency within proteins and therefore rarely represent unique modification sites, making them less suitable for selective photopharmacological targeting *in vivo*.

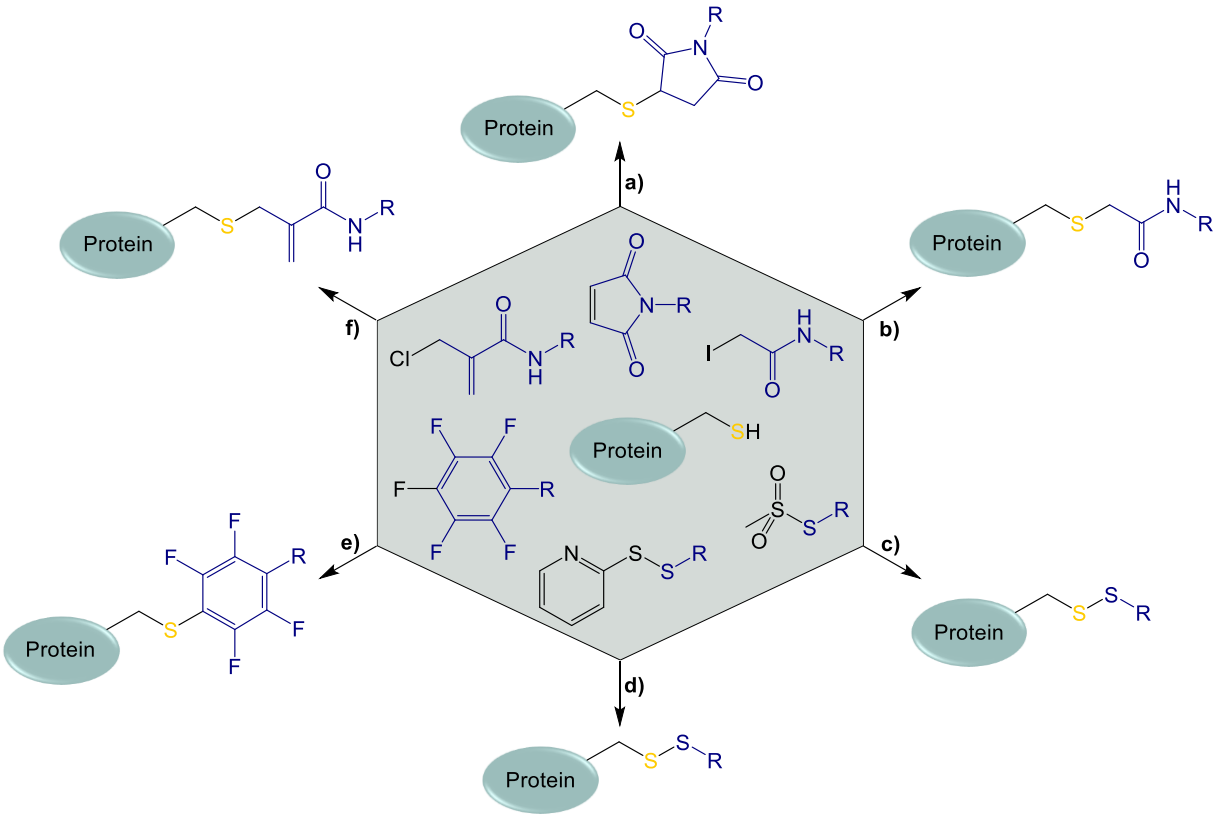


Figure 1.2-4: General structures of reagents commonly used for cysteine bioconjugation.
a) Maleimide derivatives; **b)** iodoacetamide (IAA) derivatives; **c)**; methanethiosulfonates;[221] **d)** pyridyl disulfides;[222] **e)** perfluoroarenes;[223] **f)** alternative Michael acceptors (e.g. 2-chloromethyl acrylamide derivatives).[224]

Nevertheless, cysteine modification also has its limitations and can occasionally lead to undesired side reactions, such as hydrolysis, thiazine formation as well as off-target reactivity with other cysteine residues.[225] Maleimides are furthermore prone to slow hydrolysis at physiological pH and readily form adducts with glutathione in the cytosol, which is why they are less suitable for intracellular applications.[226]. Moreover, maleimide-cysteine adducts can undergo retro-Michael reactions or other undesired side reactions under physiological conditions. An alternative bioconjugation approach relies on the introduction of unnatural amino acids, referred to as bioorthogonal ligand tethering (BOLT).[217] Residues introduced by this means may contain functional groups that do not naturally occur in biomolecules and which can thus be modified by bioorthogonal methods with exceptional selectivity.

Azides can, for instance, be converted via a (traceless) Staudinger ligation (Figure 1.2-5a),[227,228] or by bioorthogonal click chemistry such as strain-promoted azide-alkyne cycloaddition (SPAAC) that uses the ring tension of cyclooctyne derivatives as driving force of the reaction (Figure 1.2-5b).[229,230] However, the first example in which unnatural amino acids were modified in a photopharmacological approach (photo-BOLT) utilised a tetrazine-based hetero-Diels-Alder reaction with inverse electron demand (Figure 1.2-5c).[231] The reaction is strain-driven by *trans*-cyclooctene and was employed in a proof-of-principle study for *in vivo*-labelling as well as for photo-controlled inhibition of MAP (mitogen-activated protein)-kinases.[232]

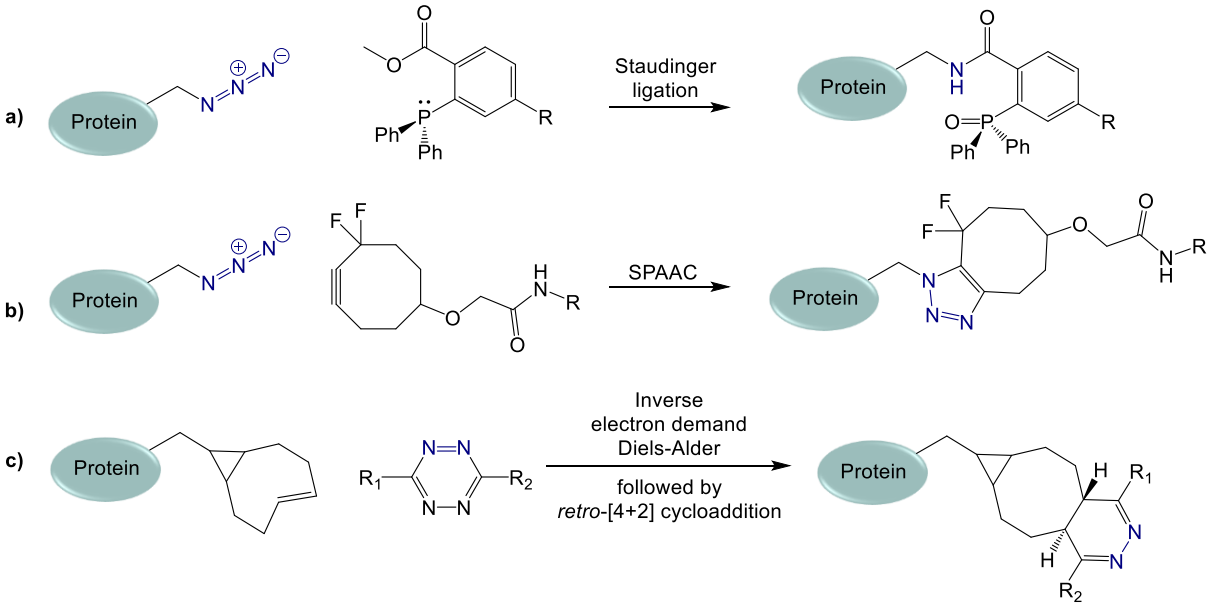


Figure 1.2-5: Chemoselective ligation strategies suitable for bioorthogonal ligand tethering (BOLT) in proteins incorporating unnatural amino acids.

Bioorthogonal labelling of azide-modified proteins using **a)** the Staudinger ligation;[227,228] **b)** strain-promoted azide-alkyne cycloaddition (SPAAC) using difluorinated cyclooctyne (DIFO) derivatives.[229,230] **c)** Bioorthogonal labelling of proteins modified with *trans*-cyclooctene via a tetrazine-based inverse electron demand Diels-Alder reaction.[231]

The use of self-labelling protein tags is a more recent strategy for bioconjugation that has frequently been used for PORTL approaches. The method is based on the introduction of engineered enzymes, usually as N- or C-terminally attached fusion proteins, into the target proteins at the genetic level. The enzyme tags convert suitably functionalised substrates (e.g. a fluorophore or a photoswitchable ligand) into covalent adducts. A common self-labelling protein is the HaloTag, a bacterial dehalogenase that covalently binds haloalkane derivatives.[233] SNAP-tags and CLIP-tags[1] are commercially available engineered variants of the human O^6-alkylguanine-DNA alkyltransferase, which covalently attach to O^6-benzylguanine and O^2-benzylcytosine substrates, respectively.[215,234] Photocages were employed by Zemelman and Miesenböck who protected capsaicin with a photocleavable 4,5-dimethoxy-2-nitrobenzyl (DMNB) group (Figure 1.2-6a).[235] The DMNB-caged capsaicin was then photoreleased for the targeted stimulation of the vanilloid receptor TRPV1 (transient receptor potential cation channel subfamily V member 1), which was expressed in mammalian hippocampal neurons.

[1] SNAP-tag® and CLIP-tag™ (New England Biolabs®) are no acronyms but trademarked brand names for the corresponding protein labelling technology based on modified O^6-alkylguanine-DNA-alkyltransferases.

TRPV1 is a non-selective cation channel that can be activated by various physical and chemical stimuli (e.g. heat, low pH, capsaicin). It mediates the sensation of pain and heat and thus represents a component of the mammalian somatosensory system.[236] By virtue of their inhibition/desensitisation effects resulting in pain relief, antagonists and agonists of TRPV1 represent promising pharmaceutical candidates for the development of analgesics.[237]

The different modes of reversible photocontrol were, *inter alia*, demonstrated using various ionotropic glutamate receptors (iGluR) as some of the first targets for photopharmacology. The iGluR family includes different subtypes of tetrameric transmembrane cation channels, which are activated upon binding of presynaptically released glutamate, thus serving as neurotransmitter.[238] iGluRs transmit a large portion of the excitation currents in the central nervous system and are essential for synaptic communication, learning, and memory.[239] Each iGluR monomer contains a clamshell-like glutamate binding domain. Agonist binding leads to a tightening of the ligand binding site, which subsequently induces the opening of the intramembrane channel. Trauner and Schiefner used an AB-equipped glutamate derivative (GluAzo, Figure 1.2-6b) as a free-diffusible photochromic ligand to target two different iGluR subtypes, which were expressed in HEK cells as well as in rat hippocampal neurons.[240] The *trans*-isomer displayed a significantly higher affinity for the receptor and induced membrane depolarisation, but reversibly lost its activity upon *trans* → *cis* photoisomerisation. The X-ray crystal structure of the iGluR-GluAzo complex revealed that binding and subsequent formation of the closed binding site conformation, essential for channel opening, is only possible in the *trans*-configuration.[241]

A similar approach was previously pursued by Trauner and Isacoff, who envisaged modification of iGluR following the PCTL approach.[242] The used ligand contained a photoswitchable glutamate moiety equipped with a short, flexible linker between the azobenzene unit and the agonist, as well as a terminal maleimide for covalent attachment to cysteine (Figure 1.2-6c). The most suitable attachment site (Cys439) was identified by screening different engineered cysteine variants. The point mutations were introduced in spatial proximity to the glutamate binding site by site-directed mutagenesis of wild type (WT) iGluR. The correspondingly modified iGluR was proven to be photo-controlled in a completely reversible and highly reproducible manner. However, in contrast to the PCL approach, the receptor became active upon photoconversion to the *cis*-isomer by irradiation with UV light. This was attributed to the linker, which invariably allowed the adoption of the closed binding site conformation in both the *trans*- and the *cis*-configuration. Importantly, bioconjugation was proven to occur only after *trans* → *cis* photoisomerisation and subsequent binding of the glutamate ligand, which actually represents a form of affinity labelling.[243] This strongly indicated that exclusively the bent geometry of the *cis*-isomer allowed simultaneous binding of the ligand and covalent attachment to the receptor. However, alteration of the attachment site or linker length resulted in the exact reversal of this photo-induced activation principle.[243] Trauner and Isacoff initially pursued the PORTL approach with the metabotropic glutamate receptor 2 (mGluR2), which belongs to the GPCR family. The receptor was N-terminally expanded with a SNAP-tag for subsequent self-labelling with a photoswitchable O^6-benzylguanine derivative (Figure 1.2-6d).[215] When expressed in mammalian hippocampal neurons, *trans* → *cis* photoconversion induced ligand binding, which in turn resulted in membrane hyperpolarisation and neuronal silencing.

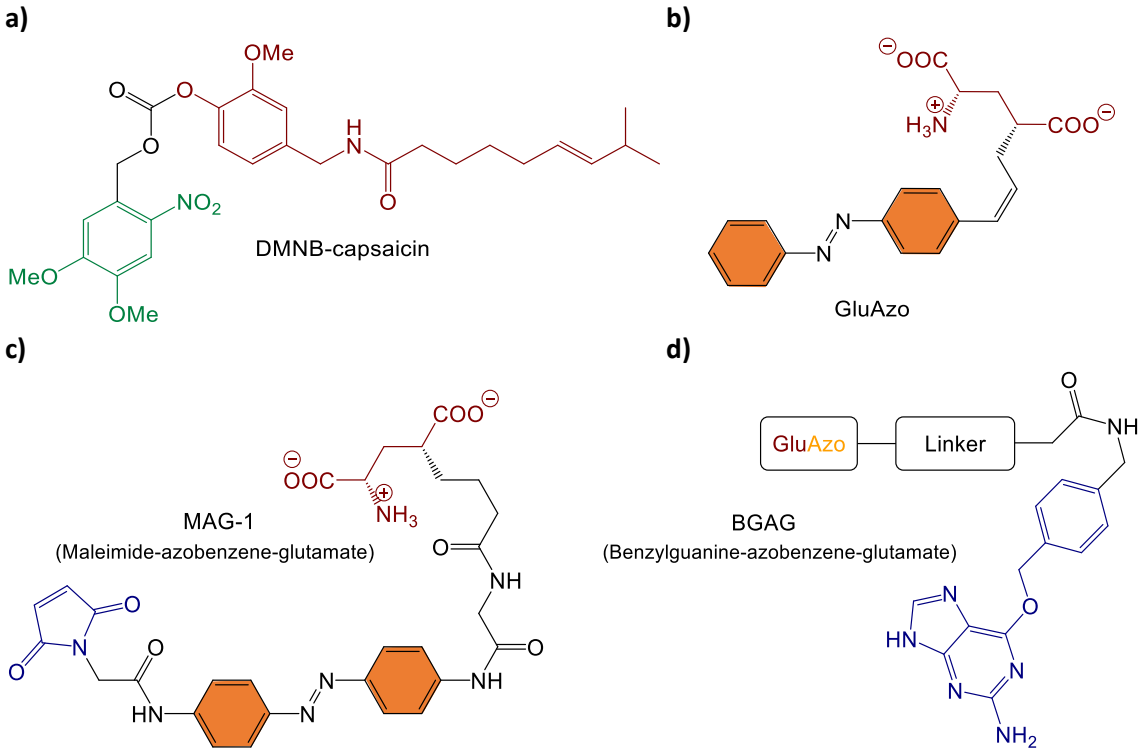


Figure 1.2-6: Photosensitive and photoswitchable ligands used for photopharmaceutical targeting of the vanilloid receptor (TRPV1) and the ionotropic/metabotropic glutamate receptors (iGluR, mGluR2). Indicated in green: photolabile DMNB group; blue: bioconjugation groups (maleimide, O^6-benzylguanine); red: agonist/ligand (capsaicin and glutamate); orange: azobenzene photoswitch.

1.3 Lectins and mannose-specific bacterial adhesion

1.3.1 The role of carbohydrates and their recognition in the field of glycobiology

After entering the cell, a large proportion of glucose is transformed into its more energetic derivative, D-glucose-6-phosphate, usually catalysed by hexokinase under ATP consumption.[244] Along the metabolic pathways, the activated glucose derivative can be converted into a range of other monosaccharides, which in turn may serve as building blocks for a plethora of di-, oligo-, and polysaccharides. Therefore, carbohydrates not only represent the structurally most diverse class of natural substances but are also the most abundant macromolecules on our planet, which we encounter literally everywhere in living nature. In addition to the aforementioned role as a primary energy source and storage medium, the intracellular glycome mainly consists of glycoproteins localised in the endoplasmic reticulum (ER) and Golgi apparatus and constitutes the structuring component of nucleic acids (RNA, DNA) as well as of small mono- and dinucleotides. The latter serve as important cofactors, for instance in enzymatic redox reactions (NAD, FAD), for acyltransferase reactions (coenzyme A), or as signalling molecules (cNMPs, GTP).

However, the vast majority of glycans in terms of quantity can be found on cell membranes and the overlying extracellular structures. The most common macromolecule in nature is cellulose, which constitutes the main element of the plant cell wall and provides the necessary strength to counteract the high turgor pressure of the protoplast.[245] Cross-linking of the cellulose fibrils with structurally more complex hemicelluloses and interweaving of the heteropolysaccharide pectin also provides additional reinforcement and the necessary flexibility for primary cell wall extension and length growth.[246] The second most common biomacromolecule is chitin, which represents the main component of the fungal cell wall and the exoskeleton of arthropods.[247] Chitin is also found in the mouthparts of molluscs and in articulata.[248] In bacteria, glycan structures play an important role as structural component of the peptidoglycan layer in both Gram-positive and Gram-negative cell walls, as antigens in lipopolysaccharides, in biofilms, capsules as well as in slime layers, which protect against dehydration and phagocytosis.[249] These “sugar coatings” in the form of capsules and slime layers are collectively subsumed as the bacterial glycocalyx, which is derived from the Greek words *glykys* (sweet) and *kályx* (husk).[250] However, the glycocalyx often refers to the entirety of glycoconjugates on eukaryotic cell surfaces, particularly on animal cells. The glycan layers can vary greatly in thickness and composition depending on the cell type and are particularly pronounced on vascular endothelial cells where the glycocalyx fulfils a pivotal role as permeability barrier.[251] Primarily responsible for the structural variability in different species, cell types and differentiation stages is the differential expression of glycosidases and glycosyltransferases, which mediate biosynthesis of glycans in the lumen of the ER and the Golgi complex.[252] Oligosaccharides are, *inter alia*, attached to glycoproteins as a post-translational modification. Classically, this modification occurs either on asparagine (*N*-glycosylation) or on serine/threonine side chains (*O*-glycosylation), and the resulting glycoproteins are subsequently transported to the lysosome or to the cell membrane via the exocytic pathway.[253,254]

N-glycans are initially synthesised as dolichol-linked precursor oligosaccharides at the cytosolic side of the ER membrane and are then further modified and finally attached to an N-X-S/T motif within the nascent protein in the ER lumen. Further processing of *N*-glycan precursors subsequently occurs in the Golgi complex. Mature *N*-glycans contain a common core-pentasaccharide (Man*α*1→3-(Man*α*1→6)-Man*β*1→4GlcNAc*β*1→4GlcNAc*β*1→Asn). The latter can be further modified either exclusively with mannosyl residues (high-mannose type), with antennae of GlcNAc-, galactosyl-, sialyl- and fucosyl residues (complex type), or a combination of both (hybrid type).[253] By contrast, *O*-glycans exhibit neither a defined recognition sequence nor a uniform core structure, although they typically share the initial GalNAc moiety, which is glycosylated to the S/T residue as the first step in the biosynthetic pathway.[254] The term glycoprotein is applied as long as the protein content is significantly larger than the attached glycan structures. However, if the glycan content predominates, as is the case with some heavily *O*-glycosylated proteins (e.g. mucins), they are referred to as proteoglycans. The long polysaccharide appendages are thus termed glycosaminoglycans (GAGs) or mucopolysaccharides, respectively.[255] Apart from classical *N*- and *O*-glycans, a variety of new glycosylation patterns were found on certain eukaryotic protein domains, which for instance include *O*-*α*-fucosylation, *O*-*β*-glucosylation, *O*-*α*-mannosylation as well as modification with *O*-*β*-GlcNAc.[256] In addition to glycoproteins and proteoglycans, mono- and oligosaccharides can be directly attached to the apical cell membrane as carbohydrate-lipid conjugates, better known as glycolipids. These are further differentiated into glyceroglyco- and glycosphingolipids depending on whether the carbohydrate moiety was esterified to glycerol or to sphingosine, respectively.[257] The eukaryotic glycocalyx plays an eminent role in various physiological and pathological processes. In addition to its function as a physical and permeability barrier, the glycocalyx is of crucial importance for various intercellular recognition, communication and adhesion events.[258] This applies to cell-cell interactions within the same organism (intrinsic recognition), which is important, for instance, for the modulation of red blood cell volume in the vascular system but also for cellular recognition during fertilisation and embryonic development.[251,259] On the other hand, the glycocalyx also facilitates the recognition of foreign cells (extrinsic recognition), including invading pathogens. Therefore, glycocalyx-mediated cell-cell recognition is also strongly involved in infection processes as well as inflammatory immune response.

Considering the widespread distribution and essential functions of glycans in nature, it is not surprising that all extant organisms feature proteins responsible for carbohydrate recognition. In the context of glycobiology, only lectins and GAG-binding proteins are considered legitimate glycan-binding proteins. In contrast to this, antibodies and enzymes that can bind or convert carbohydrates are not classified as glycan binding proteins according to the glycobiological definition.[260] Lectins were thus classified as proteins of non-immune origin that solely bind (and not enzymatically convert) carbohydrates.[261] From the first description of Ricin in 1888 by P. H. Stillmark, which was isolated from the seeds of the castor plant *Ricinus communis*, to the introduction of the designation by W. C. Boyd and E. Shapleigh in 1954, lectins were generally named and identified according to their agglutinating (i.e. cell-clumping) effect.[262] The term “lectin” for plant-derived agglutinins is derived from the Latin verb *legere* (collect, select), which alludes to their early use as specific agents in the differentiation of the human ABO blood group system.[263] In the literature, lectins are also referred to as precipitins, hemagglutinins, leukoagglutinins, adhesins and toxins depending on the initially observed effects or functions. The selective carbohydrate-binding ability of lectins was first described in 1936 by J. Sumner and S. Howell who demonstrated the precipitation of glycogen with the plant lectin concanavalin A (ConA) from jack bean (*Canavalia ensiformis*).[264]

As it became evident several decades later, the agglutinating effect of ConA (and many other, particularly plant-derived lectins) is the result of an oligomerisation of (typically two or four) lectin subunits, which enables cross-linking interactions with their (often likewise multivalent) glycan epitopes.[265] Although each individual lectin-glycan interaction is typically relatively weak, expressed by dissociation constants (K_D) in the micromolar to millimolar range, the multivalent nature of carbohydrate recognition can considerably increase the overall affinity to the receptor, which is then referred to as "avidity".[266] Lectins usually feature evolutionary conserved primary and tertiary structures, termed carbohydrate recognition domains (CRDs), that can be applied to categorise lectins into different families. The best-known lectin families include, for instance, R-type, L-type, P-type, C-type and I-type lectins, galectins (formerly S-type), and microbial as well as viral lectins.[267] In many cases, these CRDs represent independent, stable protein folds that provide a a well-defined or shallow binding cavity for the recognition of single or multiple, terminally located carbohydrate residues.[260] In this respect, lectins differ significantly from GAG-binding proteins, which usually share no sequence homologies or common protein folds and presumably have evolved through convergent evolution.[268] The interaction with the GAG chains generally occurs at internal positions and is mainly based on ionic interactions of basic amino acid residues with negatively charged functional groups on the monosaccharide units.

Although lectins occur ubiquitously in all domains of life, the structural and functional diversity is by far the highest in the realm of Animalia. In connection with the physiological significance and cell-specific composition of the mammalian glycocalyx, some lectin types are involved in important cellular adhesion processes. For this reason, many lectins are membrane-associated and located on cell surfaces. This applies, for instance, to the selectin family, which is an important class of cell adhesion molecules present on the surface of endothelial cells (E-), leukocytes (L-) and thrombocytes (P-selectins), respectively.[269] Apart from that, mammalian lectins are mainly involved in the regulation of the immune response (e.g. siglecs), apoptosis (e.g. galectins) and intracellular glycoprotein sorting in the ER (calnexin, P-type lectins).[270–273] In spite of their simple extractability from plants and their extensive use in medical research and life sciences, the physiological functions of plant-derived lectins have been poorly investigated so far. Particularly high concentrations of L-type lectins can be found in various organs of eponymous leguminous plants (e.g. ConA, phytohemagglutinin).[274] Due to the occasionally demonstrated insecticidal, antimicrobial and antifungal properties as well as their inherent toxicity and poor digestibility for herbivores, an involvement in plant immunity and protection against predators can be assumed.[273] As opposed to this, viral, bacterial and parasite lectins are very clearly associated with pathogenicity since they are usually required for the specific recognition of certain host cell tissues.[275] Carbohydrate-specific binding or adhesion often represents one of the first steps and a prerequisite for a subsequent host colonisation and infection. The inhibition of bacterial and viral lectins by potent antagonists, thus, represents an interesting therapeutic approach, as an alternative or to complement the conventional treatment of infections, in the setting of an anti-adhesion therapy.[276–278] Compared to conventional antibiotics, anti-adhesion therapy offers the key advantage of exerting minimal selective pressure, thereby significantly reducing the risk of resistance development. Moreover, the highly specific mode of action preserves the host microbiome, minimises off-target effects, and remains effective even against antibiotic-resistant bacterial strains.

In the context of infections, lectins are commonly referred to as adhesins or hemagglutinins, respectively, if erythrocytes become infected or agglutinated. The specificity of lectins for certain receptor structures can be decisive for host tropism i.e. the preference for infecting certain host species or distinct tissues. The best-known representative is certainly the hemagglutinin from the influenza A virus, which is responsible for recognition of sialyl receptors on host cell surfaces, followed by endocytosis and release of the ribonucleoproteins into the cytosol.[279] The human influenza A virus preferably binds receptors with terminal *N*-acetylneuraminic acid (Neu5Ac), which is $\alpha(2\rightarrow6)$-linked to galactose. By contrast, avian influenza A prefers $\alpha(2\rightarrow3)$-linked sialyl receptors, which are predominant in the intestinal epithelium in birds.[280] Apart from the adhesion and binding of virions and other pathogens, some secreted toxins also exert their function by specific carbohydrate recognition. Well-known examples are, for instance, the B-chains of choleratoxin and shiga toxin produced by *Vibrio cholerae* and *Shigella dysenteriae*, respectively.[281] The specific binding of the B chain to glycolipids on the host cells enables endocytotic uptake. Usually after cleavage of the toxin subunits in the ER, the A-chain enters the cytosol where it generally inhibits protein biosynthesis by depurination of the 28S rRNA, ultimately resulting in cell death.[282]

1.3.2 Type 1 fimbriae

Gram-negative as well as some Gram-positive bacteria have evolved specialised hair-like organelles, which are peritrichously distributed on the cell surface and incorporate adhesins to enable bacterial adhesion to various surfaces. These adhesins can include lectins that recognise carbohydrates on host cell receptors, but may also bind (glyco)proteins in the extracellular matrix (ECM), such as collagen or fibronectin.[283,284] The filamentous organelles are referred to as adhesive pili or fimbriae to distinguish them from other microbial cell appendages, such as type III and type IV secretion systems. The latter are conjugative (F- or P-) pili, which are involved in horizontal gene transfer, whereas type III secretion systems form microneedles for the injection of effector proteins into host cells.[285,286] Pili should also be differentiated from flagella, which are required for bacterial locomotion just as in the eukaryotic chlorophyte *C. reinhardtii*. Fimbriae are classified according to their biosynthetic pathways in Gram-negative bacteria, namely chaperone-usher (CU) pili, type IV pili, curli (assembled by the extracellular nucleation-precipitation pathway), and the alternate CU family.[287,288] Type IV pili are not only involved in adhesion but can also contract after surface attachment, resulting in a sudden movement termed twitching motility.[289] All classes of fimbriae are occasionally involved in pathogenesis by promoting bacterial adhesion, invasion and biofilm formation, and therefore represent important virulence factors that can potentially be targeted as part of an antiadhesive treatment.

Type 1 fimbriae are by far the best characterised representatives amongst CU pili, which in turn represent the most ubiquitous class of adhesive pili in Gram-negative bacteria.[290,291] The designation derives from the requirement of a periplasmic chaperone and a pore protein (“usher”), anchored in the outer membrane for fimbrial assembly. Type 1 fimbriae are highly conserved in the intestinal bacterium *Escherichia coli* (*E. coli*) including benign, commensal varieties of the gut microbiome as well as in several enteric pathotypes.[291] The latter include, for instance, enteropathogenic (EPEC), enterotoxigenic (ETEC), enterohemorrhagic (EHEC), and adherent-invasive *E. coli* (AIEC), which are connected to different forms of diarrhoea and Crohn’s disease (CD), respectively.[292] In the literature, type 1 fimbriae are classically associated with uropathogenic *E. coli* (UPEC), which are responsible for the vast majority of urinary tract infections (UTIs).[293] UTIs are among the most frequently diagnosed bacterial infections and severe disease progression can be associated with serious complications, such as pyelonephritis and urosepsis, especially in debilitated and elderly patients. Other pathogenic representatives of the Enterobacteriaceae family also express type 1 fimbriae, such as *Klebsiella spp.* and *Salmonella spp*., causing salmonellosis, UTIs, pneumonia, and other (nosocomial) infections.[294,295]

All main functional and structural elements, required for the expression and assembly of type 1 fimbriae are encoded within the shared *fim* cluster. The latter contains genes for the regulatory proteins *fimB* and *fimE* as well as the *fimAICDFGH* operon.[296,297] The reference genome from *E. coli* K12 substrain MG1655 localises the *fim* cluster between genes coding for the *N*-acetylneuraminic acid-inducible outer membrane channel (NanC) and the fructuronate transporter (GntP). The *fimB* and *fimE* genes both encode DNA recombinases that regulate the expression of the type 1 pilin subunits (also termed pilins) on a transcriptional level. This is achieved by phase variation of the *fim* switch (*fimS*) as shown in Figure 1.3-1.[298] The latter is a 314-bp invertible DNA element, located between the *fimE* and *fimA* sequences that, *inter alia*, contains the *fimA* promoter (pA).[299] It is flanked by two 9-bp inverted repeats, which enable *fim* switch inversion, mediated by the site-specific FimB and FimE recombinases. Transcription of the *fimAICDFGH* operon is thus abrogated since the pA promoter points to the opposite direction.

The recombinase FimB is able to reversibly catalyse recombination in both directions whereas FimE only mediates unidirectional inactivation of type 1 fimbriae-expression. FimE expression is in turn governed by a negative feedback regulation depending on the orientation of the *fim* switch.[300] Inverting *fimS* to the "phase-off" orientation results in Rho-dependent termination of *fimE*-transcripts, which consequently feature a significantly decreased stability compared to mRNA that was transcribed in the "phase-on"-orientation. FimB and FimE feature separate promoters (pB, pE) and are accordingly up- and down-regulated to promote increased fimbriae expression at temperatures ≥ 37 °C, presence of iron ions, neutral pH and low salt conditions.[301–303] These conditions indicate residence of the respective bacterium within the host or on distinct host tissues. Fimbriae expression can thus be upregulated when required by adjusting the equilibrium between the two recombinases in favour of FimB.

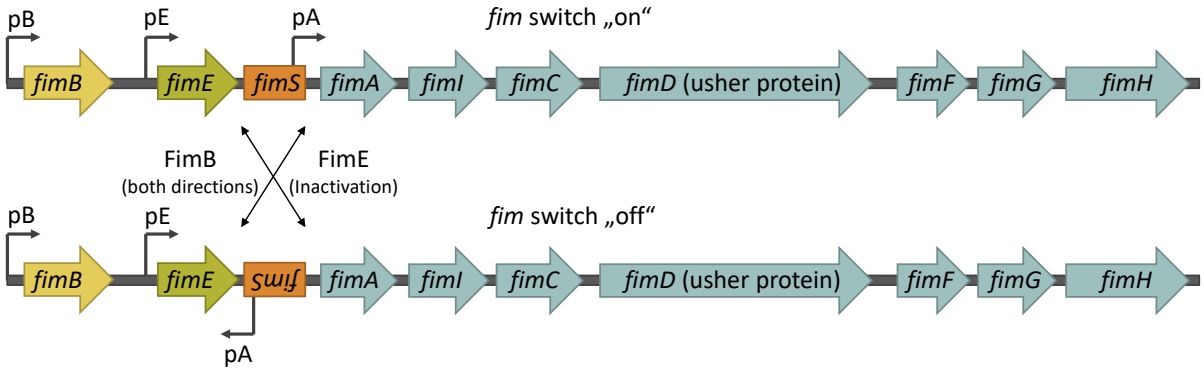


Figure 1.3-1: Structure and transcriptional control of the *fim* gene cluster.
The encoded DNA recombinases, FimB and FimE, regulate pilin expression by shifting the invertible *fimS* segment between a transcriptional "phase-on" and a non-transcriptional "phase-off"-orientation.

The *fimAICDFGH* operon encodes the chaperone (FimC), the pore-forming usher (FimD) and the structural pilin subunits, namely the major subunit FimA, the adapter proteins FimF and FimG, as well as the terminal two-domain adhesin, FimH. The *fimI* gene encodes a 16.4-kDa non-cytoplasmic protein that has been shown to be essential for fimbriae assembly, but its definite function is still unclear.[304] A role as an anchor subunit, as has been evidenced for other CU pili (e.g. PapH in P fimbriae), can be assumed due to sequence homologies.[305] The *fim* cluster further encodes a small regulatory mRNA (*fimR2*), which accumulates in the stationary growth phase and regulates several phenotypes that play an important role in pathogenesis including biofilm formation, motility, and cell morphology.[306] Apart from that, fimbriae expression is influenced by a number of other transcriptional regulators such as the integration host factor (IHF).[307] The *fimAICDFGH* transcript becomes processed by ribonuclease E and is eventually translated into pilin proteins.[308]

On electron micrographs showing type 1 fimbriae-expressing *E. coli*, piliation appears as a fine mesh formed by several hundred linear fibrils of varying lengths, protruding from all around the cell.[309] In the literature, slightly different pilus lengths between 1 – 2 µm and diameters between 7 – 8 nm at the base (pilus rod) and approximately 2 nm at the flexible tip fibrillum are mentioned.[305,310] The architecture of a mature type 1 fimbria, exhibiting this subdivision into rod and tip, is schematically illustrated in Figure 1.3-2a. All cytosolic pilus subunits contain N-terminal signal peptides that predetermine their transport via the secretory pathway. The still unfolded pre-pilins are transported across the inner membrane through the heterotrimeric SecYEG translocon complex (SecY, SecE, and SecG), whereupon the N-terminal signal peptides are removed.[311]

After translocation into the periplasm, disulfide bonds are introduced into the pilins mediated by the thiol:disulphide oxidoreductase (DsbA).[312] Disulfide-bridged pilins are rapidly and preferentially recognised by the chaperone FimC, which thus stabilises the pilin subunits until their incorporation into the growing fimbria.[313] In the absence of FimC, the misfolded Fim proteins are extremely prone to aggregation and become hydrolysed by the periplasmic serine endoprotease DegP.[314] This is mainly due to the incomplete immunoglobulin-like (Ig) fold of the majority of pilins, which consists of only six (instead of the otherwise usual seven) antiparallel β-strands forming the characteristic dual β-sheet structure.[315] The seventh C-terminal strand (also called the G β arm) is absent, thus exposing a hydrophobic groove (residues P1 – P5) in the centre of the pilin structure. On the other hand, the N-terminal extension (Nte, 10 – 20 residues) of the Fim proteins remains unstructured and flexible. The chaperone also adopts two Ig-like domains resulting in an overall L-curved shape.[315] Within the binary chaperone-pilin complex, the incomplete Ig-like fold of the Fim subunit is complemented *in trans* by the G1 β-strand of the chaperone (Figure 1.3-2b). The strand contains four alternating, hydrophobic amino acid residues (P1 – P4), which occupy the complementary positions in the pilin subunit whereas position P5 remains unallocated.[316] Unlike the canonical Ig fold, the inserted strand is aligned in parallel to the neighbouring F strand of the pilin. This interaction, referred to as donor strand complementation (DSC), protects the pilins from periplasmic aggregation as well as degradation and significantly accelerates the folding process.[313] Co-crystallisation of Fim proteins with FimC (or at least a peptide that imitates the G1 β-strand) enabled the determination of various atomic-resolution X-ray crystal structures, which would otherwise not be available due to the pronounced instability of the majority of pilin subunits.[313,317–319]

The Nte of each pilin subunit likewise features a sequence of alternating hydrophobic and bulky amino acid residues, similar to the chaperone's G1 β-strand. During fimbriae formation, the complementing G1 β-strand becomes replaced by the Nte of an incoming subunit according to a "zip-in, zip-out" mechanism.[320] The Nte enters from the unoccupied P5 position and procedes to the P1 groove, thus resulting in the canonical, antiparallel orientation towards the F strand (Figure 1.3-2b). In this way, the pilin subunits are not only stabilised, but also become interconnected to form the fimbriae. This process, termed donor strand exchange (DSE), is catalysed by the usher FimD.[321] The multidomain protein FimD forms a 24-stranded β-barrel TM pore through which the growing type 1 fimbria is passed across the outer membrane.[322] A so-called plug domain is inserted between the strands six and seven of the β-barrel, which seals the channel in its resting state (*apo*-FimD) in order to prevent unspecific membrane transport.

The usher becomes activated upon binding of an incoming FimC-FimH-complex to the N-terminal domain (NTD) of the FimD channel.[323] In the course of the subsequent conformational change, the plug tilts to the side of the NTD, and the channel widens by 100 Å from an oval to a round shape. Thereby, the tandem C-terminal domains (CTDs) of FimD become exposed, to which the FimC-FimH complex is transferred, thus allowing the NTD to recruit the next chaperone-pilin (FimC-FimG) complex.[324] The CTDs mediate DSE of the incoming FimC-FimG complex, whereby the Nte exchanges the G1 strand in the preceding FimC-FimH complex, followed by chaperone release. Simultaneously with the next handover step, the nascent fimbria is pushed through the pore towards the extracellular space.[305] The transfer of the pilins from FimC to the FimD-NTD up to the CTDs is probably driven by affinity differences.[325] Eventually, the tip fibrillum is assembled after recruitment and DSE with the next adapter protein, FimF. Henceforth, only the major subunit FimA becomes incorporated, in order to form the pilus rod. Approximately 500 to 3000 copies of FimA are arranged into a right-handed helical structure, consisting of 3.4 subunits per turn.[326]

Cryo-electron microscopy revealed a hollow tubular structure with a diameter of 69 Å, in which each FimA subunit rotates and translates around the axis by 115° and 7.7 Å, respectively.[327] Adhesive fimbriae have also been described in organisms that lack the functional genes for the tip fibrillum (FimF, FimG, FimH). In the periodontal pathogen *Porphyromonas gingivalis*, for instance, fimbriae were identified that are almost exclusively composed of FimA subunits and anchored to the outer membrane via a single FimB homologue. These FimA pili mediate binding to a variety of salivary and cell surface proteins, thus enabling recognition and colonisation of host cells in the oral cavity.[328]

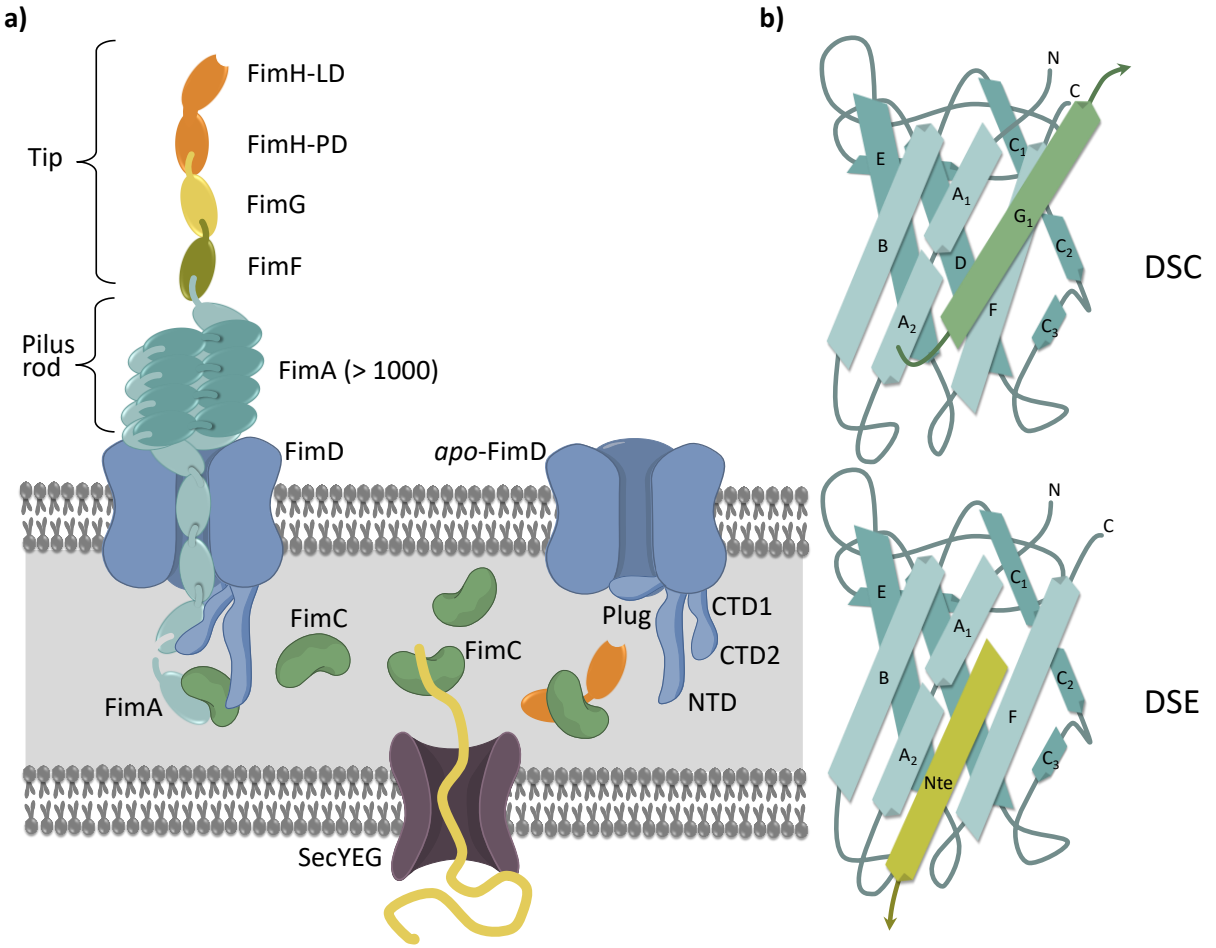


Figure 1.3-2: a) Structure and assembly of type 1 fimbriae. **b)** Donor strand complementation (DSC) and donor strand exchange (DSE) to stabilise the incomplete Ig-like fold of a pilin subunit.

The pilin structure in **b)** becomes complemented via DSC with the G1 β of the chaperone (FimC) or via DSE by the N-terminus of an incoming Fim protein. The illustration was adapted according to a model by Sauer et al.[321]

1.3.3 The bacterial lectin FimH

The two-domain adhesin FimH is located at the distal ends of canonical type 1 fimbriae.[329] Sharon and colleagues have proven that FimH is a lectin that specifically binds *α*-D-mannose and *α*-D-mannosides, thereby mediating adhesion to highly mannosylated glycan structures on host cell surfaces.[330] However, natural FimH variants (as well as those obtained by random mutagenesis) were described that featured significantly reduced or altered mannoside recognition.[331,332] The respective mutations may be pathoadaptive for the prevailing colonisation conditions of distinct host tissues or may still allow adhesion to abiotic surfaces, which can be beneficial for biofilm formation.[333] In any case, as a highly prevalent adhesin in many pathogenic members of the Enterobacteriaceae family, FimH represents an important virulence factor. Previous and current research therefore focused on the bacterial lectin in terms of an anti-adhesive treatment, particularly in the context of UTI and CD.[334]

Disregarding the N-terminal signal peptide encompassing the first 21 amino acids (aa), the protein features a molecular mass of 29.1 kDa (279 aa) after secretion into the periplasm. It is composed of two independent domains connected by a short inter-domain linker (aa 156 – 160), namely the N-terminal lectin domain (LD, aa 1 – 155, M_w 17 kDa) that comprises the mannose-specific CRD and the C-terminal pilin domain (PD, aa 161 – 279, M_w 12 kDa), which anchors the adhesin to the type 1 pilus.[315] The lectin from *E. coli* K12 is probably the best-studied FimH representative due to the high sequence similarity with many UPEC strains, such as the well-known isolate from human urine, *E. coli* J96. The lectin variant from *E. coli* isolate F18, on the other hand, frequently serves as a representative of faecal origin.[335–337] The lectin domains (FimH-LD) from these strains differ in three distinct positions (K12/J96 → F18: V27A, N70S, S78A).

The high-mannose type *N*-glycan on Uroplakin-Ia (UPK-Ia) is regarded as the main natural receptor for FimH-mediated adhesion in the urothelium.[338] Uroplakins are heterotetrameric plaque proteins that form rigid, hexagonal complexes known as asymmetric unit membranes (AUMs), which extensively cover the urothelial surface.[339,340] AUMs presumably regulate the membrane permeability and strengthen the barrier function of the urothelium. Each heterotetramer comprises the subunits UPK-Ia, UPK-Ib, UPK-II and UPK-IIIa. UPK-Ia/b contain four TM domains whereas UPK-II and UPK-IIIa only constitute a single TM helix. As shown in Figure 1.3-3a, the *N*-glycan on UPK-Ia is conjugated to Asn169 between the extracellular TM helices 3 and 4 at the interface to an adjacent UPK-Ib subunit. By contrast, UPK-Ib and UPK-IIIa feature complex *N*-glycans whereas UPK-II remains non-glycosylated.[341] Within the high-mannose type *N*-glycan, the terminal mannopyranosides are either attached to the core pentasaccharide via *α*(1–3)-, *α*(1–2)-, or *α*(1–6)-glyosidic linkages, which are recognised with descending affinities.[342] Non-native *α*(1–4)-linked mannosides are recognised with a relatively weak affinity, which is approximately in the same order of magnitude as *α*(1–6)-mannopyranosides.[343] An additional natural receptor for FimH is the glycoprotein uromodulin also known as Tamm-Horsfall protein.[344] Uromodulin is excreted in large quantities into the urine, where it forms long, filamentous aggregates.[345] As it is highly-mannosylated, uropathogenic bacteria may adhere to uromodulin rather than to UPK-Ia, thereby providing protection against UTI. An UPK-III-dependent cellular response to FimH-mediated bacterial adhesion to UPK-Ia results in apoptosis of the respective urothelial cells.[346] In addition, FimH is recognised by murine and simian antibodies that mediate immune response in the respective organisms.[347]

Apart from the often mentioned main receptors UPK-Ia and uromodulin, FimH was found to bind to various other glycoproteins in the extracellular matrix, including collagens, fibronectins, laminins as well as to the cell surface of erythrocytes and leukocytes, respectively.[305,348] Adhesion to epithelial glycoproteins, particularly members of the carcinoembryonic antigen-related cell adhesion molecule family (CEACAM6) or Toll-like receptors (TLR4s) in the small intestine, was associated with AIEC-mediated adhesion and inflammation in early stages of CD.[349]

In the past decades, FimH-mediated adhesion and recognition of type 1-fimbriated bacteria was frequently investigated with yeast cells (e.g. *Candida albicans*) and erythrocytes from guinea pigs.[350,351] As both cell types display large amounts of high-mannose type *N*-glycans on their surface, co-incubation with piliated *E. coli* results in a pronounced, macroscopically visible hemagglutination.[352,353] In their publication from 2002, Sokurenku and co-workers compared the hemagglutinating effect of diluted bacterial suspensions under static and dynamic conditions (e.g. agitation or flow-chamber setup).[354] Whereas no agglutinating effect on guinea pig erythrocytes was observed under static conditions, an enhanced hemagglutination was still detectable under dynamic (i.e. shear-mimicking) conditions despite the low bacterial density employed. In this way, an experimental proof for a so-called “catch bond” was experimentally demonstrated for the first time. According to Dembo et al., a catch bond is defined as a non-covalent interaction (e.g. between a ligand and its proteinaceous receptor), which counter-intuitively increases under application of tensile forces operating along the bond axis.[355] By contrast, if the bond lifetime considerably decreases under the impact of shear, this is termed a “slip bond”. In theory, both bond types are contrasted with the “ideal bond”, which is independent of mechanical tensile forces (Figure 1.3-3b). However, if the shear forces increase beyond measure, even the strongest receptor-ligand bond will eventually dissociate in a slip bond-like manner. As this will also be the case with a genuine catch bond, the latter is occasionally also referred to as a “catch-slip bond”.

A slip bond is usually encountered in the vast majority of ligand-receptor interactions. Apart from the FimH-associated catch bond, similar binding modes have also been reported for various cell adhesion molecules (CAMs) and other proteins interacting with the cytoskeleton or the ECM, such as integrins, cadherins or vinculin.[356–358] Well-known examples include, in particular, the catch bond mechanisms found in P-, L-, and E- selectins, which play a fundamental role in shear-controlled leukocyte rolling-adhesion and recruitment.[359] Meanwhile, shear-enhanced adhesion has been documented for other pathogenic, both Gram-negative as well as Gram-positive, bacteria indicating a catch bond behaviour.[360–362] In the case of type 1-fimbriated *E. coli*, the catch bond represents an optimised adaptation towards the variable shear conditions present in the intestines and urinary tract. While the bacterial mobility is not impaired by the short-lived lectin binding events under static conditions, the activation of catch bonds under high shear forces certainly reduces elimination through defecation and urination. The known examples from nature are generally characterised by a biphasic binding mode since the catch bond process is usually triggered above a certain minimal intensity of the mechanical tensile forces applied. Expressed as a potential curve on a free energy landscape, the binding mode can be described as an equilibrium between at least two states, S1 and S2, reflected as distinct potential wells (Figure 1.3-3c).

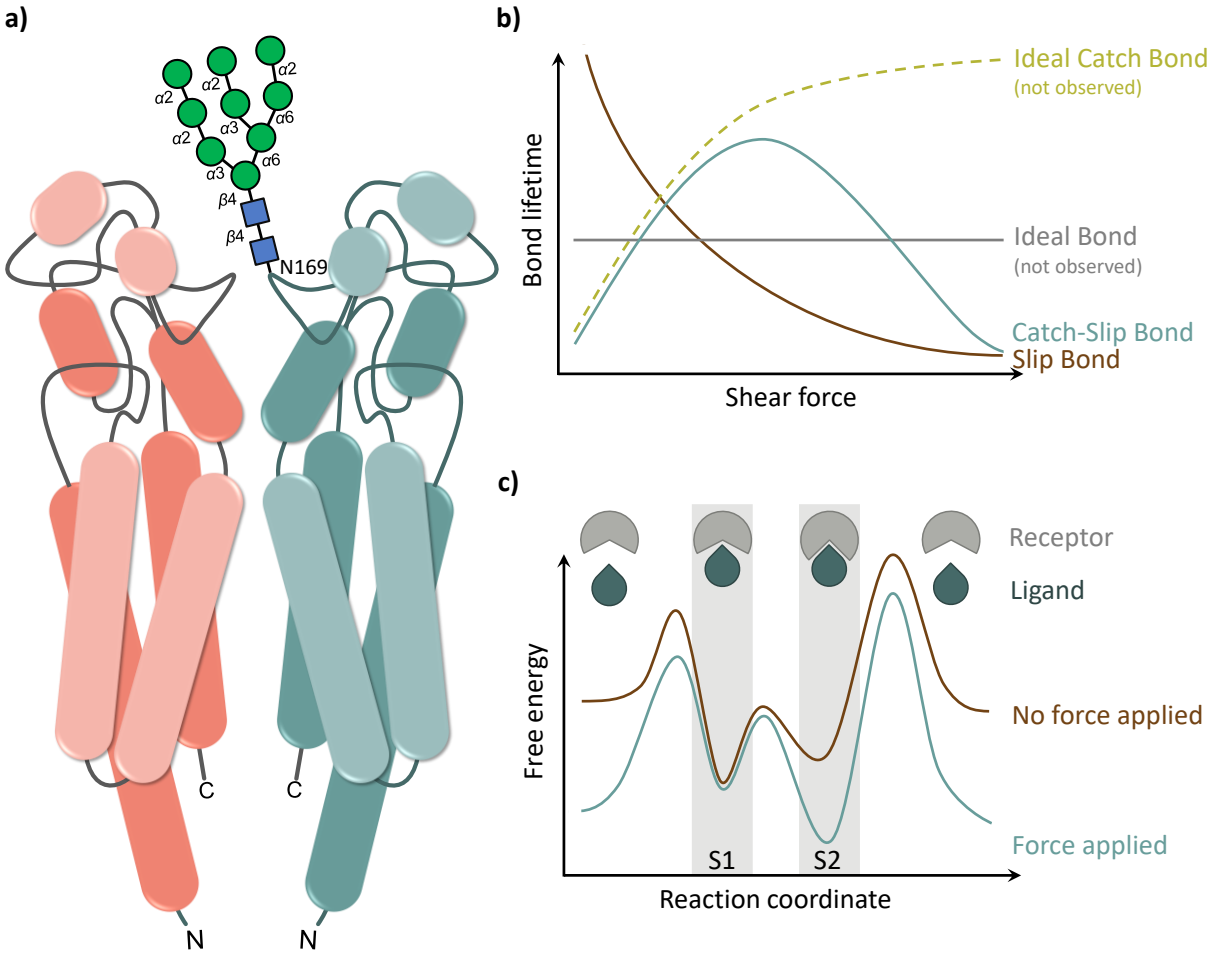


Figure 1.3-3: a) Proposed structure of a Uroplakin-Ia/Ib heterodimer with an exemplary high-mannose type *N*-glycan attached to Asn169 on UPK-Ia, which represents an important natural receptor for the type 1 fimbrial adhesin (FimH) in the urothelium. **b)** Force-lifetime plots, illustrating the different binding modes of non-covalent ligand-receptor interactions.[363] **c)** Proposed double-well free energy landscape of a genuinely observable catch(-slip) bond interaction.[364]

The illustration of a human UPK-Ia in **a)** was adapted according to a predicted structure from the AlphaFold Protein Structure Database (AF-O00322-F1-v4).

In the absence of shear, the "inactive" bound state S1 is exclusively occupied. Initiation of the catch bond process can thus be explained as a successive transition from the inactive S1 to the "activated" state S2, which becomes energetically favoured under the impact of high tensile forces.[364] In recent years, the dynamics of the FimH catch bond mechanism have been extensively studied, for instance, in flow-chamber experiments, by MD simulations, as well as constant and dynamic atomic force microscopy (AFM).[365,366] The findings were further supported with plenty of kinetic and thermodynamic data, usually obtained from isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and fluorescence polarisation (FP) assays.[337] MD simulations suggested a more linear alignment of the pilin domains compared to the relaxed pilus. Furthermore, a separation of the pilin domains throughout the fimbrial tip was predicted, particularly between the FimH lectin domain (LD) and pilin domain (PD), as well as between the adaptor proteins FimG and FimF.[365]

Extensive AFM studies revealed stretching of the pilus by a maximum of 9 – 10 µm at constant forces exceeding 60 pN, mainly due to reversible uncoiling of the helical FimA structure that forms the pilus rod.[367,368] Fimbrial extension was observed from a minimal threshold at 25 pN and from there on, across the entire physiological shear flow range up to approximately 90 pN. The individual FimH-ligand bond lifetime was proven to be extraordinarily long in this range (~ 17 min – 2.8 h) resulting in highly efficient bacterial adhesion.[369] Rupture of the FimH-ligand bond was observed at considerably higher shear forces between 150 – 190 pN, depending on the pulling velocity applied.[370]

A comprehensive insight into both the catch bond mechanism and the binding preferences for distinct FimH antagonists was provided by the multitude of structural models obtained for FimH or the lectin in complex with other pilin subunits. Currently, 62 X-ray crystal structures from *E coli* K12, F18 and J96 are available in the Protein Data Bank (PDB).[371] The first FimH crystal structure from an UPEC isolate was solved at 2.80 Å resolution by Knight and colleagues in 1999 (PDB: 1QUN). Three years later, Hultgren et al. published the structures of a non-functional Q133N variant (PDB 1KIU, 3.00 Å) and wild type (WT) FimH (PDB: 1KLF, 2.79 Å) from *E. coli* J96.[315] All these structures contained full-length (FL) FimH, comprising both the LD and PD, which were stabilised in complex with FimC during co-crystallisation. The G1 strand of FimC was therefore found to be inserted into the FimH-PD to complement its incomplete Ig-like fold. In contrast, the lectin domain exhibits a complete Ig-like fold (also referred to as a β-barrel jelly-roll fold) bearing a well-defined, polar binding cavity. The fold supports an intrachain disulfide bridge, which is formed between the only two cysteine residues in the FimH-LD sequence, Cys3 and Cys44. In the 1KLF structure, the lectin was complexed with *α*-D-mannose in the binding pocket, whereas the 1QUN structure fortuitously bound cyclohexylbutanoyl-*N*-hydroxyethyl-D-glucamide (C-HEGA), which had been added to the crystallisation solution. The *α*-D-mannose ligand in 1KLF (as in many other structures with *α*-D-mannosides) was stabilised by an extensive hydrogen bonding network involving the hydroxyl functions (except for the anomeric position), the amino acids F1, N46, D47, D54, Q133, N135, D140, and a single conserved water molecule. Among these, the residues F1, N46, D47, D54, and Q133 were proven to be essential for lectin function by mutagenesis studies, whereas alteration of D140 particularly results in reduced binding of mono-mannosides, while recognition of trimannoside glycoclusters is less affected.[372] Additionally, N135 and F142 are occasionally reported to be critical for ligand binding.[310] The amino acid residues in the periphery of the binding pocket are remarkably hydrophobic (I13, Y48, I52, Y137, F142). This feature was termed “hydrophobic ridge” and is assumed to reinforce the polar interactions of a mannopyranoside ligand with the binding cavity of FimH.[318]

As FimH-LD exhibits a complete jelly-roll fold, a correspondingly truncated FimH variant should be expectably stable. In 2004, De Greve and colleagues published a crystal structure of the lectin domain (aa 1 – 158, PDB: 1UWF) from *E. coli* J96 at 1.69 Å resolution.[373] Surprisingly, the lectin was crystallised with *n*-butyl *α*-D-mannoside within the binding pocket, which probably originated from the Lysogeny Broth (LB) medium used for bacterial cultivation. The *n*-butyl chain entered distinct hydrophobic interactions with Tyr48 and Tyr137, which adopt a parallel orientation with a small distance of only ~ 5.6 Å between the phenyl rings. Tyr48 was oriented towards Thr51 and the *n*-butyl aglycone (Figure 1.3-4a). This contrasts strongly with the 1KLF crystal structure, in which the phenyl ring of Tyr48 was rotated away from both the binding pocket and Tyr137 by ~ 3.6 Å and instead pointed in the direction of Asp47. On the other hand, Tyr137 remained largely unaffected. As the two tyrosine residues were positioned on either side of the binding pocket, Tyr48 and Tyr137 were later referred to as the “tyrosine gate”. 1KLW was accordingly assigned to the “open-gate” conformation, whereas a configuration as in 1UWF was termed “closed-gate” conformation.

The nearby positioned Ile52 is occasionally also counted as part of the tyrosine gate as it also frequently enters interactions with various aglycones.[310,374] The biological function of the structurally flexible tyrosine gate has not yet been fully elucidated. However, it seems likely that the increased rotatability of Tyr48 may reduce steric hindrance effects, thus enabling optimal adaptation to varying *N*-glycan structures.[342] This was at least supported by a third, tyrosine gate structure, in which Tyr48 adopts a "half-open" conformation (Figure 1.3-4b). The latter was identified in FimH-LD chain D, which was crystallised in complex with a rigid, *O*-propylyl-bridged biphenyl mannoside ligand (PDB: 4AV5, 1.40 Å).[375] On the basis of nuclear magnetic resonance (NMR) and chemical shift perturbation (CSP) experiments, Ernst and colleagues were able to demonstrate that the three identified Tyr48 rotamers are not merely artefacts, resulting, for instance, from crystallisation. This was supported by a transition from the open tyrosine gate to the closed-gate conformation, observed in *apo*-FimH-LD and upon ligand binding, respectively.[374] The open-gate → closed-gate transition only occurred when ligands with aliphatic or aromatic aglycones were employed.

Extensive π-π stacking with Tyr48/Tyr137 and Van der Waals interactions with the hydrophobic ridge are considered to be largely responsible for the exceptionally good binding properties of artificial mannosides that feature aryl and alkyl moieties as aglycones. Accordingly, mutation of Tyr48, and particularly of Tyr137, results in considerably reduced binding affinities to such ligands.[376] The development and optimisation of appropriate, high-affinity antagonists for FimH has been consistently pursued for decades as potential treatment against UTI.[310,334,377] An *ortho*-substitution of biphenyl mannosides relative to the monosaccharide moiety has been shown to further improve the affinity for FimH and, in the same manner, the inhibition effect. This was also thoroughly investigated by Ernst and co-workers. Using a competitive binding assay and a cell-based flow cytometry setup, the authors demonstrated a significantly enhanced inhibitory potential of biphenyl mannosides displaying *ortho*-substitution with preferably small, spherical substituents (e.g. chlorine, fluorine, methyl, methoxy, and trifluoromethyl groups) compared to the unsubstituted aglycones.[378]. This was explained by a decrease in desolvation enthalpy as well as with the improved Van der Waals-contacts between the substituent and a free cavity next to the CRD, which becomes exposed between Tyr137, Asn138, and Asp140 (Figure 1.3-5). Substitution with EWGs on the second phenyl ring, such as perfluorination or *para*-cyano groups, results in a complementary charge distribution in relation to the tyrosine gate side chains and thus optimised π-π stacking and K_D values.[379,380]

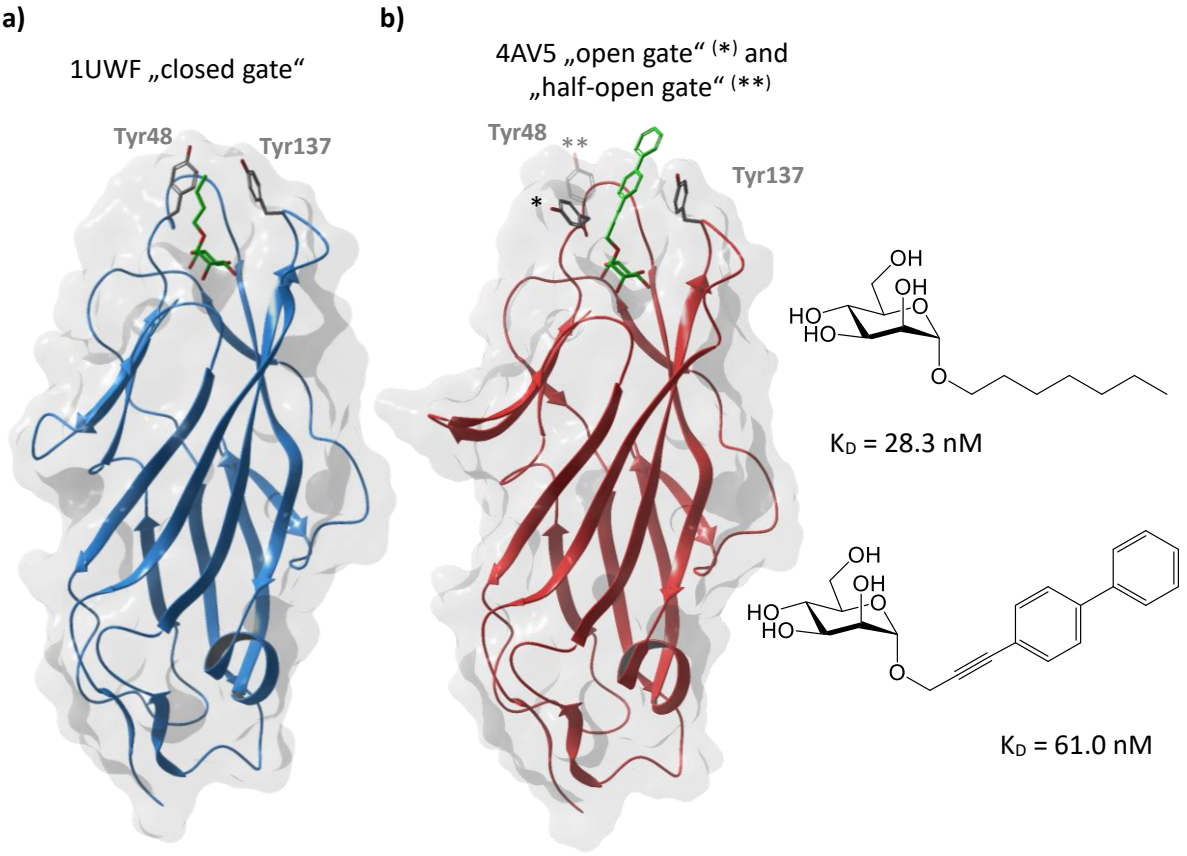


Figure 1.3-4: X-Ray crystal structures of FimH-LD illustrating the different rotamers of Tyr48.
a) 1UWF was solved at 1.69 Å resolution and with *n*-butyl α-D-mannoside within the binding pocket. The tyrosine gate is in the closed conformation. **b)** 4AV5 was solved at 1.4 Å and a bound propynyl biphenyl α-D-mannoside ligand. The tyrosine gate adopts the open-gate conformation in chain A, B, and C, whereas chain D reveals the half-open conformation. Visualised using the Maestro suite (Schrödinger Inc.). K_D values of FimH-LD determined in a competitive FP assay and by ITC, as published by Ernst et al. and Bouckaert et al.[375,379]

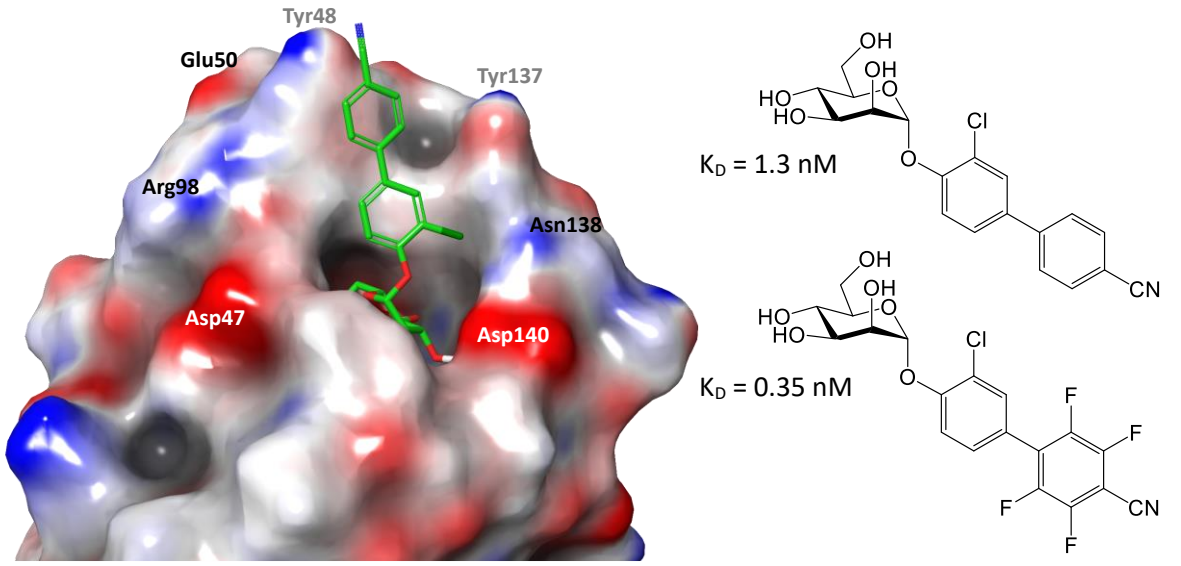


Figure 1.3-5: Binding mode of a *ortho*-chloro-substituted biphenyl mannoside bound in the FimH-CRD.
The crystal structure discloses a shallow cavity at the edge of the CRD formed between Tyr137, Asn138 and Asp140, which can be occupied by substituents in *ortho*-position relative to the mannopyranoside moiety. Surface depiction: Electrostatic potential ranging from -0.3 e/A² (red) to +0.3 e/A² (blue). X-ray crystal structure of recombinant FimH-LD in complex with 3'-chloro-4'-(α-D-mannopyranosyloxy)biphenyl-4-carbonitrile (PDB: 4CST, 1.10 Å).[379] Visualised using the Maestro suite (Schrödinger Inc.). K_D values of FimH-LD determined in a competitive FP assay by Ernst et al.[379,380]

Apart from the already mentioned biphenyl and butyl mannosides, numerous suitable FimH candidates with K_D values down to the low double-digit and single-digit nanomolar range have been extensively studied. These include, for instance, a variety of heptyl-, phenyl-, umbelliferyl-, as well as correspondingly functionalised *C*- and *S*-linked mannopyranosides.[381,382] Two promising *C*-mannoside derivatives (GSK-3882347 and NCT-03709628, better known as Sibofimloc) have been evaluated in clinical trials with good prospects for approval as orally administrable FimH antagonists for the treatment of UTIs and CD, respectively.[383–385] Highly polyvalent, mannosylated glycoclusters and glycodendrimers attain affinities in the high sub-nanomolar range.[377,386,387] However, the poor oral bioavailability for such antagonists in the urinary tract remains an obstacle to the antiadhesive treatment of UTI mainly due to poor intestinal absorption or renal excretion. The conformation of the mannopyranoside moiety is highly conserved throughout the various crystal structures. Nevertheless, Peczuh and colleagues recently introduced septanoses as FimH antagonists with great inhibitory potential.[388] The authors have documented this novel type of FimH inhibitors with a series of heptyl-, phenyl-, and biphenyl-functionalised septanoses, which revealed clearly superior affinities to FimH-LD compared to their respective mannopyranoside analogues. Similar to *C*- and *S*-mannosides, septanoses cannot be recognised or cleaved by mannose-specific glycosidases, representing a decisive advantage for their use as therapeutic antiadhesives.[389]

The isolated FimH-LD features a significantly enhanced affinity towards mannoside ligands compared to full-length FimH.[336] Based on site-directed mutagenesis of the inter-domain region, an allosteric auto-inhibition mechanism was proposed, which occurs as a result of the domain interactions between the FimH-PD and LD.[390] The first structural evidence was provided by Thomas and co-workers who presented a crystal structure of the entire fimbrial tip from *E. coli* K12, with the proximal adapter protein in DSC with the chaperone, FimC (PDB: 3JWN, 2.69 Å).[391] The crystallisation approach was later facilitated by Sauer et al. who stabilised FL-FimH variants from *E. coli* K12 and F18 with the donor strand peptide of FimG (DsG) rather than with the FimH-PD and/or FimC.[337] These structures (PDB: 4XOA, 2.54 Å and 4XOD, 1.14 Å) revealed striking differences in the FimH-LD conformation compared to all previous structural models whereas the other pilin subunits (including the FimH-PD) feature only minor conformational alterations: The fimbriae-associated FimH-LD adopts a broadened, "compressed" version of the characteristic jelly-roll fold, associated with a more pronounced twisting of the β-sandwich (Figure 1.3-6a,b).[392] Moreover, the three CRD loops and the fourth, so-called "clamp loop", adopt a loose conformation resulting in an "open" binding site. The distal and proximal end-to-end distance is about 11 Å shorter than the "elongated" conformation with the "narrow" or "closed" binding pocket, adopted by the isolated FimH-LD. The stretching of the inter-domain linker in the elongated conformation results in a "hinge motion" and an increased distance between the LD and PD. As the respective fimbriae-associated FimH exhibits a reduced affinity by about two orders of magnitude compared to the isolated FimH-LD, the two conformations are generally referred to as "low-affinity" (LA) and "high-affinity" (HA) states, respectively.[391] As the LA → HA transition is closely associated with catch bond activation, these main conformations of FimH can be assigned to the two potential wells in the free energy landscape in Figure 1.3-3c (Low-affinity: S1; High-affinity: S2).[393]

In the literature, the terms "A_{free}" and "S_{bound}" are occasionally used to describe FimH in the low-affinity state with narrow domain interactions (A, "associated") and the domain-separated, high-affinity state (S, "separated"). The assignments (A and S) are regarded independently from the "open" or "closed" binding site conformations, which are either adopted by the unliganted (i.e. ligand-free) *apo*-lectin (free) or in the presence of a mannoside ligand (bound).[337,342,394] Binding of heptyl *α*-D-mannopyranoside (PDB: 4XOB, 3 Å and 4XOE, 2.4 Å) was associated with a narrowing of the three CRD loops, as well as an upward motion of the clamp loop towards the binding pocket.[337] Apart from that, the LD remained in its compressed, LA conformation and the corresponding structure can thus be assigned to the "A_{bound}" state according to the nomenclature implemented by Sauer and colleagues. Indications for a (short-lived) HA conformation with an open binding site (S_{free} state) have so far only been delivered by MD simulations, but were not experimentally confirmed.[394] In addition to the empirical conformations (A_{free}, A_{bound}, S_{bound}), two strongly bent FL-FimH crystal structures have been solved, which cannot be categorised within this system (PDB: 5JR4, 2.60 Å and 4XOB chain G).[371] Altogether, a total of five discrete FimH conformations have beenexperimentally verified, which significantly differ in their root mean square deviations (RMSDs).[371]

During the transition from the LA to the HA state and vice versa, certain areas and amino acid residues of the lectin are particularly affected by the conformational rearrangement (Figure 1.3-6). These "toggle switches" include, for instance, the pocket zipper (aa 1 – 11) and the clamp loop (aa 8 – 16).[395,396] Apart from the already mentioned upwards and downwards tilt of the latter, the entire segment between the residue positions 3 – 11 forms a two-strand β-hairpin, which is either in a parallel ("zipped") or tilted ("unzipped") conformation in the HA and LA states, respectively (Figure 1.3-6a and b). The β-bulge region is also noteworthy, as it forms part of the front β-sheet in the HA conformation but folds into a loop in the LA state, possibly to reduce tension in the more twisted jelly-roll fold.[395] Similarly, the subsequent α-switch (aa 64 – 71) adopts a clearly recognisable α-helix in the HA state whereas a single-turn 3_{10}-helix is formed in the LA conformation. Further conformational shifts are observed in the three segments of the inter-domain region, which encompasses the swing loop (aa 25 – 34), the insertion loop (aa 111 – 119), and the linker loop that connects FimH-LD to the PD.

The exact mechanism of allosteric coupling of LD-PD interaction and FimH activation, i.e. how the allosteric signal is transferred from the inter-domain loops to the remote binding site, has not been completely elucidated. Based on thorough MD simulations, Thomas and co-workers primarily attributed this to the pronounced flexibility and twisting motion of the large β-sheet.[392] Nevertheless, the relevance of some toggle switches for the allosteric pathway was thoroughly investigated by kinetic measurements, enzyme-linked immunosorbent assays (ELISA), MD simulations, NMR spectroscopy as well as X-ray crystallography.[395–397] Interesting findings were for instance obtained with cysteine double mutations, such as in the V27C/L34C FimH variant.[397] Although expressed as isolated LD, the kinetic parameters of this double variant mostly resembled those of the fimbriae associated lectin (Table 1.3-1). The cysteine residues were designed to form a disulfide bond, which was successfully confirmed and which results in permanent stabilisation of the narrow swing loop conformation observed in the LA state, thereby impairing the transition to the HA state.[395,397] Introduction of a proline residue in the β-bulge (R60P) was performed to impede the formation of secondary structures in this segment. As illustrated in Figure 1.3-6c, the crystal structure of the *apo*-R60P variant (PDB: 5MCA, 1.6 Å) reveals a strong agreement with the LA conformation as found in fimbriae-associated FimH (RMSD = 0.6 Å compared to PDB: 3JWN).[397] Only the clamp loop adopts a slightly upward tilted conformation, similar to – but not as close to the binding pocket as in the HA conformation.

MD simulations of the R60P variant further indicated backbone alterations and a pronounced fluctuation in the inter-domain region.[398] However, both the R60P variant and the V27C/L34C double variant seem to be able to fully adopt the HA conformation upon ligand binding, as indicated by circular dichroism analysis as well as NMR spectroscopy. As the transition from the LA to the HA conformation occurred in the absence of the FimH-PD, these variants can be regarded as "minimalistic allosteric systems".[397] A mutation in the α-switch segment (V67K) likewise resulted in a LA-stabilised variant, whereas variants with amino acid exchange in the pocket zipper (A10P) and the insertion loop (A119L) preferred the HA conformation.[397]

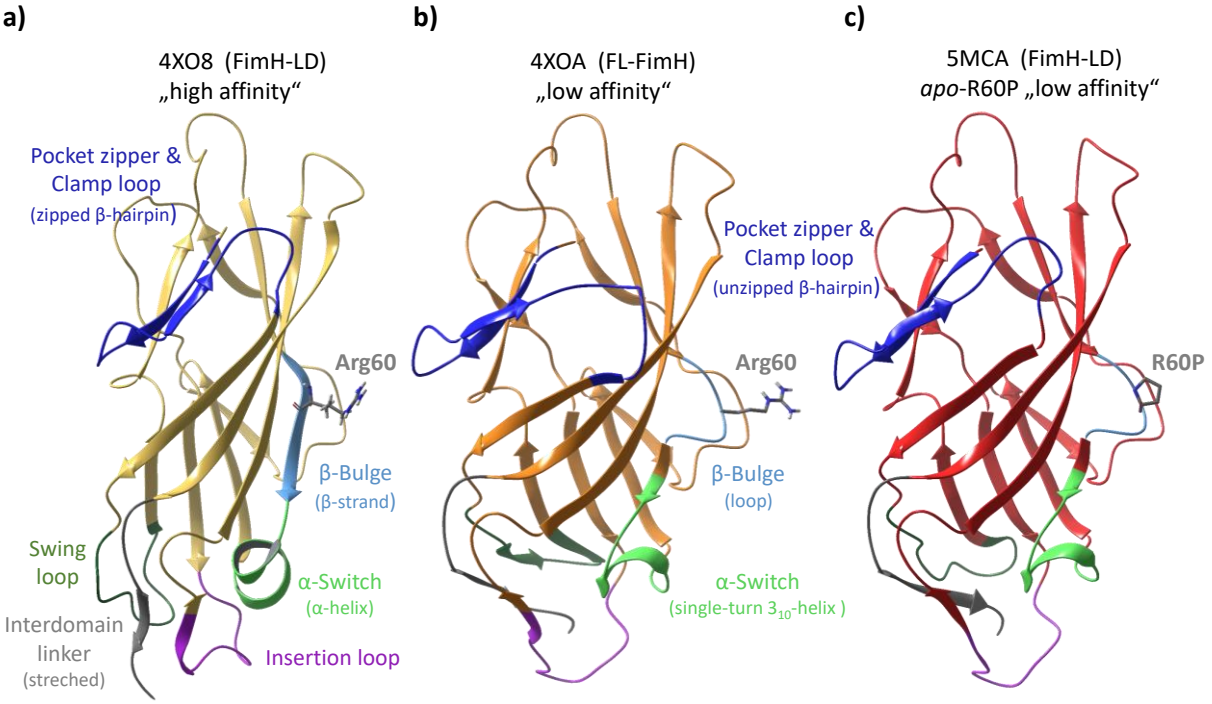


Figure 1.3-6: Comparison of high-affinity (HA) and low-affinity (LA) conformations observed in crystal structures of different FimH variants.

PDB entry 4XO8 represents the FimH-LD from *E. coli* K12 in complex with heptyl *α*-D-mannopyranoside, which was solved at a resolution of 1.70 Å.[337] 4XOA was solved at 2.54 Å resolution and represents the FL-FimH · DsG complex from *E. coli* K12, which adopts the LA conformation.[337] 5MCA (1.6 Å) displays the *apo*-FimH-LD R60P variant, which is stabilised in the LA conformation but is still able to undergo conformational transition to the HA state upon ligand binding.[397] Visualised using the Maestro suite (Schrödinger Inc.).

Table 1.3-1: K_D values and rate constants determined for the isolated FimH lectin domain, the full-length lectin, the R60P variant and the V27C/L34C double variant.

FimH variant	K_D [nM][a]	K_D [nM][b,c]	k_{on} [M^{-1} s^{-1}][b]	k_{off} [s^{-1}][b]
WT FimH-LD	19	28.8	$4.8 \cdot 10^3$	$1.4 \cdot 10^{-4}$
WT FL-FimH[d]	2300	6373	$33 \cdot 10^3$	$5116 \cdot 10^{-4}$
FimH-LD R60P	322	1656	$2.4 \cdot 10^3$	$41 \cdot 10^{-4}$
FimH-LD V27C/L34C	661	4361	$92 \cdot 10^3$	$3998 \cdot 10^{-4}$

a) K_D determined in a FP assay; **b)** kinetic data obtained by SPR; **c)** kinetic K_D, calculated using the following equation: $K_D = k_{off}/k_{on}$; **d)** FL-FimH · DsG-complex.[397]

Among other things, the available LA-FimH variants were intended for the development of monoclonal anti-FimH antibodies. Sime of these antibodies exhibit intriguing allosteric properties, which concurrently enable convenient control and monitoring of lectin function.[399] These include, for instance, the HA-specific and HA-stabilising antibody mAb21 obtained by immunisation of mice with WT FimH-LD, which is permanently trapped in the HA conformation.[400] The largely inactivated *apo*-V27C/L34C double variant, on the other hand, was exploited to obtain the orthosteric (i.e. binding directly at the CRD) antibody mAb475, which acts as a competitive inhibitor due to substantial epitope overlap with the binding site.[401] Another example, the dynasteric (i.e. modulating conformational dynamics without binding to the CRD) antibody mAb824, stabilises the respective pre-existing FimH conformation and hampers the allosteric transition.[396] With use of these antibodies, HA and LA states of FimH can thus be selectively detected, induced, or stabilised in a straightforward manner.[395,396,400] LA variants are also of interest for the testing and optimisation of FimH antagonists, which have previously been evaluated primarily using the permanently activated FimH-LD. The latter may not represent the most suitable therapeutic target for *in vivo*-use, since it is still debated in which state the lectin actually binds its natural receptor (*vide infra*).[397,402]

The considerably higher affinity (expressed by a lower K_D) of the HA-stabilised FimH-LD compared to the fimbriae-associated lectin is mainly due to a significant deceleration of ligand dissociation. This in turn is manifested by a disproportionately large reduction in the k_{off} rate by three orders of magnitude (Table 1.3-1).[397] The same applies to the LA-stabilised variants R60P and V27C/L34C, which feature K_D and k_{off} values between those of the isolated FimH-LD and the full-length constructs. Only the k_{on} rate of the double variant is higher than that of FL-FimH, indicating a slight deceleration of the ligand binding velocity. Interestingly, the difference in the conformational equilibrium between the high- and low-affinity conformations is significantly less pronounced with septanose ligands, especially for those with flexible aglycones.[388]

The obtained structural and kinetic data have contributed substantially to the understanding of the shear-induced catch bond mechanism of FimH and have contributed to the concept of dynamic allosteric regulation: Under static conditions, the FimH-PD is in close contact with the LD, on which the former acts as a negative allosteric regulator. It is generally accepted that application of tensile forces on adhering bacteria leads to fimbriae-stretching and thus coaxial domain separation. In this way, the auto-inhibitory effect of the FimH-PD on the lectin domain is overcome, resulting in conformational rearrangement, which is associated with an affinity increase. The latter has been clearly demonstrated by site-directed mutagenesis in the inter-domain region of the FimH-PD.[390] However, the available crystal structures may not optimally reflect the conformational dynamics involved in the FimH binding and catch bond activation under physiological conditions. Certain unanswered questions remain, for instance, why the largely identical binding site structures in the A_{bound} and S_{bound} states are associated with such drastically different ligand affinities. Another major problem is the complete lack of experimentally obtained structural data regarding the lectin under the influence of mechanical tensile forces. Instead, only MD simulations can be employed to simulate the conformational dynamics of FimH under shear.

In recent years, several suggestions have been made and frequently discussed concerning how to unify the partly contradictory kinetic, thermodynamic and structural data in a comprehensible model that most accurately describes the conformational dynamics *in vivo*.[403] Sauer et al. supported a “thermo-dynamic cycle model” (Figure 1.3-7a), which considers the respective positions of the conformational equilibrium of the identified (LA) A-states and (HA) S-states in the presence/absence of a ligand.[342]

The authors supported the assumption that the experimentally unverified S_{free} state coexists in an equilibrium with the other states (A_{free}, A_{bound}, S_{bound}). They furthermore anticipate that FimH preferentially binds its natural receptor in the A_{free}-state and then allows for subsequent domain separation. The thermodynamic cycle model features large similarities with the "kinetic selection model" previously introduced by Thomas et al. (Figure 1.3-7b).[403] The latter further emphasises a kinetically hindered and therefore an extremely slow binding and dissociation process of domain-separated FimH under static conditions. Accordingly, ligand unbinding rather proceeds via conformational transition to the A_{bound} state. Mechanical force however, favours a shift to a non-equilibrium cycle, which allows ligand dissociation from the S_{bound} state instead of transition to the A_{bound} state.[403] The alternative "conformational selection model", which is endorsed by Hultgren and colleagues, even takes it one step further and proposes that receptor binding occurs primarily in the HA state (Figure 1.3-7c) *in vivo*.[404] As the authors investigated different natural FimH variants, the conformational selection was particularly proposed for those variants that favour the HA conformation.[404] On the other hand, the authors supported an "induced fit" model, particularly for FimH variants that prefer the LA state (Figure 1.3-7d). The induced fit model assumes that the experimentally unverified S_{free} state does not exist and that domain separation can exclusively be induced upon ligand binding.[403,404]

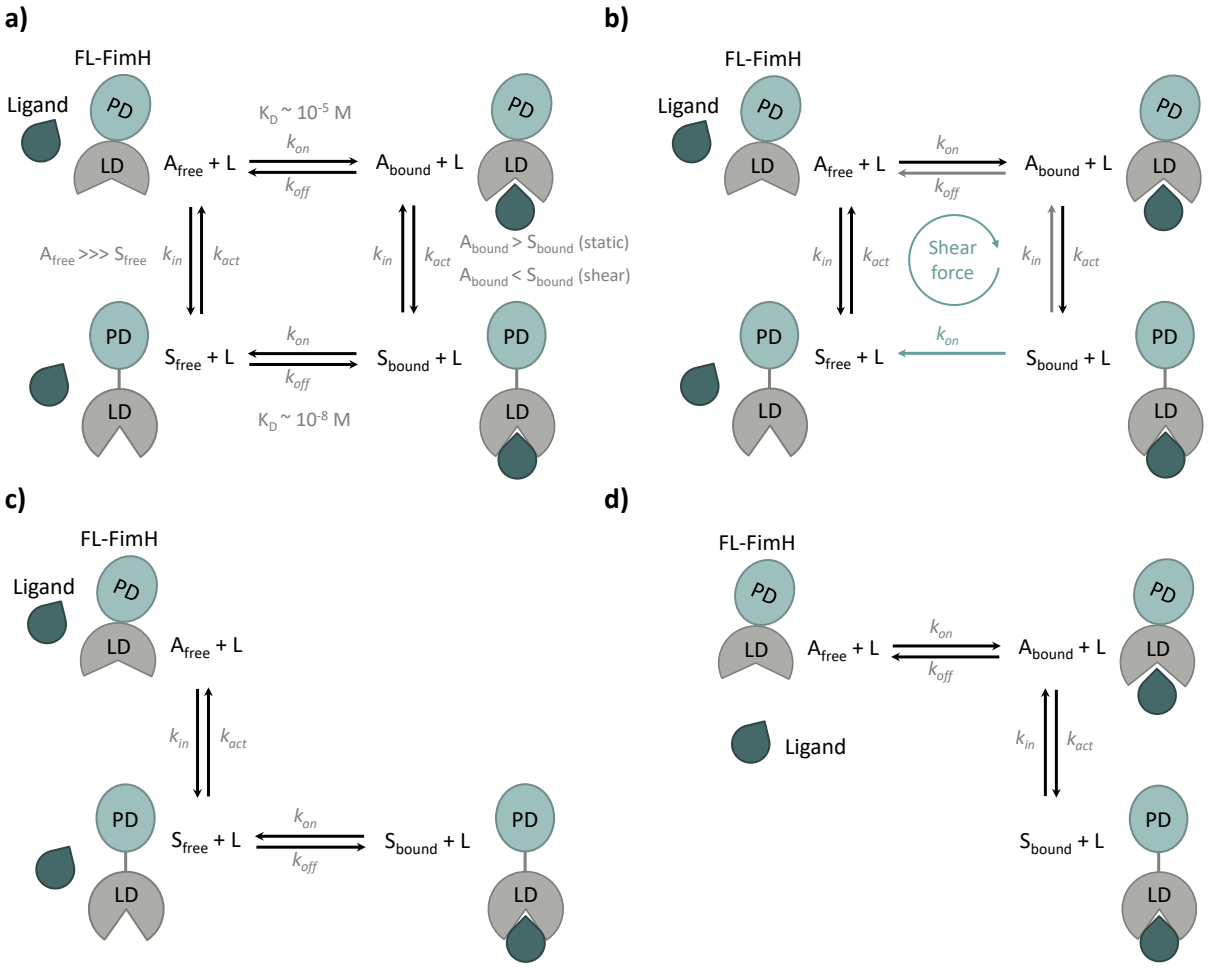


Figure 1.3-7: Models to describe the conformational dynamics of ligand binding and FimH catch bond activation *in vivo*.

a) Thermodynamic cycle;[342] **b)** Kinetic selection; **c)** Conformational selection, and **d)** Induced fit model.[403]

Even though it is usually not explicitly emphasised, the assumption of a thermodynamic equilibrium between the A_{bound} and S_{bound} state does not entirely exclude the possibility of a domain separation in the absence of shear. Indications for a shear-independent FimH activation were, for instance, provided by evaluation of the naturally occurring A27V variant. Thomas et al. demonstrated in ELISA experiments with the HA-specific antibody mAb21 and in presence/absence of methyl *α*-D-mannoside that the FL-FimH A27V variant becomes approximately 60% pre-activated solely upon ligand binding under static conditions.[393] Amino acid position 27 is located in the inter-domain region and, thus, A27V mutation obviously results in a rather shear-independent phenotype with facilitated transition to the HA state. The A27V mutation is rarely found in faecal and commensal strains, but it is predominant in many uropathogenic species, including the UPEC strain J96.[405] As already mentioned at the beginning of this chapter, the A27V mutation is likewise present in the common laboratory strain *E. coli* K12. Overall, A27V can be regarded as a pathoadaptive mutation that facilitates bladder colonisation in both static and high-shear conditions, which only occur temporarily during urination after all.[393]

An induced fit mechanism was recently also favoured by Stirnemann and colleagues who performed extensive MD simulations in combination with enhanced-sampling techniques in a state-of-the-art force field to explore the conformational flexibility of the lectin.[394] The authors substantiated the induced fit model by a corresponding population shift in favour of the closed binding site conformation. This was observed for both S- and A states, respectively, and occurred in consequence to a calculated stabilisation effect of approximately $-10\ kJ \cdot mol^{-1}$ upon ligand binding. On the other hand, the authors found indications of the existence of a S_{free} state, which actually contradicts the induced fit model. However, simulation of the FimH conformation under application of tensile force in the physiological, catch bond inducing range (50 pN) revealed a rather small but completely unexpected effect, which was never considered before: Besides the relative stabilisation of domain separation ($-15\ kJ \cdot mol^{-1}$) and binding site closure ($-10\ kJ \cdot mol^{-1}$) an additional force-induced stabilisation of the bound state was calculated ($4 - 7\ kJ \cdot mol^{-1}$). Remarkably, the stabilisation effect likewise proved to be significantly pronounced for the A_{bound} state under shear conditions. The authors termed this a “local” catch bond effect and argued that it may temporarily prevent the lectin in the A_{bound} state from premature unbinding until the comparably slow domain separation and allosteric transition is completed. Once the activated S_{bound} state has been adopted, the “local“ catch bond may further reinforce binding to the receptor and thus contribute to the “overall“ catch bond effect.[394]

2. Objectives

The diverse examples of PRPs impressively show how nature exploits light to directly or indirectly control biological function in a sophisticated manner. Light also represents an ideal stimulus for medical and bioscientific research as it can be applied with extraordinarily high spatial and temporal resolution that coincides with the time scale of cellular processes. Moreover, the intensity and wavelength of light can be accurately modulated. Decades of highly fruitful research in the aspiring field of optogenetics have focused on the photostimulation of various cell types, achieved by the introduction of photosensor proteins at the genetic level.[406] Only a few years later, the optogenetic approach was extended to the use of free-diffusible, photochromic ligands (PCLs) or alternatively to chemical modification of biological targets with artificial photoswitchable tethered ligands (PTLs). Both approaches belong to the interdisciplinary field of photopharmacology, which impressively combines photochemistry and pharmacology with the aim to exert photocontrol over the respective biological targets.[208] Inspired by the pioneering work of Trauner and colleagues who laid the foundation for the growing success of photopharmacological research, the Lindhorst group pursues similar strategies to control adhesion of type 1 fimbriae-expressing *E. coli* bacteria with light. In our previous publications, photoswitchable azobenzene mannoside derivatives were predominantly used as free-diffusing inhibitors in solution,[407–411] or immobilised on surfaces,[412–415] in order to modulate bacterial adhesion. In the latter approach, the steric accessibility of the respective ligands to the lectin can be photochemically steered. The former approach basically coincides with the PCL approach according to Trauner's classification of photopharmacological interaction modes with the biological target.[215]

A possible alternative strategy to achieve photocontrol over lectin binding and bacterial adhesion events is the covalent modification of FimH with photosensitive moieties or ligands, according to the PTL method. Femke Beiroth was the first member of the Lindhorst group who attempted chemical labelling of the isolated FimH-LD with photoswitchable molecules as part of her dissertation.[416] In the so-called "Gate-Keeper" project, azobenzene derivatives with or without terminal glycoside moieties were intended to be positioned near the binding pocket to enable photochemical control of lectin binding and unbinding. Affinity-labelling techniques are particularly suitable to increase the probability of bioconjugation events in the vicinity of the respective binding sites.[417,418] Among other things, Femke synthesised epoxide-functionalised AB *α*-D-mannoside derivatives, which are expected to react rather unselectively with adjacent nucleophilic amino acid residues of the CRD after binding of the mannopyranoside moiety to the lectin's binding pocket. However, the suitability of these epoxide ligands was only evaluated in molecular docking simulations, but FimH modification was not attempted *in vitro*. Instead, the latter was achieved by affinity-guided protein labelling, catalysed with 4-dimethylaminopyridine (DMAP), a technique that was introduced by Hamachi and co-workers.[419] This methodology relies on two steps: (i) the use of the common acyl-transfer catalyst DMAP, tethered to an appropriate affinity ligand (e.g. *α*-D-mannopyranoside) for the pre-positioning of the catalyst near the binding site, and (ii) addition of a suitable acyl-donor (e.g. aryl thioesters) for the subsequent DMAP-mediated esterification to a proximate nucleophilic residue. The DMAP-functionalised mannopyranoside ligand can subsequently be removed, leaving behind only the labelled protein with an unoccupied binding pocket (Figure 2.1). However, despite the affinity-controlled pre-positioning of the DMAP-catalyst, the ensuing acyl-transfer reaction with nucleophilic residues proved not to be sufficiently specific and efficient.

In addition to the desired modifications near the binding site (e.g. the tyrosine gate residues Tyr48 and Tyr137), other amino acids were unintentionally modified (Ser62, Lys76, Lys101, and Lys121). These bioconjugation sites were located distant from the binding pocket, i.e. much too far to allow efficient photocontrol of ligand binding with the azobenzene derivatives employed. The labelling yield was partially affected by the poor water solubility of the thioester derivatives. Organic solvents were added to dissolve the acyl donor-derivatives, but this may severely compromise the integrity and folding of the protein. Franziska Reise later continued the Gate-Keeper project with the aim to design acyl donors with adapted properties to improve the modification result. For this purpose, Franziska designed a variety of AB thioesters and correspondingly functionalised glycoside derivatives with different steric demands, improved affinity to the tyrosine gate residues or with increased solubility in water. The respective ligands were synthesised and extensively evaluated through *in silico*-docking and MD simulations.[420]

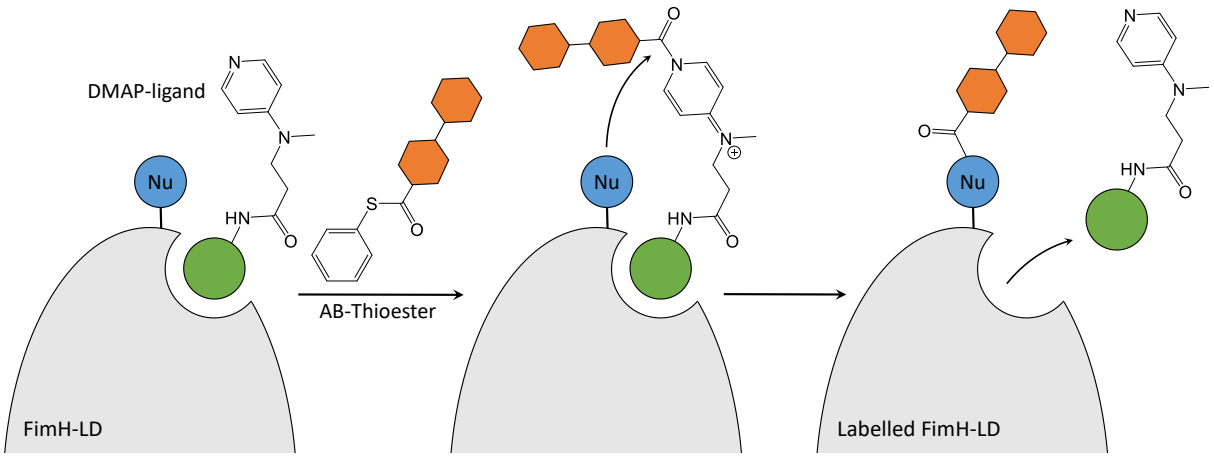


Figure 2.1: Application of the affinity-guided DMAP-labelling strategy introduced by Hamachi et al. to install azobenzene derivatives near the FimH-CRD in the context of the Gate-Keeper project.
The lectin is depicted in grey, bearing a nucleophilic residue near the binding site (blue); azobenzene units are shown as orange hexagons; the α-D-mannoside moiety is illustrated as a green circle.

The Gate-Keeper project largely pursues the same approach as the photoswitchable closely tethered ligand (PCTL) methodology, despite the different bioconjugation strategies heretofore employed.[208] In the respective publications, Trauner and colleagues chose cysteine modification as the preferred bioconjugation method, as it usually provides higher selectivity and yield compared to nonspecific targeting of proximate nucleophilic residues. The PCTL approach usually requires genetic engineering of the respective proteinaceous target, with the aim to introduce cysteine residues as bioconjugation sites at the desired positions. This work focuses on the continuation and experimental realisation of the Gate-Keeper project, but the approach is intended to follow the approach of Trauner et al. more closely (Figure 2.2a).[242] Both genetic engineering and chemical modification of the lectin are envisaged to enable photocontrol of FimH activity. Fundamentally, the following intermediate objectives must be attained to achieve this goal: (i) the design, organic synthesis, and evaluation of photochromic ligands suitable for covalent cysteine modification; (ii) engineering of appropriate FimH variants, by the introduction of cysteine residues in the vicinity of the binding site at the genetic level; (iii) expression and purification of the respective cysteine variants; and (iv) chemical modification with the synthesised photoswitchable ligands. (v) Once the desired bioconjugation is successfully achieved, appropriate detection methods must be established to assess the photochemical controllability of FimH activity.

The latter is intended to be monitored by suitable NMR methods such as saturation transfer difference (STD)-NMR. Alternatively, a lectin assay can be developed that ideally relies on fluorescence measurement as a comparably fast detection method, involving either the use of a fluorescently tagged lectin or fluorescently labelled α-D-mannoside ligands. Fluorescein isothiocyanate (FITC) can, for instance, be used as a fluorophore as it facilitates straightforward labelling due to its high reactivity with amines or sulfhydryl functions. By this means, the accessibility of the binding site can prospectively be determined either via binding of the labelled lectin to a mannosylated surface or via binding of a FITC-tagged ligand by immobilised FimH (Figure 2.2b). The functionality of the chemically unmodified FimH variants should also be evaluated in an appropriate lectin assay setup.

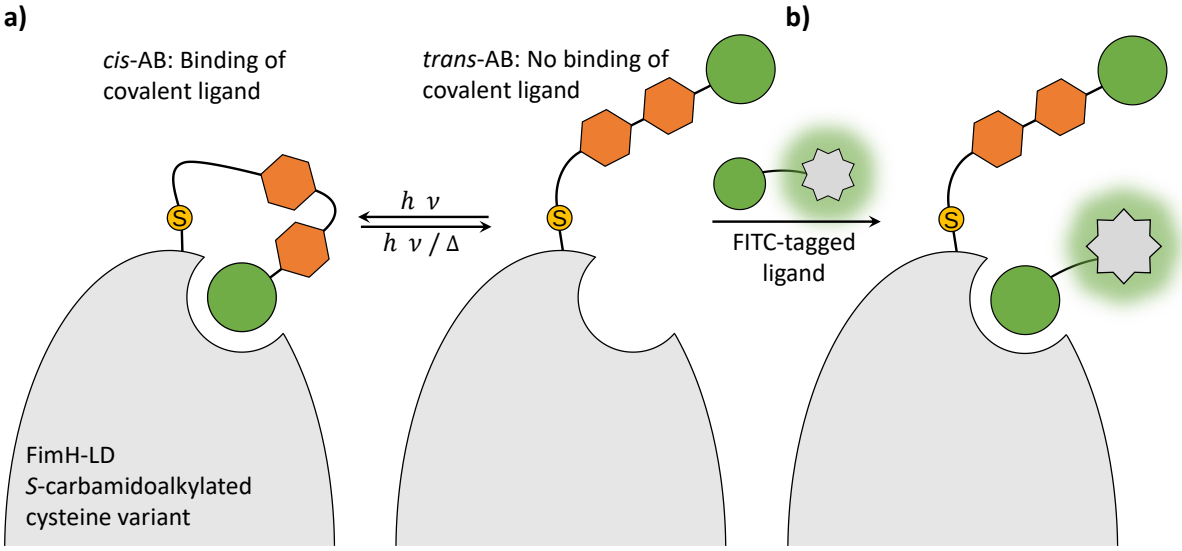


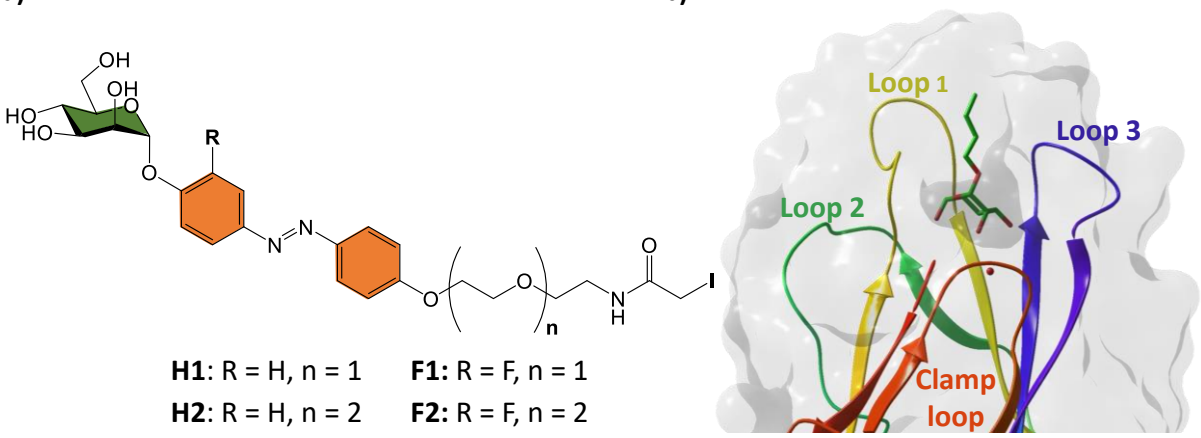
Figure 2.2: Continuation of the Gate-Keeper project: **a)** Photocontrol of FimH activity is intended to be achieved by following the PCTL approach. **b)** The accessibility to the FimH binding site may be monitored via binding or unbinding of a fluorescently tagged mannoside ligand.
The lectin is depicted in grey, bearing the introduced cysteine sulfur atom (yellow); azobenzene moieties are shown as orange hexagons in *cis*- (bent line) and *trans*-configuration (straight line); the α-D-mannoside moiety is illustrated as a green circle; the star indicates a suitable fluorophore (FITC) to enable monitoring of FimH activity.

The structure of a potentially suitable target molecule for FimH labelling is illustrated in Figure 2.3. A 4,4'-dihydroxy azobenzene moiety is chosen as reversible photoswitch. To increase the probability of a complete coverage of the binding pocket in either of the possible configurations of the AB moiety, the implementation of a terminal α-D-mannopyranoside unit into the photoswitchable tethered ligand is intended. At the opposite terminus of the molecule, an iodoacetic amide (IAA) function is attached to enable the desired *S*-carbamidoalkylation of a cysteine residue. The well-established reaction of IAA derivatives with thiols occurs with high chemoselectivity and results in bioconjugation motifs that are even smaller than those present in maleimide-cysteine adducts. In comparison to the latter, IAA-cysteine adducts are more stable in aqueous conditions, but the alkylation reaction proceeds more slowly and requires higher concentrations.[219] However, this constitutes a minor problem as the labelling of the recombinant lectin can easily be performed *in vitro*. The native disulfide bridge between Cys3 and Cys44 is located internally and is therefore presumably inaccessible for alkylating agents. A flexible oligoethylene glycol (OEG) linker is introduced between the AB and the IAA moieties, respectively, both to span the distance between the binding cavity and the respective alkylation sites and to increase the water solubility of the target molecules.

As indicated in Figure 2.3a, mono-fluorination of one phenyl ring is intended in *meta*-position to the diazene bond, or to phrase it differently, in *ortho*-position to the mannopyranosyloxy moiety. As mentioned earlier, Ernst and co-workers demonstrated a significant affinity increase towards the lectin owing to *ortho*-substitution of biphenyl mannosides.[378] A positive effect can thus also be expected for *ortho*-fluorination at the same relative position on AB mannoside derivatives. The fluorination is furthermore considered to facilitate monitoring of *cis* ↔ *trans* photoisomerisation after bioconjugation with the lectin by ^{19}F NMR spectroscopy. Identification of the respective *cis*- or *trans*-AB proton signals in ^{1}H NMR is presumably not applicable due to an expectedly pronounced spectral overlap with aromatic amino acid residues. Further benefits of fluorine detection by NMR are the large chemical shift dispersion, the high magnetogyric ratio, and consequently high sensitivity of the ^{19}F isotope as well as its occurrence as a single nuclide, which is advantageous as well.[421]

The exact number of required ethylene units within the linker moiety as well as the identification of suitable positions for the introduction of cysteine residues must be determined in molecular docking simulations *in silico*. Positioning within one of the three CRD loops, illustrated in Figure 2.3b, appears feasible. Particularly eligible seem the CRD loops 1 (Asn46 – Thr53) and 3 (Asn135 – Phe142), which are in direct vicinity to the binding cavity. As these loops are positioned slightly above the binding pocket, binding of the mannopyranoside moiety is expected to occur with the AB unit in *cis*-configuration, whereas the *trans*-configuration is purposed to preclude ligand binding.

a) **b)**



H1: R = H, n = 1 **F1:** R = F, n = 1
H2: R = H, n = 2 **F2:** R = F, n = 2

Figure 2.3: a) General structure of Gate-Keeper ligands for chemical labelling of FimH cysteine variants. **b)** Position and numbering of CRD loops 1 – 3 that are involved in forming the mannose binding site in FimH (PDB: 1UWF) and that are considered for the introduction of cysteine residues as alkylation sites.

The general capability of the synthesised target molecules to be recognised by the WT lectin is projected to be evaluated in competitive bacterial assays. However, the use of IAA-functionalised ligands in anti-adhesion assays with living *E. coli* bacteria may not be advisable, as the unintended alkylation of the fimbriae or other secreted proteins cannot be excluded with certainty. With an average occurrence of 1.1 %, the abundance of cysteine is relatively low in *E. coli*,[422] especially in secreted proteins.[423] Since cysteines usually become oxidised and enter disulfide bonds – especially upon secretion – an unintentional *S*-carbamidoalkylation in the absence of reducing agents in the employed buffer might seem unlikely.

However, modification of N- or C-terminal amines and carboxylic acids, as well as other protic side-chain functionalities, is still possible.[424] In order to prevent this, the synthesis of corresponding *N*-acetylated analogues is envisaged, which are presumably more suited for incubation in the presence of bacteria and are thus intended to be tested as photoswitchable FimH antagonists in solution.

The chemical modification with the presented target molecules will be conducted in this work as an initial proof-of-principle with two cysteine variants of the isolated FimH-LD. If successfully implemented, the concept may eventually be applied to fimbriae-associated FimH or to piliated *E. coli* in order to prospectively enable control of bacterial adhesion with light. For this consideration, the introduction of the corresponding genetic and chemical cysteine modification into the fimbriae-associated FimH would be necessary. In all previous work from the Lindhorst group, the truncated FimH-LD was kindly provided by the Knight group (Swedish University of Agricultural Sciences, Department of Molecular Biology, Uppsala). All attempts to express and isolate FimH-LD in our group were unsuccessful so far, which is why the expression and purification methods have to be fundamentally optimised. Usually, in the absence of the PD as an allosteric effector, the FimH-LD permanently adopts the HA conformation. As mentioned in Chapter 1.3.3, some literature-known FimH-LD variants are stabilised in the LA conformation as long as no ligand is bound and undergo a conformational rearrangement to the HA upon ligand binding.[397] The R60P variant is particularly interesting for the Gate-Keeper project in general as it provides more conformational variability and its engineering requires only a single amino acid exchange. By applying a chemical modification with photoswitchable ligands to the R60P-cysteine double variant, not only lectin binding and unbinding, but also the adoption of the respective FimH conformation could be governed (Figure 2.4). According to the conventional PCTL approach, the introduction of two mutations would be necessary to obtain R60P-cysteine double variants. As the expression protocol requires thorough optimisation, the expression of the R60P variant is projected first, before the attempt to obtain double variants is commenced.

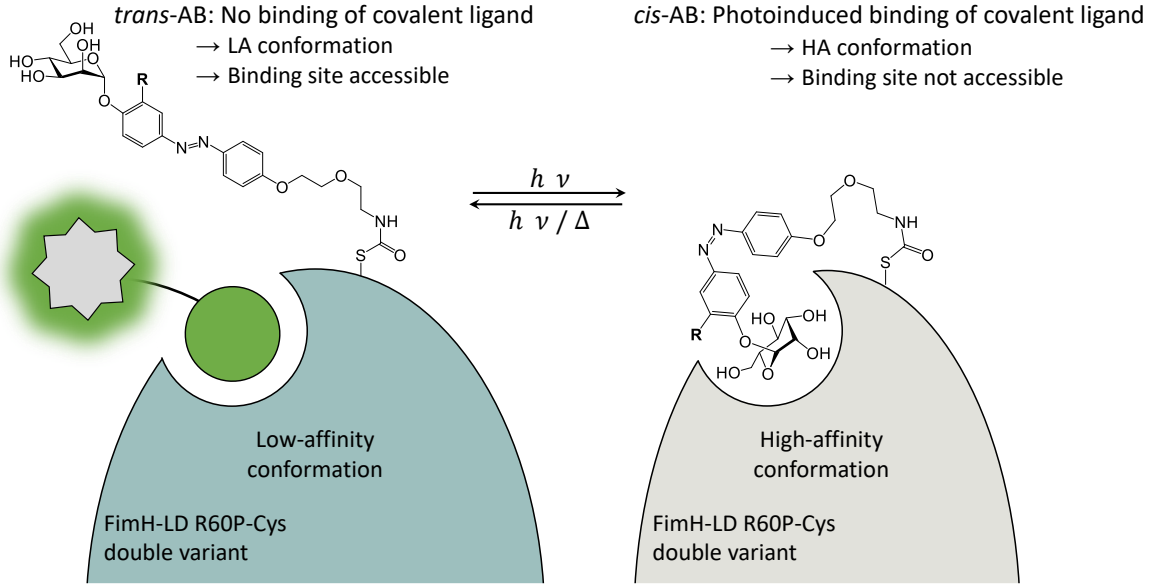


Figure 2.4: Concept for achieving photocontrol of the FimH-LD conformation following the PCL approach with R60P-cysteine double variants.

The R60P variant of FimH-LD is conformationally flexible and tends to adopt the HA conformation upon ligand binding, whereas the *apo*-lectin prefers the LA conformation. Application of the PCL approach to an R60P-cysteine double variant could thus allow not only photocontrol of lectin function, but also photoswitching of the FimH conformation. Highlighted in turquoise blue: LA conformation; in grey: HA conformation. The green circle and the star indicate a mannopyranoside ligand tagged with a suitable fluorophore (FITC) to enable monitoring of FimH activity.

So far, no structural data for FimH have been obtained under shear conditions as simultaneous structure determination under the coincident application of tensile forces is difficult to experimentally achieve. Admittedly, the here-presented approach cannot induce conformational rearrangement by domain separation or provide shear of defined strength or velocity, which would allow precise adjustment of the effective tensile forces to physiological conditions. However, proceeding from the ligand-bound state, a *cis* → *trans* photoisomerisation could still result in strained conditions that may exert shear stress on the FimH binding cavity. The envisaged setup can therefore at least be regarded as a first approach to introduce shear and simultaneously allow structural investigation of the lectin, for instance by NMR-CSP spectroscopy.[374] This could particularly be considered to explore the occurrence of structural changes in the context of the “local” catch bond effect, which was detected in MD simulations by Stirnemann and colleagues.[394] If applicable, the photopharmacological approach, presented here, could eventually be applied to other lectins, for instance to enable control of lectin binding and cellular adhesion processes for bioscientific research and therapeutic purposes. Other possibilities would be, for instance the photoregulation of drug release, in analogy to the irreversible photodecaging method, as described in Chapter 1.2.2.

3. Photocontrol of the bacterial lectin FimH

3.1 *In silico*-docking simulations

The conception of the project was developed by means of computations, performed with the molecular modelling environment Maestro from the Schrödinger company. Docking simulations were employed to analyse the influence of (i) the position of introduced cysteine residues, (ii) the *cis*/*trans*-configuration of the (fluoro-)azobenzene unit, (iii) the length and associated flexibility of the OEG linkers, and (iv) the respective conformation of the tyrosine gate on the accessibility of the FimH binding pocket in more detail. The aim was, in particular, to investigate which combination of these parameters could enable the coincident intramolecular binding of the mannoside ligand in the FimH binding pocket, after *S*-carbamidoalkylation of the introduced cysteine residues by the IAA function of the same ligand. The findings thus obtained should provide implications for the subsequent experimental realisation of the project.

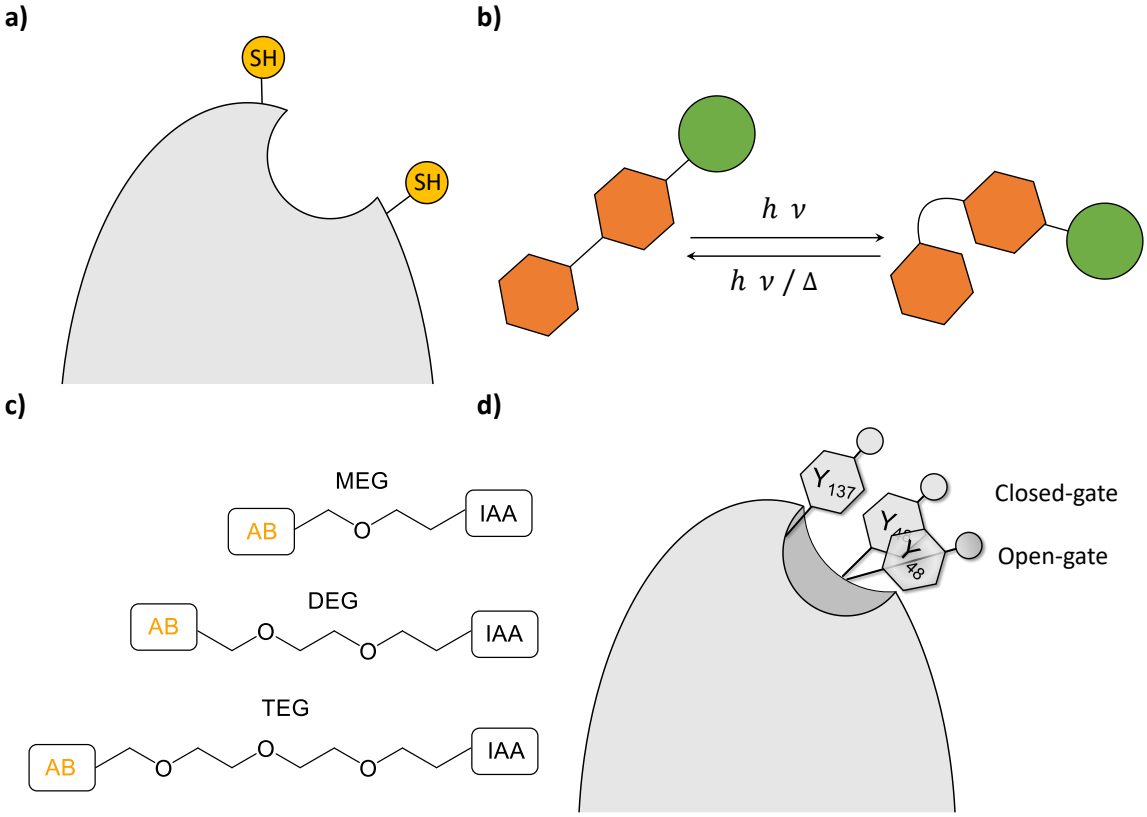


Figure 3.1-1: Parameters investigated *in silico* using Maestro (Schrödinger Inc.) to predict which combination allows for achieving photocontrol over lectin function.

a) Position of cysteine residues for bioconjugation with IAA derivatives (i); **b)** *cis*/*trans*-configuration of the azobenzene moiety (ii); **c)** length of OEG linkers, MEG: monoethylene glycol, DEG: diethylene glycol, TEG: triethylene glycol (iii); **d)** with respect to the conformation of the tyrosine gate (vi). The lectin is depicted in grey, bearing the introduced cysteine sulfur atom (yellow); azobenzene moieties are shown as orange hexagons in *cis*-(bent line) and *trans*-configuration (straight line); the α-D-mannoside moiety is illustrated as a green circle.

All simulations were performed with the representative FimH crystal structures 1UWF[373] and 1KLF.[318] Both structures were originally isolated from uropathogenic *E. coli* J96, and in both cases, the lectin domains adopt the elongated HA conformation. Therefore, both structures feature the N70, S78, and particularly the pathoadaptive V27 mutation,[425,371] the latter of which tends to cause pre-activation upon ligand binding even in the absence of shear.[393] These mutations are likewise present in the *E. coli* K12-derived FimH variants, which are intended to be expressed in the context of this work (see Chapter 3.5). These are also permanently stabilised in the HA conformation due to the absence of the pilin domain. As mentioned in Chapter 1.3.3, isolated FimH-LD was, by coincidence, complexed with *n*-butyl *α*-D-mannoside, leading to the closed-gate conformation of the tyrosine gate (PDB: 1UWF, resolution 1.69 Å, Figure 3.1-2a). The FL-FimH construct, on the other hand, (PDB: 1KLF, resolution 2.79 Å) was stabilised by complexation with the chaperone FimC and was co-crystallised with *α*-D-mannose (Figure 3.1-2b). In the absence of a hydrophobic aglycone on the ligand, the tyrosine gate adopts the open-gate conformation. As demonstrated by other crystal structures of the FimC-FimH complex (e.g. 1KIU, 1QUN and 3RFZ), the close contact of the chaperone with FimH-PD correlates with a subunit separation, which provokes FimH-LD to adopt the elongated HA conformation.[315,318,322]

The two mentioned crystal structures, 1UWF and 1KLF, were chosen in order to take the most relevant rotamers of Tyr48 into consideration even if the alternative “half-open” tyrosine gate conformation has also been described.[375] However, compared to the half-open and open-gate conformations, the closed-gate structure (1UWF) is probably more applicable for computational studies with artificial mannoside ligands containing aliphatic or aromatic aglycone moieties, since corresponding X-ray and NMR data in the presence of such ligands consistently indicate a clear preference for the closed-gate conformation.[374,375]

a) **b)**

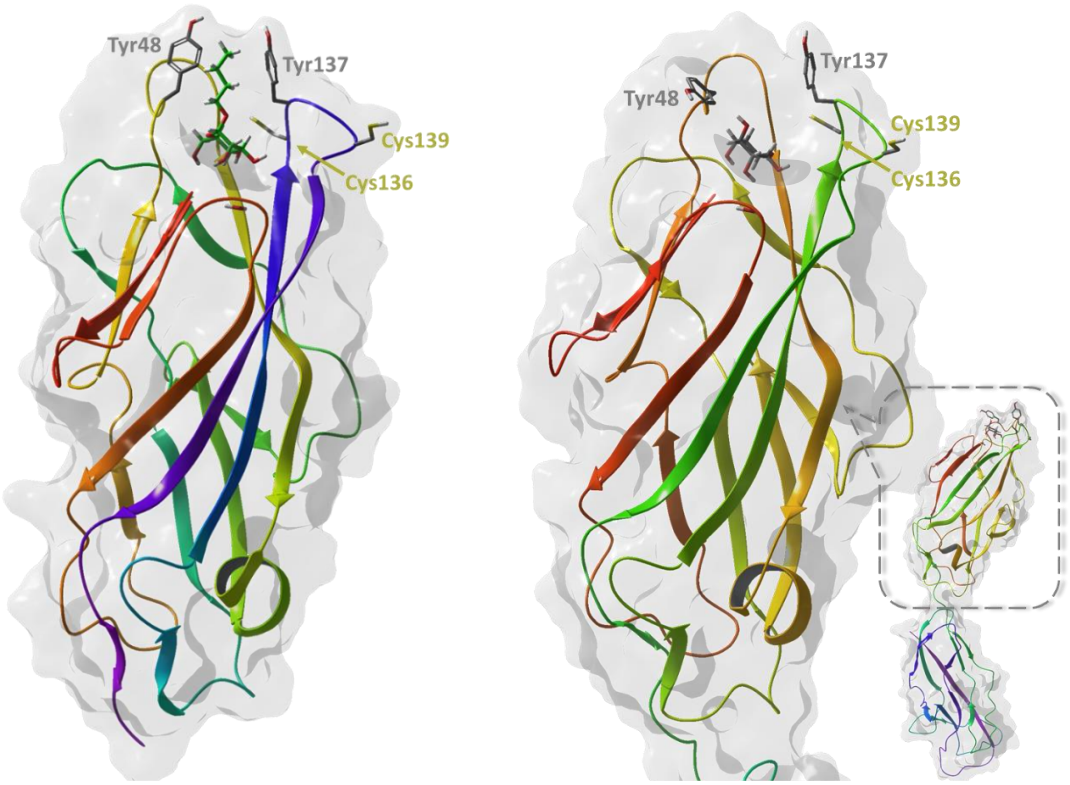
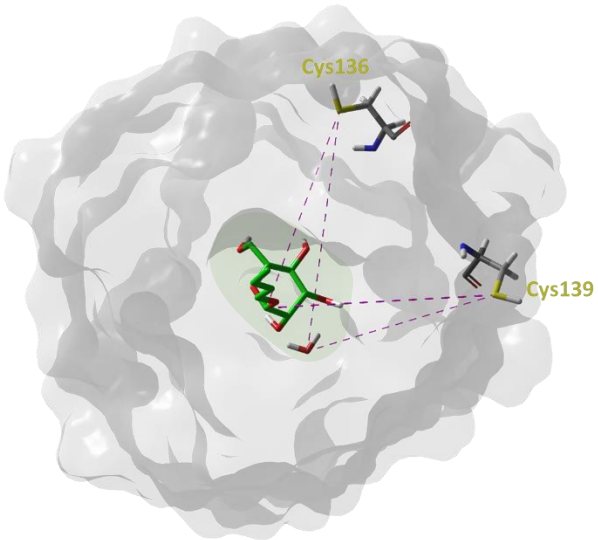


Figure 3.1-2: Crystal structures of the FimH-LD with co-crystallised ligands and the position of substituted cysteine residues (N136C and S139C) used for Glide docking and covalent docking simulations performed using Maestro (Schrödinger Inc.).

a) 1UWF, closed-gate conformation, lectin domain from *E. coli* J96, crystallised with *n*-butyl α-D-mannoside, resolution 1.69 Å;[373] **b)** 1KLF, open-gate conformation, FL-FimH from *E. coli* K12 in complex with FimC and α-D-mannose, resolution 2.79 Å.[318] The structures were energy-minimised using Maestro (Schrödinger Inc.).

a)



b)

PDB	**1UWF**	
Distance	**C_1 – S [Å]**	**H_2O – S [Å]**
N136C	11.12	14.02
S139C	12.09	12.52

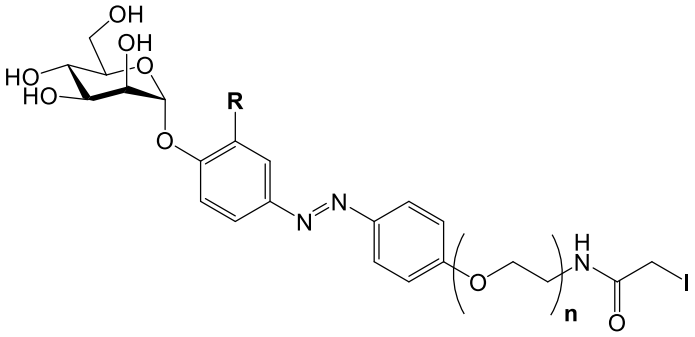
PDB	**1KLF**	
Distance	**C_1 – S [Å]**	**H_2O – S [Å]**
N136C	10.93	13.96
S139C	11.09	12.79

Figure 3.1-3: a) Top view of the FimH-CRD with co-crystallised *n*-butyl α-D-mannoside and **b)** distances between the binding site and the introduced cysteine residues Cys136 and Cys139 (PDB: 1UWF).

The distances were determined between the thiol-sulfur atom of the introduced cysteine residues and the anomeric carbon (C_1) atom of co-crystallised ligands, or the oxygen atom of a co-crystallised water molecule. The binding cavity is highlighted in green for improved visibility. The structure was energy-minimised and distances were measured using Maestro (Schrödinger Inc.).

3.1.1 Glide docking

The exact position of the cysteine residues as intended alkylation sites as well as the length of the OEG linker between the (fluoro-)AB mannoside and the terminal IAA function was initially not precisely defined. In order to identify suitable bioconjugation positions in the vicinity to the FimH binding pocket, ligand docking simulations were performed with the chosen 1UWF/1KLF crystal structures using the Glide application[426] in an OPLS-2005 force field[427,428] and in implicit water. The IAA-functionalised (fluoro-)AB mannosides, shown in Figure 3.1-4, were employed as ligands for the initial docking simulations. The ligands feature different numbers of ethylene units (mono-, di-, and triethylene glycol: MEG, DEG, and TEG) to determine the optimal linker lengths required for bringing the IAA function into spatial proximity with the possible cysteine positions.



R = H, n = 2: **H1**	R = F, n = 2: **F1**
R = H, n = 3: **H2**	R = F, n = 3: **F2**
R = H, n = 1: **3**	R = F, n = 1: **4**

Figure 3.1-4: Mono-, di-, and triethylene glycol-linked and IAA-functionalised (fluoro-)azobenzene *α*-D-mannosides used for Glide docking and covalent docking simulations performed with Maestro (Schrödinger Inc.).

A maximum output of 20 poses per ligand was calculated, and both the *cis*- and the *trans*-configuration of the AB units were considered. The dimensionless docking score is given as an approximate value for the expected binding affinity of a respective ligand in different poses.[429] A lower (i.e. a more negative) Glide docking score indicates a more favourable predicted binding affinity of a respective ligand to the given receptor. The resulting docking scores for the fluoro-AB mannoside ligands **F1** and **F2** are summarised by way of example in Table 5.6-1 (1UWF) and Table 5.6-2 (1KLF) in Chapter 5.6.2. The mono-fluorination of the AB unit proved to have little influence on the docking results, since very similar outputs were obtained with the analogous AB ligands **H1** and **H2** (data not shown).

Based on these initial docking results, the positions N136 and S139 were identified as suitable alkylation sites, both of which are located on CRD loop 3 at the outer edge of the binding cavity. The amino acids N136 and S139 are situated in close proximity to the tyrosine gate residue Tyr137 but are generally not considered essential for lectin function. Notwithstanding that the amino acid sequence of CRD loop 3 (R132 – F142) is highly conserved in a multitude of clinical isolates associated with UTI,[318] only Q133, N135, and D140 are classified as critical for mannose binding.[430] A substitution of the polar, uncharged amino acids Ser and Asn with Cys likewise does not affect the hydrophobic ridge (F1, I13, Y48, I52, Y137, F142).[318] As shown in Figure 3.1-3, the distance to the binding pocket is approximately the same from both alkylation positions.

Due to the close proximity to Tyr137, the alkylation position Cys136 is probably more strongly influenced by the presence and conformation of the tyrosine gate residues than Cys139. The docking results also indicated that DEG and TEG appear particularly suitable as linker moieties, whereas MEG-linked ligands (**3**, **4**) are probably too short to bring the IAA function into sufficient proximity to the selected alkylation sites (data not shown).

The intended cysteine modifications were inserted *in silico* into the 1UWF and 1KLF crystal structures, using the mutate residue tool, after which the models were energy-minimised using MacroModel.[431] The Glide docking simulation was then repeated with the DEG- and TEG-linked ligands **H1**, **F1**, **H2**, and **F2** in both the *cis*- and *trans*-configurations of the AB unit. The results of 10 poses with the highest docking scores for each ligand are compiled in Table 5.5-3 – 5.5-10 (Chapter 5.6.2). The S – C distances between the methylene carbon within the IAA group and the thiol-sulfur atom of the introduced cysteine residues were determined. The poses with the smallest S – C distances (preferably < 4 Å) were selected and *S*-carbamidoalkylation was performed manually with the 3D Builder tool. The respective docking illustrations and ligand interaction diagrams are likewise shown in Chapter 5.6.2 beneath the corresponding docking score tables (Figures 5.5-1 – 5.5-30). The selected *in silico*-alkylated poses were again energy-minimised. The resulting structures were assessed and categorised according to whether the mannopyranoside unit of the modelled ligand was superimposable with the co-crystallised ligand.

All outputs with correctly bound mannoside moieties and generally correct overall conformation are summarised in Table 3.1-1. Strikingly, correct mannoside binding was achieved predominantly with the S139C variant in combination with the TEG-linked ligands (**H2**, **F2**), as well as with *cis*-**H1** or *cis*-**F1**. Covalently attached DEG ligands in *trans*-configuration generally failed to bind, both in the 1UWF and in the 1KLF crystal structure. An exemplary result in which the DEG-linked ligand **F1** is correctly complexed within the binding cavity (as *cis*- but not as *trans*-isomer) is visualised in Figure 3.1-5. The only example for a successful mannoside recognition when the ligand was covalently alkylated to the Cys136 position was found in the open-gate conformation (1KLF N136C) with *cis*-**H1**. All further results concerning post-docking alkylation and final energy minimisation are provided in Table 5.6-11 (1UWF), Table 5.6-12 (1KLF) and Figures 5.5.-31 – 5.5-41 (Chapter 5.6.2).

Table 3.1-1: Selected ligand docking scores with correct binding modes of mannoside ligands obtained from Glide docking and *in silico*-alkylation of IAA ligands with FimH structures 1UWF and 1KLF.

1UWF			**1KLF**		
Cysteine position	**Alkylated ligand**	**Docking score**	**Cysteine position**	**Alkylated ligand**	**Docking score**
n. d.	n. d.	n. d.	N136C	*cis*-**H1**	-7.009
S139C	*cis*-**H1**	-6.317	n. d.	n. d.	n. d.
S139C	*trans*-**H2**	-5.013	S139C	*trans*-**H2**	-4.437
S139C	*cis*-**H2**	-5.994	n. d.	n. d.	n. d.
S139C	*cis*-**F1**	-6.310	S139C	*cis*-**F1**	-5.453
S139C	*trans*-**F2**	-4.709	S139C	*trans*-**F2**	-5.294
S139C	*cis*-**F2**	-5.269	S139C	*cis*-**F2**	-4.175

1UWF: closed-gate conformation; 1KLF: open-gate conformation of FimH; the table only summarises outputs in which the mannoside unit is correctly bound within the binding cavity (mannopyranoside ring superimposable with the respective co-crystallised ligand). n. d.: No correct docking results and/or azobenzene structures were observed. The data were obtained by Glide docking, followed by *in silico*-alkylation and energy minimisation using the Maestro suite (Schrödinger Inc.).

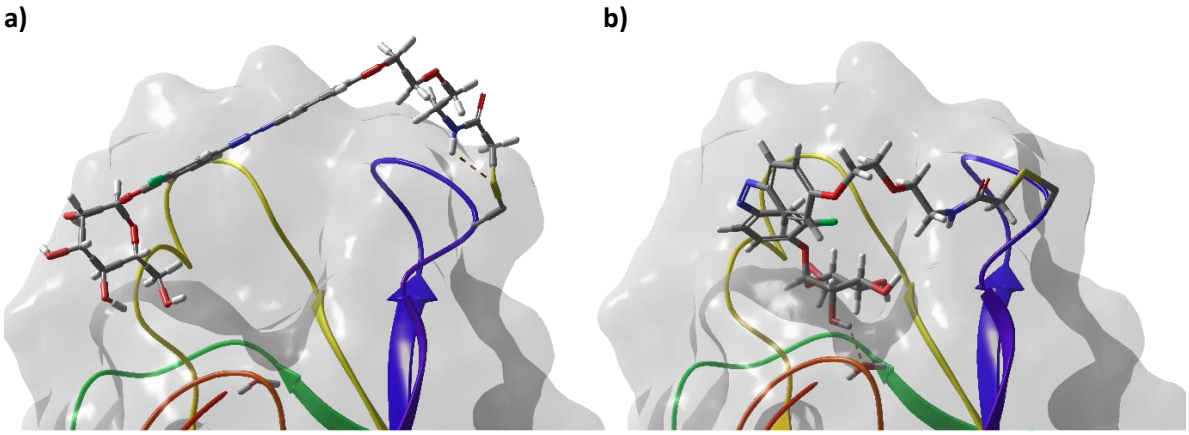


Figure 3.1-5: Exemplary outputs resulting from ligand docking simulations with **a)** *trans*-**F1** and **b)** *cis*-**F1** in the FimH variant S139C (PDB: 1UWF, closed-gate conformation).

The mannoside moiety of the covalently attached ligand is able to bind within the binding cavity in its *cis*- but not in its *trans*-configuration of the azobenzene unit. The illustrations were generated by Glide docking simulations, followed by *in silico*-alkylation and energy minimisation using the Maestro suite (Schrödinger Inc.).

3.1.2 Covalent docking

After the calculations with the established Glide docking protocol for validating the envisaged experimental setup had been completed, the scope of the computational methods was extended to include covalent docking at a later stage of the doctoral project. In recent years, the importance of covalent drug research has increased enormously and corresponding calculation tools have been significantly improved.[432–434] In contrast to the previously pursued approach for ligand docking, not only the recognition of the (mannoside) ligand by the receptor, but also the chemistry of the covalent bond formation can be taken into account. To confirm and consolidate the previous findings, covalent docking[435] was performed in the more recent OPLS4 force field[436] and in implicit water, using the FimH structures 1KLF and 1UWF and the *cis*- and *trans*-isomeric forms of the (fluoro-)AB mannoside ligands, shown in Figure 3.1-4. The introduced cysteine positions Cys136 and Cys139 were predetermined for a nucleophilic substitution reaction with an iodide ion as leaving group. Again, a maximum output of 20 poses was selected. After computation of suitable poses of ligand and receptor and introduction of the covalent bond, the poses were energy-minimised with Prime and the results were ranked according to the respective Prime energies.[437] The Prime energy indicates the lowest total energy pose of a given ligand in complex with the receptor, but it is not suitable for comparing different ligands, for which the docking score is more meaningful. In contrast to the initial Glide docking method, covalent docking produced a greater number of structurally viable poses, as only chemically and sterically feasible interactions were reported. If simultaneous alkylation and docking of the mannoside ligand were not possible, no output was generated. The resulting poses therefore only required evaluation with regard to structural plausibility, in particular, the torsion angles of the *trans*- and *cis*-azobenzene moieties. Covalent docking results, representing only the top-scoring poses based on Prime energies, are compiled in Table 3.1-2. All further data, including ligand interaction diagrams and docking poses, are illustrated in Chapter 5.6.3 (Figures 5.5-42 – 5.5-69). The respective scoring values are likewise provided in this chapter (Tables 5.5-13 – 5.5-20). Notably, no output was obtained with DEG-linked ligands in the *trans*-configuration of the azobenzene unit (*trans*-**H1**, *trans*-**F1**) when linked to the Cys136 position in both open-gate and closed-gate conformations. This finding highlights the particular suitability of DEG-linked ligands for the intended photocontrol of FimH and is consistent with the Glide docking results.

Table 3.1-2: Selected docking scores and Prime energies obtained from covalent docking of IAA ligands with FimH structures 1UWF and 1KLF.

1UWF				1KLF			
Cysteine position	**Alkylated ligand**	**Docking score**	**Prime energy [kcal · mol^{-1}]**	**Cysteine position**	**Alkylated ligand**	**Docking score**	**Prime energy [kcal · mol^{-1}]**
N136C	*cis*-**H1**	-7.124	-5841.46	N136C	*cis*-**H1**	-7.443	-10200.10
N136C	*trans*-**H2**	-7.783	-5904.70	N136C	*trans*-**H2**	-7.410	-10097.83
N136C	*cis*-**H2**	-8.231	-5924.04	N136C	*cis*-**H2**	-7.988	-10216.28
S139C	*trans*-**H1**	-7.033	-5998.62	S139C	*trans*-**H1**	-6.341	-10257.25
S139C	*cis*-**H1**	-7.150	-5992.29	S139C	*cis*-**H1**	-7.187	-10263.32
S139C	*trans*-**H2**	-7.207	-5954.12	S139C	*trans*-**H2**	-7.370	-10261.04
S139C	*cis*-**H2**	-8.000	-5995.30	S139C	*cis*-**H2**	-7.776	-10258.70
N136C	*cis*-**F1**	-7.818	-5878.78	N136C	*cis*-**F1**	-8.541	-10221.30
N136C	*trans*-**F2**	-8.353	-5913.48	N136C	*trans*-**F2**	-5.979	-10185.13
N136C	*cis*-**F2**	-8.395	-5950.48	N136C	*cis*-**F2**	-8.739	-10223.98
S139C	*trans*-**F1**	-7.37	-5969.85	S139C	*trans*-**H1**	-7.237	-10268.52
S139C	*cis*-**F1**	-8.314	-6002.09	S139C	*cis*-**H1**	-8.109	-10258.18
S139C	*trans*-**F2**	-7.319	-5955.85	S139C	*trans*-**H2**	-7.054	-10260.80
S139C	*cis*-**F2**	-8.432	-5995.12	S139C	*cis*-**H2**	-7.956	-10259.15

1UWF: closed-gate conformation; 1KLF: open-gate conformation of FimH; the table only summarises outputs with the highest Prime energies, respectively; No correct azobenzene structure and/or covalent docking result was observed with *trans*-**H1**/-**F1**. The data were obtained by covalent docking using Maestro (Schrödinger Inc.).

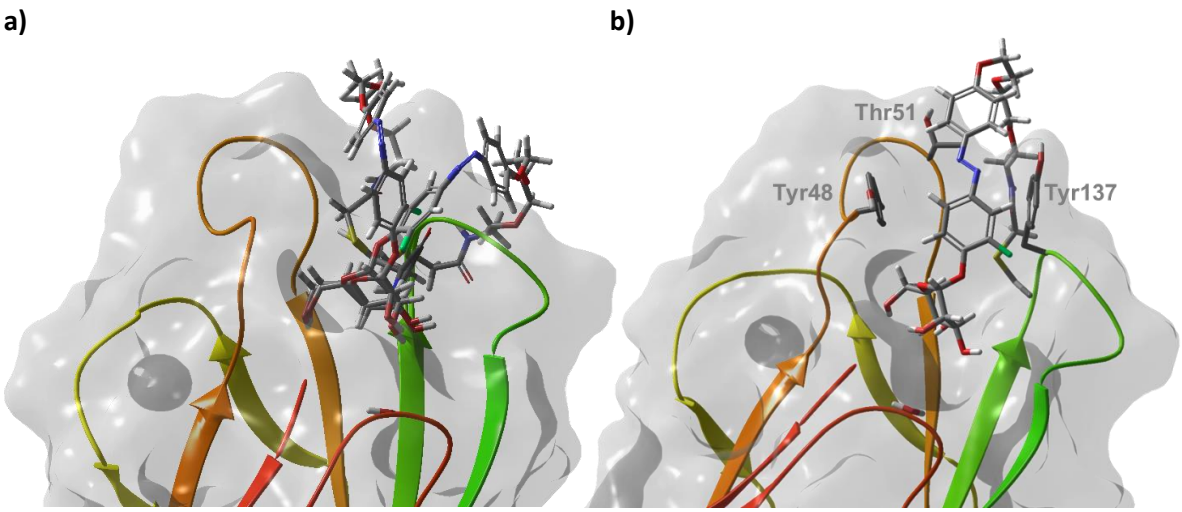


Figure 3.1-6: Exemplary outputs resulting from covalent docking simulations with **a)** superposition of two alternative conformations of *cis*-**F2** ligands and **b)** pose of a single *trans*-**F2** ligand in the FimH variant N136C (PDB: 1KLF, open-gate conformation).

The mannoside moiety of the covalently attached ligand is able to bind within the binding cavity in both *cis*- and *trans*-configurations of the azobenzene unit. The illustrations were generated by covalent docking simulations using the Maestro suite (Schrödinger Inc.).

Also in the covalent docking attempt with the MEG ligands **3** and **4**, no poses were generated (neither in the *cis*-, nor in the *trans*-configuration), confirming that the length of the MEG linker is not sufficient to bridge the distance between the binding cavity and alkylation sites. In all other combinations, simultaneous *S*-carbamidoalkylation of Cys136 and Cys139 and binding of the mannoside moiety in the binding cavity was successfully simulated. However, covalent docking of *trans*-AB mannoside ligands generally provided lower docking scores and less outputs with correct overall structure than those obtained with the respective *cis*-isomers. Frequently, docking of *trans*-configured ligands could only be achieved by inadmissible bending or twisting of the AB phenyl rings. This was also already the case with the previous Glide docking method. Even with the more flexible TEG ligands, docking as *trans*-configured isomers proved to be more difficult: As can be seen, for instance, from the simulation of *trans*-**H2** with the N136C variant of the 1KLF crystal structure, only three correct outputs (instead of more than 10 in the case of the respective *cis*-isomer) were obtained (Table 5.6-17). Only a single conformationally acceptable pose was obtained with the corresponding fluoro-AB analog *trans*-**F2**, which is depicted in Figure 3.1-6b (see also Table 5.6-19 in Chapter 5.6.3). The orientation of the ligand is restricted by the planarity of the fluoro-AB unit in *trans*-configuration and by steric hindrance, as a consequence of the direct vicinity to Tyr137. The TEG linker was placed over the small groove between Thr51 on CRD loop 2 and Tyr137 on loop 3. The corresponding ligand interaction diagram (Figure 5.6-64a) predicts π-π interactions between Tyr137 and the AB phenyl ring next to the TEG linker. If the *cis*-isomer was employed for the simulation, the planarity of the azobenzene moiety was reduced. Thus, the orientation of the linker became more flexible and positioning of the TEG unit between Tyr137 and Ser139 was also possible (see Figure 3.1-6a). The steric restriction of the ligands, when alkylated at Cys136, emerged with both the open-gate and closed-gate structures since the conformational change between both conformations primarily affects the more distant Tyr48. Fewer steric constraints apply to the Cys139 position. Therefore, more covalent docking outputs were obtained for this variant, including several poses with *trans*-**H1**/**F1** ligands.

Taken together, the chemical modification of the FimH N136C variant with DEG-linked azobenzene mannosides (**H1**, **F1**) appears to be the most promising combination for the experimental realisation of the intended photochemical control of lectin activity. Ligand docking simulations with Glide suggest that both FimH variants (N136C and S139C) are suitable candidates. According to the more reliable covalent docking simulations, however, the FimH S139C variant seems less favourable since it probably offers too much flexibility for the positioning of the ligand, especially in the *trans*-configuration of the azobenzene photoswitch. However, for DEG ligands, both *in silico*-docking approaches consistently predict that binding of the mannoside moiety in the FimH binding cavity is only possible for the *cis*-configured isomer, i.e. after UV exposure of the modified protein. MEG and TEG ligands are obviously either too short or too long, respectively, to allow efficient photocontrol of lectin activity.

3.2 Synthesis of photosensitive ligands

A linear synthesis of DEG- and TEG-linked, IAA-functionalised azobenzene *α*-D-mannosides had already been introduced in the Master thesis.[438] In the latter, the target molecules **H1** and **H2** were obtained employing the acetyl-protected mannosyl trichloroacetimidate **5**, in only five reaction steps with overall yields of 9 % and 8 %, respectively.[438] In line with the previous work, a linear synthesis route was initially followed in order to obtain the corresponding fluoro-AB mannoside derivatives **F1** and **F2**. Due to partially unsatisfying and/or non-reproducible yields, as well as the lack of permanently stable late-stage intermediates, this 8-step linear synthesis was changed in favour of a convergent strategy in the course of this doctoral thesis.

The convergent synthesis strategy offered significantly improved yields and reproducibility, and was therefore pursued for the renewed synthesis of the azobenzene mannosides **H1** and **H2**, whereby the first steps up to the DEG azobenzene **36** were carried out by Jule Bergfeld as part of her F3 practical course. Likewise, the *N*-acetylated analogues **51**, **52**, **53**, and **54** were obtained from stable intermediates taken from the convergent synthesis route towards the iodoacetamide derivatives. The retrosynthetic fragmentation of the target molecules, illustrated in Figure 3.2-1, suggests appropriate precursors for both synthesis strategies: a suitable 2,3,4,6-tetra-*O*-protected mannosyl donor, precursors for the asymmetrical fluoroazobenzene moiety, DEG and TEG linkers, as well as the terminal (iodo-)acetamidation.

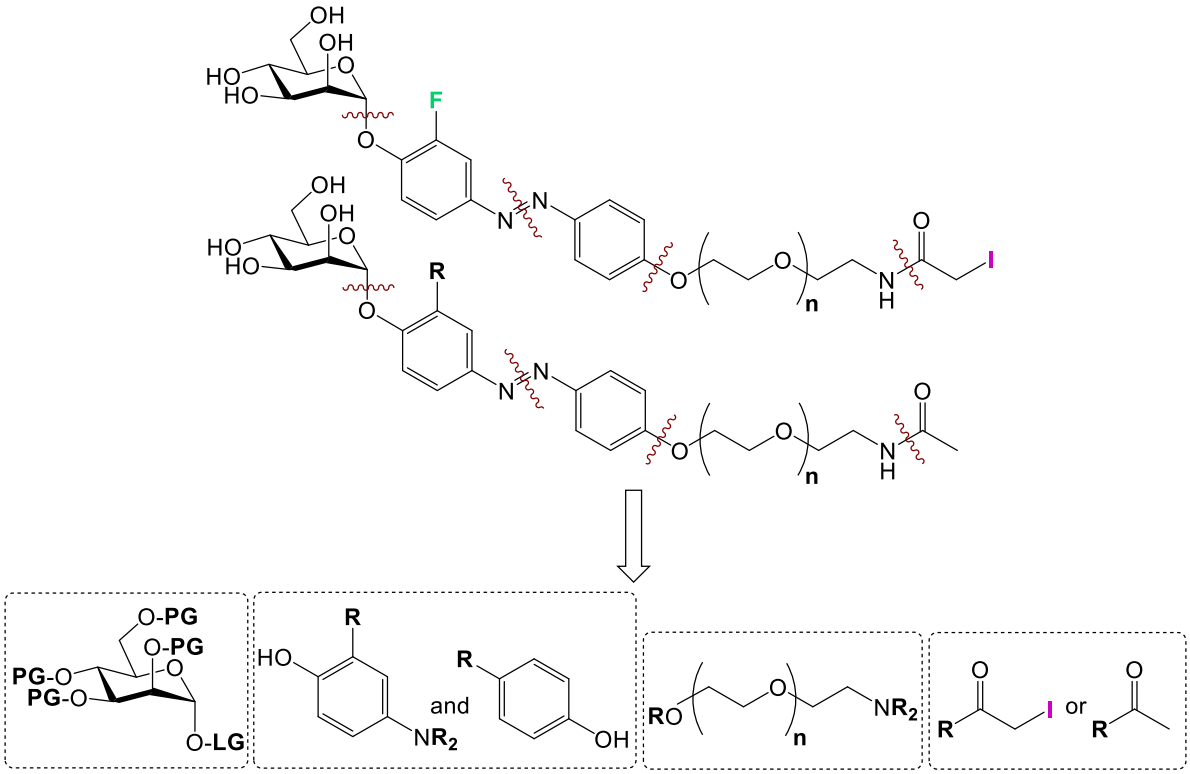
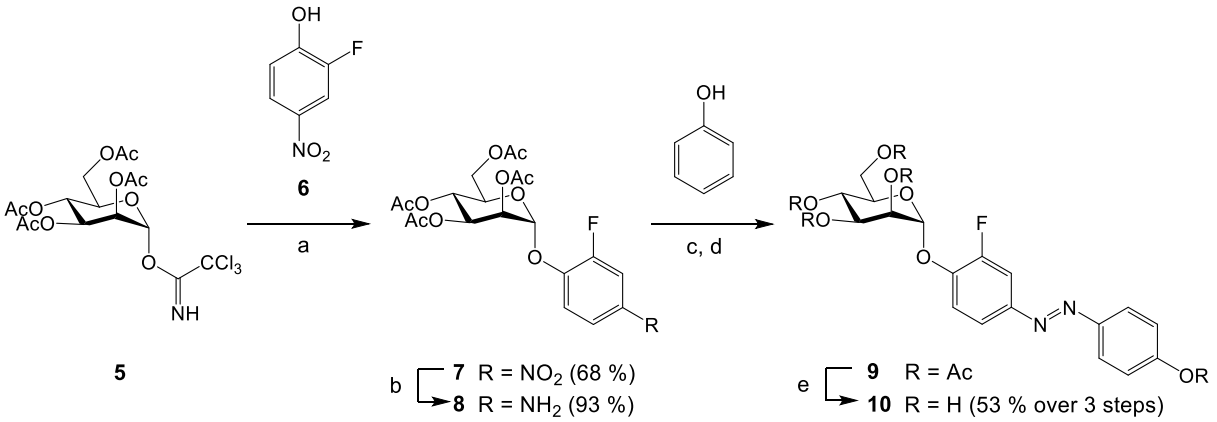


Figure 3.2-1: Retrosynthetic analysis of iodoacetamide-functionalised azobenzene mannoside derivatives and their *N*-acetylated analogues for the chemical modification of FimH cysteine variants. **PG**: Protecting group; **LG**: leaving group; **R**: unspecified substituents, **n**: 1 or 2.

3.2.1 Linear synthesis of iodoacetamide-functionalised 3-fluoroazobenzene α-D-mannosides

Both of the synthesis approaches presented here utilised the literature-known and well-established glycosylation strategy based on the 2,3,4,6-*O*-tetra-acetyl-protected mannosyl trichloroacetimidate **5** as a starting point.[439] In order to achieve the desired mono-fluorination of the dihydroxyazobenzene unit in the *ortho*-position to the mannopyranosyloxy moiety, commercially available 2-fluoro-4-nitrophenol **6** was subjected to glycosylation according to the trichloroacetimidate method. Boron trifluoride diethyl etherate ($BF_3 \cdot OEt_2$) was employed as a catalyst for the activation of the mannosyl donor (Scheme 3.2-1). The 2-fluoro-4-nitrophenyl mannoside **7** was afforded with a good 68 % yield. The approach differs in this regard from the Master thesis,[438] where dihydroxyazobenzene was mannosylated. However, the glycosylation of 3-fluoro-4,4'-dihydroxyazobenzene had been avoided, in order to supersede an otherwise required discrimination between the two hydroxyl groups on the asymmetric fluoro-azobenzene derivative.

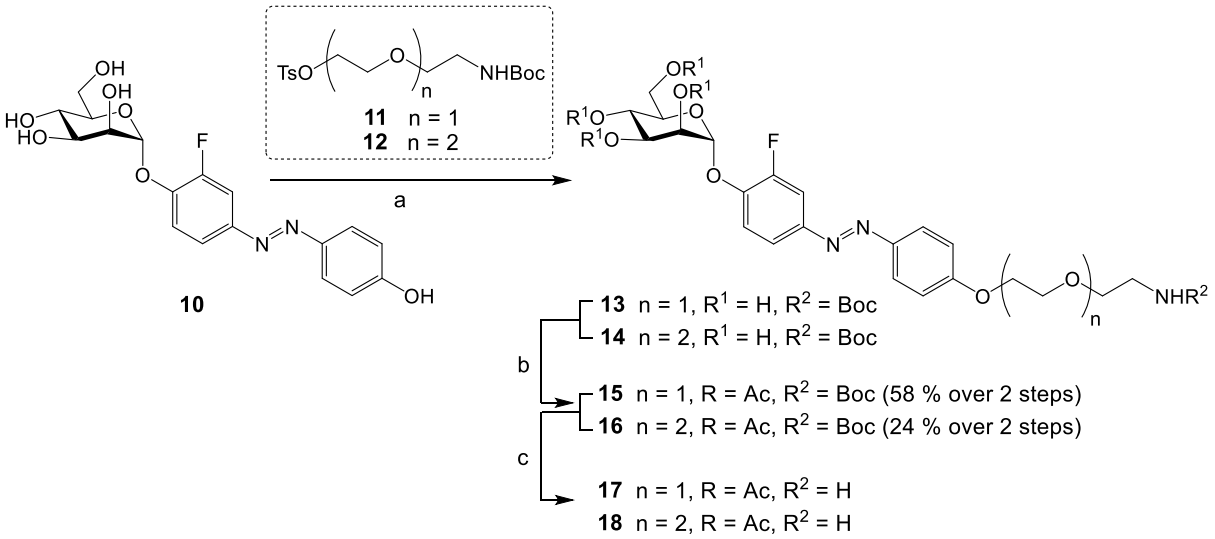


Scheme 3.2-1: Synthesis of acetyl-deprotected 3-fluoroazobenzene mannoside **10**.
a) $BF_3 \cdot OEt_2$, acetonitrile, dry DCM, 0 °C → rt, 16 h; **b)** H_2, Pd/C, dry DCM, dry methanol, rt, 16 h; **c)** 1. $NaNO_2$, HCl, MeOH, dH_2O, 0 °C, 1 h; 2. Phenol, KOH, MeOH, 0 °C → rt, 16 – 18 h; **d)** Ac_2O, pyridine, rt, 16 h; **e)** NaOMe, MeOH, rt, 2 h.

To enable conversion to an azobenzene moiety, a hydrogenation of the nitrophenol mannoside **7**, catalysed by palladium on activated charcoal (Pd/C), was performed with gaseous hydrogen. Subsequent column chromatographic purification over silica with the addition of small amounts of trimethylamine gave the corresponding amine **8** with high 93 % yield. Following this, the aromatic amine **8** was converted to the respective aryldiazonium cation with sodium nitrite in hydrochloric acid solution to achieve the subsequent *in situ*-azo coupling with phenol. Since the coupling agent is classically activated by alkalisation in the presence of methanol, deprotection of the mannosyl moiety was observed as a common side reaction. In the first instance, a quantitative deacetylation was ensured by further addition of KOH to the reaction mixture, if required. A subsequent attempt to purify the resulting, completely deacetylated fluoro-AB mannoside **10** by recrystallisation or rough flash chromatography only gave the crude compound. Also, the attempt to regioselectively remove the aryl-*O*-protecting group from **9**, following a protocol by Blay et al. using tosylic acid adsorbed on silica, was unsuccessful.[440] Alternative approaches for the synthesis of the desired azobenzene glycosides, according to methods that are likewise established in the Lindhorst group,[441] were unsuccessful. This applied at least for the reaction in a biphasic system with Oxone®,[442] as well as under acidic conditions via Mills coupling.[443]

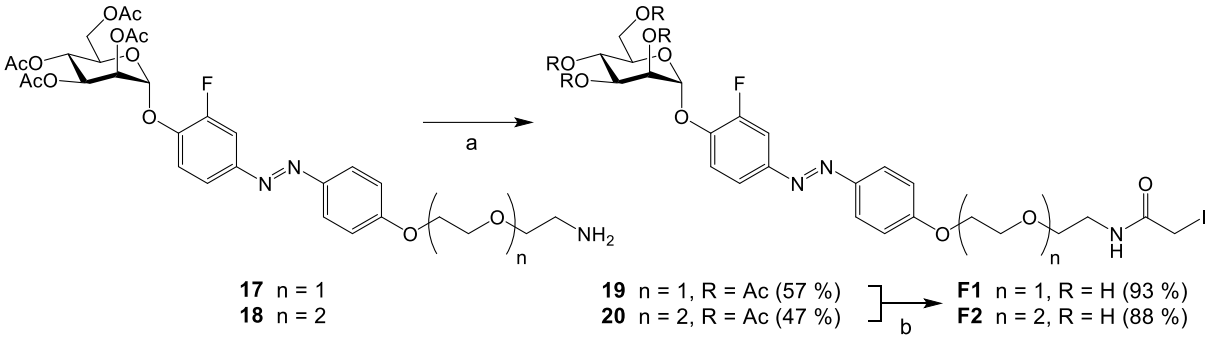
In the latter approach, the nitro compounds (**7** or *p*-nitrophenol, respectively) were re-formed instead of the actually required nitroso derivatives. Eventually, the crude compound of **10** was peracetylated, chromatographically purified, and deacetylated again under Zemplén conditions.[444] Deacetylated 3-fluoro-AB mannoside **10** was thus afforded with a 53 % yield after three reaction steps.

The higher acidity of the aryl-hydroxyl group, compared to the aliphatic OH-groups of the mannopyranoside moiety, permits the subjection of the deacetylated 3-fluoro-AB mannoside **10** to a classical Williamson ether synthesis (Scheme 3.2-2).[445] The latter was thus reacted with the asymmetrically *O*-tosyl (TsO)- and *N*-*tert*-butyloxycarbonyl (Boc)-substituted DEG and TEG linkers **11** and **12**, respectively.



Scheme 3.2-2: Williamson etherification and following protection and deprotection steps for the synthesis of amino derivatives **17** and **18**.
a) K_2CO_3, dry DMF, 2 h at 90 °C,then rt, 16 h; **b)** Ac_2O, pyridine, rt, 2 – 3 h; **c)** TFA, DCM, rt, 3 h.

The respective OEG linkers, in turn, were obtained as described in the Master thesis,[438] starting from commercially available compounds and according to literature-known procedures.[446–448] Following the Williamson reaction, the resulting crude intermediates **13** and **14** were roughly filtered over silica. Subsequently, the crude compounds were peracetylated by use the of acetic anhydride in pyridine in order to facilitate thorough purification via flash chromatography. The *O*-acetyl- and *N*-Boc-protected DEG- and TEG-linked AB mannoside derivatives **15** and **16** were obtained with 58 % and a rather low 24 % yield over two reaction steps, respectively. The terminal *N*-Boc group was removed in the conventional way with trifluoroacetic acid (TFA). After the subsequent neutralisation in an aqueous workup, the amines **17** and **18** were yielded as crude TFA salts, which were not purified further.

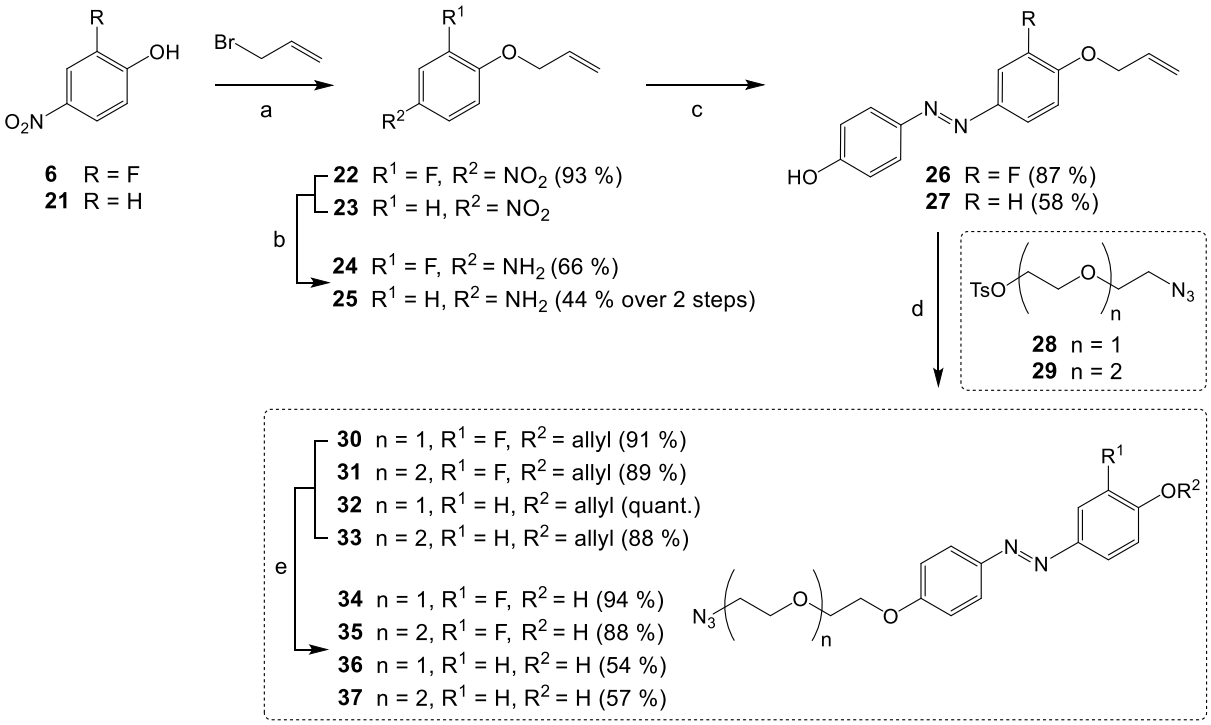


Scheme 3.2-3: Amidation and final deprotection steps for the synthesis of target molecules **F1** and **F2**. **a)** 2-Iodoacetic anhydride, dry DCM, rt, 16 h, in the dark; **b)** NaOMe, MeOH, rt, 1 h.

Following this, the conversion with iodoacetic anhydride in dry dichloromethane (DCM) and under exclusion of light gave the IAA-functionalised, tetra-*O*-acetyl-protected fluoro-AB mannosides **19** and **20** with 57 % and 47 % yield, respectively (Scheme 3.2-3). Exclusion of light not only averted the formation of hydrogen iodide during the reaction but also alleviated the subsequent column chromatographic purification by avoiding *trans* → *cis* isomerisation of the dihydroxyazobenzene moiety. Eventually, **19** and **20** were subjected to a final deacetylation reaction under Zemplén conditions. After careful purification by column chromatography in the dark and subsequent size exclusion chromatography (SEC), the target molecules **F1** and **F2** were afforded with good 93 and 88 % yield, respectively.

3.2.2 Semi-convergent synthesis of iodoacetamide- and *N*-acetyl-functionalised azobenzene α-D-mannoside derivatives

The alternative semi-convergent approach for the synthesis of 3-fluoro-AB mannoside derivatives likewise started from commercially available 2-fluoro-4-nitrophenol (**6**). To obtain the analogous non-fluorinated AB mannoside compounds, 4-nitrophenol **21** was employed as the initial starting material (Scheme 3.2-4). In order to optimise the regioselectivity for the ensuing Williamson ether synthesis with the compounds **6** and **21**, the free aryl-hydroxyl groups were each allyl-protected according to well-known literature procedures with allyl bromide.[449,450] The nitro functions of the resulting compounds **22** and **23** were reduced to the respective arylamines. This was, *inter alia*, achieved by using elemental iron and acetic acid, which afforded the fluoro-aniline **24**[450] and the non-fluoro derivative **25**[451] in 61 and 44 % yield, respectively, whereby the latter corresponds to the yield over two reaction steps. A likewise well-known strategy from the literature, suitable for the selective reduction of allyloxy anilines with tin(II)chloride,[452] was discarded in consideration of the lower yields obtained and elevated toxicity of the reducing agent.



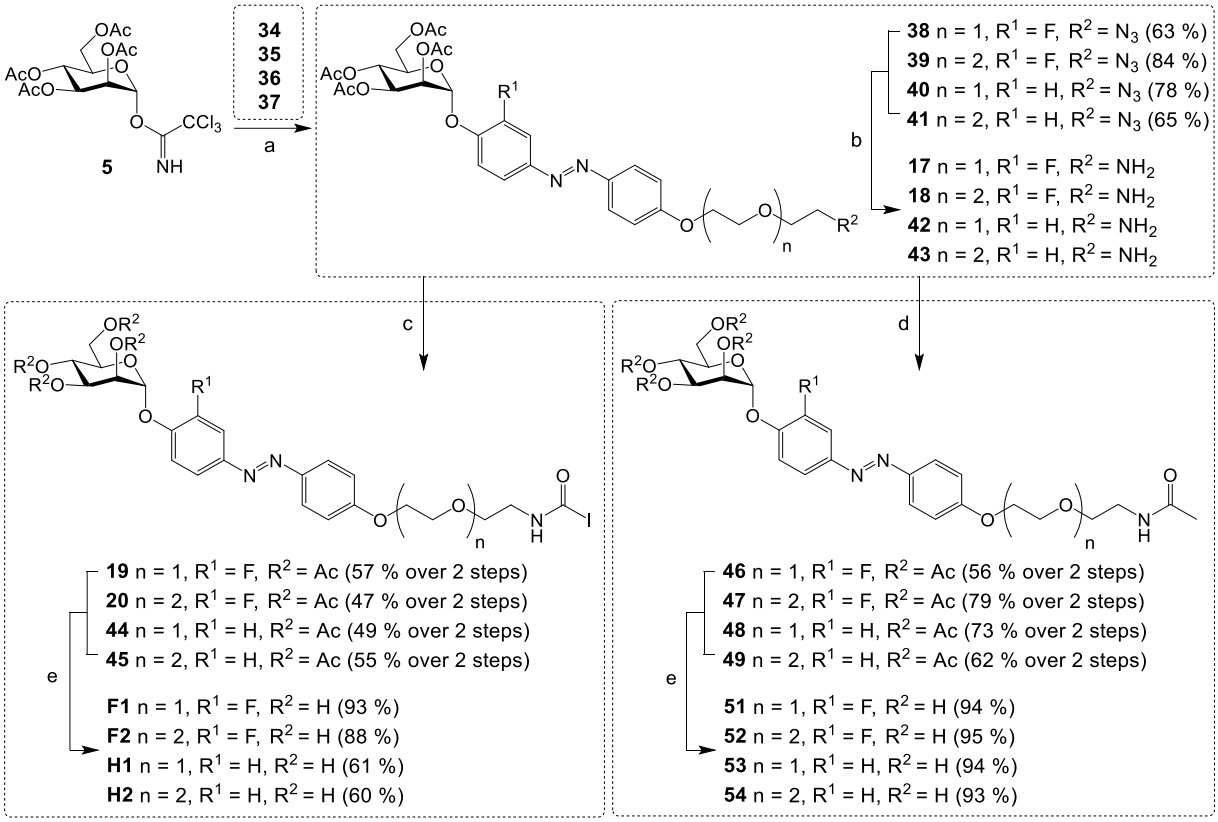
Scheme 3.2-4: Synthesis of asymmetric oligoethylene glycol-functionalised azobenzene derivatives.
a) K_2CO_3, dry acetonitrile, 80 °C, 3 h; **b)** iron powder, glacial acetic acid, dH_2O, 100 °C, 30 min; **c)** 1. $NaNO_2$, HCl, MeOH, 10 °C, 1 h; 2. Phenol, NaOH, MeOH, dH_2O, 10 °C, 1 h; **d)** K_2CO_3, dry DMF, 2 h at 90 °C,then 16 h at rt; **e)** $Pd(PPh_3)_4$, K_2CO_3, dry MeOH, rt, 5 – 48 h.

As already applied in the linear approach, the *O*-allyl-protected anilines **24** and **25** were subjected to a conventional diazotisation[453] and azo coupling reaction with phenol. The resulting asymmetric dihydroxyazobenzene compounds **26** and **27** were obtained in good (87 %) and mediocre (58 %) yield, respectively. Admittedly, the synthesis of the non-fluoro compound **27** was only performed once by Jule Bergfeld and was not reproduced thereafter. The dihydroxy-AB derivatives **26** and **27** were then reacted with the likewise asymmetric DEG and TEG tosylates **28** and **29** in a Williamson ether reaction. The allyl protection of the aryl hydroxyl groups on **26** and **27** ensures the nucleophilic attack only at one end of the linker molecules. The selective mono-allylation of the non-fluoro compound **27** was actually not necessary in this context, but the procedure was retained due to the pleasingly good yields of the etherification reaction (88 % and 100 %, respectively).

Incidentally, the asymmetric DEG and TEG monoazides **28**[454] and **29**[438,447] were obtained by conversion of the respective di-toluenesulfonates with 1 eq. of sodium azide in dry DMF at room temperature. As described in the protocol by Stütz et al.[454], a large part of the accruing salts can be easily precipitated upon addition of DCM and removed by filtration. After purification by flash chromatography, the OEG monoazides were obtained in 67 % and 52 % yield, since the respective OEG diazides are formed as side products. Subsequent to the Williamson etherification, the *O*-allyl group was removed from the OEG dihydroxyazobenzene ethers **30**, **31**, **32**, and **33** in a palladium-catalysed reaction under mildly basic conditions.[455] The thus deprotected fluoro-substituted DEG and TEG ethers **34** and **35** were gained in considerably better yields (94 and 88 %), compared to the respective non-fluoro derivatives **36** and **37** (54 and 57 %).

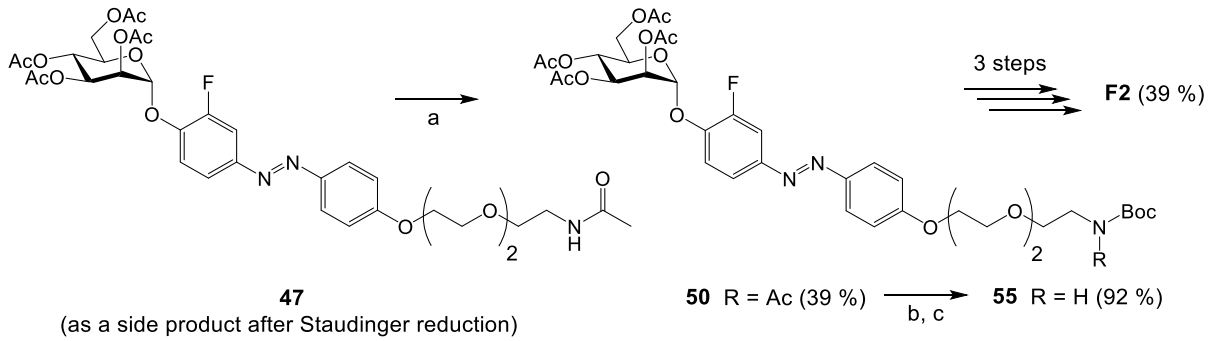
In contrast to the previously employed *N*-Boc protection strategy (Chapter 3.2.1), the terminal azido function of the OEG dihydroxyazobenzene ethers **34** – **37** is compatible with the conditions of a glycosylation reaction according to the already mentioned trichloroacetimidate method. As described in the previous Chapter 3.2.1, the reaction was catalysed by boron trifluoride etherate using the respective peracetylated mannoside **5** (Scheme 3.2-5) as the glycosyl donor. The mannosylation was achieved in relatively satisfying yields in the range of 63 – 84 %. As a positive side effect, the obtained azobenzene mannoside derivatives **38** – **41** were proven to be exceptionally stable over a prolonged time, especially when stored at –20 °C under a nitrogen atmosphere. The derivatives were therefore stocked as precursor molecules, until the target molecules were conceivably needed. In this case, the terminal azido function was reduced in a classical Staudinger reaction with triphenylphosphine (PPh_3), in a mixture of tetrahydrofuran (THF) and water.[456] The resulting amines **17**, **18**, **42**, as well as **43** were afforded as crude products after a rough filtration over silica. Alternatively, the compounds were directly converted in a subsequent amidation reaction after removal of the solvent and without any further purification. With the free amine derivatives in hand, either an IAA function was introduced, using iodoacetic anhydride in DCM under exclusion of light, or the analogous *N*-acetyl derivatives were produced with acetic anhydride in pyridine.

The IAA compounds **19**, **20**, **44**, and **45** were obtained with mediocre yields between 47 % and 57 % over two reaction steps. Here it was noticed that an intermolecular acetyl migration from the protected mannosyl moiety to the primary amino function frequently occurred as a side reaction. This migration might have already occurred during the preceding Staudinger reduction, which may have affected the yields of the subsequent iodoacetamidation. The unintentional *N*-acetylation could be clearly demonstrated by an emerging acetyl signal in the ^{1}H NMR spectrum (1.93 ppm). Furthermore, a cross peak in the HMBC spectrum was observed between the ethylene carbon next to the amine (39.5 ppm) and a carbonyl signal (170.5 ppm) assigned to the newly appearing *N*-acetyl group. However, this side reaction was comprehensibly irrelevant for the synthesis of the *N*-acetylated analogues **46** – **49**, which is why the by-products in the iodoacetamidation were possibly combined with the respective peracetylation products. If required, the *N*-acetylamide was activated under reflux in THF using 4-DMAP under strong basic conditions with sodium hydride, in order to facilitate the Boc-protection with di-*tert*-butyl dicarbonate (Di-Boc) (Scheme 3.2-6).[457] In the case of **47**, which was isolated as a side product in the synthesis of **20**, the Boc-protected *N*-acetyl compound **50** was afforded with 39 % yield. The subsequent deacetylation of the *tert*-butyl carbamate was achieved under Zemplén conditions. Also, the targeted **F1**, **F2**, **H1**, and **H2**, as well as the *N*-acetylated analogues **51**, **52**, **53**, and **54** were obtained by deacetylation of the respective *O*-acetyl protected precursors under Zemplén conditions. The target molecules were thoroughly purified by column chromatography and subsequent size exclusion chromatography and obtained in high yields. An alternative protection strategy or complete deacetylation of the mannose moiety prior to the Staudinger reduction and subsequent conversion with iodoacetic anhydride would be a viable approach to prevent the unwanted intermolecular acetyl migration.



Scheme 3.2-5: Glycosylation, amidation, and deprotection steps for the synthesis of iodoacetamide-functionalised target molecules and their *N*-acetylated analogues.

a) $BF_3 \cdot OEt_2$, acetonitrile, dry DCM, 0 °C → rt, 16 h – 18 h; **b)** PPh_3, THF, dH_2O, rt, 3 – 20 h; **c)** 2-iodoacetic anhydride, dry DCM, rt, 16 h, in the dark; **d)** Ac_2O, pyridine, rt, 16 h; **e)** NaOMe, MeOH, rt 1 – 2 h.



Scheme 3.2-6: Removal of the *N*-acetyl group from **47** after intermolecular acetyl migration by Boc protection and deacetylation under Zemplén conditions.

a) 1. NaH (60 %, 1.2 eq.), THF, rt, 1 h; 2. Di-Boc (7 eq.), DMAP (1 eq.), 3 h at 70 °C → rt, overnight; **b)** NaOMe, MeOH, rt, 5 h; **c)** Ac_2O, pyridine, rt, 16 h.

3.3 Photochromic properties of synthesised azobenzene mannosides

The photochromic properties of the target molecules, shown in Figure 3.3-1, were investigated by NMR and UV/vis spectroscopy. The respective compounds were dissolved in methanol and analysed at room temperature. *Trans* → *cis* photoisomerisation of the thermally relaxed compounds was achieved by irradiation at 365 nm (π → π* transition), whereas partial restoration of the *trans*-isomers was investigated after irradiation at 440 nm and 500 nm (n → π* transition), respectively. Among other things, UV/vis spectroscopy was used for the determination of absorption maxima $\boldsymbol{\lambda_{max}}$ of *trans*- and *cis*-isomers (conc. 0.025 – 0.035 mM). Molar extinction coefficients $\boldsymbol{\varepsilon}$ were calculated by application of the Beer-Lambert law, utilising the correlation between different sample concentrations (0.05 mM – 5.9 µM) and the absorbance of *trans*-isomers at the respective $\boldsymbol{\lambda_{max}}$ (π → π*). Photostationary states (PSSs) at the different irradiation wavelengths were also calculated by linear regression from the same plots that were employed for the calculation of molar extinction coefficients. Corresponding NMR and UV/vis spectra as well as linear regression plots are provided in the supplementary Chapter 5.5.

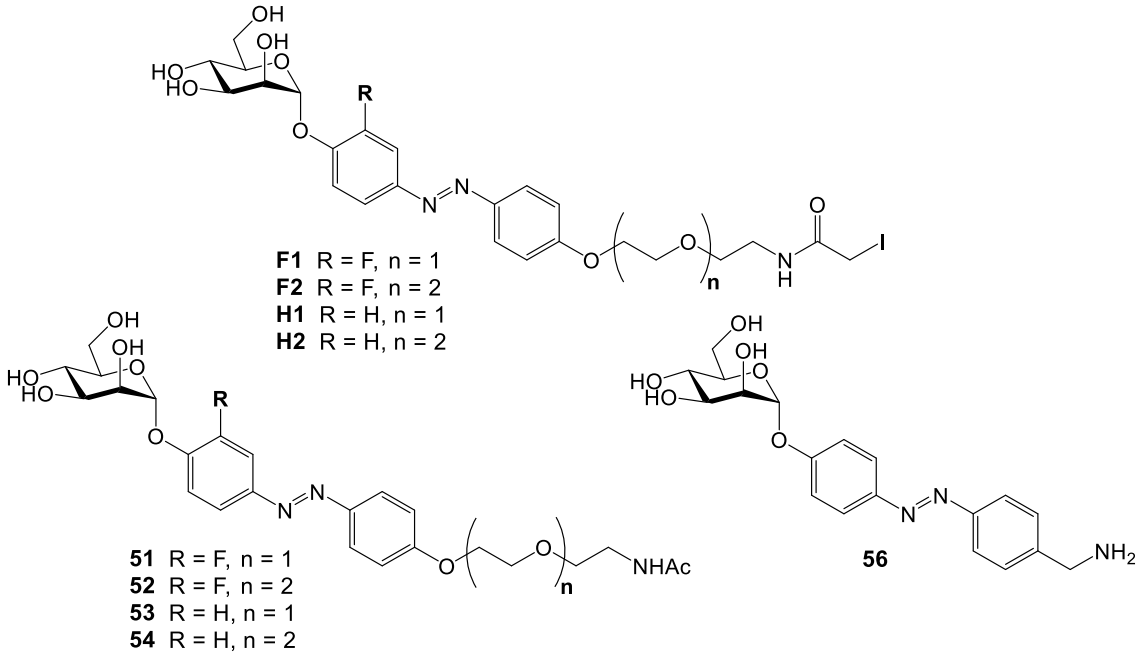


Figure 3.3-1: Structures of azobenzene and 3-fluoroazobenzene *α*-D-mannoside derivatives examined in photoirradiation studies.

The calculated PSSs after irradiation at 365 nm were further confirmed by ^{1}H and ^{19}F NMR spectroscopy either under consideration of the integrals of the H-1 signal corresponding to the mannoside unit or the aromatic fluorine signal within the azobenzene moiety, where appropriate. The anomeric proton signal was also monitored in ^{1}H NMR at 600 MHz, to investigate the kinetics of the thermal *cis* → *trans* relaxation process. The half-life $\tau_{1/2}$ was hence calculated, assuming a first-order decay of the respective *cis*-isomer. ^{1}H and ^{19}F NMR spectra for determination of the PSS_{365} and the results of thermal relaxation experiments are provided in Chapter 5.5.4 (Figures 5.4-37 – 5.4-45). The less reactive *N*-acetyl functionalised AB mannosides (**51** – **54**) were not expected to perform alkylation of cysteine residues on bacterial surface-associated proteins. Therefore, these compounds were used as inhibitors in adhesion inhibition experiments with type 1-fimbriated *E. coli* bacteria, in combination with prior photoirradiation (Chapter 3.4).

For this purpose, the determination of photochromic properties was repeated, and the parameters were accordingly adjusted to the approximate conditions during the essential incubation steps in the bacterial assays. More precisely, the designated derivatives (**51** – **54**) were gauged at 37 °C in primarily aqueous solution (PBS/methanol) on a Tecan plate reader (conc. 0.25 mM), as well as by ^{1}H NMR measurements (conc. 3.8 – 4.2 mM). The addition of 10 % v/v methanol was required to improve the poor water solubility of the AB mannoside compounds. Particularly, the *trans:cis* ratios after approximately 20 min of incubation at 37 °C were of special interest, since any relaxation, which occurs during this period, must be taken into account for the evaluation of the inhibitory properties after photoirradiation at 365 nm. In corresponding relaxation experiments, imitating the conditions in adhesion inhibition experiments as far as possible, the thermal *cis* → *trans* relaxation was therefore monitored at 37 °C by UV/vis and NMR spectroscopy. Two representative UV/vis spectra from photoirradiation and thermal relaxation experiments at 37 °C are visualised in Figure 3.3-2, using *N*-acetyl-functionalised compound **51** as an example. The corresponding results of **52**, **53**, **54**, and **56** are shown in Chapter 5.5.3 and (Figures 5.4-33 and 5.4-36) and Chapter 5.5.4 (Figures 5.4-47 – 5.4-50).

Aminomethyl azobenzene mannosides, such as compound **56** depicted in Figure 3.3-1, are known for their exceptional thermal stability of the respective *cis*-isomers. When using this AB derivative, only minimal thermal relaxation would be expected to occur during the period of the biological tests. For the optimisation and evaluation of the performed adhesion inhibition assays in combination with previous UV exposure, **56** can be regarded as an ideal test substance. Therefore, the latter compound was employed as a reference in bacterial experiments with photoirradiation (Chapter 3.4.1), and the determination of the photochromic characteristics was hence also required. The *N*-Boc- and *O*-acetyl-protected precursor molecule of **56** was synthesised according to procedures well-implemented in the Lindhorst group,[458,459] and accomplished by overall deprotection. The relevant photochemical data of all target molecules (measured in methanol and at room temperature) is summarised in the following Table 3.3-1. Additional photochromic properties (measured in PBS/MeOH, 9:1, at 37 °C) of the AB derivatives that were designated for biological inhibition assays are shown in Table 3.3-2.

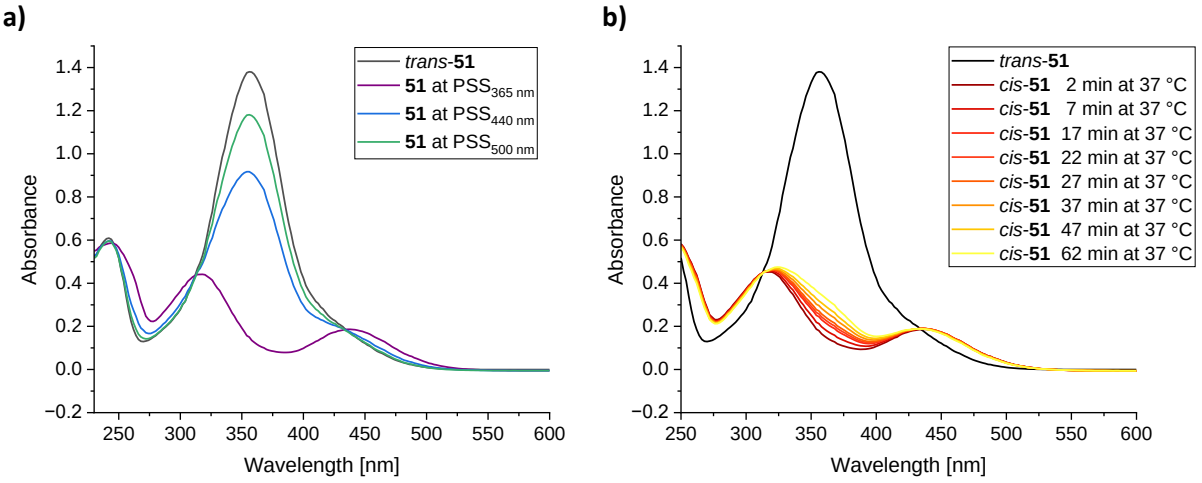


Figure 3.3-2: a) UV/vis spectrum and **b)** thermal *cis* → *trans* relaxation of **51** (0.025 mM) at 37 °C monitored by UV/vis spectroscopy on a Tecan plate reader.

The thermally relaxed sample was irradiated at 365 nm for 5 min, and the *trans*-isomer was partially restored by irradiation with 440 nm for 5 min (blue) and 500 nm for 5 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths. The thermal relaxation was monitored for 1 h at 37 °C, directly after irradiation at 365 nm for 5 min. Measured in PBS/methanol, 9:1, on a Tecan microplate reader, 100 µL/well in UV-Star® F96-well microplates (Greiner Bio-One GmbH).

Table 3.3-1: Photochemical characterisation of azobenzene mannoside derivatives.

Nº	λ_{max} **for** $\pi \rightarrow \pi^*$[a] (*trans*/*cis*) [nm]	λ_{max} **for** $n \rightarrow \pi^*$[a] (*cis*) [nm]	**PSS$_{365}$**[b,c,d] (*trans*:*cis*)	**PSS$_{440}$**[d] (*trans*:*cis*)	**PSS$_{500}$**[d] (*trans*:*cis*)	ε[e] [L · mol^{-1} · cm^{-1}]	$\tau_{1/2}$[f] [h]
F1	354/313	442	4:96 [b,d] 2:98 [c]	71:29	82:18	39257	16.57
F2	355/312	442	5:95 [b] 2:98 [c] 4:96 [d]	74:26	86:14	38294	17.71
H1	353/312	444	5:95 [b,d]	77:23	90:10	30615	17.42
H2	353/311	444	7:93 [b] 5:95 [d]	74:26	87:13	44147	9.79
51	354/312	442	5:95 [b] 4:96 [c,d]	72:28	84:16	31049	16.64
52	354/313	442	5:95 [b] 3:97 [c] 6:94 [d]	69:31	81:19	37124	17.19
53	353/312	444	5:95 [b] 6:94 [d]	72:28	84:16	33499	18.42
54	353/312	444	5:95 [b] 4:96 [d]	67:33	79:21	29924	19.17
56	341/302	437	7:93 [b] 8:92 [d]	78:22	74:26	28328	86.97

a) Wavelength of maximal absorbance (λ_{max}) for the respective $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions of *trans*/*cis*-isomers; **b)** *trans:cis* ratios at PSS$_{365}$, determined from H-1 integrals corresponding to the mannoside unit in ^{1}H NMR spectra; **c)** PSS$_{365}$, determined from the fluorine integrals in ^{19}F NMR spectra; **d)** PSS$_{365}$, PSS$_{440}$, and PSS$_{500}$, determined by linear regression from UV/vis absorbance at the respective λ_{max} vs. sample concentration of each compound; **e)** molar extinction coefficient of *trans*-isomers, determined at the respective λ_{max} of each compound; **f)** half-life of the *cis*-isomers from H-1 integrals corresponding to the mannoside in ^{1}H NMR spectra. NMR and UV/vis spectra recorded at 298 K in CD_3OD or in methanol, respectively.

Table 3.3-2: Photochemical characterisation of azobenzene mannoside derivatives designated for adhesion inhibition experiments at 37 °C.

Nº	λ_{max} **for** $\pi \rightarrow \pi^*$[a] (*trans*/*cis*) [nm]	λ_{max} **for** $n \rightarrow \pi^*$[a] (*cis*) [nm]	**PSS$_{365}$**[b] (*trans*:*cis*)	**PSS$_{440}$**[b] (*trans*:*cis*)	**PSS$_{500}$**[b] (*trans*:*cis*)	***trans*:*cis* ratio**[c,d] (∼ 20 min, 37 °C)
51	356/318	438	4:96	65:35	87:13	8:92 (17 min) [c] 9:91 (14 min) [d]
52	356/318	438	4:96	64:36	88:12	9: 91 (20 min) [c] 11:89 (40 min) [d]
53	354/314	436	4:96	66:34	84:16	7:93 (22 min) [c] 6:94 (28 min) [d]
54	356/314	439	5:95	63:37	83:17	7:93 (20 min) [c] 7:93 (25 min) [d]
56	342/306	430	8:92	65:35	75:25	8:92 (22 min) [c] n. d.[e]

a) Wavelength of maximal absorbance (λ_{max}) for the respective $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions of *trans*- and *cis*-isomers; **b)** PSS$_{365}$, PSS$_{440}$, and PSS$_{500}$, determined by linear regression from UV/vis absorbance at 365 nm vs. the sample concentration of each compound; **c)** *trans*:*cis* ratios after ∼ 20 min (accurate time in brackets) relaxation at 37 °C, monitored by UV/vis spectroscopy; **d)** *trans*:*cis* ratios after 20 min (accurate time in brackets) relaxation at 37 °C, determined from H-1 integrals corresponding to the mannoside unit in ^{1}H NMR spectra. **e)** compound **56** was not investigated in NMR due to the extraordinary slow thermal relaxation progression. NMR spectra were recorded in D_2O/CD_3OD, 9:1, and UV/vis spectra were recorded in PBS/methanol, 9:1, on a Tecan plate reader at 310 K. The addition of 10 % v/v methanol was required due to relatively poor solubility in water.

As shown in the representative example of a recorded UV/vis spectrum (Figure 3.3-2a), all main characteristic absorption bands that would be expected for AB derivatives were observed: The *trans*-isomer exhibits the intense π → π* absorption band of the diazene unit around 354 nm, accompanied by the symmetry-forbidden n → π* transition (~ 440 nm), and the π → π* band of the benzene ring electrons. The main π → π* band is red-shifted in comparison with unsubstituted AB due to the slightly electron-donating character of the *O*-ether bridges. As a result of the bathochromic π → π* shift leading to partial spectral overlap, the n → π* band was only poorly resolved, and the respective $\boldsymbol{\lambda_{max}}$ could therefore not be determined for the *trans*-isomers. The absorption band corresponding to the π → π* transition of the *cis*-isomers revealed the expected hypochromic and hypsochromic shift (in relation to the *trans*-isomers) towards ~ 312 nm. The very slightly blue-shifted n → π* transition band now became symmetry-allowed, resulting in an observable hyperchromic effect.

The obtained absorption maxima of the target molecules, shown in Figure 3.3-1, are very similar. Notably, no bathochromic shift of the π → π* transition of the *cis*-isomers was observed upon mono-fluorination in the *meta*-position to the diazene unit when comparing the structurally analogous fluoroazobenzene (**F1**, **F2**, **51**, and **52**) and non-fluoro derivatives (**H1**, **H2**, **53**, and **54**). This stands in clear contrast to the findings about mono-*m*-halogen substituted AB derivatives, published in 2020 by Das and co-workers.[204] The authors demonstrated that mono-fluorination in the *m*eta-position to the diazene bond results both in increased thermal stability and a significant bathochromic shift of the π → π* band of the *cis*-isomer.[204] As already mentioned in Chapter 1.2.1, the red-shift of about 50 nm resulted in an extensive spectral overlap of the π → π* transition bands of both photoisomers.

In contrast to (multiply) *ortho*-fluorinated AB derivatives, the n → π* transition band of the *trans*-isomers was also not bathochromically shifted. Seemingly, mono-*m*-fluorination does not sufficiently reduce the electron density around the diazene bond to result in an observable influence on the n-orbital energy. It can certainly be assumed that the negative inductive effect of the mono-fluorine substituent is overcompensated by the electron-donating effect of the *O*-ethers in the compounds investigated here. Moreover, inconsistent with the work of Das and colleagues, the thermal stability of the *cis*-fluoroazobenzene derivatives was not enhanced, as can be deduced from the decrease in $\boldsymbol{\tau_{1/2}}$ by approximately 1 h. An extension of the linker moiety from DEG to TEG, on the other hand, results in a slightly pronounced thermal stabilisation, which is consistent with earlier findings.[420] Only target compound **H2** somewhat differed from this observation, as it featured a significantly shorter half-life than the other AB mannosides investigated in this work. In order to exclude any experimental mistake in the measurement or an individual problem of the measured sample, the determination of the thermal half-life was repeated with a new synthetic batch of compound **H2**, but a very similar result was obtained.

The diverse approaches for the determination of isomer ratios at PSS_{365} yielded largely consistent results, although slightly lower ratios were determined using ^{19}F NMR spectroscopy. Nevertheless, this demonstrates the good applicability of ^{19}F NMR for this purpose. Overall, the obtained photochemical data of the OEG-linked AB derivatives – particularly the absorption maxima, the almost quantitative *trans* → *cis* photoconversion (PSS_{365}), as well as the sufficient thermal stability of the *cis*-isomers ($\boldsymbol{\tau_{1/2}}$ ~ 15 – 20 h) – are consistent with those reported for structurally similar, literature-known AB mannosides and mannobiosides, which are linked via alkyl[412] or OEG ether bonds.[409,460]

In comparison, the aminomethyl AB mannoside **56** exhibited slightly blue-shifted absorption maxima in both π → π* and n → π* transitions, a slightly lower efficiency of photoconversion at PSS_{365}, but offered a 4 – 5 times longer half-life of the *cis*-isomer at room temperature ($\boldsymbol{\tau_{1/2}}$ ~ 87 h). At 37 °C, **56** displayed almost no observable *cis* → *trans* relaxation within the period of a typical inhibition experiment (Figure 5.5-50). Due to this excellent thermal stability, **56** was ideally suited to be used as an inhibitor and as a reference for the optimisation of the inhibition assay with previous photoexposure (Chapter 3.4).

As visualised in Figure 3.3-2b, the OEG-linked AB derivatives relaxed comparably fast in the adapted thermal relaxation experiment (more data is provided in Chapter 5.5.4, Figures 5.4-46 – 5.4-49). In the case of compound **51** for instance, an advanced relaxation would be expected to occur during a usual inhibition experiment (45 min at 37 °C, expected *trans*:*cis* ratio ~ 14:86). However, the *trans*:*cis* ratios of all photoswitchable inhibitors investigated in this work were almost identical after a shortened incubation time (20 min of relaxation at 37 °C), resulting in an expected proportion of *trans* between 7 – 9 % (Table 3.3-2). Despite the faster relaxation progress of the OEG-linked AB derivatives compared to the aminomethyl AB mannoside **56**, this was compensated by the slightly poorer PSS_{365} ratio of the latter, overall resulting in very similar isomeric *trans*:*cis* ratios.

3.4 Biological evaluation of photosensitive *N*-acetylated analogues

The *N*-acetyl-functionalised AB mannoside derivatives **51** – **54** and the aminomethyl AB mannoside **56** were tested in bacterial adhesion inhibition experiments, both with and without prior photoirradiation. In this assay, the detection of an inhibitory effect indicates the recognition of the mannosides by the bacterial lectin FimH. Furthermore, the influence of photoconversion on bacterial binding was assessed by performing the assay in combination with pre-irradiation of the photoswitchable inhibitors. The non-photoswitchable 2-Fluoro-*N*-acetyl-phenylmannoside **57** and the analogous non-fluorinated H-derivative **58** were subjected to the same irradiation and assay conditions and employed as control compounds. All inhibitors used in this context are summarised in Figure 3.4-1.

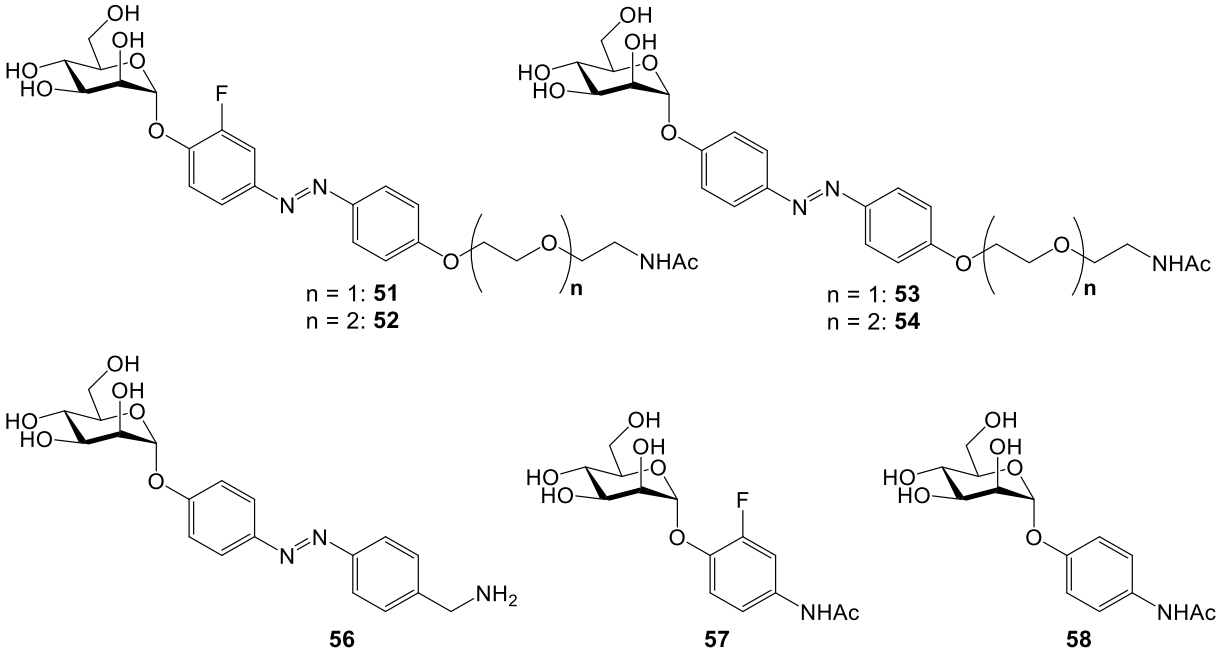


Figure 3.4-1: *N*-Acetyl-protected azobenzene mannoside derivatives and *N*-acetyl-phenyl mannoside derivatives investigated in adhesion inhibition experiments with photoirradiation.

The influence of ligand orientation on bacterial binding by exploiting the geometrical change occurring concomitantly with *trans* → *cis* photoisomerisation of azobenzene mannoside derivatives has been repeatedly demonstrated by the Lindhorst group. They have investigated the bacterial adhesion on different photoswitchable surfaces, such as AB mannoside derivatives incorporated into glyco-SAMs (self-assembled monolayers) on a gold surface,[412,413] and mono- and trivalent AB mannoside clusters on the surface of human microvascular endothelial cells (HMEC-1).[414,415] M. Hartman and V. Chandrasekaran also tested *trans*- and *cis*-azobenzene mannosides[407,408] and mannobiosides[410] as inhibitors of mannose-specific bacterial adhesion in solution. L. M. Friedrich recently evaluated orthogonally photoswitchable heterobivalent AB glycoclusters, and also considered a homobivalent analog as well as the corresponding AB glycoside antennas.[411] The consistent results of these studies were a (more or less pronounced) better adhesivity and stronger inhibition, respectively, of the *trans*-isomers compared to the *cis*-isomers. In the case of the surface-immobilised azobenzene mannosides, this was attributed to the orientation of the glycoligand and to steric effects, which affect the accessibility of the ligands for the lectin.[461]

When AB mannoside derivatives were employed as freely diffusing ligands in aqueous solution, mainly steric effects were assumed for the interpretation of the results. In addition, the poorer affinity of the *cis*-isomers could be explained by reduced π-π interactions with the FimH tyrosine gate due to the changed geometry and the higher polarity of the aglycone, which may also increase the energy barrier for partial desolvatation.[189] Recently, photoswitchable AB galactosides have also been designed and tested as photochromic inhibitors against the soluble lectin LecA from *Pseudomonas aeruginosa*.[462] However, in this case, *trans* ↔ *cis* photoisomerisation had virtually no or only a small effect on LeA binding, as clearly demonstrated in ITC experiments.

The well-implemented adhesion assays, which are frequently used in the Lindhorst group, examine the adhesion of an exclusively type 1-fimbriated *E. coli* strain (SAR18 PKL1162), to a mannosylated surface, optionally in the presence of FimH antagonists. The strain SAR18 PKL1162 expresses the green fluorescent protein (GFP) as a reporter protein. Monitoring of adhering bacteria can thus be easily achieved by fluorescence measurement ($\lambda_{ex/em}$ = 483/535 nm). The assay is usually performed in black, flat-bottomed 96-well polystyrene microplates, which are coated with the mannan-polysaccharide from *S. cerevisiae.* Yeast mannans represent a class of *N*- and *O*-glycans that mainly consist of *α*(1→6)-linked *α*-D-mannopyranosyl units that are frequently highly branched with *α*(1→2)- and *α*(1→3)-linked side chains.[463,464] Particularly the mannan from *S. cerevisiae* has demonstrated good binding properties for mannose-specific lectins and supports bacterial adhesion.

All originally genomic elements essential for the assembly of type 1 fimbriae are encoded on the plasmid pPKL174. The latter was transformed into *E. coli* SAR18, resulting in the strain PKL1162.[465] SAR18-derived strains feature constitutive expression of the gfp_{mut3} variant under the control of a strong, artificial promoter, which was chromosomally integrated using a vector-based recombination system relying on the site-specific integrase from bacteriophage λ.[465,466] In the competitive experimental setup, the adhesion step is performed under the addition of serially diluted mannoside derivatives serving as FimH inhibitors.[467] Hence, the concentration at which half-maximal inhibition (IC_{50}) occurs can be deduced as the inflection point of the sigmoidal inhibition curves resulting from semi-logarithmic plotting. Since experience has shown that the IC_{50} value can fluctuate by the order of approx. one magnitude, methyl *α*-D-mannopyranoside (MeMan) serves as an internal reference compound in each experiment. The more reproducible quotient of the IC_{50} of MeMan and the respective inhibitor employed is termed relative inhibitory potency (RIP), which is used as a general benchmark to express the inhibitory effect of each mannoside or inhibitor, respectively.

3.4.1 Adhesion and adhesion inhibition experiments with and without UV exposure

The antiadhesive properties of the mannoside derivatives mentioned above were investigated in the standard assay[467] using thermally relaxed samples. The obtained RIPs are compiled in Table 3.4-1. Exemplary results from the adapted adhesion inhibition assay with **51** and **54** used as photoswitchable inhibitors are visualised in Figure 3.4-2. Further adhesion inhibition plots of the individual antagonists and the respective inhibitory data are provided in Chapter 5.8.3 (Figures 5.7-11 – 5.7-17, and Tables 5.7-6 – 5.7-9).

In addition to the experiment under standard conditions, a protocol for an adhesion inhibition assay coupled with prior photoirradiation was implemented, in order to evaluate the influence of the respective azobenzene configuration on bacterial recognition. This approach is similar to the use of photochromic ligands (PCLs) in the context of photopharmacological applications. In contrast to the approach pursued by M. Hartman and V. Chandrasekaran, the inhibitors were not irradiated until after the transfer and serial dilution on the mannan-coated microplates. In other words, the UV exposure of the inhibitors was performed on the readily prepared microplate directly before the bacterial suspension was added. With this procedure, significantly more reproducible data and results of higher quality were obtained than with UV treatment of the inhibitors prior to their application to the microplate. In this way, both a more advanced thermal relaxation as well as a possible temperature-dependent volume change, which could lead to inaccuracies during liquid handling, can be avoided. The obtained RIPs of either thermally relaxed (i.e. not irradiated) samples (**51** – **58**) and the same compounds after UV exposure at 365 nm are compiled in Table 3.4-1.

Table 3.4-1: Antiadhesive properties of photoswitchable *N*-acetyl-protected azobenzene mannosides and of control compounds from adhesion inhibition experiments involving photoexposure.

Inhibitor	RIP[a] Standard assay	RIP *trans*-isomer [a,b] Not irradiated	RIP *cis*-isomer [a,c] Irradiated at 365 nm	IC_{50} ratio [d] relaxed/irradiated
51 [e]	2240.6 (± 499.7)	1484.9 (± 445.5)	279.6 (± 116.3)	0.23 (± $1.5 \cdot 10^{-1}$)
52	1110.6 (± 127.5)	1668.3 (± 588.6)	413.7 (± 149.9)	0.26 (± $1.1 \cdot 10^{-1}$)
53 [e]	1095.8 (± 129.0)	815.8 (± 498.9)	204.2 (± 8.0)	0.31 (± $2.0 \cdot 10^{-1}$)
54	614.9 (± 89.3)	499.8 (± 133.4)	127.6 (± 15.2)	0.27 (± $7.4 \cdot 10^{-2}$)
56	332.8 (± 40.2)	493.8 (± 126.3)	104.3 (± 23.3)	0.23 (± $1.3 \cdot 10^{-1}$)
57	522.7 (± 29.4)	437.9 (± 81)	593.7 (± 133.6)	1.42 (± $5.7 \cdot 10^{-1}$)
58	270.8 (± 91.3)	207.4 (± 4.2)	231.7 (± 56.6)	1.12 (± $2.8 \cdot 10^{-1}$)

Average IC_{50}, RIP and IC_{50} ratios from 2 – 6 individual experiments. The respective ± standard deviations are given in brackets. **a)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **b)** The respective sections of the microplate containing the diluted inhibitors were wrapped in aluminium foil to avoid UV exposure during photoirradiation. **c)** The respective parts of the microplate containing the diluted inhibitors were irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension. **d)** Mean ratio of IC_{50} values of the respective inhibitors with/without prior UV exposure $\left(IC_{50}\,\text{ratio} = \frac{IC_{50}\ \text{not irradiated}}{IC_{50}\ \text{irradiated}}\right)$. **e)** The IC_{50} and RIP values are probably systematically elevated due to poor water solubility of the DEG-linked azobenzene mannosides.

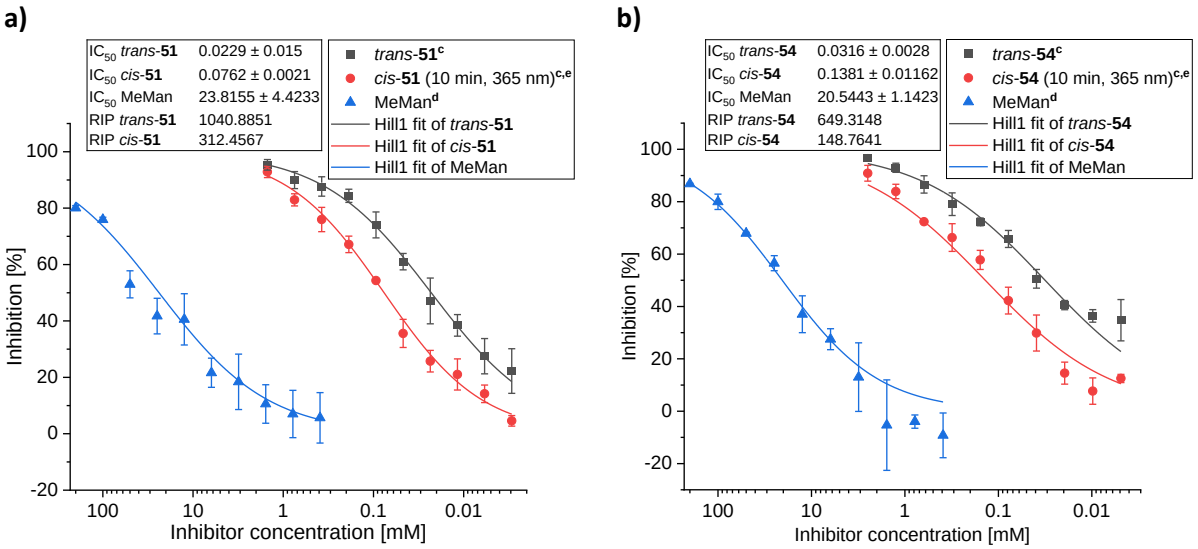


Figure 3.4-2: Exemplary results from adhesion inhibition experiments with *N*-acetyl-protected azobenzene mannoside derivatives **a) 51** and **b) 54**, following UV exposure.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **c)** triplicate or **d)** duplicate measurements. **e)** Inhibitor solutions were applied as 1:2 dilution series onto the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

Both approaches – the standard assay as well as the adapted assay with photoirradiation – revealed a highly pronounced antiadhesive behaviour for all azobenzene mannoside derivatives investigated in this work. This applied in particular to the OEG-linked AB derivatives, as can be deduced from the comparably high RIP values (RIPs > 600 in the standard assay). In the standard assay, the observable inhibitory effect of the DEG-linked AB mannosides **53** and especially **51** was extraordinarily strong (RIPs > 1000 – 2200), and both calculated RIPs were about twice as high as for the corresponding TEG-bridged analogues **54** and **52**. It should be noted that the DEG-linked derivatives were still difficult to dissolve, despite the addition of 10 % v/v methanol to the employed buffer as a solubilising additive (see also Chapter 3.4.2). It cannot be precluded with certainty that the substances partially precipitated in the course of the experiment, which may have led to a systematic error due to extensive blocking of the mannan layer. Therefore, the observable RIPs of the DEG-linked derivatives **51** and **53** may not accurately reflect the actual inhibitory effect of these compounds.

Remarkably, *meta*-fluorination of the azobenzene unit (relative to the diazene bond) also resulted in a significantly enhanced lectin affinity, which was evidenced by a roughly twofold increase in the RIP (standard assay). This observation was confirmed with the non-photoswitchable 2-fluoro-phenyl mannoside **57** and the non-fluorinated analog **58**, thereby reflecting the same gradation in their inhibitory potency. As expected, the positive impact on FimH inhibition by implementation of an *ortho*-substituent directly adjacent to the phenyl group is thus in good agreement with the findings made by Ernst and co-workers. As beforehand mentioned in the Chapters 1.3.3 and 2, these authors investigated the same relative position on biphenyl aglycones, demonstrating a substantial affinity increase in ITC experiments.[378]

The same trend was observed for the *trans*-isomers (i.e. the non-irradiated samples), in the adapted assay with UV irradiation. This is despite the fact that the obtained RIPs are not directly comparable to those of the standard assay due to the modified experimental conditions. Consistent with the literature mentioned above, UV exposure of the ligands and the associated *trans* → *cis* photoisomerisation resulted in a dramatic reduction in the inhibitory effect.[407,408,410,411] The difference between irradiated and non-irradiated samples can be recognised at first glance, for instance, in the inhibition curves obtained with **51** and **54** (Figure 3.4-2). The non-photoswitchable control compounds, **57** and **58**, on the other hand, demonstrated no – or even the opposite effect (i.e. a slight increase in inhibitory potency) as a result of photoirradiation (Figures 5.7-9 – 5.7-10 in Chapter 5.8.3). The latter can presumably be attributed to the emerging temperature increase in the exposed areas of the microplate. To illustrate the effect of UV irradiation, (or photoconversion, in the case of photoswitchable inhibitors), more simply, the quotient from the IC_{50} values of the unexposed and exposed test series on each microplate was calculated. All photoswitchable mannosides featured a mean IC_{50} ratio of smaller than 1 (between ∼ 0.2 – 0.3), indicating that UV exposure significantly reduced their inhibitory potency. In other words, for photoswitchable mannosides, an IC_{50} ratio below 1 directly reflects the stronger inhibitory effect of the unexposed *trans*-isomers compared to their respective *cis*-isomeric counterparts. In contrast, the IC_{50} ratios of the non-photoswitchable control substances were slightly larger than 1 (∼ 1.1 – 1.4), suggesting no effect or even a slight increase in inhibition after UV treatment.

During the incubation of the bacterial suspension in the presence of the irradiated inhibitors (in the standard assay: 45 min at 37 °C), a significant decrease in the *cis*-isomer fraction can be expected due to thermal relaxation. Therefore, a reduction of the incubation time was reasonably implemented in the adapted assays with UV exposure, to minimize the thermal relaxation of the *cis*-isomer during the experiment. Moreover, owing to the incomplete PSS_{365} achieved upon irradiation, a residual amount of the *trans*-isomer is invariably present. To limit the ratio of the *trans*-isomer in the irradiated samples as much as possible, incubation times between 10 – 45 min after photoexposure were tested. Based on the photochromic properties of the tested compounds (Chapter 3.3), a delay time[2] of overall 20 min at 37 °C after irradiation was selected for the determination of the antiadhesive characteristics shown in Table 3.4-1. After this period, the *cis*-ratio of all photoswitchable compounds amounts to approximately 91 – 93 %. This was considered sufficiently high to reliably reflect the inhibitory effect of the *cis*-isomer. A longer incubation time frequently resulted in higher quality results, which, however, were concurrently compromised by a slight but noticeable convergence of the inhibition curves of the *cis*-isomers to those of the non-irradiated samples (i.e. the *trans*-isomers) due to the progressive thermal relaxation (see Figures 5.7-4 – 5.7-8 in Chapter 5.8.3). The only exception was the extraordinarily stable *cis*-aminomethyl AB mannoside **56**, which, even after 20 h of relaxation at room temperature, still exerted a distinctly weaker inhibitory effect than the *trans*-isomer (Figure 5.7-8d in Chapter 5.8.3). After such a long incubation period, any influence on bacterial inhibition resulting from the UV exposure-induced temperature increase can be excluded. The observed difference in inhibition between the exposed and unexposed solutions of **56** must therefore be mainly due to the different isomeric configurations of the AB unit.

[2] The total delay time of 20 min accounts for 15 min of incubation at 37 °C, as well as ∼ 5 min, e.g. required for pipetting of the bacterial suspension and washing steps.

3.4.2 Control experiments

In order to preclude the possibility that the observed inhibition might be mainly or partly ascribable to a possible bactericidal or bacteriolytic effect, a toxicity test was performed. For this purpose, the bacterial viability was evaluated by growth on LB agar plates after co-incubation with the inhibitors under the usual conditions of an adhesion inhibition experiment (45 min, 37 °C). No toxic effects of any of the inhibitors on bacterial growth were observed within the concentration range employed in the biological assays (data not shown). A significant impact of UV irradiation (5 – 30 min at 365 nm) on the mannan-coated polystyrene surface prior to the addition of bacterial suspension could likewise be excluded by a corresponding control experiment (Figure 5.8-18 in Chapter 5.8.3).

The mediocre water solubility of the DEG-linked AB mannosides **51** and **53**, in particular, should be improved by adding small amounts of organic solvents, in order to minimise any associated interference with the experimental results. Additives such as methanol or dimethyl sulfoxide (DMSO) are commonly regarded as uncritical when used in low concentrations.[468,469] Nevertheless, the tolerance of both solvents was tested in the context of the adhesion inhibition assay with varying solvent concentrations and with and without the use of MeMan as an inhibitor (Figure 5.8-19 in Chapter 5.8.3). The synergistic impact of methanol in the presence of MeMan (Figure 5.8-19b) on the sigmoidal inhibition behaviour was less pronounced compared to that of DMSO (Figure 5.8-19a). A similar trend was observed for the bacterial adhesion in the absence of MeMan: The inhibitory effect, observed only in the presence of high methanol concentrations, was already completely attenuated at solvent concentrations below 25 % v/v. The effect of methanol was thus eliminated within two dilution steps, in contrast to DMSO, which required a concentration below 12.5 % v/v (i.e. three dilution steps) to omit the effect of the organic solvent. As further demonstrated in Figure 5.7-20, addition of 10 % v/v methanol to phosphate-buffered saline (PBS) had only a minimal impact on the observed inhibition profiles and calculated RIP values. Consequently, methanol was chosen as a solubilising additive at a concentration of 10 % v/v in all experiments presented in this chapter.

3.5 Expression and purification of FimH variants

In order to enable both the chemical modification of FimH with photoswitchable ligands and the functional validation of the engineered lectins, the expression and purification of suitable, C-terminally truncated FimH variants represented a key prerequisite. In particular, FimH constructs carrying strategically positioned cysteine residues (N136C, S139C) were required to allow site-specific bioconjugation with IAA-functionalised azobenzene mannosides. In addition, the wild type FimH (WT) and the proline variant (R60P) were expressed to serve as functional reference constructs. The former represents the permanently stabilised high-affinity conformation (HA),[373] which features a considerably elevated ligand-binding affinity compared to the fimbriae-associated lectin.[336,391] The R60P variant, on the other hand, was included to introduce conformational plasticity into the lectin domain, as the *apo*-lectin prefers the low-affinity (LA) state, whereas ligand binding induces adoption of the HA conformation.[397] For the expression of full-length (FL)-FimH, the presence of the N-terminal donor strand peptide of FimG[337] or even the chaperone FimC.[315] would be required to stabilise the incomplete immunoglobulin-fold of the pilin domain (FimH-PD), which in turn usually serves as an allosteric auto-inhibitor.[390] By contrast, the isolated lectin domain (FimH-LD) adopts a stable jelly-roll fold that allows independent expression without the requirement for donor strand complementation (DSC).

This chapter describes the homologous expression and purification of the aforementioned C-terminally truncated FimH variants (WT, N136C, S139C, and R60P), based on two different expression plasmids (pPKL241 and pFimH-His_6-NT-trc). The respective gene products are each equipped with a C-terminal hexahistidine (His_6) tag, to allow for facilitated purification of the target proteins. Both approaches are described in the following two chapters, each using the respective N136C variant as a representative example. Thanks to the kind support of Prof. A. Scheidig, a large part of the genetic engineering and the preparative-scale expression of the FimH variants could be performed in the appropriately equipped laboratory of the Structural Biology group at Kiel University.

First expression attempts were essentially based on a protocol by Schembri and colleagues.[470] The corresponding expression plasmid pPKL241 was kindly provided by Prof. S. Knight (Uppsala University, Sweden). It encodes the truncated lectin domain (aa 1 – 156) originating from *E. coli* K12, attached to a C-terminal His_6 tag to facilitate purification by affinity chromatography. The *fimH* sequence was primarily subcloned from pPKL4 into a pQE30 vector and includes a codon-optimised N-terminal signal peptide.[471] The latter is responsible for the transport and accumulation of the target protein in the periplasm, since all other elements of the *fim* gene cluster are not present. In this work, all truncated FimH variants produced with pPKL241-based expression systems are referred to as $FimH_L$.

Unfavourably, the provided construct was damaged due to an extensive 32-bp deletion within the hybrid T5/*lac*O promoter sequence (Figure 3.5-1). The promoter is partially derived from bacteriophage T5 and is provided on the pQE30 vector for high-level expression of recombinant proteins, which is induced by the addition of isopropyl *β*-D-1-thiogalactopyranoside (IPTG). As a consequence, the *σ*-subunit of the RNA polymerase is unable to bind and transcription of the target sequence is not possible. The target protein thus cannot be obtained with this kind of defective construct. The problem of partial promoter deletion, which regularly occurs within pQE30-based constructs, was first described and investigated by Kawe and co-workers.[472] The authors reported that the lack of the *lac* repressor gene (*lacI*) on the vector was closely associated with promoter deletion, which was only observable directly after transformation of the respective plasmids.

Basal expression of the target proteins seems to occur under control of the strong, viral promoter even in the absence of IPTG. During the recovery step in cultivation medium in the transformation protocol, this basal expression presumably leads to immense resource stress for the recently transformed cells. Obviously, the transformants tend to avoid this, potentially by replication slippage between tandem repeats within the promoter region during plasmid replication, resulting in partial sequence deletion.[473] It is conceivable that the deleted trasformants thus receive a clear selection advantage over the non-deleted cells as the deletion has recurred repeatedly in the course of this work.

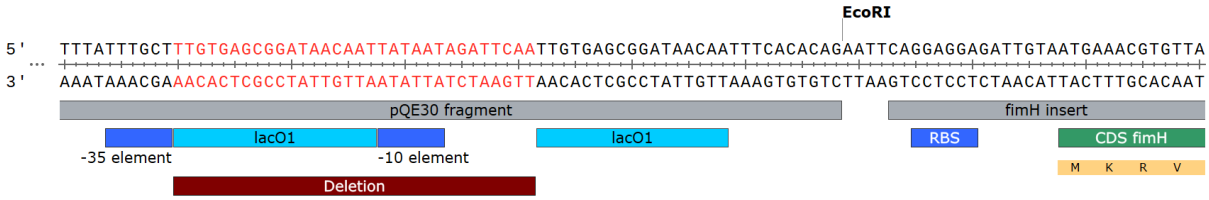


Figure 3.5-1: Partial deletion of the T5/*lac*O promoter that frequently occurred after transformation of pPKL241-based constructs used in this thesis.

The red area indicates the 32-bp deletion, including one of the *lac*O1 operator sites and the –10 element (Pribnow-Schaller Box).[474,475] RBS: ribosome binding site; CDS: coding DNA sequence; EcoRI: first restriction endonuclease isolated from *E. coli* RY13.[476] The –35 and –10 elements are highly conserved promoter sequences essential for RNA polymerase binding and transcription initiation in bacteria.[477] The designations –35 and –10 denote the positions of these promoter elements in base pairs upstream of the transcription start site.

The deletion in pPKL241 is mainly responsible for the failure of all our previous $FimH_L$ expression attempts. A reconstruction of the expression plasmid was thus required and achieved by reintegrating the *fimH-his*$_6$ sequence from the deficient construct into a new, intact pQE30 vector. A detailed description of the subcloning protocol is given in Chapter 5.8.1. Subsequent site-directed mutagenesis for the purpose of single amino acid exchange (N136C, S139C, R60P) was generally performed according to the well-known QuikChange™ method (Chapters 5.7.6 and 5.8.2, as well as Table 5.7-8).[478]

3.5.1 Expression and purification of $FimH_L$ variants with pPKL241-based expression systems

Largely following the protocol by Schembri et al., the first expression trials of the WT variant and the cysteine variants were performed in *E. coli* HB101, cultured overnight in 5 L LB medium. Induction was generally performed with the addition of 1 mM IPTG once the required optical density (OD_{600} = 0.5) was reached. Initially, the harvested cells were mechanically disrupted using a high-pressure homogeniser, followed by enrichment of the target protein via immobilised metal affinity chromatography (IMAC) and further purification by size exclusion chromatography (SEC). Due to the only moderate expression success, especially for the production of the cysteine variants, the expression strain was changed to *E. coli* SG13009 and M15, both of which are recommended for the overexpression of recombinant proteins with pQE-based plasmids by the manufacturer (Qiagen). Preliminary analysis of the obtained protein species by liquid chromatography-tandem mass spectrometry (LC-MS/MS), performed by Philipp Kaulich (Institute of Experimental Medicine, Kiel University), revealed a truncation of the C-terminal His_6 tag by 3 to 5 histidine residues, as well as the presence of numerous non-tryptic peptide fragments (Chapter 5.8.8, Table 5.8-13). Assuming a severe proteolytic degradation of the target protein during the purification process or even already during periplasmic accumulation, further expression experiments were performed in the presence of a protease inhibitor cocktail. In addition, an expression strain deficient in the outer-membrane protease OmpT (*E. coli* C43(DE3)) was employed, following the protocol by Wellens and co-workers.[479]

However, a significant improvement in the expression efficiency and purification success could ultimately only be achieved with the use of the *E. coli* strain HM125, which was kindly provided by Prof. Rudi Glockshuber (ETH Zürich, Switzerland). HM125 is optimised for high-level expression and accumulation of periplasmic proteins and has established itself as the most popular expression strain for recombinant FimH in recent years.[337,376,397,480] On the one hand, this can be attributed to a mutation in the σ^{32} subunit of RNA polymerase (*rpoH15*),[481] which affects the recognition of the conserved -35 element within the bacterial promoter sequence and results in significantly increased expression levels of secreted proteins. On the other hand, HM125 features a deficiency in the widely conserved periplasmic serine endoprotease DegP,[482] which is known to be involved in the degradation of fimbrial subunits when the corresponding periplasmic chaperones are absent.[314] At low temperatures (≤ 22 °C), DegP solely acts as a chaperone itself but undergoes an increasing functional shift towards proteolytic activity at elevated temperatures (~ 32 – 42 °C) and thus actually represents a heat shock protein.[483] As such, it is responsible for the removal of un- or misfolded proteins that accumulate within the periplasm in response to various stress conditions[484] and it is thus essential for survival at elevated temperatures.[485] Consequently, *degP* mutants grow poorly at 37 °C and are susceptible to cell lysis,[482,483] which is why *E. coli* HM125 was always cultured at a maximum temperature of 30 °C.

After expression and harvest by centrifugation, the target proteins were released from the pelleted cells by periplasmic extraction, using the membrane-disrupting antibiotic polymyxin B in the presence of ethylenediaminetetraacetic acid (EDTA).[486] The latter was removed by extensive dialysis prior to IMAC, in order to avoid undesired loss of Ni^{2+} ions from the affinity column. Compared to mechanical cell lysis, periplasmic extraction usually causes lower levels of contamination with cytosolic proteins, which is favourable for purification (Chapter 5.8.4, Figure 5.8-21). However, since the *eda*/*rpoH15* double mutation in HM125 is associated with facilitated cell lysis in the stationary growth phase,[482] the periplasmic extracts usually contained moderate amounts of cytosolic proteins.

The clarified extracts were subjected to IMAC for enrichment of the His_6-tagged target proteins, followed by gel filtration for the removal of the eluent (imidazole) and other contaminants. The respective chromatograms and the evaluation of the purification process from subsequent sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) are exemplified here by the $FimH_L$ variant N136C in Figures 3.5-2 – 3.5-4. Experimental details for the purification of other $FimH_L$ variants, including the chromatograms from IMAC and SEC, as well as the corresponding SDS-PAGE results are provided in Chapter 5.8.5.

As illustrated in Figure 3.5-2, unbound and weakly adsorbing proteins were removed from the Ni^{2+} sepharose column (column volume, CV = 5 mL) after sample application, by washing with 10 CV binding buffer and a further washing step with 59 mM imidazole (Wash I, 5 CV). The target protein was eluted in a linear imidazole gradient from 60 to 500 mM throughout 14 CV. A successful elution was indicated by a temporary increase in UV absorbance at 280 nm starting at approx. 70 mM imidazole (under consideration of the dead volume). Subjecting the respective samples from IMAC to a reducing SDS-PAGE clearly revealed an overexpression band at ~ 18 kDa, representing the target protein (17.8 kDa, indicated with the black arrowhead in Figure 3.5-3a). However, minor contaminants were also observed in the main peak fractions of the chromatogram (highlighted in grey), with migration bands at ~ 45 and 27 kDa. The lane corresponding to the periplasmic extract (**PE**) indicated only a moderate level of target protein expression compared to other protein species in the sample. In general, no overloading of the IMAC column was observed, which could result in yield loss, although this is only marginally visible in the example illustrated here (cf. Figures 5.7-23 and 5.7-27 in Chapter 5.8.5 for a clearer staining and visibility of the flow-through samples). The fractions containing the target protein (fractions 4 – 8) were collected and concentrated to a small volume in a centrifugal filter. A minimal sample volume is beneficial for the success in the final purification step via gel filtration (CV = 120 mL, equilibrated with PBS). The respective SEC chromatogram is depicted in Figure 3.5-4. The main elution peak (highlighted in grey) emerged at a retention volume between ~ 70 – 80 mL, which corresponds to the successfully purified target protein. A subsequent SDS-PAGE (Figure 3.5-3b) confirmed the efficient removal of contaminants from the pooled fraction (**P**). The respective fractions (9 – 15) were once again collected, concentrated to a small volume, aliquoted and finally stored at −80 °C. The target protein $FimH_L$ N136C was obtained with a total yield of 2.01 mg (0.4 mg/L expression volume). The result from an SDS-PAGE with the purified WT, N136C and S139C variants (Figure 5.8-30) and the yields of all target proteins expressed with the pPKL241-based expression systems (Table 5.8-10) are provided in Chapter 5.8.5. Individual aliquots containing the $FimH_L$ cysteine variants were hereafter subjected to alkylation with IAA derivatives (Chapters 3.6 and 5.7.7). Both the unmodified and chemically labelled protein samples were subjected to mass spectrometric analysis via LC-MS/MS (Chapters 3.7 and 5.7.8).

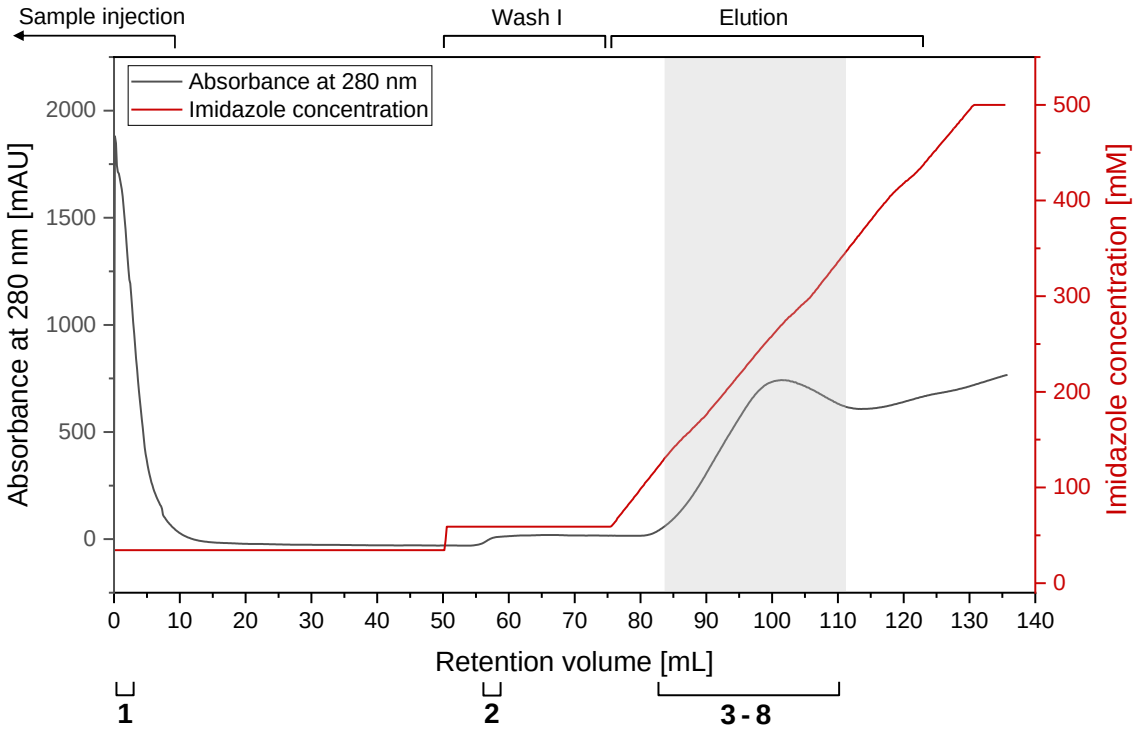


Figure 3.5-2: Chromatogram from the IMAC purification of FimH$_L$ N136C using a linear imidazole gradient.

FimH$_L$ N136C was expressed in *E. coli* HM125 pPKL241 in 5 L LB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare) equilibrated with binding buffer. The column was washed with 59 mM imidazole (Wash I, 5 CV) prior to elution of the target protein via a linear imidazole gradient from 60 to 500 mM imidazole throughout 14 CV. Fractions **1** – **8** were analysed by SDS-PAGE (Figure 3.5-2a). The main peak (fractions **4** – **8**), containing the target protein, is highlighted in grey.

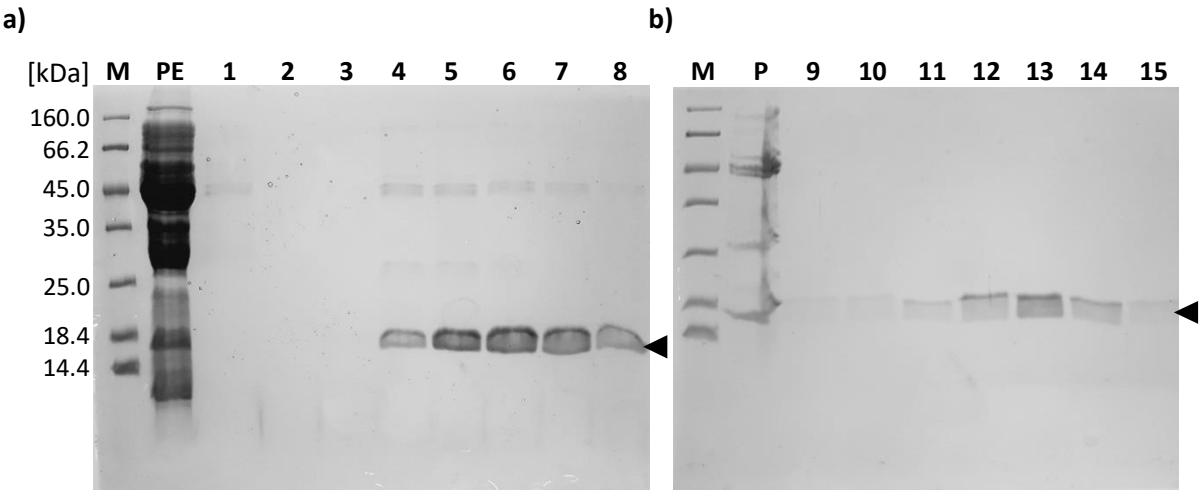


Figure 3.5-3: SDS-PAGE analysis to evaluate the **a)** IMAC and **b)** SEC purification of FimH$_L$ N136C.

M: Pierce™ Unstained Protein M$_W$ Marker (cat. no. 26610, Thermo Scientific™); **PE**: periplasmic extract; **1** – **8**: collected fractions from the IMAC purification shown in Figure 3.5-2; **P**: pooled fractions from the IMAC purification shown in Figure 3.5-2; **9** – **15**: collected fractions from the SEC purification shown in Figure 3.5-4. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 30/40 min at constant currents (I = 30/25 mA/gel). Black arrowheads indicate the migration bands corresponding to the overexpressed target protein FimH$_L$ N136C (17.8 kDa).

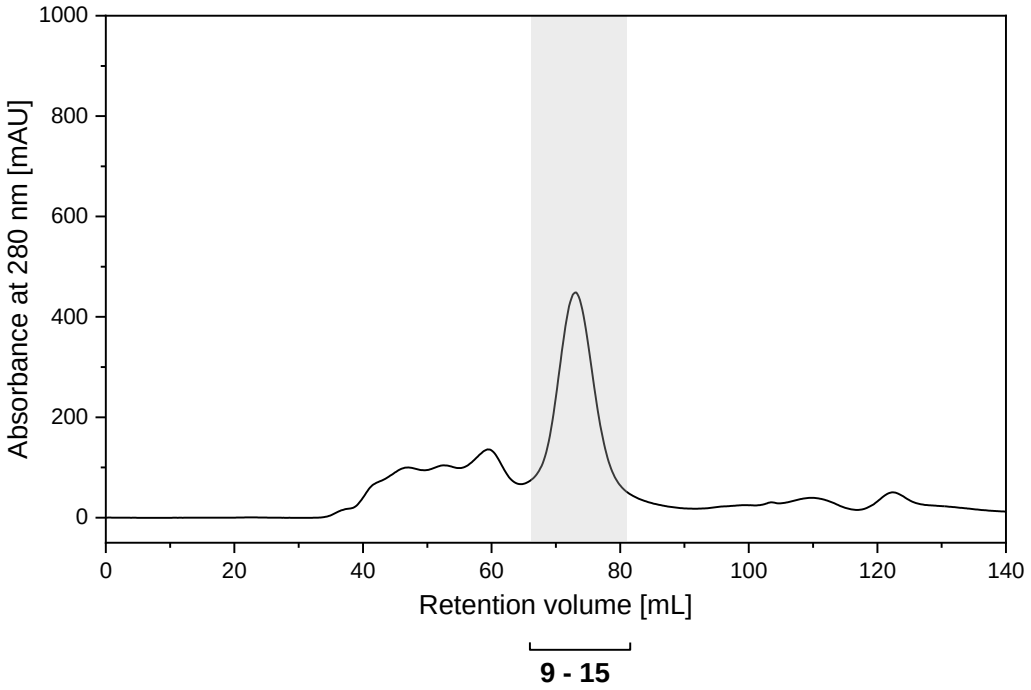


Figure 3.5-4: Chromatogram from the SEC purification of $FimH_L$ N136C.

The collected eluate fractions from the IMAC purification of $FimH_L$ N136C shown in Figure 3.5-2 were concentrated to a final volume of 1.5 mL in a centrifugal filter (Vivaspin® Turbo, MWCO = 3 kDa), and applied onto a 120 mL SEC column (HiLoad® 16/600 Superdex® prep-grade column, GE Healthcare, cat. no. 28-9893-33) that had previously been equilibrated with PBS. The indicated fractions **9** – **15** were further evaluated by SDS-PAGE (Figure 3.5-3b). All fractions of the main peak (highlighted in grey), which contained the target protein, were pooled and concentrated to a final volume of 1.25 mL. Individual aliquots were used for chemical modification (Chapter 3.6) and mass spectrometric analysis (Chapter 3.7).

Although the expression success of different $FimH_L$ variants could be slightly increased by the use of HM125 as the expression strain, the protein yields remained remarkably low compared to the literature (approximately 11 mg/L in minimal medium).[486] The hybrid T5/*lacO* promoter present on the expression plasmid is partially derived from the T5 bacteriophage and, like many viral or phage-derived promoters, is generally associated with high promoter strength. However, a high expression rate is often accompanied by the risk of aggregation of the overexpressed proteins and their accumulation in the cytoplasm or periplasm in the form of inclusion bodies. This could be especially relevant in the context of periplasmic protein expression, possibly due to spatial limitations within the periplasm.

For the evaluation of expression and purification, the cell debris was occasionally applied to the SDS-PAA gels (e.g. in Figure 5.8-27). No evidence of target protein was detected in these samples that would indicate the presence of inclusion bodies, even though the protein migration behaviour was strongly impaired and appeared smeared, despite the addition of DNase I (1 U/mL) during cell lysis. In order to reduce the expression rate and the associated probability of inclusion body formation, extensive optimisation experiments were performed at smaller scale (50 mL) including lower expression temperatures and/or reduced IPTG concentrations. Unfortunately, these attempts did not lead to the desired increase in expression levels wthin the periplasmic extracts (data not shown).

The avoidance of Sec-pathway-dependent translocation was thus considered a rational strategy to potentially improve expression success for $FimH_L$ varants in this manner. Under my supervision, Sofie Todt thus tested the suitability of cytosolic expression and accumulation of $FimH_L$ N136C as part of her Bachelor thesis.[487] However, the removal of the N-terminal signal sequence from pPKL241 at the genetic level, followed by expression and purification attempts in *E. coli* Origami B, afforded only an extremely poor yield of the respective gene product, $FimH_L$ N136C ΔN-sigpep (Chapter 5.8.5, Table 5.8-10). Since cytosolic expression did not provide the desired improvement in expression success, the conventional periplasmic expression approach had to be further pursued and optimised.

3.5.2 Expression and purification of $FimH_{trc}$ variants in HM125 pFimH-His_6-NT-trc

For the investigation of lectin activity, for example in the context of a lectin assay (Chapter 3.8) and for NMR studies in particular, a significantly higher protein yield was required than previously achieved (2 – 2.8 mg total yield). For this reason, as well as due to the recurring problem of partial T5/*lacO* promoter deletions in pPKL241, the latter was replaced by an alternative expression plasmid. The new plasmid (pFimHs-NT-trc) was also kindly provided by Prof. Rudi Glockshuber and contained the lectin domain of $FimH_{1-155}$. The total of seven mutations, when compared to the $fimH_L$ sequence on pPKL241, are silent, while the codon for Gly156 and the C-terminal His_6 tag are not provided on this construct. Transcription of the target sequence occurs under the control of an inducible *lac*/*trp* hybrid (*trc*) promoter.[488] The coding sequence for the *lac* repressor (*lacI*) is likewise present on the plasmid. Hence, resource stress after transformation events and associated promoter deletions (as in the case of pQE30-derived constructs) are rather unlikely. As the *trc* promoter is moreover exclusively derived from bacteria, it probably exhibits a reduced promoter strength compared to the previously employed viral T5/*lacO* promoter, thereby decreasing the probability of inclusion body formation resulting from excessive overexpression.

For this work, the missing codon for Gly156, along with the C-terminal His_6 tag sequence, were introduced by site-directed mutagenesis into the *fimH* sequence on the precursor plasmid pFimHs-NT-trc. Likewise, the cysteine (N136C, S139C) and proline (R60P) point mutations were introduced using the QuikChange™ method.[478] A more detailed description of the preparation of the target plasmid is provided in Chapters 5.7.6 and 5.8.2, as well as in Table 5.7-8. The gene products of the new construct (pFimH-His_6-NT-trc) are identical to those encoded on pPKL241. An alignment of the respective coding sequences is provided in the appendix. The new expression plasmid was transformed into the suitable expression strain HM125. For purposes of differentiation from the pPKL241-encoded gene products ($FimH_L$), the target proteins expressed from pFimH-His_6-NT-trc are designated as $FimH_{trc}$ in this work (trc: trc promoter).

Owing to the introduced C-terminal His_6 tag, a purification of $FimH_{trc}$ variants by IMAC was enabled. However, the first attempts to purify $FimH_{trc}$ variants according to the same protocol as used for the pPKL241-based expression system repeatedly resulted in a complete yield loss of the target proteins. Although the protein amounts during purification by IMAC were at first quite promising, a massive, macroscopically visible protein aggregation subsequently occurred. The latter was particularly pronounced in the case of the cysteine variants, especially in the IMAC elution fractions with the highest protein concentrations. The high amount of aggregates was rather difficult to reconcile with the concentration of the target protein by centrifugal filtration, which would be necessary for a subsequent SEC. Consequently, the protocol was adapted in a way that the purification could preferably be accomplished in a single chromatography step by IMAC, as performed earlier by Bouckaert and co-workers.[376] This was partially achieved by direct periplasmic cell disruption after harvesting (without prior freezing of the harvested cells), in order to further minimise contamination by cytoplasmic proteins. An intended relative overloading of the affinity column was also quite beneficial, to improve the purity of target proteins. This was, for instance, achieved by markedly increasing the cell mass (≥ 40 g), which was mainly obtained by expression in more nutritious Terrific Broth (TB) medium. Alternatively, a smaller IMAC column volume (1 mL) was used for the enrichment and purification of the target proteins. In addition, a step gradient was chosen instead of the previously employed linear imidazole gradient, in order to obtain higher protein concentrations and, thus, to avoid subsequent concentration steps.

The expression and purification of the target proteins using plasmid pFimHs-His_6-NT-trc are hereafter once more exemplified by the N136C variant. The respective chromatograms and SDS-PAA gels for the other $FimH_{trc}$ variants (WT, R60P and S139C) are provided in Chapter 5.8.6 (Figures 5.7-31 – 5.7-36). Protein yields of all $FimH_{trc}$ variants used in this thesis are summarised in Table 5.8-11, and the purity of the respective expression products was confirmed by SDS-PAGE (Figure 3.5-9, at the end of this chapter). The identity of all $FimH_{trc}$ variants was verified by matrix-assisted laser desorption/ ionisation-time of flight mass spectrometry (MALDI-TOF MS). Corresponding *m*/*z* ratios and MALDI mass spectra are provided in Chapter 5.8.8, Table 5.8-14 and the appendix.

Pleasingly, the optimised expression and purification protocol enabled satisfactory protein yields of WT $FimH_{trc}$ (15 mg/L of expression volume on average), which are comparable to those reported in the literature. The R60P variant likewise yielded sufficient amounts that could certainly be further increased by simply upscaling the expression volume. In addition, the occurrence of protein aggregation was less pronounced for the WT and R60P variants. However, the cysteine variants of $FimH_{trc}$ could not be obtained in sufficient yield and/or purity via a single IMAC step adequate for subsequent chemical modification and/or analysis of lectin activity. Therefore, the protocol for the purification of $FimH_{trc}$ cysteine variants required further optimisation.

An exemplary chromatogram and SDS-PAGE result from the IMAC purification of $FimH_{trc}$ N136C is depicted in Figures 3.5-5 – 3.5-6. After application of the dialysed, clarified periplasmic extract to the IMAC column, the imidazole concentration was stepwise increased to 25 mM (Wash I, 10 CV) and then to 50 mM (Wash II, 5 CV), in order to remove weakly bound contaminants. The target protein was subsequently eluted by further increasing the eluent concentration to 250 mM (15 CV). Glycerol (10 % v/v) was added to both the binding and elution buffers, in order to reduce protein aggregation. The distinctly higher UV absorbance of the elution peak in the resulting IMAC chromatogram (Figure 3.5-5) already indicated a satisfactory protein yield in a smaller volume compared to the linear gradient protocol. SDS-PAGE of the corresponding fractions, (Figure 3.5-6, lanes **3** – **17**), however, revealed the presence of other protein species with higher apparent mass in addition to the target protein (17.8 kDa, labelled with a black arrowhead). The fractions with the highest absorbance at 280 nm (and thus the highest protein concentration) exhibited a remarkably gel-like consistency after denaturation in Laemmli buffer. The electrophoretic migration behaviour (lanes **7** – **10**) was considerably affected, presumably due to protein aggregation. The same phenomenon was also observed with the highly concentrated IMAC fractions of the other $FimH_{trc}$ variants. However, dilution of the respective SDS-PAGE samples significantly reduced this gelation effect and improved the migration behaviour, which again illustrates the pleasingly high protein concentration in the elution fractions.

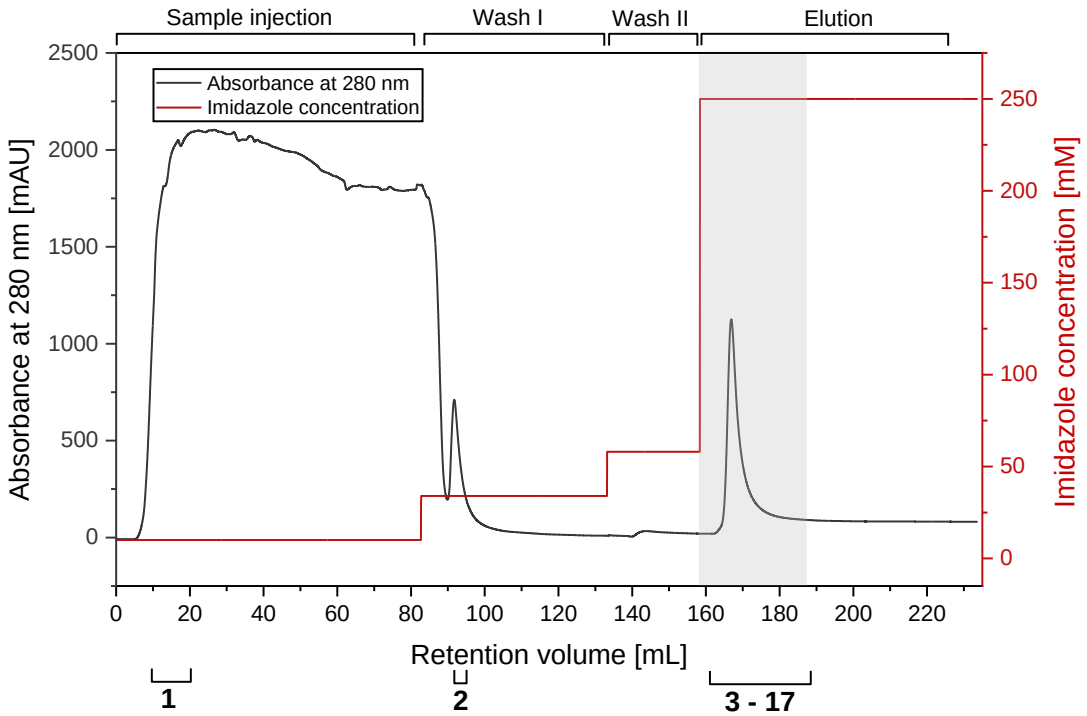


Figure 3.5-5: Chromatogram from the IMAC purification of $FimH_{trc}$ N136C using the optimised step-gradient protocol.

$FimH_{trc}$ N136C was expressed in *E. coli* HM125 pFimH-His_6-NT-trc in 3 L TB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare) equilibrated with binding buffer + 10 % v/v glycerol. The column was washed with 25 mM imidazole (Wash I, 10 CV) and 50 mM imidazole (Wash II, 5 CV) prior to elution of the target protein with 250 mM imidazole (step gradient, 15 CV). Fractions **1** – **17** were analysed by SDS-PAGE (Figure 3.5-6). The main peak (fractions **3** – **17**), containing the target protein, is highlighted in grey.

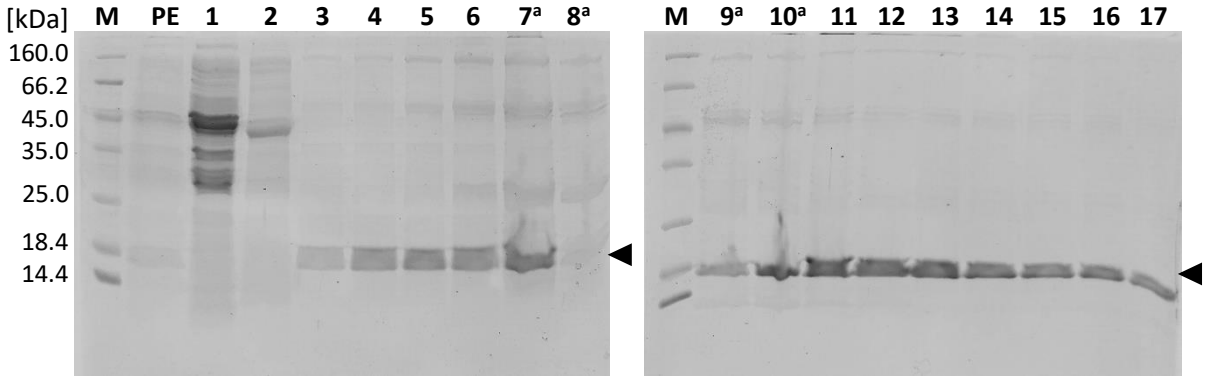


Figure 3.5-6: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_{trc}$ N136C.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **17**: collected fractions from the IMAC purification shown in Figure 3.5-5. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 27.5 mA/gel). Black arrowheads indicate the expression bands of the target protein $FimH_{trc}$ N136C (17.8 kDa). **a)** The smeary migration behaviour results from the gel-like consistency of the SDS-PAGE samples due to the high concentration of target protein.

Since a single IMAC step was not sufficient to remove all contaminants from the eluted target protein fractions, an attempt was made to purify $FimH_{trc}$ N136C by anion exchange chromatography (AEC), which is, among other applications, frequently used for the purification of untagged FimH variants.[480] In contrast to gel filtration, ion exchange chromatography generally causes less dilution of the protein samples and does not require a preliminary concentration step. Following another IMAC attempt (Figures 5.7-37 – 5.7-38, Chapter 5.8.6), the combined fractions were dialysed against AEC low-salt buffer (10 mM NaCl, pH 8.0) and applied onto the AEC column (MonoQ®, 1 mL, Cytiva). After extensive washing with AEC low-salt buffer, interacting protein species were eluted using a linear, increasing NaCl gradient from 10 to 400 mM throughout 45 CV (Figure 5.8-39). Although the SDS-polyacrylamide (SDS-PAA) gel (Figure 5.8-40) indicated largely pure target protein in the main peak fractions, the AEC chromatogram revealed a rather unusual elution behaviour, as these fractions were eluted in several sequential portions. This observation, together with the previously mentioned tendency of the cysteine variants to aggregate, suggested that the contaminants in the main peak fractions from IMAC might (at least partly) represent different oligomeric states of the target protein. Another plausible explanation might be biological cysteine modifications, which were detected in the $FimH_L$ N136C and S139C variants by tandem mass spectrometry, as will be discussed in Chapter 3.7. Since the cysteine modifications were proven to be fully reversible under reducing conditions, a small amount of the combined IMAC eluate fractions was subjected to an analytical SEC (CV = 24 mL) after treatment with 10 µmol dithiothreitol (DTT), or without prior reduction. The chromatogram revealed several protein species, regardless of whether the sample was previously reduced or not, which was therefore considered to be due to oligomerisation and/or contaminants rather than biological modification (data not shown).

In order to obtain pure, monomeric target protein, the SEC was repeated on a preparative scale: The remaining IMAC fractions containing $FimH_{trc}$ N136C (Figure 3.5-6a) were carefully concentrated to a smaller volume by centrifugal filtration, during which the samples were repeatedly rinsed with PBS and 10 % v/v glycerol. The concentrate was reduced with ∼ 5 mM DTT and finally applied onto a preparative SEC column (CV = 120 mL), equilibrated with the same buffer/glycerol mixture. The corresponding chromatogram, shown in Figure 3.5-7, manifests at least four major peaks that appeared at retention volumes between 40 – 80 mL. All the respective fractions contained the target protein in larger quantities according to analysis by reducing SDS-PAGE (Figure 3.5-8a). As expected, other protein species of higher apparent mass were observed in the Peaks I – III, which could represent impurities or oligomeric states of $FimH_{trc}$ N136C. Only the fractions corresponding to Peak IV revealed clearly purified, apparently monomeric target protein (Figure 5.8-42 in Chapter 5.8.6). The latter was eluted from the SEC column at a retention volume of 70 – 80 mL, as expected from the previous purification attempts of $FimH_L$ variants, discussed earlier (Chapter 3.5.1). When combining all fractions from Peak IV, $FimH_{trc}$ N136C was obtained with a total yield of 2.42 mg (≙ 0.81 mg/L expression volume).

Individual fractions from SEC peaks I – IV (lanes **2** – **5** and **7** in the result from the reducing SDS-PAGE illustrated in Figure 3.5-8a) were tested for lectin activity in a FITC-*p*AP mannoside-binding assay (for details, view Chapter 3.8 and Chapter 5.7.9), whereby the same amount of protein was employed in each case. The by far highest FITC-*p*AP mannoside-binding activity was measured in Peak IV (Fraction **7** in Figure 5.8-70, Chapter 5.7.9): This supports the assumption regarding the putative monomeric state of the FimH species contained in Peak IV, since only the monomeric form is expected to maintain full lectin activity, whereas aggregation or oligomerisation most likely would result in decreased binding capability.

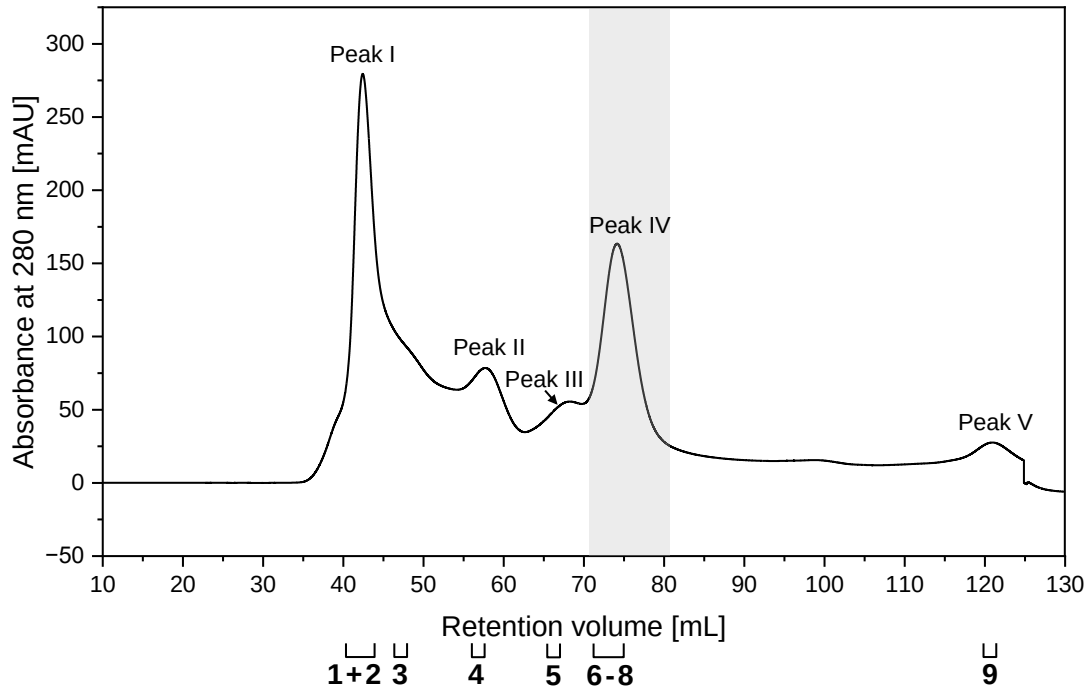


Figure 3.5-7: Chromatogram from the SEC purification of $FimH_{trc}$ N136C.

The collected eluate fractions from the IMAC purification of $FimH_{trc}$ N136C, shown in Figure 3.5-5, were carefully concentrated to a final volume of 2 mL in a centrifugal filter (Amicon®, MWCO = 3 kDa), reduced with ~ 5 mM DTT, and applied onto a 120 mL SEC column (HiLoad® 16/600 Superdex® prep-grade column, GE Healthcare, cat. no. 28-9893-33) that had previously been equilibrated with PBS and 10 % v/v glycerol. The indicated fractions **1** – **9** were further evaluated by SDS-PAGE (Figure 3.5-8). All fractions in Peak IV (highlighted in grey), containing the monomeric (and covalent-dimeric) target protein, were proven to be without impurities by SDS-PAGE (Figure 5.8-42). Individual fractions were used for chemical modification (Chapter 3.6) and in FITC-*p*AP mannoside-binding assays (Chapter 3.8).

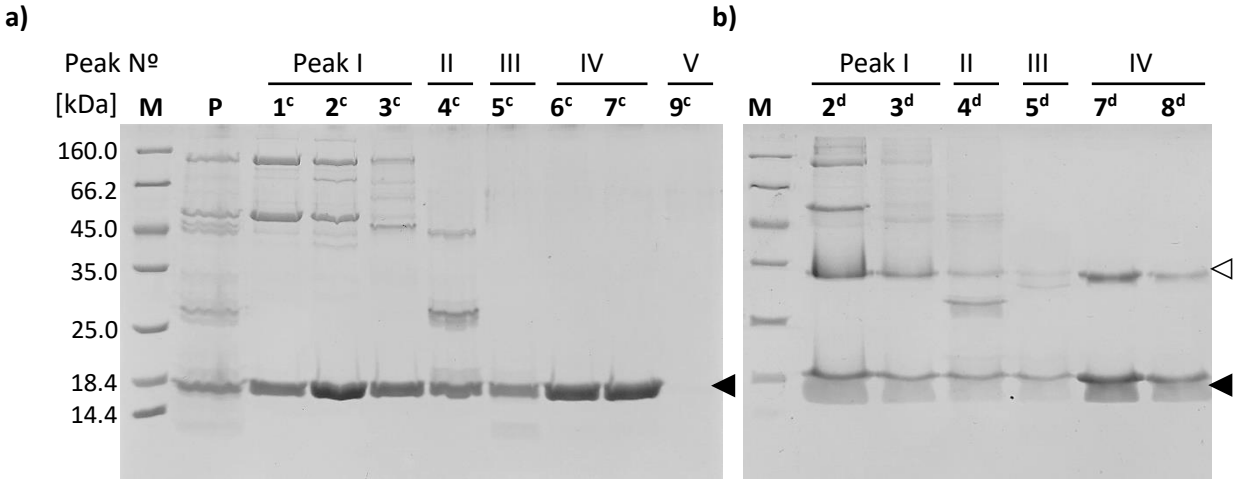


Figure 3.5-8: Results from **a)** reducing and **b)** non-reducing SDS-PAGE validating the SEC purification of $FimH_{trc}$ N136C.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **P:** pooled fractions from the IMAC purification shown in Figure 3.5-5; **1** – **9**: collected fractions from the SEC purification shown in Figure 3.5-7. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 1 h at constant current (I = 25 mA). Black arrowheads indicate the migration bands corresponding to the monomeric target protein $FimH_{trc}$ N136C (17.8 kDa). The white arrowhead indicates the migration bands corresponding to the disulfide-linked dimers of the target protein (35.6 kDa). **c)** Samples were denatured at 95 °C for 5 min in the presence of β-mercaptoethanol (10 % v/v, reducing SDS-PAGE). **d)** Samples were not heated and dissolved in non-reducing Laemmli buffer (non-reducing SDS-PAGE).

Unfavourably, Peak IV did not show the highest UV absorption in the SEC chromatogram, indicating a further, significant yield loss of the target protein ($FimH_{trc}$ N136C). A similar result was obtained by analytical SEC of $FimH_{trc}$ S139C after single-step purification with IMAC (Figure 5.8-41c in Chapter 5.8.6). The peak containing the target protein, which is expected after a retention volume between 17 – 19 mL, was accompanied by other eluting protein species with partly higher UV absorbance. In contrast, the amount of other protein spieces and/or the occurrence of putative oligomeric states was comparatively lower in the case of the WT protein and R60P variants (Figure 5.8-41a and b).

Non-reducing SDS-PAGE (i.e. SDS-PAGE without the use of β-mercaptoethanol in the Laemmli buffer and without denaturation of the SDS-PAGE samples at 95 °C) was conducted to examine the purified protein samples for covalent oligomerisation. As clearly visible in the corresponding gel in Figure 3.5-8b, an additional protein species appeared in the corresponding gel under non-reducing conditions in all samples from the SEC peaks I – IV, exhibiting an apparent mass about 34 kDa (indicated with a white arrowhead). These protein species most likely represent covalent $FimH_{trc}$ N136C homodimers, linked via a disulfide bond between the introduced cysteine residues. Even the highly pure SEC fractions (Peak IV) contained detectable amounts of the covalent dimer. A corresponding species was also observed in cysteine variant S139C, but not in the WT nor the R60P variant (cf. the non-reducing SDS-PAGE result in Figure 5.8-43, Chapter 5.8.6). No additional protein species with a molecular weight corresponding to higher covalent oligomeric states of the cysteine variants were observed under non-reducing conditions that were not already present in the reducing SDS-PAA gel (Figure 3.5-8a). Nevertheless, this clearly underlines the importance of reducing the concentrated protein samples prior to gel filtration when purifying cysteine variants, to prevent unnecessary yield loss.

The reducing agent could possibly be added to the SEC buffer, but this could not be tested due to time constraints. Likewise, reducing agents could also be useful at earlier stages of the purification process of the cysteine variants, particularly to prevent protein aggregation. It should be noted that reducing agents with a less negative redox potential than DTT would have to be used in this case, such as tris(2-carboxyethyl)phosphine (TCEP) or β-mercaptoethanol (1 – 5 mM), in order to limit the risk of unwanted reduction of immobilised Ni^{2+} ions on the IMAC column. For the same reason, the reducing agent would possibly have to be removed again after purification, for instance, prior to the immobilisation of the lectin on the Ni^{2+}-chelate surface in the FITC-*p*AP mannoside-binding assay (Chapter 3.8).

Based on the results from the *in silico*-docking simulations (Chapter 3.1), suggesting that the $FimH_{trc}$ N136C variant was more promising for an effective photocontrol by bioconjugation to an AB mannoside as the Gate-Keeper than the S139C variant, the focus was initially put on the successful expression and purification of the former cysteine variant. Due to the enduring problems with macroscopic protein aggregation and the associated yield loss, the purification of the S139C variant was not optimised further. In other words, the aim to obtain a sufficient amount of $FimH_{trc}$ S139C for the chemical modification and the subsequent investigation of photoswitching the lectin activity was not pursued further. Visible protein aggregation was continuously observed during and after IMAC/SEC as well as dialysis in different buffers and after storage at 4 °C/ −80 °C, even when glycerol (10 % v/v) was added. It should be mentioned that, according to the literature, truncated FimH variants are explicitly stored at these temperatures in water or in conventional buffers (pH 7.4) without additives such as glycerol.[337,342,376,397] However, FimH is not longer purified via IMAC according to the majority of publications, which might favour protein aggregation due to unfavourable elution conditions and/or the high concentration of target protein. In contrast, recent publications increasingly employ sequential cation and anion exchange chromatography, which does not require an affinity tag.[337]

Unfortunately, it could not be proven without doubt by SDS-PAGE, that the frequently observed macroscopic aggregates were particularly composed of the target proteins, since the precipitates could not be resolubilised under reducing condition (5 mM DTT in PBS:glycerol, 9:1, 37 °C, 2 h). However, a primary aggregation of the target proteins can be considered very likely, as the extent of aggregation negatively correlated with the final protein yield obtained from the different $FimH_{trc}$ variants. Since increasing the protein concentration by centrifugal filtration did not appear to be a viable option, the IMAC or SEC fractions, respectively, with the each highest protein concentrations were eventually dialysed and employed for the subsequent experiments. The corresponding fractions were applied onto the SDS-PAA gel illustrated in Figure 3.5-9.

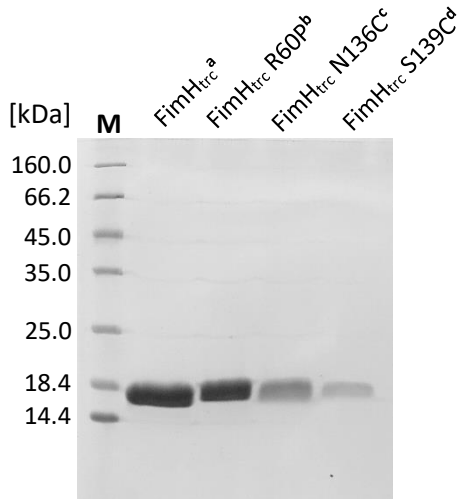
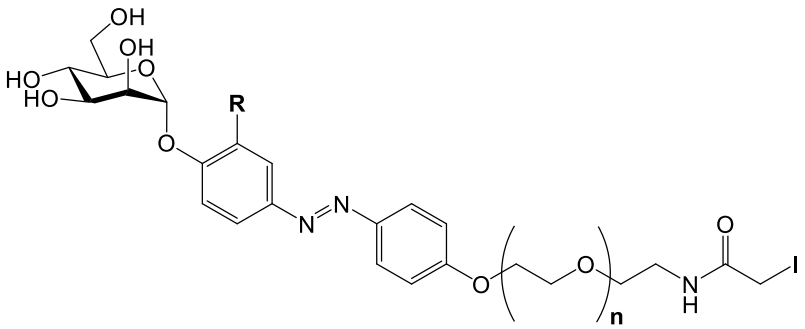


Figure 3.5-9: SDS-PAGE analysis to evaluate the purity of $FimH_{trc}$ variants used in this thesis.
M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™). Protein concentrations: **a)** $FimH_{trc}$, 1.0 mg/mL (assay buffer); **b)** $FimH_{trc}$ R60P, 0.45 mg/mL (PBS); **c)** $FimH_{trc}$ N136C, 0.30 mg/mL (PBS); **d)** $FimH_{trc}$ S139C, 0.20 mg/mL (PBS). Expression and purification of the here visualised target proteins is documented in Chapter 5.8.6. Individual aliquots were used for chemical modification (Chapter 3.6) and in FITC-*p*AP mannoside-binding assays (Chapter 3.8). SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 30 mA).

Despite the problems associated with oligomerisation and/or protein aggregation and the resulting yield loss, adequate amounts of all four $FimH_{trc}$ variants (WT, R60P, N136C, and S139C) were obtained to successfully evaluate their lectin activity in the FITC-*p*AP-Man binding assay, which will be introduced in Chapter 3.8.2. Particular attention was directed towards optimising the expression and purification of the N136C variant, in order to achieve the best possible yield and concentration. Individual fractions obtained from SEC (Peak IV in Figure 3.5-7) contained the sufficiently concentrated, pure and predominantly monomeric target protein, which exhibited a pleasingly high lectin activity. The respective fractions with the highest concentrations of $FimH_{trc}$ N136C were thus subjected to site-specific *S*-carbamidoalkylation with the IAA-functionalised AB mannosides **F1** and **H1** (Chapter 3.6). After successful verification of the chemical modification by MALDI-TOF MS (Chapter 5.8.8), the photocontrol of lectin activity was finally investigated within the scope of the Gate-Keeper project (Chapter 3.8.3).

3.6 Chemical modification of FimH cysteine variants

With the $FimH_L$ variants N136C and S139C in hand, a classical *S*-carbamidoalkylation was performed using the IAA-functionalised AB mannosides **H1** and **H2**, as well as the 3-fluoro-AB derivatives **F1** and **F2** (Figure 3.6-1). Largely following the basic protocol described by Crankshaw et al.,[218] an amount of approximately 250 µg protein (100 µL, in PBS) was reduced using a tenfold molar excess of DTT per protein disulfide. Given that the intrachain disulfide bridge between Cys3 and Cys44 is internally located and thus largely inaccessible to the external medium, a fivefold molar excess of DTT per FimH molecule (equivalent to 0.5 protein disulfide units per molecule) was considered sufficient. The reaction tubes were purged with nitrogen and incubated for 1 h at 37 °C. Subsequently, a fivefold molar excess of the respective IAA derivatives was added, calculated over the total amount of thiol groups present (including those from DTT, as well as of cysteine residues, regardless of their redox state). The reaction mixtures were again incubated for 1 h at 37 °C, under a nitrogen atmosphere and protected from light.



R = H, n = 1: **H1** (M = 631.42 g/mol)
R = H, n = 2: **H2** (M = 675.47 g/mol)
R = F, n = 1: **F1** (M = 649.41 g/mol)
R = F, n = 2: **F2** (M = 693.46 g/mol)

Figure 3.6-1: Structure and molar masses of iodoacetamide (IAA)-functionalised azobenzene and 3-fluoroazobenzene mannoside derivatives used for chemical labelling of the truncated FimH cysteine variants N136C and S139C.

Where the alkylation mixtures were intended for a subsequent analysis by LC-MS/MS, no additional treatment was required. The respective samples were stored on ice and immediately consigned to the Institute of Experimental Medicine (Kiel University). In all other cases, the reaction was quenched by the addition of a fivefold molar excess of β-mercaptoethanol relative to the previously added alkylating reagent. Eventually, the removal of salts and side products was achieved by small-scale gel filtration using Zeba™ Spin desalting columns with an appropriate volume capacity and molecular weight cut-off (MWCO).

The DEG-linked AB derivatives **H1** and **F1** were not completely water soluble under the selected alkylation conditions. Furthermore, the reaction mixtures occasionally dried out in the course of the alkylation experiment if the Eppendorf tubes were not completely sealed under nitrogen purging through a cannula. Nevertheless, the chemical modification of the $FimH_L$ cysteine variants was successful overall, which was clearly demonstrated by LC-MS/MS analysis. The corresponding results are described in more detail in the following chapter.

3.7 Mass spectrometric analysis of alkylated and unmodified FimH variants

3.7.1 Analysis of $FimH_L$ variants by LC-MS/MS

A thorough characterisation of modified and unmodified $FimH_L$ cysteine variants was enabled by liquid chromatography-tandem mass spectrometry (LC-MS/MS), which was performed in cooperation with the working group of Prof. Dr. Tholey at the Institute of Experimental Medicine (Kiel University). In this way, a reliable identification of the target proteins could be achieved, as well as a more precise evaluation of the protein integrity and chemical cysteine modifications. The protocols for both approaches used for top-down and bottom-up proteomics were developed, experimentally conducted, and evaluated by Philipp Kaulich.[489,490] In contrast to the bottom-up method, in which the analysed proteins are first proteolytically digested,[491] the top-down approach involves the ionisation and measurement of intact proteins by tandem mass spectrometry.[492] Bottom-up proteomics is generally the more accessible and most widely used approach for the characterisation of proteins. However, in our previous attempts to analyse alkylated $FimH_L$ according to the bottom-up approach, no complete sequence coverage could be achieved. Moreover, the respective peptide fragments containing the *S*-carbamidoalkylated cysteine residues could not be identified. In contrast to this, the top-down method proved to be particularly suitable for the analysis of (modified) $FimH_L$ cysteine variants, since specific proteoforms could be completely characterised and quantified. A selection of the corresponding results for the $FimH_L$ N136C variant, with and without bioconjugation to IAA-functionalised 3-fluoro-AB derivative **F1**, is presented here by way of example. All other top-down proteomics results, including experimental details, mass spectra, and fragment maps of the N136C variant after alkylation with the other photoswitchable IAA derivatives (**H1**, **F2**, **H2**), as well as the complete analysis of the modified S139C variant, are summarised in Chapter 5.8.8. The latter also contains a compilation of *m/z* data obtained with modified $FimH_L$ variants, which were produced in strains that were not deficient in *degP*. The most important data compiled from the mass analysis of the modified cysteine variants that were expressed in *E. coli* HM125 are also shown in Table 3.7-1.

The protein samples (30 µg) were precipitated, resuspended in LC-MS/MS loading buffer[493] and an amount of 50 – 100 ng was subjected to a Dionex UltiMate™ 3000 UHPLC system, equipped with a reversed-phase C4 column for proteoform separation and a pre-column for sample loading. The *m/z* ratios were identified in an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer using the multi-compensation voltage (CV) field asymmetric ion mobility spectrometry (FAIMS) method for ion separation, which was introduced by Kaulich et al.[489] Fragmentation of ions was induced with different methods, including collision-induced dissociation (CID),[494] higher-energy collisional dissociation (HCD),[495] and electron-transfer/higher-energy collision dissociation (EThcD).[496]

Table 3.7-1: Most abundant masses found in LC-MS/MS analysis of chemically labelled and unmodified cysteine variants of $FimH_L$.

FimH variant	Ligand	Theoretical mass[a] [Da]	Identified mass[b] [Da]	Accuracy [ppm]
$FimH_L$-His_6 N136C	unmodified	17 820.83	17 820.85	1.12
$FimH_L$-His_6 N136C	**H1**	18 324.02	18 324.03	0.55
$FimH_L$-His_6 N136C	**H2**	18 386.04	18 368.06	0.54
$FimH_L$-His_6 N136C	**F1**	18 342.02	18 342.03	0.55
$FimH_L$-His_6 N136C	**F2**	18 368.05	18 386.05	0.54
$FimH_L$-His_6 S139C	unmodified	17 847.84	18 152.93 17 847.86[c]	305.09 1.12[c]
$FimH_L$-His_6 S139C	**H1**	18 351.03	18 351.04	0.54
$FimH_L$-His_6 S139C	**H2**	18 395.06	18 395.07	0.54
$FimH_L$-His_6 S139C	**F1**	18 369.03	18 369.03	0.03
$FimH_L$-His_6 S139C	**F2**	18 413.05	18 413.07	1.09

a) Monoisotopic mass including one disulfide bridge. **b)** Monoisotopic mass of the most abundant mass feature calculated by FLASHDeconv v2.0. **c)** Lower abundant identified mass with corresponding accuracy, obtained after reduction with 20 mM TCEP, 10 min at 60 °C. Raw files were converted to centroid mzML files and deconvoluted using FLASHDeconv (v2.0, default settings). The theoretical monoisotopic mass was calculated based on the amino acid sequence of the protein and the chemical formula of the alkylated ligands. All data were obtained, analysed and kindly provided by Dr. Philipp T. Kaulich (Kiel University).

In order to assess the integrity of the target proteins, the unmodified cysteine variants were analysed. The most intense mass feature (17 820.68 Da) could be assigned quite accurately to the theoretical intact mass of the unmodified $FimH_L$ variant, containing the native disulfide bridge between Cys3 and Cys44 (17 820.83 Da). However, many other *m*/*z* signals were also detected that could not initially be allocated to the target protein (Figure 3.7-1 and Figure 5.8-44 in Chapter 5.8.8). As the corresponding fragmentation pattern (Figure 5.8-45) indicated a partial, actually unintended modification of the introduced Cys136, a portion of unmodified protein was therefore reduced with TCEP (20 mM) at 60 °C for 10 min[497] and the LS-MS/MS measurement was subsequently repeated (without FAIMS). As illustrated in Figure 3.7-1, predominantly the signal of the intact, unmodified protein was observed after reduction, confirming a reversible cysteine modification. Intriguingly, several of the mass differences between reduced and non-reduced proteoforms could be attributed to distinct biological modifications, for instance, adducts with homocysteine (+ 133.02 Da), hexuronic acid (+ 176.04 Da) and coenzyme A (+ 765.12 Da). The same mass shifts were also identified in the case of the S139C variant (Figure 5.8-66). Additionally, a mass difference of +305.10 Da was observed that could be attributed to a modification with glutathione disulfide. The corresponding proteoform (18 152.93 Da) even represented the most abundant mass feature in the respective spectrum (Figure 5.8-45), and only a low-intensity signal of the intact unmodified $FimH_L$ S139C (17 847.86 Da) was observed. However, the latter was again fully recovered after reduction with TCEP.

The intact-mass spectrum (Figure 3.7-2a) illustrates the charge envelope (i.e. the distribution of different charge states) of $FimH_L$ N136C after alkylation with IAA-functionalised fluoro-AB mannoside **F1**. The most intense *m*/*z* signal (in this case, the + 15 peak) was zoomed in to reveal the isotope distribution in Figure 3.7-2b, which is in good agreement with the theoretical data.

Deconvolution of the raw data yielded a mass of 18 342.03 Da, which can be attributed to the monoisotopic mass of successfully labelled $FimH_L$ N136C exhibiting a disulfide bridge. The fragment map of the corresponding proteoform, shown in Figure 3.7-3, reveals a high sequence coverage and validates both the natural disulfide bridge and the position of the *S*-carbamidoalkylation at Cys136. The amino acid sequence, including the introduced cysteine residues and the position of the disulfide bridge, was also confirmed in all other $FimH_L$ samples (alkylated and unmodified) by CID and EThcD fragmentation, respectively.

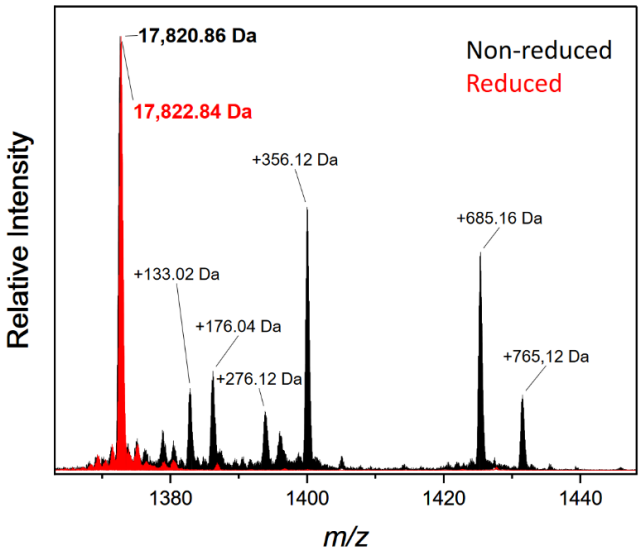


Figure 3.7-1: Overview of the intact-mass spectrum of reduced $FimH_L$ N136C (red) versus the non-reduced variant (black) obtained by top-down LC-MS/MS analysis of the unmodified protein prior to alkylation with photoswitchable IAA derivatives.

The additional, mass-shifted proteoforms of the non-reduced N136C variant presumably result from the following biological modifications: + 133.02 Da, *S*-homocysteine; + 176.04 Da, hexuronic acid; + 765.12 Da, coenzyme A. These modifications were efficiently removed after reduction with 20 mM TCEP for 10 min at 60 °C.[497] The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University).

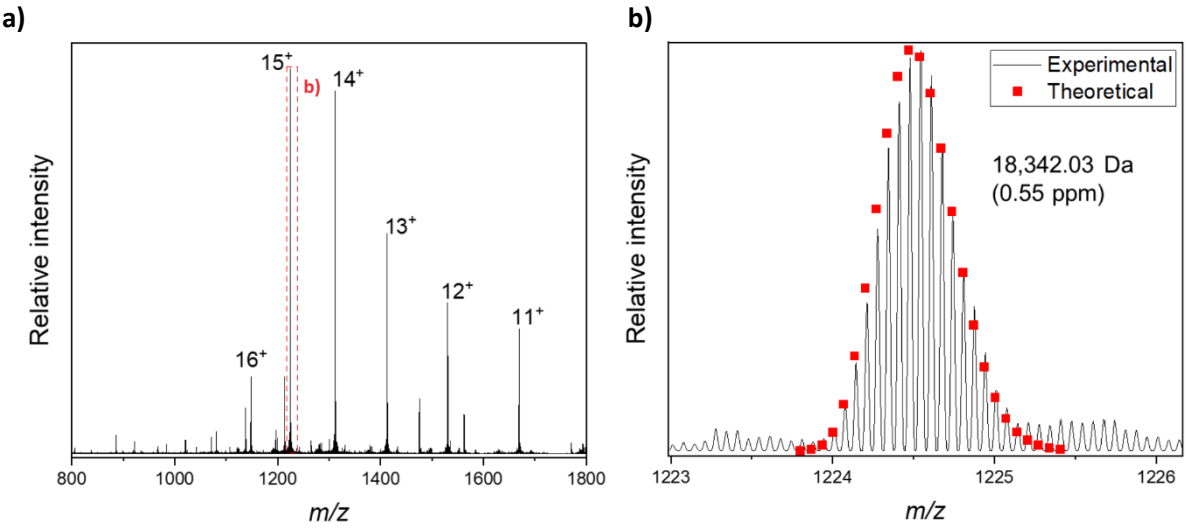


Figure 3.7-2: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ N136C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F1**.

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

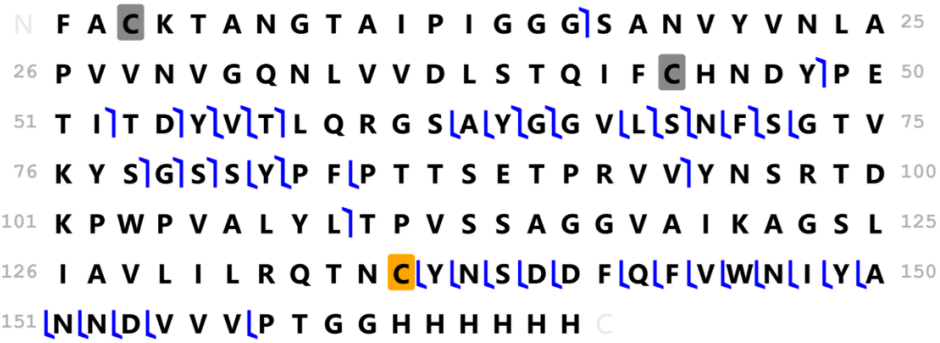


Figure 3.7-3: Fragment map of the identified proteoform of $FimH_L$ N136C after alkylation with the iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F1**.

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (collision-induced dissociation fragmentation, CID); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis and alkylated with **F1**.

No overlabelling was observed, neither on any of the cysteine residues that form the intrachain disulfide bridge nor on other nucleophilic amino acid side chains or the N- and C-termini. On the other hand, chemical modification at most of the desired positions proved to be largely quantitative. The only exception was represented by the reaction mixtures in which the IAA derivative **F2** was used as the alkylating agent. In these cases, a small number of unmodified proteoforms were identified with both cysteine variants, indicating a non-quantitative but still satisfactory conversion. Likewise, no biological modifications were detected in the alkylated samples, certainly due to the treatment with a large molar excess of DTT. Only very few impurities, and most notably, almost no truncated proteoforms of $FimH_L$ were observed. This emphasises the superiority of the *degP*-deficient *E. coli* strain HM125 for the expression of FimH variants. In a previous LC-MS/MS measurement, many non-tryptic peptide fragments and exclusively truncated His_6 tags were identified, strongly indicating severe peptidase activity. The respective $FimH_L$ variants were expressed in *E. coli* strains that still produced the periplasmic endoprotease DegP (Table 5.8-13). However, alkylation was also very successful with these $FimH_L$ N136C samples, including the modification with IAA derivative **F2** (Table 5.8-12 in Chapter 5.8.8).

Taken together, the integrity and purity of the analysed $FimH_L$ variants were highly satisfactory and met the expectations regarding the sequence, the natural disulfide bridge, and the presence of the cysteine residues 136 and 139, respectively, which were introduced by site-directed mutagenesis. The *S*-carbamidomethylation of the engineered cysteine residues was overall highly successful and predominantly quantitative in most cases. The unmodified cysteine variants may exhibit biological modifications to a considerable extent, if not treated with a suitable reducing agent. This had implications for the use of these variants, for instance, for further purification attempts or the subsequent determination of lectin activity (Chapter 3.8), and may have necessitated a reduction step with a suitable agent such as DTT. The high proportion of biological cysteine modifications is most likely attributable to the oxidative conditions that typically prevail within the periplasm. These are largely maintained by representatives of the bacterial Dsb enzyme family, which are responsible for disulfide bond (Dsb) formation in many cell envelope proteins.[498–500]

3.7.2 Analysis of $FimH_{trc}$ variants by SDS-PAGE and MALDI-TOF MS

After verifying the suitability of the alkylation protocol by LC-MS/MS, the modification of the $FimH_{trc}$ N136C variant was performed on a slightly larger scale (500 µL), using the Gate-Keeper ligands **F1** and **H1** as the alkylating agents. As previously mentioned, the bioconjugation of the N136C variant with DEG-linked AB mannoside derivatives represented the most promising combination to achieve effective photocontrol of the lectin, according to the *in silico*-docking results (Chapter 3.1).

Since the expression of $FimH_{trc}$ variants was only established at a later stage of the doctoral project and the alkylation method had already been proven highy successful, the chemical modification was, on this occasion, confirmed by MALDI-TOF MS. The respective results obtained for both the unmodified and alkylated N136C variant are compiled in Table 5.8-14 (Chapter 5.8.8). The observed *m/z* ratios exhibited appropriate mass shifts within the expected range, although the isotopic masses in the MALDI mass spectra cannot be very highly resolved. However, the mass difference of approximately 504 – 522 Da between alkylated and unmodified $FimH_{trc}$ N136C was sufficient to be visualised by a small migration shift in the SDS-PAA gel (Figure 3.7-4). The successfully **F1**- and **H1**-modified $FimH_{trc}$ N136C samples were available in sufficient amounts to allow for the investigation of their lectin activity and its control by UV irradiation (Chapter 3.8.3). The macroscopic protein aggregation, however, which was described in the previous chapter and which was particularly observed in the $FimH_{trc}$ cysteine variants, persisted to a certain extent even after chemical modification and storage of the alkylated protein at 4 °C or −80 °C, respectively. The identities of $FimH_{trc}$ R60P, the WT variant, as well as the variant lacking the signal peptide ($FimH_L$ N136C ΔN-sigpep), could also be reliably confirmed by MALDI-TOF MS (Table 5.8-14). Representative MALDI mass spectra of these variants are provided in the appendix.

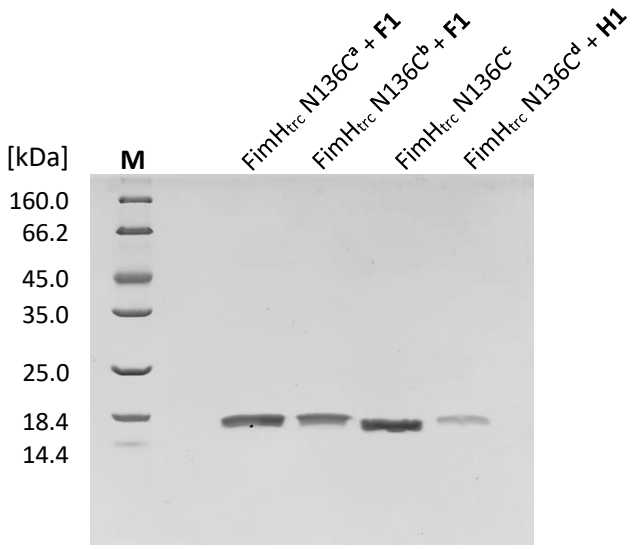


Figure 3.7-4: SDS-PAGE analysis to evaluate the purity of $FimH_{trc}$ N136C after alkylation with iodoacetamide-functionalised azobenzene mannoside derivatives **H1** and **F1.**

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™). The protein concentrations were determined in a BCA assay (Pierce™ Thermo Scientific™): **a)** $FimH_{trc}$ N136C + **F1**, 0.45 mg/mL; **b)** $FimH_{trc}$ N136C + **F1**, 0.43 mg/mL; **c)** $FimH_{trc}$ N136C (unlabelled), 0.50 mg/mL; **d)** $FimH_{trc}$ N136C + **H1**, 0.30 mg/mL. Molar masses: $FimH_{trc}$ N136C, 17831.94 g/mol; labelled with **H1**, + 503. 52 g/mol; labelled with **F1**, + 521.51 g/mol. Expression and purification of the here visualised $FimH_{trc}$ variant is documented in Chapter 3.5.1. SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 65 min at constant current (I = 20 mA).

3.8 Evaluation of FimH activity

In the course of this work, four truncated FimH variants (WT, R60P, N136C, and S139C) were produced in satisfactory yields. At least for the two cysteine variants, lectin activity has not been demonstrated thus far. Some mutations in the immediate vicinity of the substituted positions 136 and 139, such as at Asn135, Tyr137, and Asp140, are known to result in a dramatic loss of FimH activity.[318,376,501] Although the residues substituted here in favour of cysteine are not considered essential for adhesin function,[430] the lectin activity of the respective FimH variants should nonetheless be confirmed using an appropriate assay system. This also applies to the fractions of the N136C variant, which were subjected to covalent modification with IAA-functionalised AB mannosides **F1** and **H2**. The availability of a suitable assay system for investigating lectin activity was therefore crucial for achieving the central aim of this work, namely the continuation and experimental realisation of the Gate-Keeper project. Accordingly, the lectin assay employed in this context should ideally be compatible with UV irradiation.

The original concept of this work envisaged the analysis of lectin activity using saturation transfer difference (STD)-NMR spectroscopy. The latter is a powerful technique for investigating transient protein-ligand interactions. It is also generally well suited for the investigation of lectins that recognise small carbohydrate ligands with weak to moderate affinity.[502,503] As a ligand-based method, STD-NMR focuses primarily on the ^{1}H NMR signal of the ligand and does not require isotope labelling of the receptor or multinuclear and multidimensional NMR methods. The STD-NMR technique is widely used for ligand screening, but can also be exploited for determination of dissociation constants (K_D) and for epitope mapping (i.e. identification of ligand protons, which come into closest contact with the receptor). STD-NMR is based on the nuclear Overhauser effect (NOE)[504] and relies on selective proton saturation in a spectral region that only affects the macromolecular receptor. The saturation disseminates over the whole protein by spin diffusion (intramolecular NOE) and can even be transferred to a transiently bound ligand, via intermolecular ^{1}H–^{1}H cross-relaxation (intermolecular NOE). On the experimental level, the STD-NMR includes the recording of two ^{1}H NMR spectra of the ligand: one without previous saturation (off-resonance) and the other after selective saturation of the receptor (on-resonance). The resulting difference spectrum exclusively shows signals of those molecules that were temporarily in close contact with the macromolecule and thereby received saturation transfer via NOE.

However, several STD-NMR experiments that were performed in the spectroscopy department of the Otto Diels Institute of Organic Chemistry (Kiel University), in collaboration with Prof. Frank Sönnichsen and Ingo vom Sondern from the Lindhorst group, did not provide any meaningful results. No reliable STD signals of methyl α-D-mannoside could be detected when employed as ligand with the WT variant of $FimH_{trc}$ under various experimental conditions, including different spectrometer settings, buffers, temperatures and varying concentrations of the ligand or lectin, respectively. The isolated lectin domain is obviously not well suited for STD-NMR, probably due to its relatively small molecular mass (18 kDa), which may result in inefficient spin diffusion. Moreover, dissociation is decelerated for the isolated FimH lectin domain, which is therefore associated with a significantly higher ligand affinity compared to the FL-construct or fimbrial FimH. The dissociation constants of the isolated FimH-LD are usually in the low to moderate nanomolar range with deviations depending on the individual method and ligands used.[337,373,397,505] Consequently, the amount of saturated ligands that dissociate and accumulate in solution is insufficient in order to receive any STD signal. Weak- and medium-affinity interactions with K_D values in the millimolar to micromolar range are much more favourable for STD-NMR due to the faster dissociation in the timescale of the experiment.

In fact, no STD-NMR experiments with recombinant FimH been reported in the literature thus far, with the exception of a single publication, written by Palmioli and colleagues.[506] However, the authors there present an on-cell STD-NMR method using type 1-fimbriated *E. coli* bacteria. Of course, these are considerably larger than recombinant FimH and feature fimbriae-associated adhesins with affinities in the low micromolar range.[337]

A variety of methods have been published over the last decades to demonstrate and investigate FimH activity. Apart from the GFP assay[467] also employed in this work, many other, often competitive, experiments with type 1-fimbriated *E. coli* bacteria, have been described and were used for screening for suitable FimH antagonists. These include, for instance, hemagglutination,[507] ELISA,[508] flow-cytometry,[509] aggregometry,[510] and the biotin-streptavidin assay,[467] which was introduced by the Lindhorst group. Biophysical and spectroscopic techniques, particularly ITC,[511,512] SPR,[513,514] and FP[515,516] are considered the methods of choice for the investigation of recombinant FimH.[379,517] In general, these applications are well-established for measuring lectin-carbohydrate interactions and provide high sensitivity as well as the possibility to determine thermodynamic and/or kinetic data.[518,519] However, appropriate equipment and expertise are required to use these methods, which are occasionally not readily available.

In recent years, ELISA-based, cell-free *in vitro*-assays have also become increasingly popular for analysing the lectin activity of recombinant FimH or purified type 1 fimbriae. This is not least owing to the development of novel, monoclonal FimH antibodies that might simultaneously act on the lectin as allosteric effectors.[396,400,401] Horseradish peroxidase (HRP) is widely used for indirect immuno-detection in several ELISA setups, whereby the reporter enzyme may be coupled to a secondary antibody[400] or be employed as a streptavidin conjugate.[520] The peroxidase is heavily glycosylated, including many *N*-glycans with terminal α-D-mannopyranosyl residues.[521,522] Consequently, FimH activity can also be detected in a more direct manner by HRP binding and subsequent substrate conversion.[393] The immobilisation of recombinant FimH or isolated and purified type 1 fimbriae is either achieved on a highly adsorbent polystyrene surface, or indirectly, via immobilised, likewise naturally mannosylated bovine serum albumin (BSA).[520,523]

3.8.1 Conception of the FITC-*p*AP mannoside binding assay

Inspired by the literature-known FimH assay systems, we wanted to implement a fast and simple method for the analysis of FimH activity, which ideally could be combined with UV exposure. It seemed appropriate to exploit the affinity of the C-terminal His_6 tag on the expressed $FimH_{trc}$ variants to divalent cations for immobilisation, in analogy to protein purification by IMAC. Ni^{2+}-chelate-functionalised polystyrene surfaces, which are commercially available in the form of 96-well microplates plates, have proven to be an excellent platform for the immobilisation of the hexahistidine labelled lectin. Such microplates can be combined with various detection methods (colorimetric, chemiluminescent, and fluorescent) and only small amounts of protein are required in comparison to unselective surfaces (detection limit ~ 1 ng). Instead of an indirect and thus time-consuming detection method, such as by a colour reaction mediated by a reporter enzyme, the detection should be as fast as possible, in order to avoid overly advanced *cis* → *trans* relaxation of the FimH-conjugated AB units in the course of the experiment. Therefore, the use of fluorescently-labelled mannoside derivatives seemed eligible, which are also used in FP assays.[337,524] Fluorescein derivatives, for instance, are among the most common dyes employed in medical and bioscientific research due to their excellent fluorescence properties and good water solubility. Several derivatives are available with functionalisation for straightforward labelling of biomolecules, such as fluorescein *N*-succinimidyl esters (FAM-OSu/NHS-fluorescein) or fluorescein isothiocyanates (FITC).

Han and colleagues synthesised a high-affinity fluorescein-labelled FimH ligand through conversion of an aminoalkyl mannopyranoside with FAM-OSu under ambient conditions and in the presence of triethylamine.[524] In analogy to this, FITC (isomer I, fluorescein 5-isothiocyanate) was used here to obtain a suitable, fluorescent FimH ligand by conversion with *p*-aminophenyl mannoside under formation of a thiourea bridge. The aminophenyl group is supposed to both ensure a high affinity to the lectin and concurrently serve as a linker moiety, providing a larger distance between the outer residues of the binding pocket and the sterically demanding fluorophore. The reaction product **59**, here also referred to as FITC-*p*AP mannoside, is shown in Figure 3.8-1. The depiction likewise illustrates the conception of the lectin assay used in this work. In fact, the FITC-*p*AP mannoside (FITC-*p*PAP-Man) has proven to be recognised by the lectin FimH with high affinity, which was demonstrated in a competitive assay setup using red-fluorescent, type 1-fimbriated *E. coli* bacteria. The latter will be described in more detail in Chapter 4.4. A recorded UV/vis absorbance spectrum of the fluorescent mannoside **59** is provided in Figure 5.8-71 (Chapter 5.7-9).

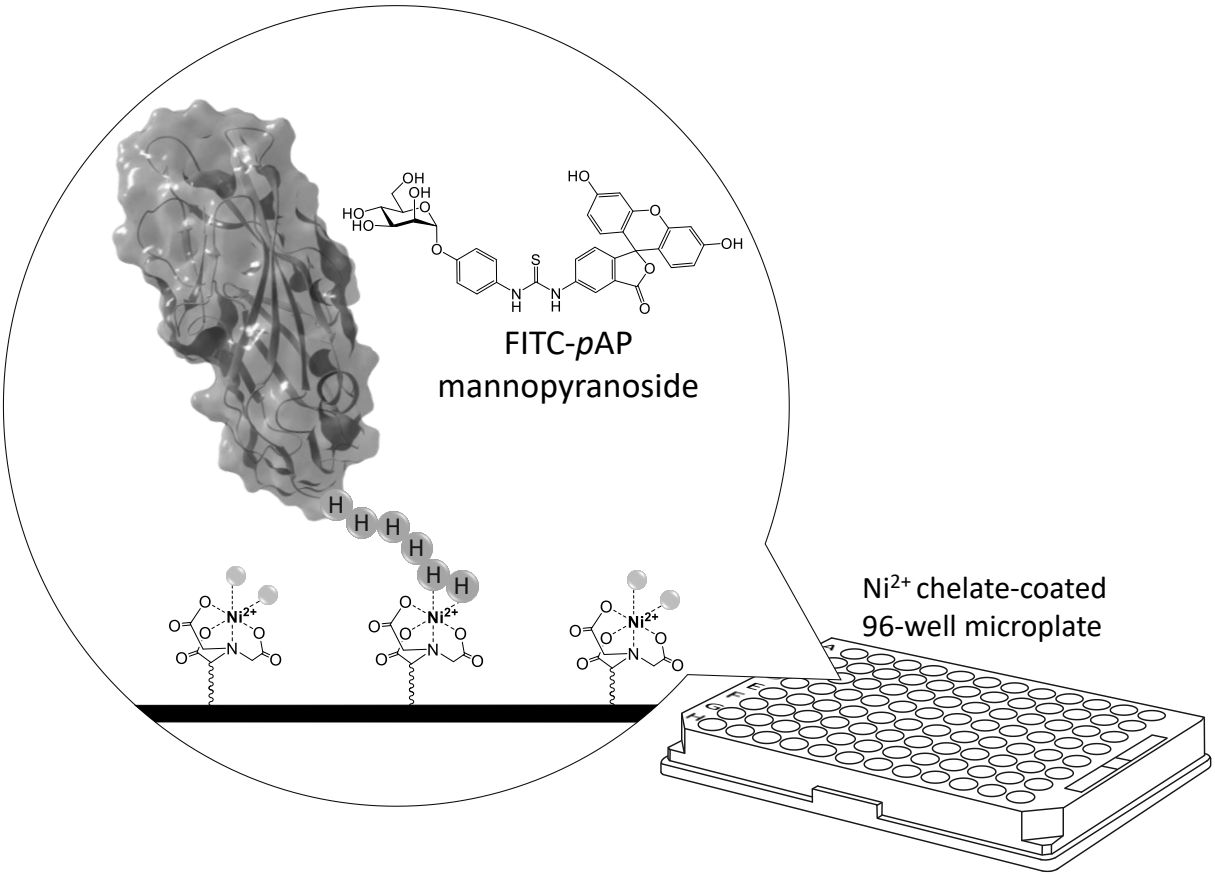


Figure 3.8-1: Design of the FITC-*p*AP mannopyranoside binding assay to investigate the lectin activity of His_6-tagged $FimH_{trc}$ variants immobilised on a Ni^{2+}-chelate-coated polystyrene surface.

An alternative setup for a lectin assay with hexahistidine-tagged FimH could include the use of commercially available Ni^{2+}-nitrilotriacetic acid (NTA) labelled fluorophores. By this means, C-terminally labelled lectins are supposed to bind to a mannan surface, similar to the bacterial inhibition assay with GFP-expressing *E. coli*. Suitable FimH antagonists could likewise be employed in order to inhibit lectin binding and to evaluate the respective inhibitory potential. Such a setup could be particularly suitable for antagonists that cannot be analysed in the conventional bacterial inhibition assay, for instance due to toxic effects arising at the commonly used concentration range (≥ 1 mM). A main disadvantage would be the high experimental and monetary effort required for the comparatively large quantities of labelled FimH and the procurement of the Ni^{2+}-NTA-conjugated fluorophores. However, a more affordable, non-selective labelling of the lectin, such as with FITC, would evidently also be possible.

3.8.2 FITC-*p*AP mannoside binding activity of unmodified $FimH_{trc}$ variants

To verify the general functionality and suitability of the FITC-*p*AP mannoside assay to evaluate lecin activity, the four chemically unmodified $FimH_{trc}$ variants (WT, R60P, N136C and S139C) were initially tested. Lectin activity was confirmed by a concentration-dependent decrease in the fluorescence signal of the FITC aglycone moiety upon serial dilution of the respective $FimH_{trc}$ variant. The results were further validated through competitive inhibition of FITC-pAP-Man binding, using methyl *α*-D-mannoside (MeMan) as a mannose-specific antagonist.

The following protocol proved suitable for successfully demonstrating lectin activity in all tested variants: An equal amount of unmodified hexahistidine-tagged $FimH_{trc}$ variants (10 µg/well) was immobilised in duplicates or triplicates on black Ni^{2+}-chelate F96-well microplates and incubated for 2 h under ambient conditions and with gentle agitation. Likewise, beforehand prepared 1:10 dilution series of the respective lectins were applied as triplicates or duplicates. PBS (100 µL/well) or BSA (10 µg/well) served as controls, to determine the baseline fluorescence (background signal) of the microplate. Subsequently, the plates were only washed with PBST since no further blocking of the microplates was necessary, according to the manufacturer's specifications. Per experiment, only a predefined part of the F96-well plate was employed.[3] The unused areas were covered with Parafilm®in the meantime. Afterwards, the microplate was hermetically sealed, enclosed in a secondary container with the supplied desiccant, and stored in the refrigerator until further use. Two equivalents of fluorescein-labelled mannoside **59** (FITC-*p*AP-Man) were added (calculation based on 10 µg of WT $FimH_{trc}$) and incubated for 1 h at ambient conditions. After removal of unbound FITC-*p*AP-Man, its residual fluorescence was measured at $\lambda_{ex/em}$ = 485/535 nm using the Tecan plate reader.

The dilution series of the $FimH_{trc}$ N136C variant, exemplarily shown in Figure 3.8-2a, clearly demonstrates the correlation between lectin concentration and measured fluorescence. A significant FITC-*p*AP-Man signal was detectable from a minimum lectin quantity of 1 ng (≙ 5.61 nM) and increased with ascending protein concentration. A similar result was achieved with the other FimH variants WT, R60P, and S139C (Chapter 5.7.9, Figures 5.7-67a – 5.7-69a). The immobilisation of the His_6-tagged lectins could be reduced by more than half through the addition of imidazole (100 mM) during the initial incubation step. In a similar manner, the lectin activity could be reduced even more significantly by the addition of MeMan (50 mM) during co-incubation with the fluorescein-labelled mannoside (Figure 3.8-2b). The mannoside-specific recognition of the lectin was thus confirmed, and the same result was achieved with the other FimH variants (Figures 5.7-67b – 5.7-69b).

[3] Pierce™ Nickel-coated 8-well strips (Thermo Scientific™) were only available for colourimetric detection but not for fluorescence measurements.

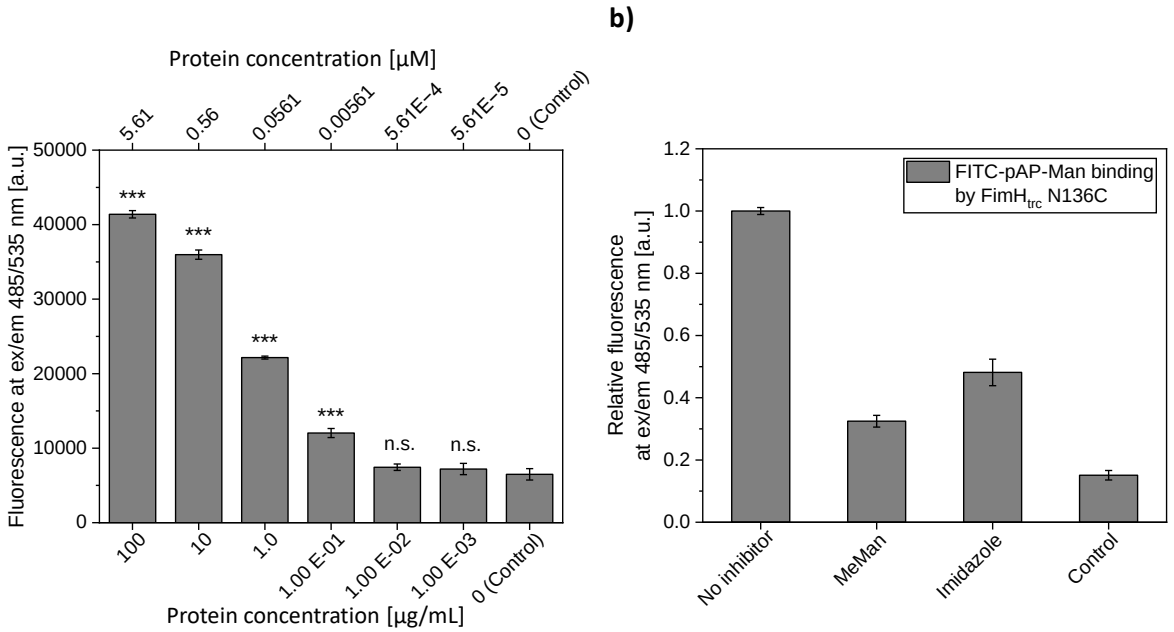


Figure 3.8-2: a) Concentration dependence and **b)** inhibition of FITC-*p*AP mannopyranoside binding activity of unmodified $FimH_{trc}$ N136C.

$FimH_{trc}$ N136C solutions in PBS (100 µg/mL; 5.61 µM as well as a premixed 1:10 dilution series, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and the lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well). The immobilisation was blocked by addition of imidazole (100 mM) and the lectin activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). Error bars represent the ± standard deviation of triplicate measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (***$p < 0.001$, two-sample t-test) compared to the control (0 µg/mL $FimH_{trc}$ N136C).

Figure 3.8-3a illustrates the corresponding FITC-*p*AP-Man binding activity of the different $FimH_{trc}$ variants in direct comparison. Although the same amount of protein was applied, the variants featured different binding activity, albeit in all cases significantly higher than BSA and the negative control. The $FimH_{trc}$ cysteine variants exhibited a significantly reduced activity by ~ 25 – 30 %, compared to the WT and proline variants. This can reasonably be explained by the formation of covalent dimers as well as possibly oligomeric states and/or lower purity. The latter may apply particularly to the S139C variant, as previously discussed in Chapter 3.5.2.[4] The N136C variant used in this experiment was purified by AEC. By this approach, any present oligomeric states could potentially not have been removed, as would have been possible by purification using SEC. Incidentally, those SEC fractions, which contained only pure, predominantly monomeric lectin, exhibited a significantly higher binding activity than the other SEC fractions (possibly containing impurities and/or FimH N136C oligomers). This was clearly demonstrated in a corresponding FITC-*p*AP-Man binding experiment (Figure 5.8-70 in Chapter 5.7.9). The relative FITC-*p*AP-Man binding activity of four different $FimH_{trc}$ variants (WT, R60P, N136C, S139C) is compiled in Figure 3.8-3b, with the respective inhibition controls (imidazole, MeMan) from various individual experiments (Figures 5.7-67b – 5.7-69b). The inhibition controls revealed an observable, graduated difference in relative FITC-*p*AP-Man binding, which correlated negatively with the basic lectin activity of the respective variants. However, this was probably due to the normalisation of the fluorescence values for comparative purposes as well as the intrinsic signal amplification of the plate reader. It does not necessarily represent an actual correlation, concerning the inhibition effects on different $FimH_{trc}$ variants.

[4] The addition of reducing agents should be avoided during immobilisation according to the manufacturer´s instructions.

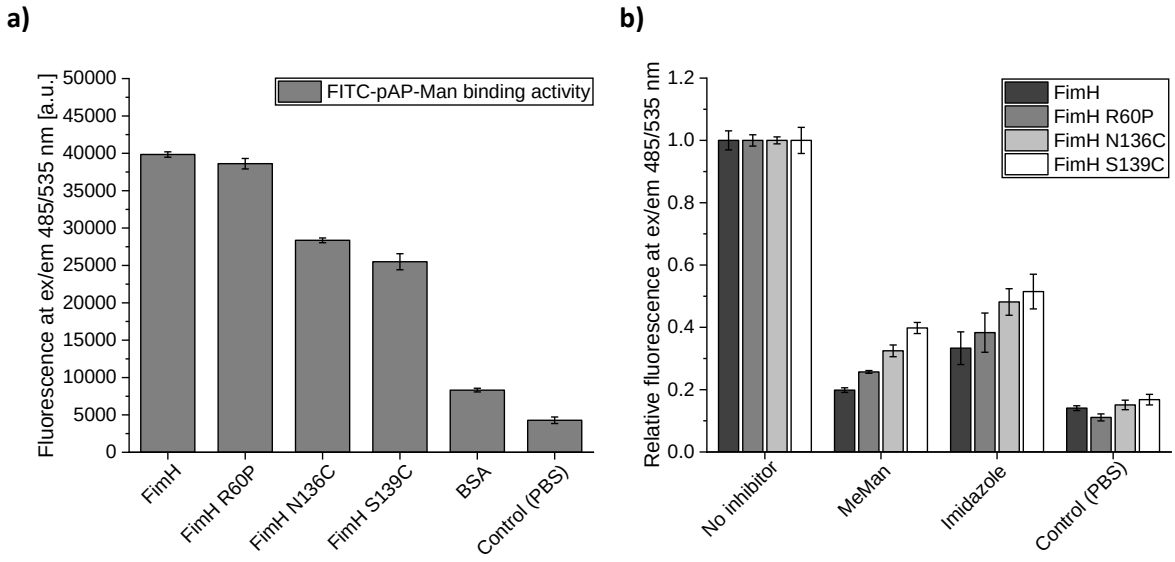


Figure 3.8-3: a) Binding of FITC-*p*AP mannopyranoside **59** and **b)** inhibition of lectin immobilisation and FITC-*p*AP-Man binding of unlabelled $FimH_{trc}$ variants.

Protein solutions (100 µg/mL in PBS, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™; 2 h, at 300 rpm, 20 °C) and incubated with 2 eq. of FITC-*p*AP mannoside (**59**, 11.2 µM, in PBS, 100 µL/well, 1 h at 300 rpm, 20 °C). The immobilisation was blocked by addition of imidazole (100 mM) and the lectin activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM) during the respective incubation steps. BSA (100 µg/mL, and 100 µL/well) and PBS (100 µL/well) were used as controls. Error bars represent the ± standard deviation of triplicate measurements. Please note that the individual results in **b)** were compiled from different experiments, shown in Figure 3.8-2b and in Figures 5.7-67b – 5.7-69b and normalised to the respective fluorescence values measured without inhibitor.

3.8.3 FITC-*p*AP mannoside binding activity of $FimH_{trc}$ variants with UV exposure

Originally, the activity of FimH variants (and its photocontrol) was intended to be monitored using NMR spectroscopic methods. For instance, it was envisaged to examine the influence of the conformation of the photoswitchable ligand on lectin activity by STD-NMR. Additionally, the photo-induced *trans* → *cis* conversion of the covalently attached AB moiety itself was actually planned to be investigated by NMR methods. More precisely, this was to be achieved by observing the fluorine signal in $FimH_{trc}$ cysteine variants that were alkylated with 3-fluoro AB derivatives (**F1**, **F2**) in ^{19}F NMR. As described earlier (Chapter 3.3), a generally good distinguishability of the aromatic ^{19}F signal, before and after UV irradiation at 365 nm, was previously demonstrated using the soluble IAA-functionalised fluoro-AB mannoside derivatives. However, as can be gathered from the spectrum in Figure 3.8-4b, no fluorine signal could be obtained in a corresponding ^{19}F NMR experiment (2000 scans) with $FimH_{trc}$ N136C that was covalently labelled with IAA-fluoro-AB derivative **F1**. As expected, also no aromatic signal could be detected in the ^{1}H NMR experiment, which could be assigned to the AB unit. Only NMR signals of residual glycerol, which was used as an additive during protein purification, are visible in the corresponding ^{1}H NMR spectrum (Figure 3.8-4a).

Evidently, the concentration of the modified lectin in the sample was too low to obtain any NMR signals. Measurement of UV/vis absorption therefore also proved to be unsuitable. For the NMR measurement, the most concentrated alkylated $FimH_{trc}$ sample was already employed and due to the progressing macroscopic aggregation, as discussed earlier, a further concentration of the labelled protein was not possible. Therefore, neither information regarding the photostationary state (PSS) of the AB unit after UV exposure of the labelled FimH nor on the kinetics of thermal *cis* → *trans* relaxation could unfavourably be obtained by NMR spectroscopy.

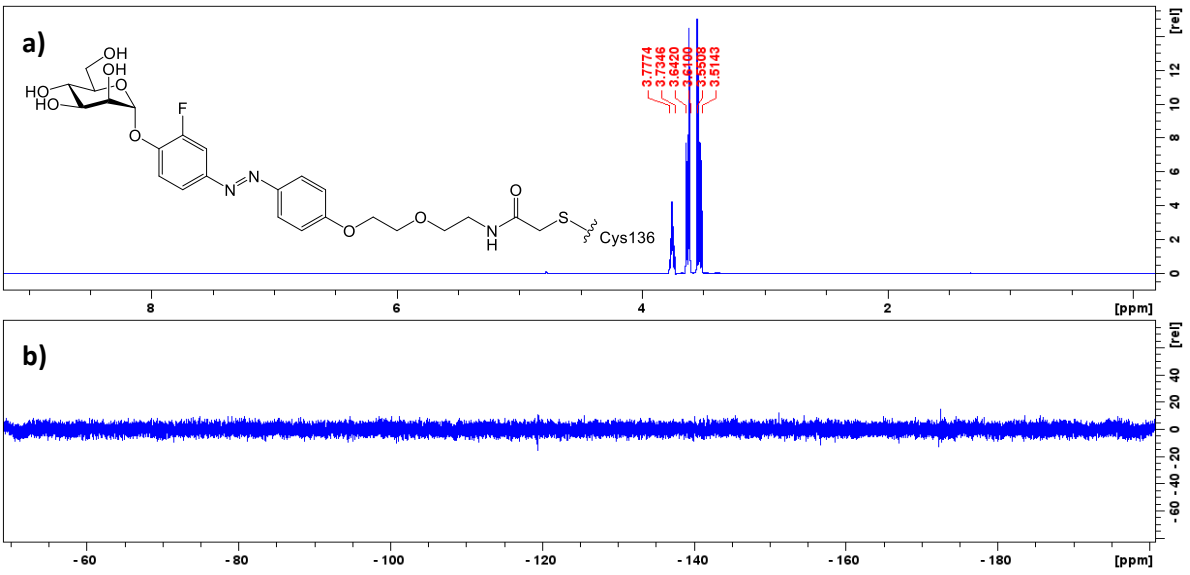


Figure 3.8-4: Attempts to investigate modified $FimH_{trc}$ N136C, which was covalently labelled with 3-fluoro-AB derivative **F1**, using **a)** ^{1}H NMR (500 MHz) and **b)** ^{19}F NMR spectroscopy (471 MHz).
No protein or azobenzene signals were obtained. The ^{1}H NMR signals between 3.77 and 3.51 ppm can be assigned to glycerol, which was added during protein purification. Concentration of labelled lectin: 0.224 mg/550 µL; Measured in PBS/D_2O, 9:1, 298 K.

Nevertheless, photocontrol of FimH activity could still be investigated in the context of the FITC-*p*AP-Man binding assay. Therefore, the latter required optimisation with regard to the implementation of UV irradiation. As a little reminder, *trans* → *cis* photoisomerisation of the AB unit was expected to facilitate binding of the covalently attached mannose moiety to the lectin's binding site. Based on the results from the covalent docking simulations (Chapter 3.1.2), particularly the investigation of the N136C variant was prioritised. This variant, especially in combination with the DEG-linked AB mannosides **F1** and **H1**, appeared to be most promising for achieving the desired photocontrol of lectin activity.

Initially, the FITC-*p*AP-Man binding activity of unmodified N136C was tested after exposure to UV light (365 nm for 5 min), following its immobilisation onto the Ni^{2+}-chelate-coated microwell plate. The WT variant was likewise analysed for comparison purposes, and the result is provided in Figure 5.8-72 (Chapter 5.7.9). Unfavourably, UV exposure of the immobilised proteins resulted in a significant reduction in the observable FITC-*p*AP-Man binding signal. This effect was observed both when irradiating the lectin-containing immobilisation solution directly on the plate, and after subsequent washing with PBST. Although it was initially unclear whether the reduced signal was due to impaired lectin activity or impaired immobilisation, UV irradiation ideally should not affect unmodified protein variants. In particular, FITC-*p*AP-Man binding must not decrease upon UV, since a comparable decrease is expected for the modified variants. Consequently, the lectin samples had to be irradiated prior to surface immobilisation, as on-plate UV exposure proved unsuitable.

As previously mentioned, neither the PSSs nor the kinetics of *cis* → *trans* relaxation of the AB moiety, covalently attached to WT $FimH_{trc}$, could be analysed by ^{19}F NMR spectroscopy or by UV/vis spectroscopy. Both techniques could have provided useful implications for subsequent lectin assays involving UV exposure. Characterisation of the synthesised IAA-functionalised AB derivatives **F1** and **H1** by NMR spectroscopy (Chapter 3.3) revealed a very efficient photoconversion (PSS_{365}: *trans* ratio 2 – 5 %). Additionally, based on the thermal relaxation experiments, a *cis* ratio of approximately 89 % can be expected after 2 h of incubation at room temperature. However, estimating whether and to what extent protein alkylation affects the photochemical properties of the covalently attached azobenzene moiety remains difficult.

The incubation times for immobilisation and binding of the FITC-*p*AP mannoside should be significantly reduced in order avoid an overly advanced relaxation in the course of the experiment. To this end, corresponding control experiments with shortened incubation times (without UV exposure) were performed with unmodified $FimH_{trc}$ N136C. Immobilisation appeared to be already completed after 10 – 15 min, as the fluorescein signal began to fluctuate and showed no further increase with longer incubation times (Figure 5.8-73). Nevertheless, the control experiment with FITC-*p*AP mannoside **59** revealed a more distinct correlation between measured fluorescence signal and incubation time. As shown in Figure 5.8-74, FITC-*p*AP-Man binding gradually decreased with shorter incubation, but remained significantly above the PBS control. In a further control experiment, UV pre-irradiation of the lectin was combined with surface immobilisation and FITC-*p*AP-Man binding, both combined with shortened incubation times (10 min each). For this purpose, unmodified N136C and WT variants were transferred into autoclaved 2 mL glass vials and pre-irradiated at 365 nm for 5 min on ice at varying intensities (10, 30 and 50 %). Encouragingly, this treatment did not reduce lectin activity, as no decrease in the fluorescence signal was detected, unlike in previous on-plate irradiation experiments (Figure 5.8-75). On the contrary, UV pre-exposure resulted in a moderate increase in FITC-*p*AP-Man binding, which was particularly pronounced for the N136C variant.

Since the opposite tendency would be expected for the alkylated FimH variants, this slight increase was not considered problematic for the final experiments: Accordingly, the $FimH_{trc}$ N136C variant covalently modified with either of the DEG-linked AB mannosides (**F1** or **H1**, 20 µg/well, in triplicates) was tested in the lectin assay following UV irradiation (5 min at 365 nm, 30 % intensity, on ice). The resulting FITC-*p*AP-Man binding signal was compared to that of the same alkylated protein without previous UV treatment. The unmodified protein (irradiated vs. not irradiated) and PBS served as controls. The results of two independently performed experiments are shown in Figure 3.8-5.

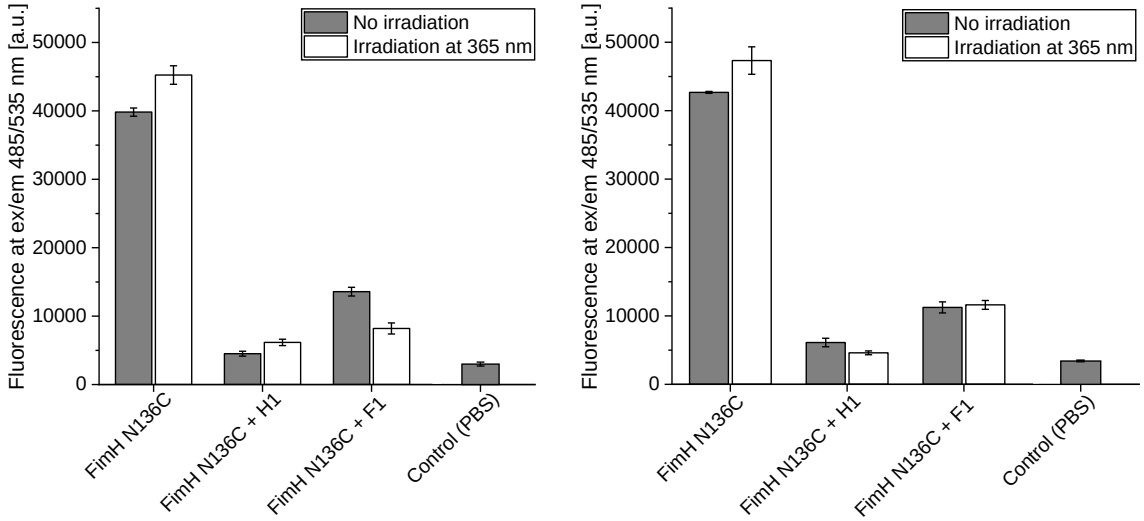


Figure 3.8-5: Binding of FITC-*p*AP mannopyranoside by modified and unlabelled $FimH_{trc}$ N136C with and without prior irradiation at 365 nm.

Unmodified $FimH_{trc}$ N136C and the lectin labelled with azobenzene mannoside derivatives **H1** and **F1** (200 µg/mL, 11.24 µM in PBS/glycerol, 9 :1; 100 µL/well) were exposed to UV light at 365 nm, 5 min (40 % intensity, LED NC4U133AT, Nichia Corporation) and immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™; 10 min, at 300 rpm, 20 °C). The lectin activity was tested using 1 eq. of FITC-*p*AP mannopyranoside (**59**, 11.2 µM, in PBS, 100 µL/well, 10 min at 300 rpm, 20 °C). PBS (100 µL/well) was used as control. The protein concentrations were determined in a BCA assay (Pierce™ Thermo Scientific™). Both figures show the results of two independent experiments. Error bars represent the ± standard deviation of triplicate measurements.

Disappointingly, the anticipated reduction in FITC-*p*AP-Man binding activity, induced by photoirradiation and the associated *trans* → *cis* photoconversion, was not observed. Thus, the desired photocontrol of lectin activity could not be realised within the implemented assay setup. Instead, the measured binding signals of irradiated and non-irradiated lectins fluctuated without any meaningful tendency. Notably, regardless of UV exposure, the observable lectin activity of the modified $FimH_{trc}$ N136C variants was significantly diminished compared to the unmodified protein. In the alkylated samples, the FITC-*p*AP-Man binding was reduced by up to 75 – 90 %, in some cases barely exceeding the background signal of the PBS control. The substantial loss of FITC-*p*AP-Man binding activity of $FimH_{trc}$ N136C after modification with the ligands **F1** and **H1** was confirmed in a further experiment with UV exposure (Figure 5.8-76). The same result was observed in two corresponding assays without photoirradiation (Figure 5.8-77). In these latter assays, FITC-*p*AP-Man bindin by the modified FimH variants remained drastically impaired (up to 50 % reduction), even though incubation times for both immobilisation and mannoside binding had been increased again. Nevertheless, the addition of methyl *α*-D-mannopyranoside (MeMan) still resulted in a further decline in fluorescence signal. This indicated that the chemically modified lectin retained residual activity and that its CRD remained at least partially accessible for (small) mannoside ligands, albeit to a much lesser extent than in the unmodified protein.

3.8.4 Discussion

In summary of Chapter 3.8, a comparably fast and reproducible lectin assay was established for the investigation of C-terminally His_6-tagged *α*-D-mannoside-specific lectins, which was successfully applied to analyse the activity of different $FimH_{trc}$ variants. However, the actually intended photocontrol of lectin activity could not be achieved, neither by STD-NMR experiments nor within the framework of the FITC-*p*AP-Man binding assay. Instead, it was demonstrated that chemical modification of the $FimH_{trc}$ N136C variant with DEG-linked AB mannosides resulted in a drastic, albeit not complete, abrogation of lectin activity.

The rationale for the extensive loss of lectin activity subsequent to alkylation cannot be derived from the findings obtained from the FITC-*p*AP-Man assay. The results from the LC-MS/MS measurements (Chapter 3.7) provided no evidence for any incorrect folding or degradation of the $FimH_L$ cysteine variants caused by the employed alkylation protocol. This applies at least to the $FimH_L$ variants that were expressed in the *degP*-deficient expression strain *E. coli* HM125. The formation of covalent homodimers of $FimH_{trc}$ cysteine variants (Chapter 3.5.2), which could result in less efficient immobilisation and possibly ligand binding, should not be expected any longer after modification with IAA derivatives. It should be noted that only the highly active, predominantly monomeric SEC fractions of $FimH_{trc}$ N136C were used for the alkylation reaction as well as in the corresponding FITC-*p*AP-Man assays with modified and unmodified $FimH_{trc}$. As already mentioned in Chapter 3.6, the modified $FimH_{trc}$ variants tend to undergo pronounced macroscopic aggregation, but this also applied to the unmodified cysteine variants of $FimH_{trc}$ (Chapter 3.5.2).

A plausible explanation for the substantial loss of lectin activity could be a blockage or occupation of the CRD (i) either by an excess of IAA ligands, which had not been converted in the alkylation reaction, or (ii) by the covalently bound ligand itself, regardless of the photoisomeric state of its AB unit. Both options would probably prevent the FITC-*p*AP mannopyranoside from efficient binding. Once a high-affinity ligand (such as an excess of not covalently attached IAA derivative) is bound to the recombinant lectin, a complete removal may be difficult to achieve. In the more recent literature, recombinant FimH is therefore usually expressed in supplemented minimal medium. This is done to avoid unintended binding of mannosides as in the case of the FimH crystal structure 1UWF, for instance.[373] The modification sites (Cys136, Cys139) are located quite deliberately in the immediate proximity of the mannoside-binding pocket. As a consequence, the lectin activity could easily be impaired. As already mentioned, the most promising combination of cysteine variant and ligands (FimH N136C, alkylated with DEG-linked AB mannosides **H1** and **F1**) was investigated in the lectin assay, described here. According to the molecular docking simulations (Chapter 3.1.2), the DEG linker appeared to be exactly long enough to bridge the distance between the alkylated cysteine residue 136 and the mannoside binding site of the lectin. A binding event was predicted to occur only if the azobenzene moiety adopts the *cis*-configuration. A shortening of the linker by one ethylene glycol unit to result in an insufficient length to permit effective binding of the covalently attached mannoside. However, the explicit binding of a ligand within the binding pocket is not absolutely necessary to obstruct access for other mannosides such as the FITC-*p*AP mannoside. Recognition of the latter can simply be impaired by steric hindrance, especially since the fluorophore within the aglycone is sterically demanding. Nevertheless, the connecting *p*-aminophenyl group was actually inserted to prevent steric stress between the CRD and the fluorophore.

It is also conceivable that the tyrosine gate of FimH might play a role in hindering access to the CRD. π-π Interactions may occur between the AB unit of the covalently bound ligand and the phenyl group of Tyr48 and/or Tyr137, which could stabilise the steric blocking of the binding pocket by the ligand. In fact, such interactions were observed in some of the molecular docking simulations, even if these were generally exceptions (Figure 3.8-6). However, it can be argued that a comparatively weak inhibitor such as MeMan was able to further reduce the FITC-Man binding activity of the modified FimH variants (Figure 5.8-77, Chapter 5.7.9), indicating that access to the binding pocket was possible (at least partially, and/or for antagonists with small aglycone moieties).

a) **b)**

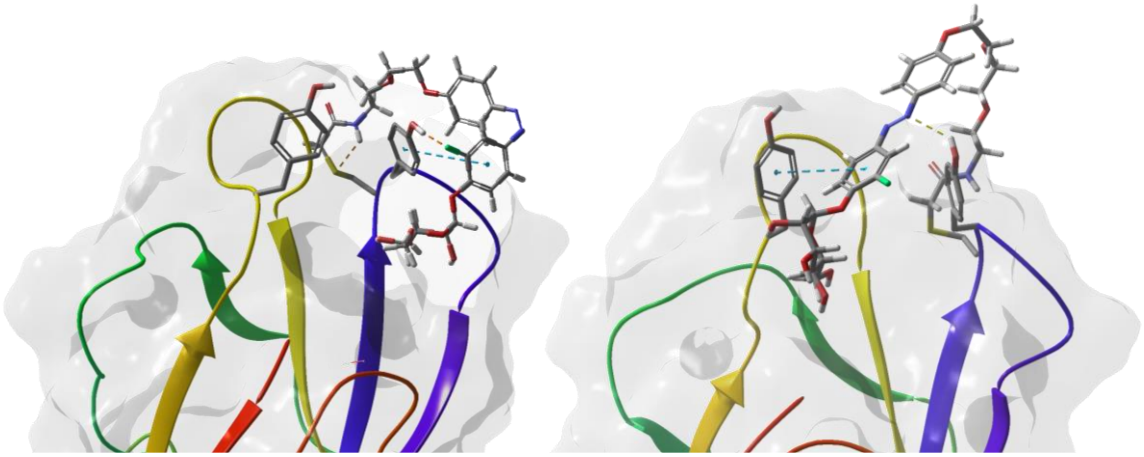


Figure 3.8-6: π-π Interactions of covalently attached fluoroazobenzene mannosides **a) F1** and **b) F2** with the FimH tyrosine gate residues observed in docking simulations.

a) π-π Interactions between Tyr137 and the fluoroazobenzene moiety of **F1**, covalently bound to the closed-gate structure (PDB: 1UWF) of FimH variant N136C. **b)** π-π Interactions between Tyr48 and the fluoroazobenzene moiety of *trans*-**F2**, covalently ligated to the closed-gate structure (PDB: 1UWF) of FimH variant N136C. The mannoside ligands are not located within the FimH binding pocket but could nevertheless hinder access to the CRD for other mannosides. Intramolecular interactions are depicted as dashed lines: π-π interactions (cyan blue); hydrogen bonds (orange). Illustrations were generated by Glide docking simulations, followed by *in silico*-alkylation and energy minimisation using the Maestro suite (Schrödinger Inc.).

In fact, FimH activity can generally be inhibited by antagonists that are not recognised by the lectin in the classical sense. This was at least coincidentally observed in a standard bacterial adhesion inhibition test with *p*-nitrophenyl *β*-D-mannoside (*β*-*p*NPMan), conducted as part of a side-project during the doctorate. As clearly demonstrated in Figure 3.8-7, a distinct inhibition effect on bacterial adhesion was achieved in a 1:2-dilution series. The observable inhibitory potential corresponded approximately to that of the universal reference inhibitor MeMan ($RIP_{\beta\text{-}p\text{NPMan}}$ = 1.025). The resulting inhibition curve is not perfectly sigmoidal, but a further increase in concentration (> 40 mM) could not be achieved due to the poor water solubility of *β*-*p*NPMan. Needless to say, FimH is demonstrably unable to bind *β*-D-mannosides under the conditions of the bacterial adhesion inhibition assay.[329] However, extensive interaction of the aromatic aglycone with one of the tyrosine gate residues is quite possible even without direct binding of the mannoside moiety within the binding pocket. This might nevertheless result in steric blocking of the CRD, thereby preventing bacterial adhesion to the mannan surface. Consequently, *p*-nitrophenyl *β*-D-mannoside could also be regarded as non-orthosteric inhibitor of the bacterial lectin FimH (i.e. an inhibitor that does not bind to the active site but impairs function through steric effects).

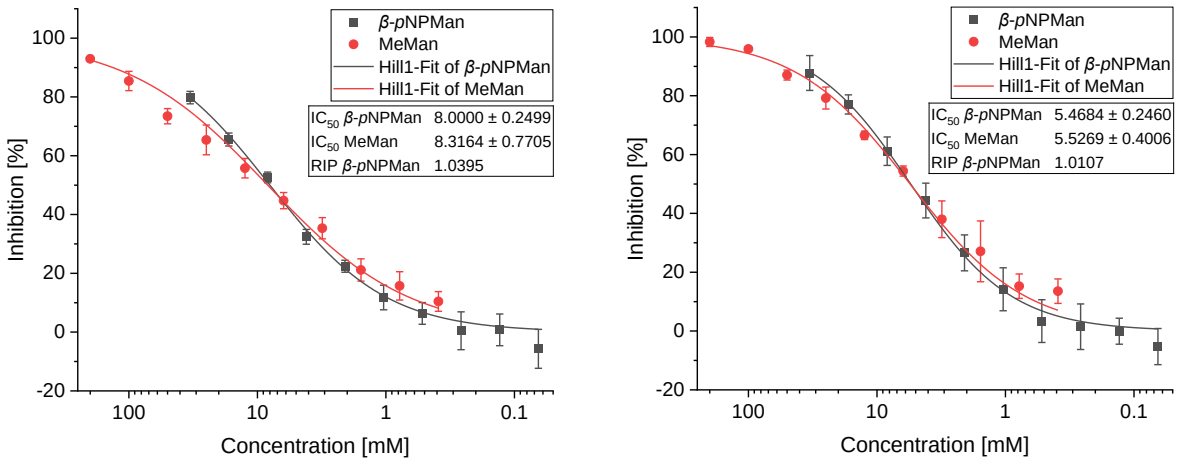


Figure 3.8-7: Results from two bacterial adhesion inhibition experiments using *p*-nitrophenyl *β*-D-mannoside (*β*-*p*NPMan) as a non-orthosteric antagonist.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation from quadruplicate measurements. An average RIP = 1.0205 ± 0.0204 was calculated from the shown experimental results. RIP values represent the inhibitory potency of each test compound compared to the reference inhibitor methyl *α*-D-mannopyranoside (MeMan) tested on the same microplate (IP MeMan ≡ 1).

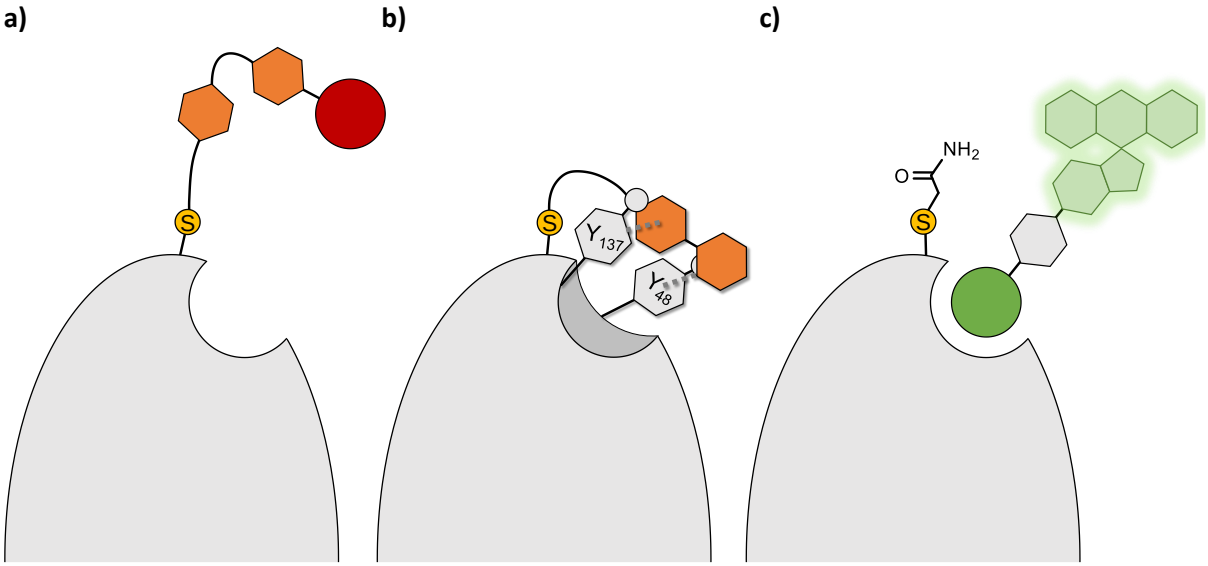


Figure 3.8-8: Alternative cysteine modifications of FimH N136C for evaluation purposes and to explore possible causes for the observed loss of lectin function.

Replacement of the mannoside with **a)** an *α*- or *β*-D-glucoside or **b)** no glycoside moiety, in order to prevent lectin binding events. **c)** *S*-Carbamidomethylation with iodoacetamide, to demonstrate the suitability of the used protocol regarding the maintenance of lectin function. After modification according to **a)** – **c)** FimH, lectin activity could be assessed in a FITC-*p*AP-Man binding assay. The lectin is depicted in grey, with the introduced cysteine sulfur atom (yellow) and the tyrosine gate residues (grey hexagons). Azobenzene units are shown as orange hexagons in *cis*- (bent line) and *trans*-configuration (straight line). FITC-*p*AP mannoside is illustrated in green and D-glucoside as a red circle.

In order to evaluate whether the covalently bound mannoside ligand reaches into the binding pocket depending on the respective photoisomeric state of its AB unit, alternative modification strategies could be considered. As illustrated in Figure 3.8-8a, for instance, glucose analogues could be prepared and alkylated to FimH for comparative purposes and to avoid the actual lectin binding event. Alternatively, the lectin could be modified with an IAA-functionalised AB derivative without a glycoside moiety (Figure 3.8-8b). However, the implementation of a corresponding FimH cysteine modification would presumably be hampered by the poor water solubility of IAA-functionalised azobenzene derivatives, which can be expected in the absence of a hydrophilic glycoside moiety. Eventually, the *S*-carbamidomethylation of FimH N136C with the much less sterically demanding iodoacetamide (Figure 3.8-8c) could be considered. Bioconjugation with such a small molecule would possibly enable better assessment of whether the alkylation reaction generally leads to lectin inactivation, or whether this observation is more likely due to steric hindrance of access to the CRD.

3.9 Summary and conclusion

The main project presented in this thesis involved the homologous expression of two recombinant cysteine variants (N136C and S139C) of the truncated lectin domain (FimH-LD), and the chemical synthesis of photoswitchable azobenzene (AB) mannoside ligands for the modification of these FimH variants. To this end, the mannoside ligands were functionalised with a terminal iodoacetamide (IAA) group, to achieve the *S*-carbamidoalkylation of the introduced cysteine residues with the synthesised ligands. In this way, an experimental setup was established, which was intended to enable the control of FimH activity by UV exposure. According to the Gate-Kepper concept, binding of the covalently attached ligand at the lectin's CRD should only be possible when the AB moiety adopts the *cis*-configuration, whereas the *trans*-configuration should prevent binding of the covalent ligand. In other words, access to the binding pocket for other mannosides is already blocked by the binding of the covalently attached, *cis*-configured ligand, whereas access to the CRD remains possible in the *trans*-configuration.

The conditions for accessibility or the prevention of access to the FimH binding pocket were analysed in detail by *in silico*-docking simulations using the Maestro software from Schrödinger. The positions N136 and S139 were identified as potentially suitable sites for amino acid substitution and chemical modification in close proximity to the binding pocket. The use of tri- and diethyleneglycol (TEG and DEG) units appeared suitable as linkers between the AB moiety and the IAA function. Based on the number and quality of simulations according to which binding of the mannoside by the lectin through the covalently attached ligand is possible, the modification of FimH N136C with DEG-linked AB mannosides (**H1**, **F1**) seemed the most promising combination to realise the intended photocontrol of lectin activity in line with the Gate-Keeper project.

Two alternative approaches for the chemical synthesis of (fluoro-)azobenzene mannosides with DEG and TEG linkers and the introduction of the terminal IAA function were pursued. For the non-fluorinated AB derivatives (**H1**, **H2**), the linear approach, which was already presented in the Master thesis,[438] proved to be more recommendable. The semi-convergent approach was particularly convenient for the synthesis of the fluoro-AB derivatives (**F1**, **F2**). The corresponding route began with the preparation of the mono-*O*-allyl-protected (fluoro-)AB unit, followed by etherification with asymmetrically azido-functionalised OEG linkers. Subsequently, the *O*-allyl group was removed and the phenolic hydroxyl group was glycosylated with a suitable acetyl-protected mannosyl donor. After Staudinger reduction, disclosing the free amine, the terminal IAA function was introduced. Apart from IAA derivatives, less reactive *N*-acetylated analogues (**51** – **54**) were synthesised, which were intended as soluble FimH antagonists for bacterial adhesion inhibition experiments.

The photochemical characteristics of all synthesised AB mannoside derivatives were determined by NMR and UV/vis spectroscopy, elucidating an almost quantitative *trans* → *cis* photoconversion and good thermal stability. The photochromic properties of the *N*-acetylated analogues (**51** – **54**) as well as those of the aminomethyl AB mannoside derivative (**56**) were also analysed at 37 °C and in aqueous solution in order to simulate the conditions of bacterial inhibition experiments. The respective mannosides were employed as inhibitors to prevent the adhesion of type 1-fimbriated *E. coli* bacteria to a mannan surface, with the aim of demonstrating their general ability to be recognised by the lectin FimH. A quite high inhibitory potential (RIP values between 333 and 2241) was observed for all tested AB mannoside derivatives in these experiments. In particular, the mono-fluorination of AB in the *ortho*-position to the mannosidepyranoside unit further increased the inhibitory effect of the respective inhibitors.

In agreement with our earlier publications, UV irradiation of the inhibitors and the associated photoconversion of the AB hinge resulted in a reduction in inhibition of bacterial adhesion when compared to the corresponding unexposed *trans*-isomers.[407,408] In contrast to the previously used pre-irradiation of the inhibitors in solution, on-plate exposure allowed results of comparatively better quality and reliability. A negative impairment of the results, for instance due to a temperature increase as a consequence of photoirradiation, could be excluded in appropriate control experiments.

The expression and purification of the FimH-LD variants was kindly supported by Prof. Axel Scheidig and were performed in the laboratory of the Structural Biology group (Kiel University). Initially, the His_6-tagged WT $FimH_L$ and cysteine variants (N136C, S139C) were expressed using a pPKL241-based expression system and purified by IMAC according to the protocol of Schembri and colleagues.[470] The preparation was completed by an additional purification step via SEC. Despite the relatively low expression levels and correspondingly poor $FimH_L$ yields, the cysteine variants could be *S*-carbamidoalkylated with the IAA-functionalised (fluoro-) AB mannosides (**H1**, **H2**, **F1**, and **F2**) following the procedure by Crankshaw and Grant.[218] The integrity of the target proteins as well as the success of the amino acid exchange and chemical cysteine modifications was confirmed by detailed LC-MS/MS analysis. The latter was kindly performed by Dr. P. Kaulich from the working group of Prof. A. Tholey at the Institute of Experimental Medicine (Kiel University). Notably, a number of biological cysteine modifications could be detected, which are favoured under the oxidative conditions in the periplasm and were proven to be completely reversible under reducing conditions. The expression level of His_6-tagged FimH-LD was significantly improved by changing the expression system in favour of *E. coli* HM125 pFimH-His_6-NT-trc. The expression host was optimised for this purpose,[337,482] particularly in combination with periplasmic cell disruption.[480] The C-terminal coding *his_6*-sequence needed to be introduced on the precursor plasmid, which was kindly provided by Prof. R. Glockshuber (ETH Zürich, Switzerland). Although the subsequent purification was considerably impeded by, in some cases, excessive protein aggregation, the aforementioned $FimH_{trc}$ variants could be obtained in acceptable purity and quantity. These difficulties applied in particular to the expression of the cysteine variants, and were further complicated by covalent dimerisation as well as oligomerisation and/or impurities. In addition to the aforementioned $FimH_{trc}$ truncates, the R60P variant was expressed. The proline mutation affects folding of the β-bulge and thus stabilises the "compressed" low-affinity (LA) conformation. The latter is associated with fast ligand dissociation and thus comparably low K_D values, but the R60P variant is capable of conformational change to the high-affinity (HA) conformation upon mannoside binding.[397]

The activity of the expressed $FimH_{trc}$ variants was successfully demonstrated in the FITC-*p*AP-Man binding assay, which includes fluorescence detection of the eponymous mannoside after immobilisation of the lectins on a black, Ni^{2+}-chelate-functionalised microplate. The reliability of the assay was clearly demonstrated by several dose-response relationships as well as the inhibition of lectin activity or surface immobilisation. The activity of $FimH_{trc}$ N136C, modified with DEG-linked AB mannosides (**H1**, **F1**), was also investigated in the FITC-*p*AP-Man binding assay. Unfortunately, the lectin immobilisation was not compatible with direct on-plate UV exposure. Accordingly, the alkylated lectin had to be pre-irradiated, and incubation times were therefore significantly reduced. However, the modified $FimH_{trc}$ N136C variants revealed a dramatic, almost complete loss of observable lectin activity, regardless of whether UV exposure was previously applied or not. The residual FITC-*p*AP-Man binding activity could be further reduced by the addition of methyl *α*-D-mannoside as an inhibitor.

The alternative approach with STD-NMR was not suitable for the investigation of FimH activity due to the unfavourably low mass and nanomolar K_D of the isolated lectin domain. As no other appropriate detection methods (e.g. ITC, SPR, FP) were readily available, the envisaged photocontrol of FimH activity could not be demonstrated on the basis of the introduced lectin assay. This applied at least to the Cys136 modification performed here with DEG-linked AB mannosides.

On the condition that the lectin is still intact and/or correctly folded after the chemical modification procedure, the following explanations (Figure 3.9-1) for the loss of observable FimH activity appear plausible: (i) The covalently attached mannoside ligand may be permanently bound within the binding cavity, regardless of the *cis*- or *trans*-configuration of its AB unit. This could particularly be the case when the OEG linker is too long (Figure 3.9-1a). (ii) The binding pocket may be occupied with an unreacted, not covalently attached ligand molecule. This is likewise plausible since the IAA-functionalised ligands were added in excess for the alkylation reaction and may not become entirely removed by the ensuing small-scale gel filtration (Figure 3.9-1b). Eventually, the access to the binding cavity may be sterically hindered, especially for mannosides with sterically demanding aglycone moieties (e.g. *p*AP-FITC). However, this must not necessarily apply to smaller molecules such as MeMan, which was proven to even further reduce the already strongly affected activity of the alkylated lectin. (iii) This conformational state could be stabilised by π-π interactions of the AB unit with the tyrosine gate residues, without requiring the actual binding of the mannoside moiety by the CRD (Figure 3.9-1c).

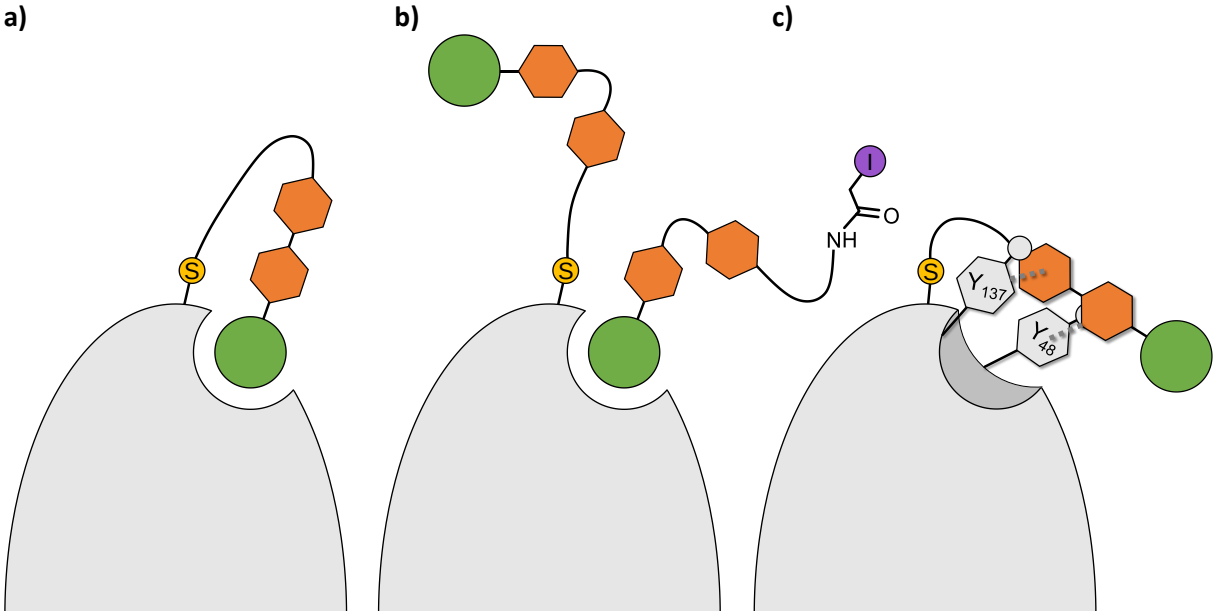


Figure 3.9-1: Possible explanations for the observed extensive loss of FITC-*p*AP-Man binding activity after *S*-carbamidoalkylation of $FimH_{trc}$ N136C with DEG-linked AB mannoside ligands.

The lectin is depicted in grey, with the introduced cysteine sulfur atom (yellow) and the tyrosine gate residues (grey hexagons). Azobenzene units are shown as orange hexagons in *cis*- (bent line) and *trans*-configuration (straight line). The α-D-mannoside moiety is illustrated as a green circle.

Apart from alternative approaches of FimH cysteine modification to examine the loss of lectin activity, (Chapter 3.8.3, Figure 3.8-8), a structural analysis, for instance by X-ray crystallography, could provide more valuable and profound insights into the structural and/or mechanistic basis underlying the observed loss of activity. Obviously, FimH can be crystallised quite easily, as demonstrated by the high number of X-ray crystal structures deposited in the PDB.[371] Among them are many representatives of the isolated FimH-LD and various crystallisation conditions have been published. A crystal structure of the modified protein could unambiguously clarify whether the lectin becomes inactive due to a permanent complexation of the mannoside in the FimH binding pocket. If this is the case despite the *trans*-configuration of the AB unit, and the bond is only formed under a certain tension, this could possibly occur as a result of conformational strain within the FimH binding site.

Binding of the *trans*-configured ligand could therefore have a noticeable conformational impact on the FimH structure. Given these circumstances, the applied FimH modification would represent a first step towards experimentally investigating the "local" catch bond effect, which was predicted in MD simulations under tensile force by Stirnemann et al. and that was shown to be entirely independent from domain separation.[394] Favourably, the *trans*-configuration is permanently adopted under exclusion of light, which is easily compatible with normal crystallisation conditions. The use of certain NMR techniques, such as chemical shift perturbation (CSP) or relaxation experiments in particular, would also be highly suitable in this context as these can conveniently be combined with UV exposure. Moreover, these techniques allow the monitoring of even subtle conformational changes under dynamic conditions in solution.[374,525,526] However, such experiments would require ^{13}C-/^{15}N-isotope labelling of the target protein and, thus, optimisation of expression in minimal medium, especially with regard to the FimH cysteine variants.

In conclusion, although photocontrol of FimH activity could not be demonstrated yet, this work nonetheless achieved significant progress and important methodological advances over previous approaches. A straightforward and efficient synthetic strategy was developed for photoresponsive mannoside ligands, including novel fluorinated AB derivatives, that feature excellent photochemical properties. Their biological relevance was demonstrated by strong inhibitory effects in bacterial adhesion assays, confirming specific lectin recognition and the influence of the AB configuration on the inhibitory effect. For the first time in our group, expression of recombinant FimH truncates, including cysteine variants, was achieved, enabling a nearly quantitative covalent modification of the lectin via selective *S*-carbamidoalkylation using IAA-functionalised ligands. In terms of bioconjugation efficiency and selectivity, this represents a significant improvement over previous strategies, such as affinity-guided DMAP labelling. In addition, a sensitive and robust lectin assay was developed, which directly detects FimH activity via a fluorophore-labelled mannoside, without relying on the reaction of secondary reporter enzymes. The immobilisation strategy, based on the affinity of the His_6-tagged protein to a Ni^{2+}-chelate surface, proved particularly efficient and reproducible, even though more recent literature favours untagged variants over His_6-tagged constructs. Together, these developments provide a solid basis for future studies on the activity and light-dependent control of FimH and similar lectins.

Although the precise reasons for the observed loss of lectin activity remain unclear, the result demonstrates that targeted covalent inhibition of FimH is worth to be pursued further. This work has also highlighted the critical importance of employing site-specific and highly selective bioconjugation methods. The next chapters will explore new ideas and approaches, based on these insights, that were previously unconsidered, opening new perspectives for future research.

3.10 Future prospects

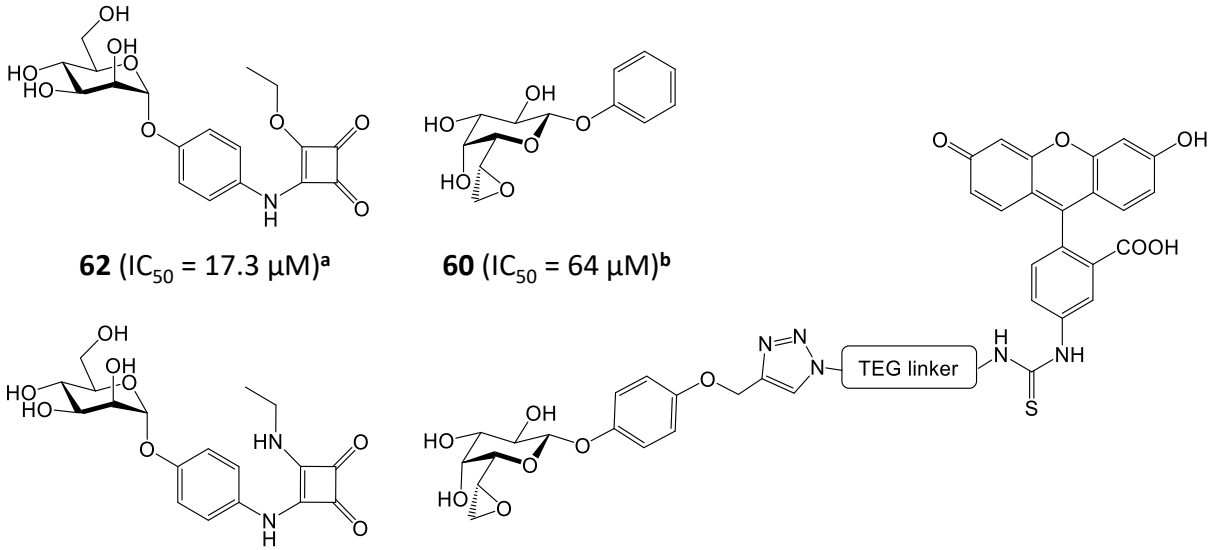
3.10.1 Covalent inhibition of FimH

Although actually unintended, the approach employed here presumably represents one of the few examples of permanent inhibition of a lectin by covalent modification of a FimH cysteine variant. For some decades, the attention to covalent inhibition in clinical research increased only slowly, due to concerns about selectivity and potential toxicity.[527] Nevertheless, in addition to some famous historical representatives (e.g. aspirin and penicillin), a growing number of covalent drugs have been developed in recent years. Among them are some kinases and proteases that were successfully approved for cancer treatment.[528] Although covalent inhibition offers some invaluable advantages, glycomimetics for the covalent targeting of lectins are still entirely underrepresented.[529] Conveniently, the carbohydrate specificity of lectins can be ideally utilised for affinity-guided, covalent modification of the CRD, which could significantly counteract potential selectivity problems. Under equilibrium conditions, lectins generally feature only a low to moderate affinity for natural and artificial carbohydrate ligands, respectively. Lectin recognition is furthermore usually characterised by rapid ligand dissociation, which is often compensated by multivalent interactions. Covalent targeting, by contrast, enables inhibition in a virtually irreversible manner.

In 2017, Titz and colleagues first published a promising covalent inhibitor against the bacterial lectin LecA from *P. aeruginosa*, a clinically relevant, Gram-negative pathogen.[530,531] It can cause severe wound infections as well as nosocomial infections in humans, and is associated with high morbidity and mortality, especially for immunocompromised or chronically debilitated patients.[532,533] Of particular concern is the increasing global spread of isolates with resistance to various antibiotics.[534,535] *P. aeruginosa* secretes the soluble, homotetrameric lectins LecA (PA-IL, D-galactose-specific) and LecB (PA-IIL, L-fucose-specific). These represent important virulence factors, which are particularly essential for bacterial adhesion, host cell invasion and biofilm formation.[536–539] Since the latter plays a significant role in multidrug resistance, inhibition of LecA and LecB bears great potential for the development of alternative treatment of *P. aeruginosa* infections.[540,541] Titz et al. exploited the presence of the Cys62 residue within the LecA CRD as a target for two diastereoisomeric phenyl *β*-D-galactoside derivatives, each equipped with an electrophilic epoxide at the C-6 position, for site-specific, covalent cysteine modification.[530] The lectin revealed a pronounced diastereoselectivity towards the 6_D-epoxygalactoheptoside isomer **60** (Figure 3.10-1). Efficient inhibition was demonstrated with an IC_{50} of 64 µM in a competitive fluorescence polarisation (FP)-based binding assay,[542] and decreased even further over time, attributable to ongoing covalent bond formation. Although ring opening of the epoxide could not be observed in the solved crystal structure (PDB: 5MIH), covalent modification was confirmed by mass spectrometry. The covalent labelling strategy was proven to be well suited for biological applications, and the FITC-labelled epoxy galactoside **61** was successfully used for LecA-specific imaging of *P. aeruginosa* biofilms.

The covalent inhibition of the bacterial lectin FimH has already been discussed by the Lindhorst group, for instance in the context of *α*-D-mannosides featuring squaric acid (SA) monoamide as part of the aglycone.[543] The latter can undergo an amidation reaction with primary amines, resulting in cross-linked diamide derivatives, which can be exploited for bioconjugation.[544,545] The SA-monoamide-functionalised *α*-D-mannoside **62** (Figure 3.10-1) revealed an excellent inhibitory potential against FimH in an ELISA-based adhesion inhibition assay as well as in inhibition experiments with GFP-expressing bacteria.[543,546]

Due to the N-terminus of Phe1 being in direct proximity to the CRD, bioconjugation of the SA-monoester with the lectin at this position seemed possible. This would result in inhibition of bacterial adhesion just in the same way as recognition of the mannose moiety in the sense of an antagonist would do. However, after a more extensive investigation involving docking studies and further inhibition experiments with suitable test substances, a covalent inhibition was considered rather unlikely. For instance, the SA-diamide derivative **63** can definitely not be covalently attached to the lectin, but nonetheless exhibited a significantly stronger inhibitory effect than the SA-monoamide derivative **62**, which is theoretically able to form a FimH-SA-diamide conjugate.



62 (IC_{50} = 17.3 µM)[a] **60** (IC_{50} = 64 µM)[b]

63 (IC_{50} = 6.4 µM)[a] **61** (IC_{50} ~ 109 µM)[b,c]

Figure 3.10-1: Literature-known glycomimetics tested in the context of covalent lectin inhibition.[530,543] **a)** Bacterial inhibition was tested in a competitive assay with GFP-expressing, type 1-fimbriated *E. coli* PKL1162.[467] **b)** Lectin inhibition was tested in a competitive binding assay with recombinant LecA, monitored by fluorescence polarisation.[542] **c)** The shown IC_{50} value does not actually refer to the depicted FITC-labelled molecule **61**, but to the alkyne precursor, prior to the click reaction with the azide-functionalised FITC derivative.

A targeted chemical modification of FimH was already attempted a few years ago by the Lindhorst group, but not with the aim of covalent inhibition. Instead, protein labelling for the investigation of lectin-ligand interactions was intended. In this context, mannosides with photoactive groups, such as aryl azides, benzophenone and diazirines have been employed to facilitate photoaffinity labelling of the lectin.[547–549] Upon photolysis, these form highly reactive nitrenes, ketyl radicals, and carbene intermediates, which thus can insert into C – H, N – H, or O – H bonds of diverse biomacromolecules, including peptides and proteins.[550–552] When tested with different model peptides, diazirines initially appeared promising and revealed a preference for modifying tyrosine residues.[548] With the intention of primarily targeting the tyrosine gate residues, the FimH-LD was exposed to UV exposure in the presence of high-affinity diazirine-functionalised mannoside ligands. However, the labelling yield was extremely low (< 1.6 %), and the modified amino acid residues could therefore not be identified. Consequently, this approach was not pursued further.[549]

The example of this work convincingly demonstrates how effectively FimH activity can be inhibited by covalent modification. In addition to the envisaged photocontrol of lectin activity, a renewed pursuit of covalent FimH inhibition could be a promising objective for future research projects. Now that a functional expression system has been established in the Lindhorst group, FimH is henceforth available as the conformationally flexible R60P variant as well as in the form of the WT variant, which permanently adopts the HA conformation. With the FITC-*p*AP-Man binding assay, a suitable test system for the investigation of FimH activity is also readily available. Nevertheless, an additional spectroscopic method (e.g. ITC, SPR) would be useful in order to enable the determination of kinetic and thermodynamic data. Instead of cysteine modification, a different ligation strategy would have to be chosen, since usually no other cysteine residues are present within the lectin domain (apart from Cys3 and Cys44, forming the intrachain disulfide bond). Considering the previous attempts at cysteine-independent FimH modification by the Lindhorst group, the selectivity of the labelling strategy should be significantly increased. All other amino acid residues that could be selectively addressed are either not present within the FimH primary structure (Met), occur more than once (Thr, Ser, Tyr, Arg), or are not positioned in the vicinity of the CRD (Trp).[553,554]

An interesting approach could therefore be the selective modification of the terminal amino function on Phe1. The N-terminus is both sterically and chemically accessible, and is positioned relatively close to the binding pocket, between the CRD loop 1 and the clamp loop. Favourably for covalent inhibition, the amino group of Phe1 is critical for lectin function, since it is involved in mannoside complexation.[318] Some bioconjugation strategies with enhanced N-terminal regioselectivity have been published, for instance imine formation using benzaldehyde derivatives,[555] the formation of isoquinolinium salts with 2-ethynylbenzaldehydes,[556] and azolation using oxazolines.[557] These methods exploit the lower basicity of the N-terminal amine compared to the ε-amine of lysine, and therefore require strict pH control to ensure discrimination between both functional groups. 2-Pyridinecarboxaldehydes (2-PCA)[558] and triazolecarbaldehydes (TA4C)[559] are regarded as specific reagents for N-terminal bioconjugation. After initial imine formation, the adjacent amide nitrogen is able to nucleophilically attack the electrophilic imine carbon, resulting in the formation of a 4-imidazolinone (Figure 3.10-2). The corresponding imine formed with lysine residues is unstable under physiological conditions, since cyclisation is unlikely in the absence of a contiguous amide. The reaction of amines with N-terminally selective and N-terminally specific agents proceeds under mild conditions. However, the reactivity and long-term stability of the resulting conjugates can vary considerably depending on the respective target protein.[563]

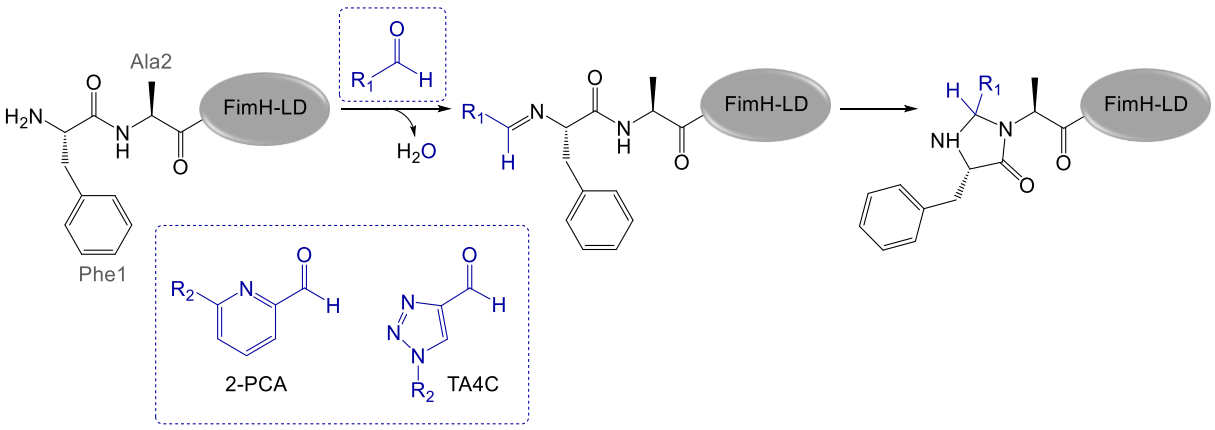


Figure 3.10-2: Principle of N-teriminal specific modification of the FimH lectin domain (FimH-LD) using 2-pyridinecarboxaldehydes (2-PCA) and triazolecarbaldehydes (TA4C).

3.10.2 Modification of FimH using an expanded genetic code

Certainly, selective N-terminal modification could be not only used to ligate covalent inhibitors to FimH. Likewise, photoswitches with or without a mannoside moiety could be conjugated to the lectin via N-terminal labelling in order to ultimately realise photocontrol of FimH activity. Especially for the continuation of the Gate-Keeper project, the introduction of unnatural amino acids would be an interesting alternative, which could be achieved by expanding the genetic code in *E. coli*. This approach would essentially correspond to the photo-bioorthogonal ligand tethering (photo-BOLT) method mentioned in Chapter 1.2.2.[217,231]

The "amber codon" (TAG/UAG) is the least abundant codon in *E. coli,* and thus also the rarest of the three termination codons.[560] The redundancy of the genetic code can be exploited to incorporate non-canonical amino acids into proteins. For example, in organisms that utilise the 21st proteinogenic amino acid, selenocysteine, the latter is encoded by the "opal" stop codon (TGA/UGA).[561] The reassignment of the amber stop codon for the incorporation of unnatural amino acids *in vivo* is the aim of the so-called amber codon suppression technology. Besides the unnatural amino acid to be introduced, the complementary tRNA, as well as the corresponding aminoacyl-tRNA synthetase are required at the genetic level.[562] The synthetase is responsible for specific recognition and adenylation of a certain amino acid, the subsequent transesterification of the resulting aminoacyl-adenylate onto the tRNA, and final release of the charged aminoacyl-tRNA. Since its introduction by Wang et al., amber codon suppression has become the most popular method for the introduction of various unnatural amino acids into a range of different organisms.[563]

A variety of aminoacyl-tRNA synthetases had been evolved, originating from the archaebacterial species *Methanococcus jannaschii*, which features an unusual amber codon assignment for the incorporation of tyrosine. Some orthogonal tRNA/aminoacyl-tRNA synthetase pairs can be acquired, for instance, from non-profit plasmid repositories. The plasmid AzoRS-4, for example, can be obtained from Addgene (Plasmid no. 182884) and enables arabinose-inducible expression for the incorporation of amino acids with azobenzene side chains. As demonstrated by Amiram and colleagues, the encoded synthetase not only mediates incorporation of *p*-phenylazo-phenylalanine, but also of red-shifted *ortho*-fluoro-derivatives, which are responsive to visible light.[564] Photoswitchable amino acids with additional groups, suitable for additional functionalisation, have been incorporated into proteins by Hoppmann and co-workers,[565] but the tRNA/synthetase-encoding plasmids were regrettably not deposited on Addgene. However, this does not apply to the plasmid pAzF (Plasmid no. 31186). This plasmid mediates *in vivo*-incorporation of *p*-azido-L-phenylalanine, which is both synthetically and commercially available.[566] The introduced azido function could be addressed by click chemistry with arbitrary AB (mannoside) derivatives. A copper(I)-catalysed azide-alkyne cycloaddition (CuAAC), for instance, might be particularly suitable for selective *in vitro*-modification of recombinant FimH.[567–569] Bioorthogonal click reactions, such as strain-promoted azide-alkyne cycloaddition (SPAAC) or traceless Staudinger ligation can be considered if the labelling of living *E. coli* bacteria is projected.[227,229]

The mentioned unnatural amino acids could preferably be introduced as replacements for aromatic amino acids that are positioned in proximity to the FimH-CRD, such as Phe1, Tyr48 and Tyr137 (Figure 3.10-3). Particularly, the tyrosine gate residues could represent a promising target for this approach due to their direct location on the boundary of the binding pocket. The structural similarity between tyrosine and *p*-azido-phenylalanine, as well as the conformational flexibility of Tyr48, could be advantageous in this context.

As discussed earlier (Figure 3.9-1), the tyrosine gate could be involved in stabilising a conformation of an AB mannoside ligand that would potentially result in steric blockage of the CRD without actual mannoside binding. Such an undesirable interaction could expectedly be avoided by modification of the tyrosine gate residues themselves instead of other closely positioned amino acids. The amber codon only rarely terminates essential genes in *E. coli*.[570] In any case, amber codon suppression would inevitably lead to an accumulation of non-functional proteins within the host cells. Consequently, adverse effects on the fidelity of the bacterial proteome are to be expected.[571] Hence, Lajoie and colleagues took up the challenge to replace all (in total 321) amber codons with the "ochre" termination codon (TAA/UAA) in *Escherichia coli* substrain MG1655.[572] The 3rd generation of the resulting strains, optimised for amber codon suppression, can be ordered from Addgene (*E. coli* C321.ΔA.exp, strain N° 49018). In contrast to the previous generations, the strain C321.ΔA.exp can be cultured at 37 °C and features a normal mutation frequency.

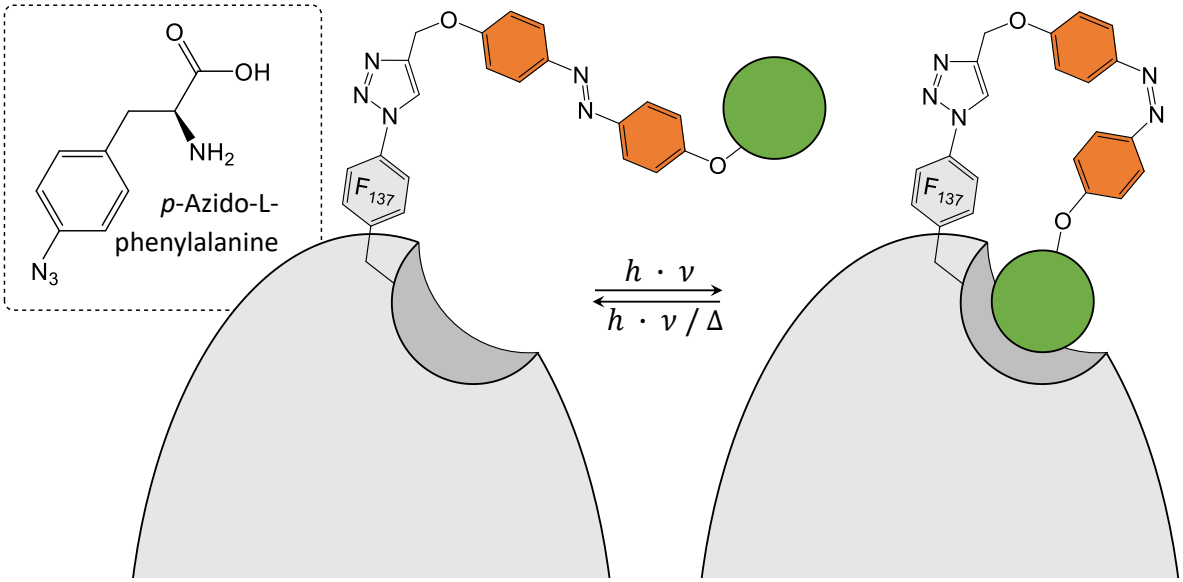


Figure 3.10-3: Alternative approach to achieve photocontrol of FimH activity using the amber codon suppression method for the incorporation of selectively addressable, artificial amino acids.

By using the expanded genetic code, *p*-azido-L-phenylalanine (F137) could be incorporated instead of Tyr137 to possibly reduce unfavourable π-π interactions between the AB unit and the tyrosine residue, which may contribute to steric blockage of the CRD. The lectin is depicted in grey, with F137 (grey hexagon), which could be converted, e.g. via a click reaction (CuAAC), using suitably functionalised azobenzene derivatives. The *α*-D-mannoside moiety is indicated as a green circle.

The expression plasmids pPKL241 and pFimH-His_6-NT-trc do not possess any UAG stop codon. Hence, only one base triplet exchange, thus encoding the respective unnatural amino acid, would be required. Overexpression of the FimH truncates in C321.ΔA.exp would presumably be impaired due to the secretion of the protease DegP into the periplasm. If this is the case, a chromosomal gene knockout of *degP* in the expression strain could be considered. However, using the extended genetic code offers great potential, particularly for FimH modification in living, type 1-fimbriated *E. coli* bacteria.

In contrast to cysteine-based modification, which potentially can cause undesirable side reactions on other secreted or membrane-associated proteins, no severe selectivity problems should be expected with bioorthogonal alkyne-azide click chemistry. In fact, bioorthogonal reactions are especially popular for chemical modification on cell surfaces and, as such, have immensely contributed to the success of metabolic oligosaccharide engineering (MOE).[573,574]

According to the *E. coli* K12 genome sequence, deposited on EcoCyc.org, only a single amber codon is present within the type 1 fimbrial gene cluster, terminating the *fimB* gene. The plasmid pPKL174, which mediates expression and assembly of type 1 fimbriae in *E. coli* PKL1162, frequently employed in our bacterial adhesion inhibition assays,[467] is not compatible with the aforementioned plasmids (pAzF and AzoRS-4), since all three possess the same antibiotic resistance gene for selection (chloramphenicol acetyltransferase, *cat*). The precursor plasmid pPKL4, however, also encodes the entire *fim* gene cluster, features an ampicillin resistance gene as a selectable marker (β-lactamase, *bla*), and contains no UAG-termination codons on its pBR322 backbone.[296,297] A transduction with fluorescent reporter protein-encoding genes, as described in the following chapter, could be envisaged to facilitate monitoring of bacterial adhesion. It should be mentioned that the λ prophage was removed in *E. coli* C321.ΔA.exp. Therefore, the λ *att*B site is presumably not available for recombination, which would be required for chromosomal integration of the *mcherry* gene using pOSIP-KL-*mcherry* (for details, see Chapter 4). However, the precursor strain C321.ΔA could be used instead, or the transformation with readily available plasmids for the expression of reporter genes could be a suitable alternative (e.g. pETM-G, which encodes a modified GFP variant from *Aequorea coerulescens*, Chapter 4).[575,576]

4. Red-fluorescent bacteria for adhesion studies

4.1 Introduction

Fimbriae have always been of great scientific and medical interest, as they are essential factors for bacterial adhesion, host colonisation, and biofilm formation, and thus substantially contribute to the virulence of many pathogenic bacteria.[577–579] The chaperone-usher (CU) pathway is a highly conserved secretion system in Poteobacteria and Cyanobacteria.[580] Although, an impressive diversity of CU fimbriae has been evolved, particularly within the family of Enterobacteriaceae.[283] At least 38 different types of CU fimbriae have been identified in the *E. coli* pangenome, many of which are associated with distinct pathotypes.[291,581] In uropathogenic *E. coli* (UPEC) for instance, 15 different types of adhesive pili have been described, and among these, type 1, type 3, type 9, S, P, and F1C fimbriae are probably the most abundant.[582,583] Statistical evaluation of the CU fimbriae distribution revealed that pathogenic *E. coli* often present more different fimbrial types (UPEC: $\sim 6 - 7$) than commensal isolates ($\sim 3 - 4$).[584]

In addition to type 1 fimbriae, P pili represent another well-studied class of CU fimbriae, which are exclusively expressed in pathogenic *E. coli* bacteria. The designation of P fimbriae derives from the ability of corresponding UPECs to bind antigens of the eponymous human blood group system P1PK (formerly referred to as blood group system P).[585] P fimbriae are strongly associated with complicated urinary tract infections (UTIs), particularly acute pyelonephritis, and are consequently also referred to as Pap pili (pyelonephritis-associated pili).[206–208] The coding gene cluster, accordingly termed the *pap* operon, contains 11 genes, of which *papI*, *B*, *A*, *H*, *C*, *D*, *E*, *F*, *G*, and *K* were assigned to functional Pap proteins.[588,589]

As typical representatives of CU pili, P fimbriae possess the characteristic structure and assembly illustrated in Figure 4.1-1a.[290,590] As soon as the Pap-proteins are released into the periplasm by the SecYEG machinery and the intrachain disulfides are formed, the pilins become initially stabilised via donor strand complementation (DSC) with the chaperone (PapD) and are then supplied to the pore protein (PapC).[591–593] In analogy to FimD, the pore protein PapC forms a 24-stranded transmembrane (TM) β-barrel and likewise features four periplasmic domains.[592,594] These include the internal plug domain, which closes the inactive *apo*-channel and becomes displaced into the periplasm upon PapC activation. Similar to the processes during the formation of type 1 fimbriae, the N-terminal domain (NTD) of PapC recruits the incoming PapD-pilin complex. The tandem C-terminal domains (CTDs) catalyse the assembly of the pilin subunits into the linear Pap pilus through donor strand exchange (DSE) with the N-terminal extension (Nte) of the subsequent Pap protein, thereby stabilising the incomplete Ig-like fold of the pilin subunits.[316] The pilus rod is formed by more than 1000 copies of PapA, which wind up helically with 3.3 subunits per turn to a width of 6.8 nm.[595] The tip fibrillum is continued by the adapter proteins PapK, PapE (5 – 10 copies), and PapF. Located at the distal end of the P fimbrial tip is the two-domain adhesin PapG.[596,597] In contrast to type 1 fimbriae, the P pilus rod is terminated on the proximal side by a further subunit (PapH), which remains within the pore due to its inability to undergo DSE.[598,599] The P5 pocket is absent in the hydrophobic groove of the PapH subunit, whereas P1 – P4 are intact. The G1 strand of the chaperone can therefore easily stabilise the folding of PapH by DSC. The termination subunit can likewise be incorporated into the pilus, but the Nte of a subsequent pilin cannot bind to the unoccupied groove of PapH. Consequently, another DSE event is not possible, and further pilus growth is effectively abrogated.[599]

Similar to FimB, the paralogous PapB serves as a transcriptional regulator of the *pap* gene cluster, together with papI.[600] In contrast to type 1 fimbriae, P pilus expression is not regulated by phase variation, but by methylation of two GATC sites upstream of the *pap* promoter.[601] Interestingly, PapB is likewise able to repress the type 1 fimbriae expression by inversion of the *fim* switch.[602] Distribution analysis of CU-fimbrial operons in different pathotypes of *E. coli* revealed that the *pap* cluster is usually located in pathogenicity-associated islands (PAIs), which are prone to undergo recombination events and horizontal gene transfer.[291,603,604]

The 34.5-kDa adhesin PapG is conventionally subdivided into three classes (PapG-I, PapG-II, and PapG-III), which exhibit slightly varying receptor specificities and thus different host tropisms.[603,605,606] All three PapG variants recognise glycosphingolipids (GSLs) with a common minimal galabioside receptor motif (*α*-D-galactopyranosyl-*β*-D-galactopyranoside). PapG-I preferentially binds to globotriaosylceramide receptors (GB_3Cer: Gal(*α*1→4)Gal(*β*1→4)GlcCer, also known as the P^k antigen), which are abundant on the surface of different mammalian tissues, including human erythrocytes, neurons and renal cells.[607] However, the *papG*-I allele is relatively rare in UPEC, and its role in pathogenesis is therefore poorly understood. In contrast, PapG classes II and III are clinically highly relevant in the context of UTIs,[608–610] and feature increased binding preferences towards further representatives of the globo-series glycosphingolipids (Figure 4.1-1b). These particularly include globotetraosylceramide (GB_4Cer, also known as the P antigen), globopentaosylceramide (GB_5Cer, also referred to as stage-specific embryonic antigen 3, SSEA-3), and the Forssman antigen, respectively, but also other iso-receptors in the uroepithelium.[611–615] Although *E. coli* carrying P fimbriae with the clinically more relevant PapG classes II and III preferentially adhere to GSLs with longer carbohydrate chain lengths, the known PapG inhibitors are generally based on derivatives of the minimal galabioside unit (Gal(*α* 1→4)Gal).[616–618]

In analogy to various FimH antagonists, aromatic substituents at the reducing end (*O*-1) or the non-reducing position (*O*-3') of the galabioside, as well as multivalent galabioside-bearing glycoclusters, greatly enhance the inhibition effect on P fimbriae-mediated bacterial adhesion. This was demonstrated in corresponding hemagglutination inhibition assays as well as by surface plasmon resonance.[619,620] Classical methods for the detection of P-piliated *E. coli* encompass the just-mentioned hemagglutination of human erythrocytes (blood group A^+), ELISA-based assays as well as the P-specific particle agglutination test (PPA test) developed for clinical applications.[621–624] Screening for *papG* genes in a library of more than 300 UPEC isolates have revealed a fourth *papG* allele (PapG class IV), but the preferred isoreceptors of this variant are still unknown.[625,626] Force-spectroscopic measurements of P-fimbriated *E.* coli using optical tweezers revealed a transitory compensation of tensile forces (≥ 28 pN).[627–629] However, this was exclusively attributed to the elastic stretching of the P fimbriae up to a maximum extension of approximately 5 µm. The fimbrial elongation was associated with extensive unwinding of the helically arranged PapA subunits within the pilus rod. Unlike FimH, a catch bond mechanism of the adhesin PapG could not be observed.

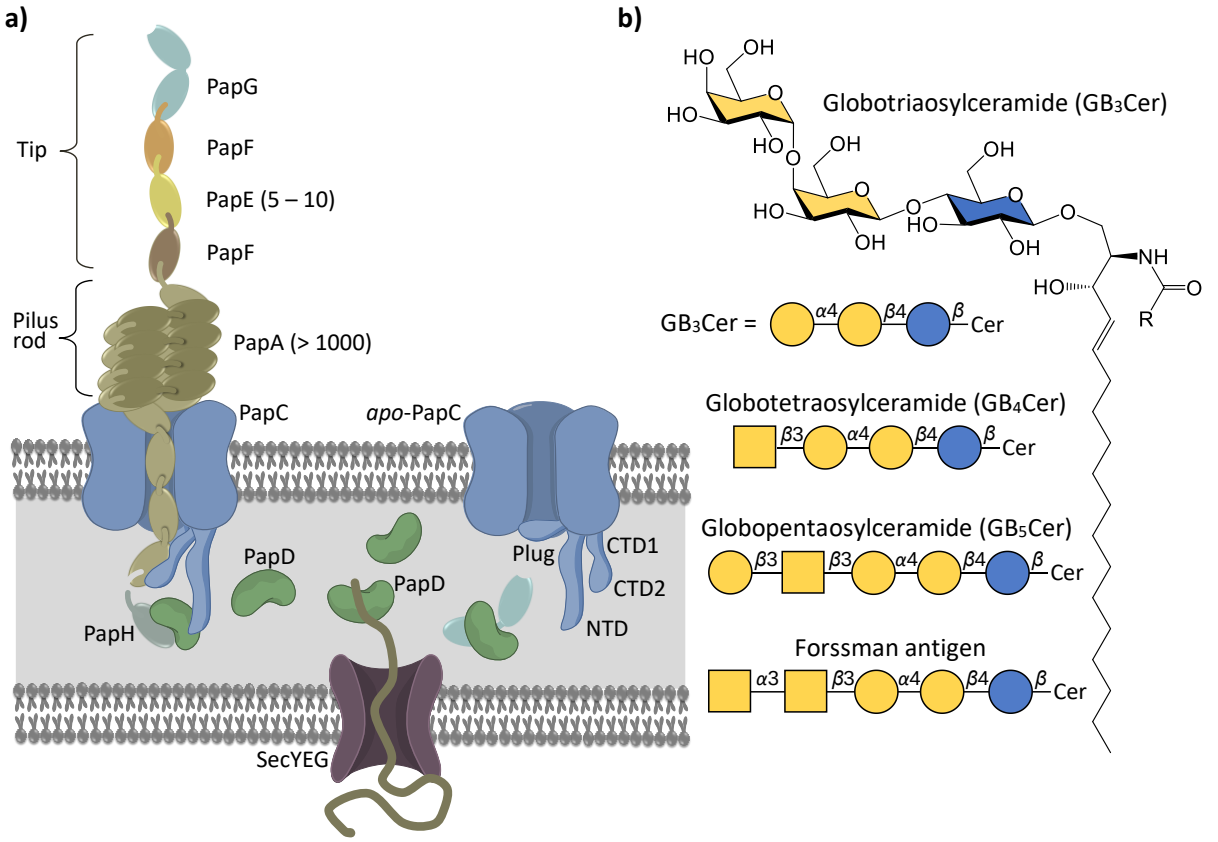


Figure 4.1-1: a) Assembly and structure of P pili as representatives of CU fimbriae and **b)** their preferred receptors from the globoseries glycosphingolipids.
GalNac; Gal; Glc; R = fatty acid alkyl chain.

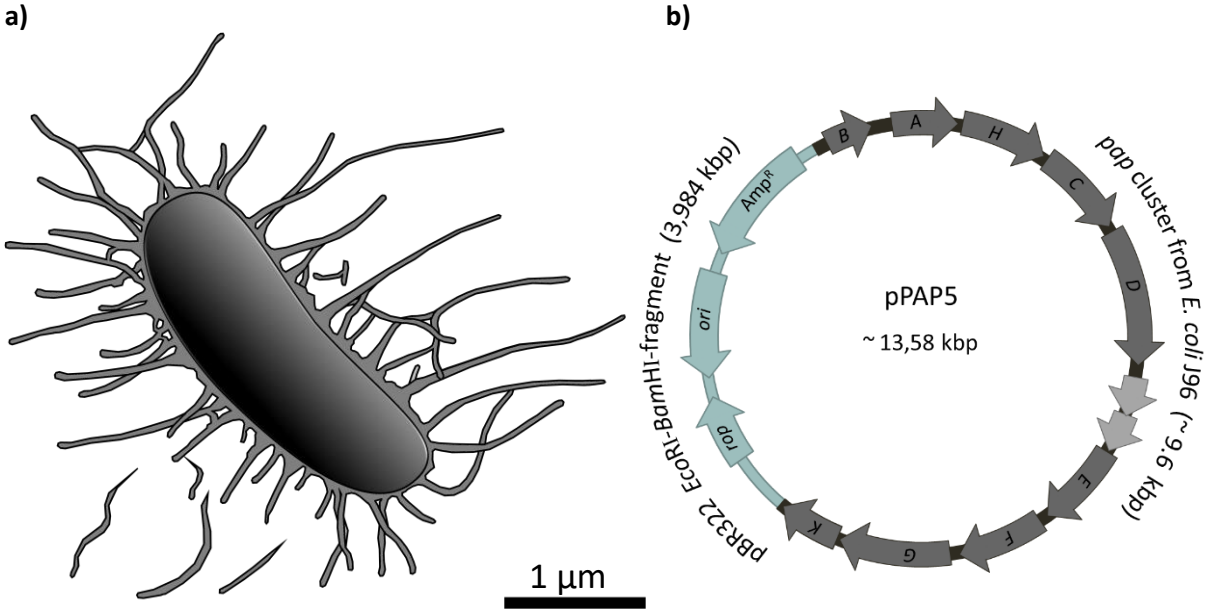


Figure 4.1-2: a) Illustration of a P pili-expressing *E. coli* and **b)** plasmid map of pPAP5, including the *pap* operon encoding P fimbriae.
The depiction in **a)** is adapted from an AFM image of HB101 pPAP5, published by Klinth and co-workers.[630]

Our available P-fimbriated strains harbour the plasmid pPAP5, which was kindly provided by the group of Prof. Aarestrup (Technical University of Denmark). The plasmid contains the entire *pap* gene cluster from *E. coli* J96, including all elements required for transcriptional regulation, expression, and assembly of P fimbriae (Figure 4.1-2). The cluster includes a class I *papG* allele and was subcloned at the *Eco*RI and *Bam*HI restriction sites on a pBR322 backbone, which enables selection under utiisation of the encoded ampicillin-resistance.[588] The plasmid pPAP5 was transformed into *E. coli* HB101 (resulting in VR123), the predecessor for *E. coli* VR123-G. The latter, in turn, contains a second plasmid for heterologous expression of the green fluorescent protein (GFP), namely pETM-G. This plasmid was generated by Claudia Fessele via subcloning of the *gfp* gene from the jellyfish *Aequorea coerulescens* into a commercially available pETM-14 vector as part of her doctoral thesis.[576] The plasmid pETM-G can be selected by utilising the encoded kanamycin resistance (Kan^R, aminoglycoside 3'phospho-transferase). However, GFP transcription on pETM-G is under control of a T7/*lacO* promoter and therefore requires induction, for instance by addition of IPTG. This, in turn, implies that a careful adjustment of the bacterial growth phase is necessary prior to induction, which is associated with longer incubation periods and repeated determination of the bacterial density. Strict compliance with the bacterial growth phase is particularly required when VR123-G is used in the context of bacterial adhesion assays to ensure the reproducibility of the *ac*GFP expression level. In brief, the bacterial assay is significantly more time-consuming and experimentally elaborate when VR123-G is employed, which is unsuitable, particularly for rather untrained experimenters.

In collaboration with Prof. Andreas Terfort and Dr. Sebastian Balser (Institute of Inorganic and Analytical chemistry, Goethe University in Frankfurt am Main) we have recently presented a nanolayer-based sensor system that facilitates rapid and selective quantification of adhesive bacteria.[631] The conductive polypyrrole (PPy) nanolayer can be modularly functionalised with small carbohydrates or glycan receptors as recognition sites for bacterial adhesion, which is gauged by conductivity measurements based on electrochemical impedance spectroscopy.[632,633] We polymerised a mannosylated nanolayer on top of the conductive PPy layer to allow specific adhesion of type 1-fimbriated bacteria. For that purpose, we chose the *av*GFP-expressing *E. coli* strain SAR18 PKL1162 to enable confirmation of the performed conductivity measurements via fluorescence monitoring. To demonstrate the suitability of the system as a selective detection method, for nstance, for pathogenic bacteria from microbial mixtures, we quantified SAR18 PKL1162 in the presence of a non-fimbriated *E. coli* strain, which constitutively expressed mOrange as a fluorescent reporter protein.[634] Actually, we would have preferred to use a P fimbriae-expressing *E. coli* as a non-adhesive comparative strain but we only had the corresponding strain available (VR123-G), which likewise expresses *ac*GFP and was therefore unsuitable.

As already mentioned, a single copy of the gfp_{mut3} gene in SAR18 PKL1162 was integrated into the bacterial chromosome by use of a site-specific recombinase system, which relies on gene products from the bacteriophage λ.[465] In this manner, the use of (inducible) plasmids can be avoided, which can be accompanied by various side effects. These are particularly plasmid loss or fluctuation in plasmid copy number, which inevitably impairs the stability and reproducibility of fluorophore expression. The well-investigated *Escherichia virus λ* belongs to the viral class of Caudoviricetes and represents a tailed bacteriophage that exclusively infects *E. coli*.[635] As a temperate phage, it is capable of entering both a lytic cycle (i.e. a reproductive cycle, resulting in host cell lysis), as well as a lysogenic life cycle (i.e. integration into the bacterial genome).

The lysogenic cycle can be induced, for instance, if the growth conditions are unfavourable. In this state, the prophage can persist within the genome for a prolonged time and even be inherited by daughter cells. Some lambda genes, essential for its lysogeny are widely used in various classical cloning and gene editing techniques, of which Gateway cloning,[636] *λ* Red recombineering,[637,638] and plasmid-based integration methods, such as CRIM[639] (conditional-replication, integration, and modular) are probably the most popular. CRIM and gateway cloning both rely on the site-specific *λ* integrase (*λint*), which catalyses the decisive homologous recombination step that results in prophage integration (Figure 4.1-3).[640]

The integrase recognises distinct 15-bp core type sites at which the recombination occurs. Two sites (P, P') are embedded within a viral 234-bp DNA fragment referred to as attachment site (*att*P). Two further core-type sites (B, B') are located on the bacterial host chromosome (*att*B). In *E. coli K12* substrain MG1655, the *att*B site is embedded between the *ybhC* and *ybh*B genes, encoding the outer membrane lipoprotein YbhC and a putative cytoplasmic kinase inhibitor, respectively.[641–643] Each pair of core-type sites represents two imperfect inverted repeats, interrupted by a non-palindromic, 7-bp overlap region (O).[644] The latter is identical in both recombination sites, where the integrase cleaves distinct phosphodiester bonds, exchanges viral and bacterial DNA strands and rejoins them under formation of an intermediate Holliday junction.[645] As a result, the entire phage genome, which circularises within the host cytoplasm,[646] becomes linearised and inserted into the bacterial chromosome. The inserted sequence is flanked by two recombined attachment sites (*att*L and *att*R). In analogy to this, a CRIM plasmid can be genomically integrated, carrying the gene(s) of interest (GOI) as well as the essential *att*P site. The *λ* integrase gene is usually provided on a helper plasmid. A second transformation step is therefore required to achieve integration of the CRIM plasmid (Figure 4.1-4).[639]

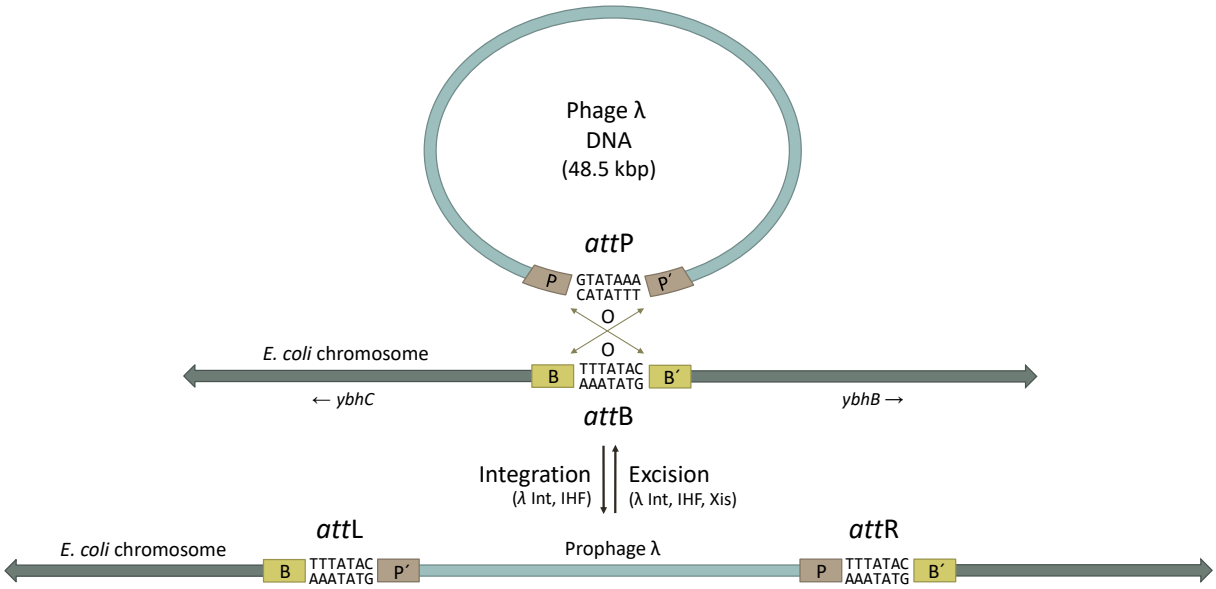


Figure 4.1-3: Site-specific recombination of *Escherichia virus Lambda* at the chromosomal *att*B site.

λ Int: lambda integrase, *att*P and *att*B: attachment sites in the viral and bacterial genomes; *att*L and *att*R: attachment sites resulting from the recombination event; P, P' and B, B': core-type sites within the attachment sites; O: overlap region; *ybhC* and *ybhB* genes located next to the *att*B site in the bacterial genome, IHF: integration host factor; Xis: excisionase.

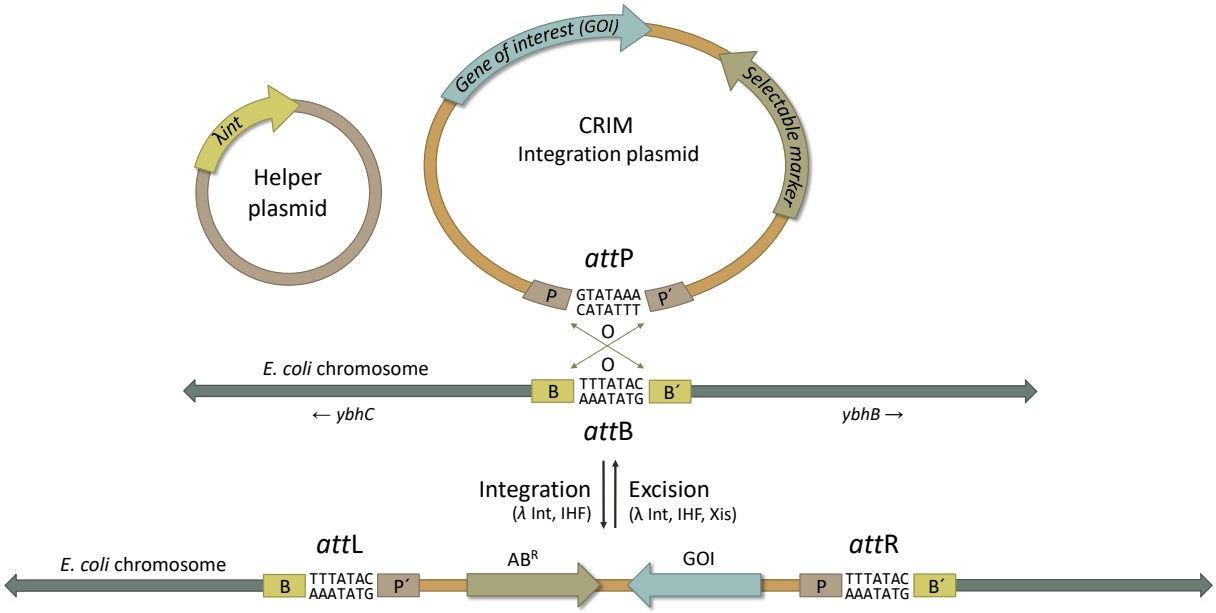


Figure 4.1-4: λ*Int*-dependent chromosomal integration of a CRIM plasmid.

CRIM: conditional-replication, integration, and modular; *λ* Int: lambda integrase, *att*P and *att*B: attachment sites in the viral and bacterial genomes; *att*L and *att*R: attachment sites resulting from the recombination event; P, P' and B, B': core-type sites within the attachment sites; O: overlap region; *ybhC* and *ybhB* genes located next to the *att*B site in the bacterial genome, IHF: integration host factor; Xis: excisionase; AB^R: selectable marker (mediates antibiotic resistance); GOI: gene(s) of interest.

4.2 Objectives

Although our well-established adhesion inhibition assay with SAR18 PKL1162 does not require induction of fluorophore expression, it is nevertheless not suitable for mannoside inhibitors with fluorescent or coloured aglycones that strongly absorb (or emit) light in the same wavelength range as GFP. For instance, the spectral profiles of GFP and fluorescein isothiocyanate (FITC) are nearly identical. Consequently, the anti-adhesive properties of FITC-*p*AP-Man, which was used to detect FimH activity in the presented lectin assay (Chapter 3.8), could initially not be determined. An attempt was made to perform a similar assay with type 1-fimbriated bacteria, as well as with the plant lectin concanavalin A (ConA). For this assay, both the bacterial cells and ConA were labelled with TAMRA-X, a red-fluorescent, amine-reactive tetramethylrhodamine derivative. However, both approaches were unsuccessful, presumably due to insufficient labelling yield and thus weak fluorescence intensity of the modified bacteria or lectin, respectively. Due to the identical reporter proteins expressed in our P- and type 1-fimbriated bacteria, we have previously not been able to perform adhesion experiments with mixtures of both microbial strains, neither in collaboration with Prof. Terfort nor in the context of our conventional assay systems.

The attainment of fimbriated *E. coli* strains, capable of constitutive expression of a fluorophore other than GFP, was therefore desirable. Red fluorescent protein (RFP) variants were considered ideal for this purpose, since they provide minimal spectral overlap with GFP.[647] A variety of monomeric RFP variants with improved solubility, maturation, and optimised spectral properties were evolved from dsRed. The latter is a tetrameric orange/red-fluorescent GFP homologue, which occurs naturally in different *Discosoma spp.* from the marine order of Corallimorpharia.[634,648] In analogy to the approach pursued by Reisner et al., the reporter gene should preferably be integrated at the *att*B site in the bacterial chromosome. A suitable λ*int*-based system that fulfils all necessary requirements for this purpose is the plasmid pOSIP-KL-*mcherry*, which is available from Addgene (N° 116947). One-Step Integration Plasmids (pOSIP) combine all elements required for cloning and chromosomal integration ("clonetegration") on a single hybrid vector.[649] In contrast to the CRIM protocol, clonetegration can be achieved after only one transformation event and thus OSIPs represent a faster, yet highly efficient, alternative to conventional genome integration techniques. The clonetegration method was introduced by Shearwin and co-workers who constructed pOSIPs with integrase genes from six different bacteriophages. Each *int* gene was combined with one of three resistance genes (positive selection marker) and a multiple cloning site, in which a lethal gene (*ccdB*) can be exchanged with a gene of interest (negative selection).[650,651] The *mcherry* gene was subcloned by Bikard and colleagues, and is under the control of the constitutive promoter Antoine R1.[652] The monomeric mCherry protein (27 kDa) belongs to the "mFruits" descendant series, which was originally derived from dsRed.[634,653] Compared to mOrange, mCherry features lower brightness (product of quantum yield and extinction coefficient) but in return, requires a significantly shorter maturation time. The excitation/emission maxima of mCherry are in the yellow and red wavelength range, respectively ($\lambda_{ex/em}$ = 587/610 nm). As illustrated in Figure 4.2-1, mCherry therefore features only minimal spectral overlap with GFP.

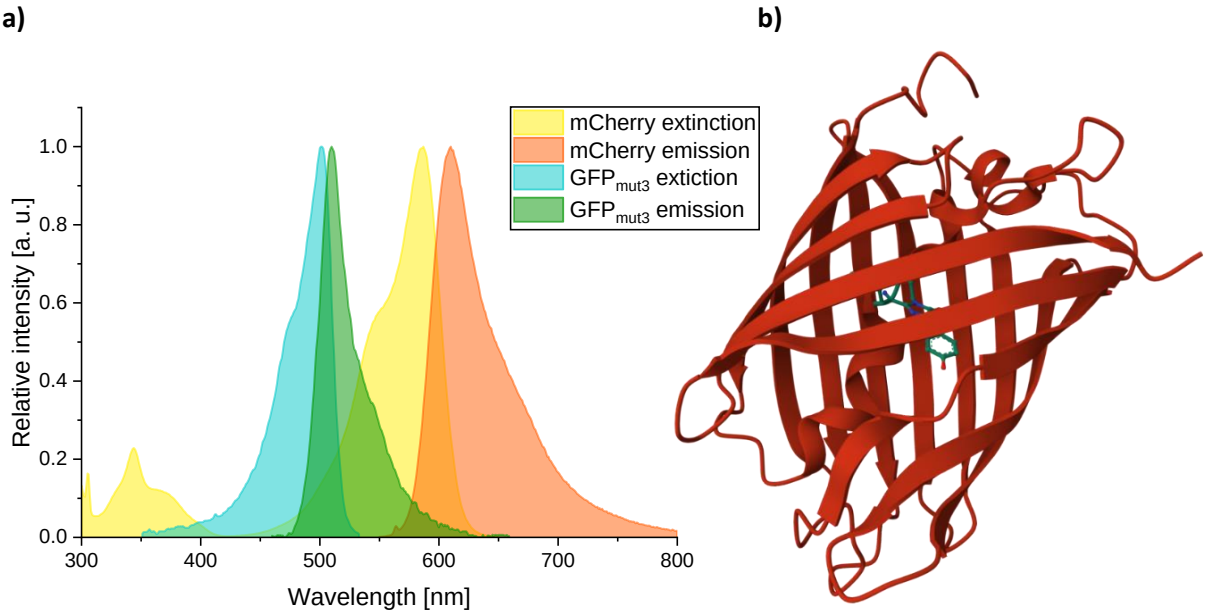


Figure 4.2-1: a) Fluorescence spectrum and **b)** crystal structure of the red fluorescent protein mCherry from *Discosoma sp.*[634] (PDB: 2H5Q)[654]

The spectrum in **a)** depicts the spectral overlap of mCherry with GFP$_{mut3}$ from *Aequorea victoria.*[655] Both structural and spectral data were obtained from the open-source Fluorescent Protein Database (FPbase).

In the following chapters, the construction and use of red-fluorescent, piliated *E. coli* bacteria are described. The *mcherry* reporter gene was supposed to be introduced at the chromosomal *att*B site in HB101 strains by using the corresponding OSIP plasmid. As a hybrid of *E. coli* K12 and *E. coli* B, whose genome does not contain any fimbrial clusters, HB101 is generally well suited as a cloning and transformation host in the context of bacterial adhesion experiments.[656] The genetic information for the expression of P pili and type 1 fimbriae is provided by transformation with the plasmids pPAP5 (*pap* cluster, including the *papG*-I allele) and with pPKL4/pPKL174 (*fim* cluster), respectively.

With the new strains in hand, the following objectives were projected: (i) The possibility of a more applicable and reproducible fluorophore expression, also in our P-fimbriated strains, which does not require induction and the associated time-consuming cultivation and incubation steps. The red-fluorescent *E. coli* could likewise be employed as a host for new fimbriae-encoding plasmids that were previously altered or subjected to site-directed mutagenesis. (ii) Extension of the available spectral application range for bacterial monitoring in competitive adhesion experiments, which, among other things, enables the use of inhibitors that possess green (-fluorescing) aglycones. (iii) Facilitation of experiments under selective conditions (e.g. by using appropriately functionalised surfaces) for the specific adhesion of either P-piliated or type 1 fimbriae-expressing *E. coli* bacteria. The differently fimbriated bacteria can prospectively be monitored by the diverging expression of GFP and mCherry in each strain, which can be independently excited even in the presence of the other fluorophore.

4.3 Clonetegration of *mcherry* into the *E. coli* chromosome

The acquired pOSIP-KL-*mcherry* construct mediates resistance to kanamycin and contains genes coding for the viral Lambda integrase (λ*int*) as well as the fluorophore mCherry. To achieve chromosomal integration of the entire pOSIP sequence, the attachment sites on both plasmid and bacterial chromosome (*att*B and *att*P), as well as the integration host factor (IHF) are required.[657–659] The λ*int* gene is under the control of a temperature-sensitive repressor (λcI^{ts}). The repressor protein likewise originates from the *Escherichia virus*, where it is responsible for maintaining the lysogenic life cycle of the latter.[660] On the experimental level, only one transformation event is required to obtain clonetegration of the plasmid. Subsequent to the short heat-shock at 42 °C, the OSIP transformation protocol prescribes the conventional recovery step at 37 °C. At this temperature, the λ repressor becomes largely inactive, upon which the integrase is expressed and chromosomal integration can occur. The also encoded $R6K_\gamma$-based replicon does not allow for propagation of the pOSIP within the host as long as the replication initiation protein (pir) is not provided.[661] Consequently, excess plasmids or non-integrated pOSIPs are gradually eliminated after a few growth cycles of the resulting integrants.

Following the protocol described in great detail by Cui and Shearwin,[662] the bacterial strains VR123 (HB101 pPAP5), HB101 pPKL4, as well as an unaltered HB101, were transformed with the plasmid pOSIP-KL-*mcherry*, with the aim to achieve chromosomal integration. The transformants were selected via the encoded kanamycin resistance, and successful integration was demonstrated by a duplex polymerase chain reaction (PCR). The duplex PCR involves the use of four primers (P1 – 4), arranged as two primer pairs that inversely flank either the *att*P (P2, P3) or the *att*B site (P1, P4), respectively (Figure 4.3-1, and Table 5.9-1). Successful integration of a single pOSIP fragment leads to the amplification of two PCR products (P1 – P2: 631 bp; P3 – P4: 1156 bp), whereas an unaltered *att*B site (P1 – P4: 741 bp) would result in a single amplicon. As indicated in Figure 4.3-2, there is a certain possibility for the insertion of more than one pOSIP fragment (tandem integration). In this case, a third PCR fragment (P2 – P3: 1046 bp) would occur, in addition to the two previously mentioned amplicons (P1 – P2: 631 bp and P3 – P4: 1156 bp).

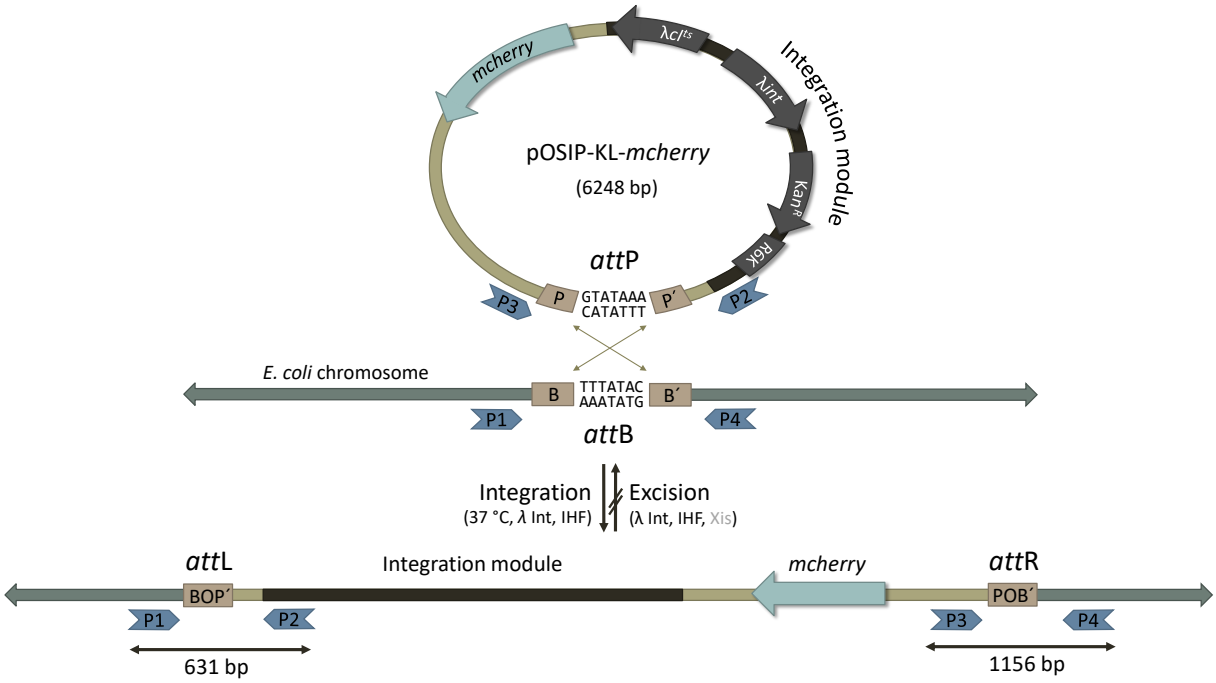


Figure 4.3-1: Integration of pOSIP-KL-*mcherry* at the *att*B site in the *E. coli* chromosome.
λ*cI*^{ts}: repressor gene; λ*int*: viral integrase gene; Kan^{R}: kanamycin resistance marker; R6K$_{\gamma}$: replicon; *att*: recombination sites of λ integrase; IHF: integration host factor; Xis: excisionase; primers P1 – 4 are indicated by blue arrows.

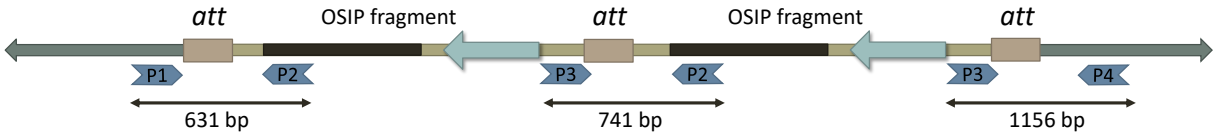


Figure 4.3-2: Result of a potential tandem integration event of pOSIP-KL-*mcherry*.

After a few days of cultivating the transformants on LB agar, a slightly red colouration of the mature colonies was observed, which further intensified over time. This already indicated an increased expression and accumulation of the fluorophore and proved to be eminently helpful in selecting the most promising clones for the subsequent evaluation via duplex PCR (Figure 5.9-1 in Chapter 5.9.1). In this context, extraction of the chromosomal DNA prior to duplex PCR significantly improved the reliability and reproducibility of the amplification result in comparison to the use of bacterial colonies. The latter was actually suggested in the pOSIP protocol in order to supply the required DNA template. According to the duplex PCR results, pOSIP clonetegration was successful with HB101 and the P fimbriae-expressing strain VR123 (now referred to as HB101-mCherry and HB101-mCherry pPAP5, respectively). Duplex PCR with the chromosomal DNA from HB101 pPKL4, that was previously transformed with the pOSIP construct, however, provided only a single PCR fragment (~ 741 bp) indicating that no clonetegration at the *att*B site had occurred. The negative result was later confirmed in a second clonetegration attempt (data not shown). Further experimental details and the respective agarose gels to illustrate the PCR results corresponding to this chapter are provided in Chapters 5.9.1 to 5.9.3.

Since the temperature-sensitive λcI repressor becomes inactive at 37 °C, it is suggested to culture the resulting integrants at a lower temperature (30 °C) in order to maintain the repression of λ*int* transcription. Aside from the presence of the viral integrase and the bacterial IHF protein, a potential excision event would actually require the presence of a third protein, the excisionase, encoded by the *xis* gene.[663,664] The latter is usually not present in the host genome, and loss of the pOSIP sequence is therefore considered rather unlikely. However, to sustain the applicability of the integrants particularly for adhesion experiments, which are usually performed at 37 °C, the ability to cultivate integrants safely under these conditions would be highly advantageous. The viral λ*int* gene, its repressor λ*cI* gene, the R6K$_\gamma$-replicon, as well as the kanamycin acetyltransferase gene, are localised next to each other and are referred to as the "integration module". Shearwin et al. introduced two further recombination sites on the pOSIP constructs, each flanking both ends of the integration module, in order to enable convenient removal of the latter by use of the well-known Flp/FRT recombination system. The flippase (Flp) is a site-specific recombinase originating from *S. cerevisiae*. Within the ancestral organism, Flp is encoded on the 2 microns cycle plasmid where it ensures the maintenance of a stable copy number.[665,666] The Flp recombinase mediates homologous recombination between so-called flippase-recognition targets (FRTs). Similar to the attachment sites recognised by λ integrase, the minimal 34-bp FRT sites each contain two 13-bp palindromic repeats disrupted by an 8-bp core spacer.[667] Depending on the relative position and orientation of two respective FRT sites to each other, the flippase can be used to achieve either recombination, inversion or even removal of the corresponding, enclosed or contiguous, sequences.[668–670] As the FRT sites on the pOSIP constructs were conceived for the removal of the interjacent integration module, they feature the same orientation. Shearwin and colleagues constructed the plasmid pE-Flp, containing the gene coding for a thermostable flippase variant, as well as a *bla* selectable marker, which was also deposited on Addgene (Plasmid no. 45978). The recombinase is expressed under the control of a strong, constitutive promoter (pE from phage P2).[662] As was already the case with the pOSIP construct, pE-Flp is also equipped with a temperature-sensitive replicon (Rep101^{ts}) and plasmid propagation thus requires a lower temperature (30 °C). Hence, the transformants were initially incubated at 30 °C under selective conditions (ampicillin), in order to enable plasmid replication and Flp-mediated excision. Subsequently, a prolonged incubation step at a higher temperature (initially at 42 °C for 2 h, then at 37 °C overnight) is necessary to initiate plasmid loss.

Due to the identical selection conditions for pE-Flp and pPAP5 (ampicillin), only HB101-mCherry was suitable for transformation with pE-Flp to achieve removal of the integration module. Transformants that successfully underwent Flp-mediated excision and subsequent plasmid loss were identified by negative selection, i.e. by exploiting the associated loss of the antibiotic resistance to both ampicillin (indicating loss of pE-Flp) and kanamycin (indicating loss of the integration module). Even if considered unnecessary by Shearwin and colleagues due to the exceptionally good excision rate usually achieved with pE-Flp, the recombination event was confirmed by PCR: The binding site of primer P2 is located within the integration module. In the case of successful removal of the latter, the respective amplicon between P1 and P2 (631 bp) would no longer be observed. In contrast, the 1156-bp fragment, amplified by P3 and P4 would be retained. As a consequence of the recombination event, the binding site of P1 would get sufficiently close to the *mcherry* gene to allow amplification of another PCR fragment (1422 bp) when the complementary primer *mcherry*20-orf is used (Figure 4.3-3).[652] The corresponding PCR result, shown in Figure 5.9-2, revealed the P3 – P4 as well as the P1 – *mcherry*20-orf amplicon, but not the P1 – P2 fragment, confirming that Flp-mediated excision was successful.

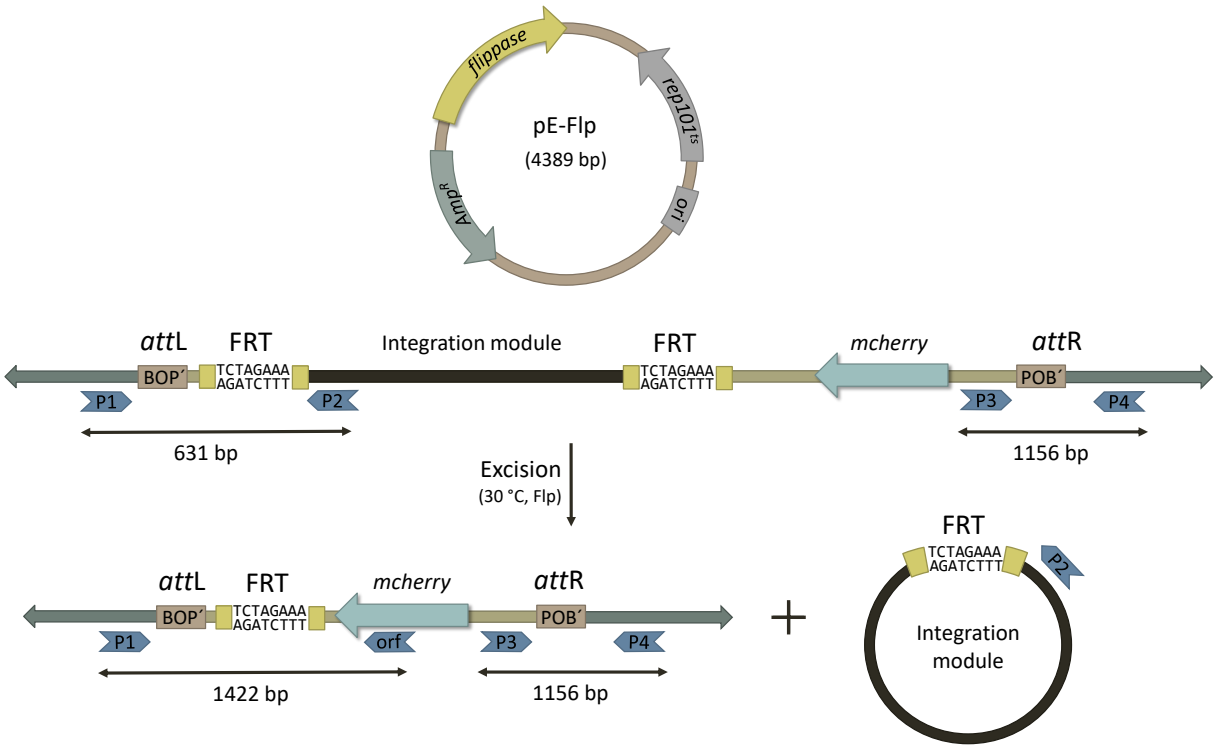


Figure 4.3-3: Flippase-catalysed excision of the integration module from the chromosomally integrated pOSIP fragment.

FRT: flippase recognition target; Amp^R: ampicillin resistance marker; *rep101*^{ts} + *ori*: genes encoding the temperature-sensitive replication initiation protein and origin or replication from pSC101; primers P1 – 4 and orf (*mcherry*20-orf) are indicated as blue arrows.

Prior to transformation with the fimbriae-encoding plasmids into the resulting strain, the loss of the pE-Flp plasmid was likewise verified by PCR (Figure 5.9-3). Despite the low copy number, maintained by the origin of replication (ori) derived from plasmid pSC101, elimination of pE-Flp required three cycles of streaking on agar plates and overnight incubation at 37 °C.[671,672] The loss of pE-Flp was eventually proven successful and the resulting strain was designated HB101-mCherry^{Flp}.

The fimbriae-encoding plasmids pPKL4, pPKL174 (*fim* cluster), and pPAP5 (*pap* cluster) were transformed into HB101-mCherry^{Flp}. This was achieved by electroporation in order to increase the transformation probability of the comparatively large plasmid size (~ 12 kbp and ~ 14 kbp). The pBR322-derived plasmids (pPKL4 and pPAP5) were selected with ampicillin, whereas pPKL174 was selected via the encoded resistance to chloramphenicol. Experimental details and PCR results confirming the presence of fimbriae-encoding plasmids (more specifically, the adhesion genes *fimH* and *papG,* class I) are provided in Chapter 5.9.3. PCR-positive clones were propagated and are here referred to as HB101-mCherry^{Flp} pPKL4/pPKL174/pPAP5, respectively.

4.4 Bacterial fluorescence and adhesion inhibition experiments

For optimal applicability in adhesion assays under both competitive and uncompetitive conditions, the obtained fluorescence signal should generally be as intense as possible. This is particularly required as only the residual signal is detected, which is emitted by the remaining bacteria that still adhere to the carbohydrate-coated surface after repeated washing steps. Furthermore, the fluorophore expression level – and thus the observable fluorescence – should be as reproducible as possible in order to ensure the comparability of results from different bacterial adhesion experiments. The new strains were thus supposed to be characterised with regard to fluorescence intensity and reproducibility. As the green-fluorescent, type 1 fimbriae-expressing strain SAR18 PKL1162 optimally fulfils these requirements, it was employed here for comparison purposes.

The bacterial fluorescence of serially diluted suspensions of HB101-mCherry^{Flp}, as well as of those strains that were transformed with either of the fimbriae-encoding plasmids, was measured at $\lambda_{ex/em}$ = 568/630 nm on the Tecan microplate reader. It should be noted that the bacterial suspensions were applied onto microplates, which had not been coated beforehand with polysaccharides as an adhesive surface. The suspensions were furthermore not removed by washing prior to the fluorescence measurement, i.e. the experiment did not represent an adhesion assay. The observed fluorescence signals were statistically evaluated in two-sample t-tests in order to determine at which bacterial densities a significant difference was still detectable in comparison to the negative control (PBS, 100 µL/well). In Figure 4.4-1, the relative fluorescence intensities observed at different OD_{600} of HB101-mCherry^{Flp} pPKL4 are given as an example. The results were compared to the relative fluorescence signal of the serially diluted SAR18 PKL1162 ($\lambda_{ex/em}$ = 485/535 nm). Further fluorescence data are provided in Chapter 5.9.4, including measurements with serial dilutions of HB101-mCherry^{Flp} pPKL174 (Figure 5.9-8) and HB101-mCherry^{Flp} pPAP5 (Figure 5.9-9).

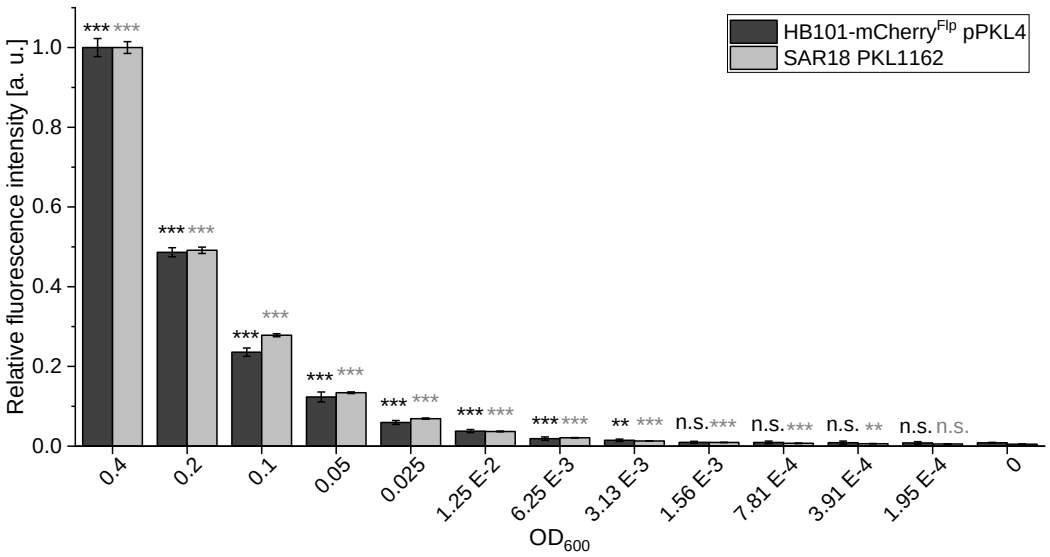


Figure 4.4-1: Fluorescence measurements with serially diluted suspensions of *E. coli* HB101-mCherry^{Flp} pPKL4 and SAR18 PKL1162.

Fluorescence of bacterial 1:2 dilution series of HB101-mCherry^{Flp} pPKL4 (at $\lambda_{ex/em}$ = 568/630 nm) and SAR18 PKL1162 (at $\lambda_{ex/em}$.= 485/535 nm) starting from OD_{600} = 0.4. Error bars represent the ± standard deviation of five replicates. n.s. denotes statistically non-significant signal differences while the asterisks (*) indicate statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (PBS, 100 µL/well, OD_{600} = 0).

As shown in Figure 4.4-1, the minimal still detectable fluorescence signal of HB101-mCherry^{Flp} pPKL4 was observed at an OD_{600} = $3.13 \cdot 10^{-3}$. The same result was consistently obtained in all other fluorescence measurements with HB101-mCherry^{Flp}-derived strains, suggesting a well-reproducible fluorophore expression. However, the smallest measurable fluorescence signal was one order of magnitude higher than that obtained with the GFP-expressing SAR18 PKL1162 (OD_{600} = $3.91 \cdot 10^{-4}$). A linear correlation between bacterial density and observed fluorescence signal was determined with all *E. coli* strains tested (Figures 5.7-7 – 5.8-10). Accordingly, a reproducible and adequate signal intensity can be assumed, which is certainly still suitable for detection of residually adhering bacteria.

The HB101-mCherry^{Flp} pPKL4/pPKL174 strains, which thus harboured the gene cluster for expression of type 1 fimbriae on their respective plasmids, were subsequently tested in adhesion inhibition assays. In analogy to the conventional protocol, bacterial suspensions were incubated on mannan-coated microplates, and adhesion was hampered by the addition of serially diluted mannoside inhibitors serving as FimH antagonists. Initially, the most common reference inhibitors methyl α-D-mannopyranoside (MeMan), as well as *p*-nitrophenyl α-D-mannopyranoside (*p*NPMan) were tested. Once the functionality of the adhesion inhibition experiment with the mCherry-expressing *E. coli* was proven, the fluorescein-labelled mannoside **59** (FITC-*p*AP-Man) could eventually be used as an inhibitor. In doing so, the recognisability by the lectin, manifested as the antiadhesive behaviour, could be determined, which was previously not possible with the PKL1162-based assay. Two exemplary results obtained in competitive adhesion experiments with HB101-mCherry^{Flp} pPKL4 are illustrated in Figure 4.4-2, using *p*NPMan (a) and FITC-*p*AP-Man (b) as inhibitors. The corresponding IC_{50} and RIP data for *p*NPMan and FITC-*p*AP-Man are compiled in Table 4.4-1, below. Further results from bacterial adhesion assays are provided in Chapter 5.9.5.

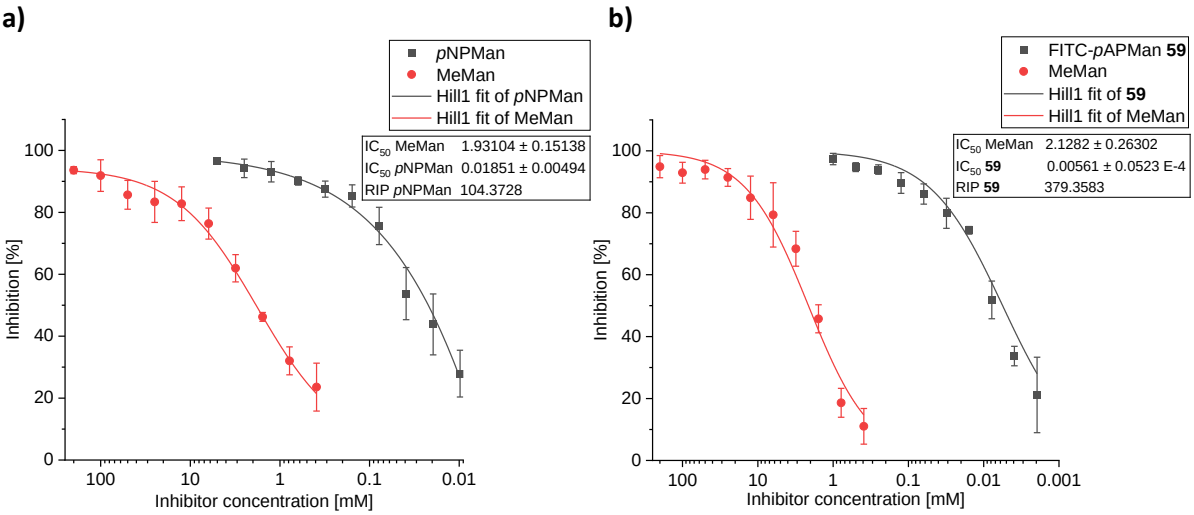


Figure 4.4-2: Representative results from two adhesion inhibition experiments with red-fluorescent HB101-mCherry^{Flp} pPKL4.

a) Inhibition experiment with *p*-nitrophenyl α-D-mannopyranoside (*p*NPMan). **b)** Inhibition experiment with fluorescein-labelled α-D-mannopyranoside **59** (FITC-*p*AP-Man). Bacterial suspensions were adjusted to OD_{600} = 0.8 and the residual fluorescence was monitored at $\lambda_{ex/em}$ = 568/630 nm using a microplate reader (Tecan). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

Table 4.4-1: Antiadhesive properties of methyl *α*-D-mannopyranoside (MeMan), *p*-nitrophenyl *α*-D-mannopyranoside (*p*NPMan), and FITC-labelled mannoside **59**, determined in adhesion inhibition experiments with HB101-mCherry^{Flp} pPKL4.

Value	MeMan[a]	*p*NPMan[b]	FITC-*p*APMan (59)[c]
IC_{50} [mM]	2.39 (± 0.49)	$2.67 \cdot 10^{-2}$ (± $8.78 \cdot 10^{-3}$)	$8.43 \cdot 10^{-3}$ (± $3.35 \cdot 10^{-3}$)
RIP [d]	≡ 1	79.52 (± 21.69)	324.55 (± 71.92)

a) Average IC_{50} value from 10 individual experiments. **b)** Average IC_{50} and RIP values from 4 individual experiments. **c)** Average IC_{50} and RIP values from 6 individual experiments. **d)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl *α*-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). The respective ± standard deviations are given in brackets. Literature values for RIP_{pNPMan}= 57 – 91, determined in GFP assays by Hartman et al.[467,673]

The determined RIP for *p*NPMan was in good agreement with the literature values obtained in the conventional PKL1162-based assay.[467,673] The fluorescein-labelled mannoside derivative **59** revealed a very satisfactory antiadhesive potential (RIP ∼ 325), which was certainly favourable for its use in the FITC-*p*AP-Man-based lectin assay. To obtain the desired sigmoidal dose-response relationships at the usual inhibitor concentrations, the suspension of HB101-mCherry^{Flp} pPKL4 had to be adjusted to at least OD_{600} = 0.8. This represents a considerable increase compared to the PKL1162-based assay, which is usually performed with OD_{600} = 0.4. The higher bacterial density required for HB101-mCherry^{Flp} pPKL4 may be partially attributable to the slightly lower detectability of the fluorophore mCherry compared to GFP. On the other hand, higher bacterial densities of pPKL4-based fimbriae expression systems have already been employed in previous competitive adhesion experiments by the Lindhorst group (e.g. the biotin assay).[467] This could imply a graduated adhesiveness, potentially resulting from the use of the two different plasmids or vector backbones in PKL1162 (equipped with pPKL174) and HB101 pPKL4, respectively. Likewise, this might explain slightly lower adhesiveness observed for HB101-mCherry^{Flp} pPKL4 in comparison to PKL1162.

The plasmid pPKL4 is a pBR322-derived plasmid with a medium copy number (approximately 20 – 50 copies/cell).[674,675] It contains the entire *fim* cluster from *E. coli* K12 with all elements necessary for assembly and regulation of type 1 fimbriae.[296,297] These include natural promoters (*fim*A_p, *fim*B_{p1-3}, *fim*E_p), binding sites for transcription factors, small regulatory RNA sequences (*fim*R2), as well as the recombinases FimB and FimE, which catalyse phase variation of the invertible *fim* switch (*fim*S).[299,301,306,676] Plasmid pPKL174 was previously isolated from PKL116 and presumably contains the same *fim* cluster as that present on pPKL4.[467]

However, the respective HB101-mCherry^{Flp} transformed with pPKL174 did not exhibit any bacterial adhesiveness in either adhesion experiments or competitive assays performed (Figure 5.9-13). This was the case even though the strain demonstrably featured strong red fluorescence (Figure 5.9-8) and the respective colonies were clearly confirmed to be PCR-positive for *fim* cluster-associated *fimH* (Figure 5.9-5). Admittedly, it should be mentioned that the exact vector used for the construction of plasmid pPKL174 is currently unknown and potential incompatibilities with the host can therefore not be excluded. Likewise, it is not known with certainty whether any modifications may have been made to the encoded *fim* sequences on pPKL174 compared to the original cluster on pPKL4, if any. On the common *pap*-encoding plasmid pHMG93 for instance, an insertion of several strong, constitutive promoters has previously been performed, intentionally resulting in upregulation of P fimbrial assembly.[629]

Nonetheless, the disadvantage of the lower detectability and/or adhesiveness of HB101-mCherry^{Flp} pPKL4, compared to PKL1162, can be compensated by the possibility of using lower inhibitor concentrations. In particular, this could be beneficial for mannoside derivatives that are synthetically elaborate to produce or that may feature poor water solubility.

Unfortunately, it was not possible to test the recognition of galabioside-functionalised surfaces (or corresponding inhibitors) with any of our P-fimbriated *E. coli* strains (HB101-mCherry^{Flp} pPAP5 and VR123-G, Figure 5.9-14). The formation of the glycosidic Gal(α1→4)Gal bond is synthetically particularly demanding due to the low reactivity of the axial 4-position and the poor α-selectivity of the galactosyl acceptor. At the time, neither small galabioside derivatives nor corresponding polysaccharides were commercially available, which could have been used as suitable adhesive platform. This was not surprising, either, since galactopyranoside moieties are usually β-glycosidically linked – especially in plant-derived glycans. Nevertheless, in a quick attempt to let P-piliated *E. coli* adhere to a plant-derived hemicellulose, commercially acquired galactomannan (ROTH®) was tested. However, no adhesion of VR123-G or HB101-mCherry^{Flp} pPAP5 to galactomannan was observed (Figure 5.9-14). Galactomannans of the carubin type are primarily composed of linear β(1→4) linked mannan chains, regularly branched with single β(1→6)-D-galactopyranosyl moieties.[677–679]

Despite all, the use of mCherry-expressing bacteria in microbial mixtures with our previously existing green fluorescent strains appeared to be promising. The spectral properties of the bacterially expressed fluorophores proved to be different enough to detect the green-fluorescent, type 1-fimbriated *E. coli* strain (PKL1162) in the presence of the red-fluorescent, P-fimbriated strain (HB101-mCherryFlp pPAP5) under selective adhesion conditions (Figure 5.9-15a). It is mentioned in the literature that the expression of the unaltered *pap* operon on the plasmid pPAP5 can result in the formation of relatively few P fimbriae, which can be problematic for certain experiments.[629] As the adhesiveness of P-piliated HB101-mCherry^{Flp} was thus assumed to be rather low, the use of higher bacterial densities appeared advisable. For non-competitive adhesion experiments, bacterial suspensions of HB101-mCherry^{Flp} pPAP5 were consequently adjusted to a higher bacterial density than usual (OD_{600} = 0.4) as had been previously necessary with pPKL4-based fimbrial expression systems (OD_{600} = 0.2).

4.5 Conclusion and perspectives

Here, the construction of red-fluorescent *E. coli* bacteria was described, that were intended for application in adhesion experiments under both competitive and uncompetitive conditions. The chromosomal integration of a reporter gene, coding for the red-fluorescent protein mCherry, was achieved by transforming a single plasmid (pOSIP-KL-*mcherry*) into a selection of HB101 strains. Catalysed by the encoded viral λ integrase, the plasmid mediates its own integration at the genomic *att*B site via recombination with the homologous *att*P site on pOSIP-KL-*mcherry*. As clearly demonstrated by duplex PCR, the chromosomal integration was successful with unaltered *E. coli* HB101, as well as with HB101 pPAP5 harbouring a plasmid for the expression and assembly of P pili. The resulting strains, here referred to as HB101-mCherry (pPAP5), should preferably be cultured at 30 °C in order to maintain the repression of the λ integrase. The latter, as well as its temperature-sensitive repressor were therefore removed. This was achieved by straightforward transformation of the plasmid pE-Flp into HB101-mCherry. The encoded flippase mediates the highly efficient excision of the "integration module", which was conveniently flanked by the appropriate FRT recognition sites. Both plasmids, pOSIP-KL-*mcherry* and pE-Flp, cannot be propagated within the new hosts and were therefore efficiently removed within a few growth cycles.

The resulting strain HB101-mCherry^{Flp} exhibited a strong red fluorescence, which was promising for the detection of residually adhering bacteria. By electroporation, it was transformed with plasmids that encoded either type 1 fimbriae (pPKL4) or P fimbriae (pPAP5). HB101-mCherry^{Flp} pPKL4 was successfully employed in a variety of adhesion assays and inhibiton experiments. In this way, the anti-adhesive properties of FITC-*p*AP-Man could be determined, which had previously not been suitable using the corresponding experiment with GFP-expressing bacteria. It was furthermore possible to independently detect the residual adhesion of the green fluorescent, type 1-fimbriated PKL1162 to a mannan-coated surface after co-incubation with non-adherent, mCherry-expressing *E. coli.* This clearly demonstrates the general applicability of HB101-mCherry^{Flp} in selective adhesion experiments and in microbial mixtures with green-fluorescent *E. coli*.

Unfortunately, no functionalised platform appropriate for PapG recognition and adhesion of P-piliated bacteria was readily available. As a consequence, the adhesiveness of the respective strain (HB101-mCherry^{Flp} pPAP5) could not be demonstrated in the scope of this work. However, this can prospectively be achieved in the future by continuing the previous efforts made by Claudia Fessele who focused on the chemical synthesis of α(1-4)-linked digalactosides in her doctoral thesis.[576] Recently, Sönke Sdunnus, a current member from the Lindhorst group, has succeeded in optimising the challenging synthesis of galabioside derivatives, including those featuring azobenzene and pyrrole-based aglycones. With the former molecules, photoswitching of bacterial adhesion on eukaryotic cell surfaces is envisaged in cooperation with the group for Physical Glycosciences led by Prof. Leonhard Möckl (Max Plank Institute for the Science of Light, Erlangen). The collaboration with the research group of Prof. Terfort (Goethe University, Frankfurt) is also planned to be continued. In this context, among other things, it is intended to equip the polypyrrole-based biorecognition sensor with pyrrole-functionalised Gal(α1→4)Gal disaccharides to enable rapid and straightforward detection and quantification of P-fimbriated bacteria.

5. Experimental and supplemental section

5.1 Chemical material and methods

5.1.1 General methods

Unless otherwise stated, all reactions were performed under atmospheric conditions. Moisture-sensitive reactions were carried out under a nitrogen atmosphere in dry glassware, which had been dried at 50 °C overnight prior to use. Synthesised target products and stable intermediates were stored for later use at –20 °C under a nitrogen atmosphere.

5.1.2 Chemicals and solvents

All reagents were purchased from commercial suppliers (Abrc GmbH, Acros Organics™, Alfa Aesar™, Sigma-Aldrich GmbH, Merck, TCI GmbH, VWR Chemicals) and used without further purification. Ion exchange resins (Amberlite® IR 120) were washed with methanol directly before use. For extractions and column chromatography, technical-grade solvents were used, which were purified by distillation beforehand. Dry solvents were either directly purchased from the manufacturer or dried by an Inert PureSolv MD5 solvent purification system:

Acetonitrile	dried over calcium hydride
Dichloromethane	dried over aluminium oxide
Dimethylformamide	dried over molecular sieves (Acros Organics™)
Ethanol	dried over molecular sieves (Acros Organics™)
Methanol	dried over magnesium
Pyridine	dried over potassium hydroxide

5.1.3 Thin-layer chromatography

Analytical thin-layer chromatography (TLC) was performed on aluminium plates, coated with silica gel (GF_{254}, Merck). Visualisation was achieved by UV irradiation using a CAMAG® UV lamp (λ = 254 nm) and/or by staining with one of the following solutions, followed by heat treatment:

Vanillin stain:	Vanillin (3.0 g), H_2SO_4 (40 mL) ethanol (250 mL), dH_2O (300 mL)
Ninhydrin stain:	Ninhydrin (300 mg), butanol (100 mL,) glacial acetic acid (3 mL)
Phosphomolybdic acid:	$H_3[Mo_{12}PO_{40}] \cdot 12\ H_2O$ (10 g), ethanol (100 mL)

5.1.4 Preparative column chromatography

Preparative flash chromatography was performed on silica gel 60 (230 – 400 mesh, particle size 0.040 – 0.063 mm, Merck) or by automated flash chromatography over a reversed phase cartridge (ChromaBond Flash RS25 C18 ec, Macherey-Nagel™) with a puriFlash® 5.020 device (Interchim® Inc.), using distilled solvents.

5.1.5 IR spectroscopy

Infrared spectra were recorded on a FT-IR Paragon 1000 (ATR) spectrometer (PerkinElmer®). The absorption signals are denoted as wavenumbers (cm^{-1}). The signal shape and intensity was furthermore characterised as strong (s), weak (w), moderate (m), or broad (br), respectively.

5.1.6 Mass spectrometry

The measurement of EI mass spectra was performed on an AccuTOF 4GCV spectrometer (Jeol GmbH). High-resolution ESI mass spectra (ESI-HRMS) were recorded on a Q Exactive™ Plus mass spectrometer (Thermo Scientific™). MALDI mass spectra were measured on a MALDI-TOF AutoFlex® instrument (Bruker Corporation).

5.1.7 NMR spectroscopy

NMR spectroscopy was used in order to ensure the identity and purity of synthesised compounds. Full assignment of the signals was achieved by using 2D NMR techniques. All NMR spectra (^{1}H, ^{13}C, ^{19}F, ^{1}H–^{1}H COSY, ^{1}H–^{13}C HSQC, ^{1}H–^{13}C HMBC, and STD) were recorded on one of the following instruments:

Bruker AC 200	^{1}H NMR (200 MHz)	^{13}C NMR (50 MHz)
Bruker DRX 500	^{1}H NMR (500 MHz)	^{13}C NMR (126 MHz)
	^{19}F NMR (471 MHz)	COSY (500 MHz)
	HSQC (500 MHz)	HMBC (500 MHz)
Bruker AV 600	^{1}H NMR (600 MHz)	^{13}C NMR (151 MHz)
	COSY (600 MHz)	HSQC (600 MHz)
	HMBC (600 MHz)	STD-NMR (600 MHz)

Chemical shifts are referenced to the internal standard tetramethylsilane, or to the residual proton, or carbon signal of the NMR solvent, respectively:

$CDCl_3$	^{1}H NMR: 7.26 ppm (s)	^{13}C NMR: 77.2 ppm (t)
D_2O	^{1}H NMR: 4.79 ppm (s)	
CD_3OD	^{1}H NMR: 3.31 ppm (quint)	^{13}C NMR: 49.0 ppm (sept)

NMR data are presented in the following format: chemical shift δ in parts per million (ppm), (multiplicity: s = singlet, d = doublet, dd = doublet of doublets, t = triplet, dd ~t = apparent triplet, br. s = broad singlet, q = quartet, and m = multiplet; coupling constant *J* in Hertz (Hz), integration, assignment to the respective proton or carbon). NMR spectra of *trans*-isomeric azobenzene derivatives were recorded after thermal relaxation in darkness for at least 18 h at 45 °C. NMR spectra of the respective *cis*-isomers were recorded after irradiation with UV light (λ = 365 nm) for 10 – 15 min. Processing, analysis and visualisation of NMR spectra was carried out using the NMR data analysis software TopSpin (Bruker).

5.1.8 Polarimetry

Polarimetric measurements were performed on a PerkinElmer® 241 spectropolarimeter (sodium D-line: 589 nm, cell length: 1 dm) in the given solvents. The specific rotations $[\alpha]_{20}^{D}$ are indicated in [° · mL · g^{-1} · dm^{-1}] and the concentration c in [mol · mL^{-1}] (unless otherwise stated). Solutions of azobenzene derivatives were thermally relaxed by storage at 45 °C, in darkness for 18 h, in order to prevent potential electromagnetic absorption at 589 nm.

5.1.9 Melting point determination

Melting points were determined on a Melting Point M-560 instrument (BÜCHI).

5.1.10 Fluorescence and UV/vis spectroscopy

Fluorescence and UV/vis absorbance spectra were measured on an Infinite® 200 Pro M Nano+ dual mode microplate reader (Tecan) in MaxiSorp™ F96 microwell plates (Nunc™ Thermo Scientific) or transparent UV-Star® F96-well plates (Greiner Bio-One GmbH), respectively, at the given temperature. UV/vis absorption spectra were also recorded on a UV-2600i UV/vis spectrophotometer (Shimadzu) in quartz cuvettes with a diameter of 1 cm at a temperature of 25 °C (+/−0.1 °C).

5.3 Chemical syntheses

Compounds **5**,[439,680] **64**[681] and **58**[524] were prepared starting from D-mannose or 4-nitrophenyl *α*-D-mannopyranoside, respectively, according to literature-known procedures. *tert*-Butyl (4-aminobenzyl)carbamate (**65**) was obtained by conventional *N*-Boc-protection of the methylamino group of 4-aminobenzylamine.[682]

In the case of both symmetric and asymmetric azobenzene derivatives, the following numbering was used for the assignment of NMR proton and carbon signals:

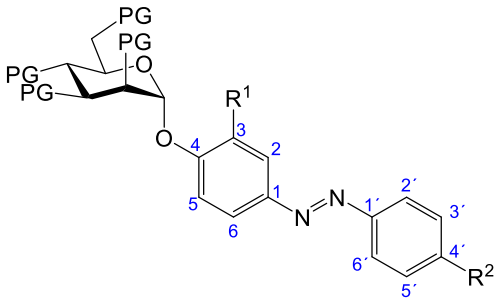
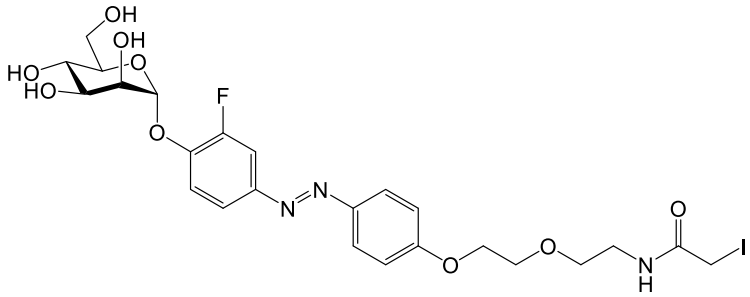


Figure 5.3-1: Numbering of azobenzene derivatives for the assignment of NMR signals.
PG: Protecting group; **R**: unspecified substituents.

{4-[3-(4'-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-*α*-D-mannopyranoside (F1)



Compound **19** (312 mg; 0.38 mmol) was dissolved in methanol (15 mL) and a freshly prepared sodium methoxide solution (20 µL, 5.4 M in methanol) was added. The reaction mixture was stirred at ambient conditions for 1 h and was subsequently neutralised with Amberlite® IR 120. The resin was filtered off and the solvent was removed under reduced pressure. The residue was purified by column chromatography on silica gel (DCM/methanol, 10:1) under avoidance of light and a subsequent size exclusion chromatography (LH-20, methanol) to afford compound **F1** as an orange solid.

Yield: 189 mg (0.29 mmol, 93 %);

TLC: R_f = 0.16 (DCM/methanol, 10:1),
trans-isomer: 0.50 (ethyl acetate/methanol/dH_20, 20:2:1),
cis-isomer: 0.15 (ethyl acetate/methanol/dH_20, 20:2:1);

Melting point: 131.1 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 96.6 (c = 0.008 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.87 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.7 Hz, $^4J_{Ar\text{-}H6,Ar\text{-}H2}$ = 2.3 Hz, $^5J_{Ar\text{-}H6\text{-}F}$ = 1.2 Hz, 1H, Ar-H_6), 7.64 (dd, $^3J_{Ar\text{-}H2,F}$ = 11.9 Hz, $^4J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.3 Hz, 1H, Ar-H_2), 7.54 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.09 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.22 (dd∼t, $^3J_{Ar\text{-}OCH2,OCH2CH2O}$ = 4.6 Hz, 2H, Ar-OCH_2), 4.10 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.94 (dd, $^3J_{3,4}$ = 9.4 Hz, $^3J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.86 (dd∼t, $^3J_{Ar\text{-}OCH2CH2O,Ar\text{-}OCH2}$ = 4.6 Hz, 2H, Ar-OCH_2CH_2O), 3.80 (dd, $^2J_{6,6'}$ = 12 Hz, $^3J_{6,5}$ = 2.4 Hz, 1H, H-6), 3.77-3.71 (m, 2H, H-4, H-6'), 3.69 (s, 2H, $COCH_2I$), 3.65 (ddd, $^3J_{5,4}$ = 9.4 Hz, $^3J_{5,6}$ = 5.4 Hz, $^3J_{5,6'}$ = 2.1 Hz, 1H, H-5), 3.63 (t, $^3J_{OCH2CH2NH,CH2NH}$ = 5.6 Hz, 2H, OCH_2CH_2NH), 3.39 (t, $^3J_{CH2NH,OCH2CH2NH}$ = 5.6 Hz, 2H, CH_2NH) ppm;

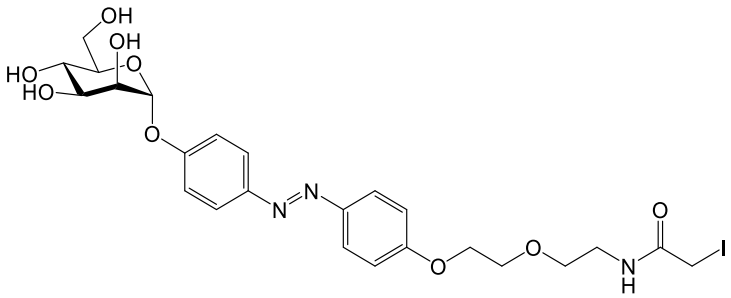
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 169.7 ($NHCOCH_2I$), 161.2 (Ar-$C_{4'}$), 155.9 (d, $^1J_{Ar\text{-}C3,F}$ = 247 Hz, Ar-C_3), 147.7 ($^3J_{Ar\text{-}C1,F}$ = 5.3 Hz, Ar-C_1), 146.3 (Ar-$C_{1'}$), 145.7 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.0 (2C, Ar-$H_{2'/6'}$), 121.0 (Ar-C_6), 117.4 (Ar-C_5), 114.2 (2C, Ar-$C_{3'/5'}$), 107.0 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 99.4 (C-1), 74.3 (C-5), 70.5 (C-3), 69.9 (C-2), 68.7 (OCH_2CH_2O), 68.6 (OCH_2CH_2NH), 67.2 (Ar-OCH_2), 66.4 (C-4), 62.8 (C-6), 39.2 (CH_2NH), – 3.9 ($COCH_2I$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 132.38 (s, Ar-$F_{trans\text{-isomer}}$), – 131.52 (d, Ar-$F_{cis\text{-isomer}}$) ppm;

IR (ATR): $\tilde{\nu}$ = 3307 (br), 2929 (m), 2360 (w), 2308 (w), 2191 (w), 2010 (w), 1648 (s), 1601 (s), 1501 (s), 1425 (m), 1317 (m), 1258 (s), 1211 (m), 1128 (s), 1097 (s), 1003 (s), 976 (s), 837 (m), 673 (m), 662 (m), 522 (m), 484 (m), 460 (w), 429 (w) cm^{-1};

ESI-HRMS: m/z = 650.09942 $[M + H]^+$, (calc. 650.10053 for $C_{24}H_{29}FIN_3O_9 + H^+$).

[4-(4'-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}phenyl)azophenoxy]-*α*-D-mannopyranoside (H1)[438]



Acetyl protected, iodoacetamido-functionalised azobenzene mannoside derivative **44** (93 mg, 0.12 mmol) was dissolved in methanol (5 mL) and sodium methoxide solution (2.5 µL, 5.4 M in methanol) was added. The reaction mixture was stirred at ambient conditions for 1 h, until TLC indicated full consumption of the starting material and was subsequently neutralised with Amberlite® IR 120. The resin was filtered off and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol/dH_20, 10:2:1) and a subsequent size exclusion chromatography (LH-20, methanol) to afford the title compound **H1** as an orange solid.

Yield: 44.8 mg (0.07 mmol, 61 %);

TLC: R_f = 0 (cyclohexane/ethyl acetate/methanol, 5:5:1),
trans-isomer: 0.36 (cyclohexane/ethyl acetate/methanol, 20:2:1),
cis-isomer: 0.27 (cyclohexane/ethyl acetate/methanol, 20:2:1);

Melting point: 131.0 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 105.5 (c = 0.06, methanol);

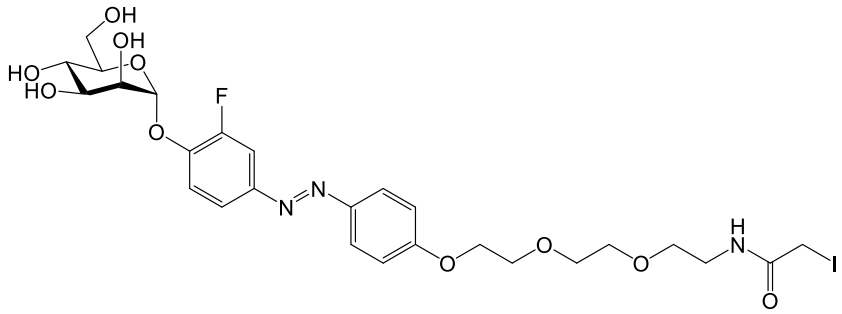
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.86 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.85 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.26 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.08 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.59 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.23 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.6 Hz, 2H, Ar-OCH_2), 4.04 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.4 Hz, 1H, H-2), 3.92 (dd, $^{3}J_{2,3}$ = 3.4 Hz, $^{3}J_{3,4}$ = 9.4 Hz, 1H, H-3), 3.86 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.6 Hz, 2H, Ar-OCH_2CH_2), 3.79-3.72 (m, 3H, H-4, H-6, H-6'), 3.69 (s, 2H, $NHCOCH_2I$), 3.63 (dd∼t, $^{3}J_{CH2CH2NH,CH2NH}$ = 5.5 Hz, 2H, CH_2CH_2NH), 3.59 (ddd, 1H, $^{3}J_{5,4}$ = 9.8 Hz, $^{3}J_{5,6}$ = 5.3 Hz, $^{3}J_{5,6'}$ = 2.4 Hz, H-5), 3.39 (t, $^{3}J_{CH2CH2NH,CH2NH}$ = 5.5 Hz, 2H, CH_2NH), ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 162.7 ($NHCOCH_2I$), 159.9 (Ar-C_4), 149.3 (Ar-$C_{4'}$), 148.4 (Ar-$C_{1'}$), 147.1 (Ar-C_1), 125.5 (2C, Ar-$C_{2/6}$), 125.2 (2C, Ar-$C_{2'/6'}$), 118 (2C, Ar-$C_{3'/5'}$), 116 (2C, Ar-$C_{3/5}$), 100.2 (C-1), 75.7 (C-5), 72.4 (C-3), 71.9 (C-2). 70.6 (CH_2CH_2NH), 70.4 (Ar-OCH_2OCH_2), 69.1 (Ar-OCH_2), 68.3 (C-4), 62.7 (C-6), 41.1 (CH_2NH), – 2.1 ($NHCOCH_2I$) ppm;

IR (ATR): $\tilde{\nu}$ = 3314 (br), 2924 (m), 2349 (w), 2050 (w), 1647(m), 1598 (m), 1582 (m), 1498 (s), 1229 (s), 1120 (s), 1107 (s), 838 (s), 672 (m) cm^{-1};

ESI-HRMS: *m/z* = 632.10901 $[M + H]^+$, (calc. 632.10995 for $C_{24}H_{30}IN_3O_9 + H^+$).

(4-{3-[4'-(2-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)]-*α*-D-mannopyranoside (F2)



Compound **20** (45 mg, 0.064 mmol) was dissolved in methanol (5 mL) and sodium methoxide solution (2.5 µL, 5.4 M in methanol) was added. The reaction mixture was stirred at ambient conditions for 1 h, until TLC indicated full consumption of the starting material and was subsequently neutralised with Amberlite® IR 120. The resin was filtered off and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol/dH_20, 20:2:1) under avoidance of light and a subsequent size exclusion chromatography (LH-20, methanol) to afford compound **F2** as an orange solid.

Yield: 39 mg (0.047 mmol, 88 %);

TLC: R_f = 0.38 (DCM/methanol, 10:1),
trans-isomer: 0.35 (ethyl acetate/methanol/dH_20, 20:2:1),
cis-isomer: 0.23 (ethyl acetate/methanol/dH_20, 20:2:1);

Melting point: 112.7 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 70.8 (c = 0.007 mmol · mL^{-1}, mehanol);

^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.87 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^{3}J_{Ar\text{-}H6,\ Ar\text{-}H5}$= 8.7 Hz, $^{3}J_{Ar\text{-}H6,Ar\text{-}H2}$= 2.3 Hz, $^{5}J_{Ar\text{-}H6\text{-}F}$ = 1.2 Hz, 1H, Ar-H_6), 7.64 (dd, $^{3}J_{Ar\text{-}H\text{-}3,F}$ = 12 Hz, $^{4}J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.3 Hz, 1H, Ar-H_2), 7.54 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.09 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^{3}J_{1,2}$ = 1.7 Hz, 1H, H-1), 4.22 (dd∼t, $^{3}J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.7 Hz, 2H, Ar-O$C\underline{H}_2$), 4.10 (dd, $^{3}J_{2,3}$ = 3.4 Hz, , $^{3}J_{2,1}$ = 1.7 Hz, 1H, H-2), 3.94 (dd, $^{3}J_{3,4}$ = 9.4 Hz, $^{3}J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.86 (dd∼t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.7 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.86 (s, 2H, $CO\underline{C}H_2I$), 3.79-3.63 (m, 8H, H-4, H-5, H-6, H-6', $OC\underline{H}_2CH_2O$, $OCH_2C\underline{H}_2O$), 3.59-3.54 (m, 2H, $OC\underline{H}_2CH_2NH$), 3.44-3.34 (m, 2H, $C\underline{H}_2NH$) ppm;

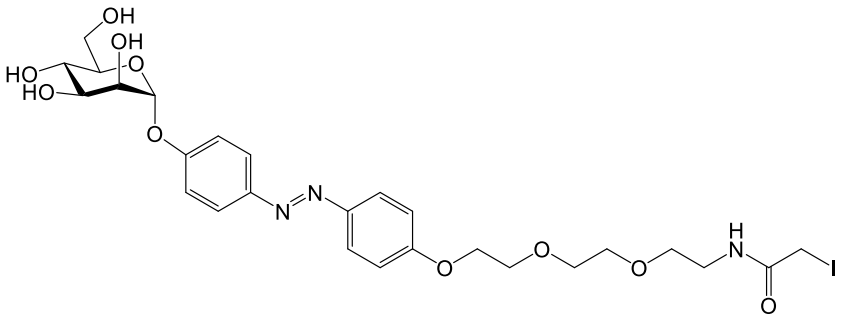
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 172.6 ($NH\underline{C}OCH_2I$), 163.1 (Ar-$C_{4'}$), 154.7 (d, $^{1}J_{Ar\text{-}C3,F}$ = 247.6 Hz, Ar-C_3), 149.5 (d, $^{3}J_{Ar\text{-}C1,F}$ = 5.5 Hz, Ar-C_1), 148.1 (Ar-$C_{1'}$), 147.5 (d, $^{2}J_{Ar\text{-}C4,F}$ = 11.8 Hz, Ar-C_4), 125.8 (2C, Ar-$C_{2'/6'}$), 122.8 (d, $^{4}J_{Ar\text{-}C6,F}$ = 3.1 Hz Ar-C_6), 119.3 (Ar-H_5), 116.1 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^{2}J_{Ar\text{-}C2,F}$ = 20.6 Hz, Ar-H_2), 101.2 (C-1), 76.1 (C-5), 72.6 (C-3), 72.3 ($O\underline{C}H_2CH_2O$), 71.8 (C-2), 71.77 ($OCH_2\underline{C}H_2O$), 71.3 (Ar-$OCH_2\underline{C}H_2O$), 70.8 ($O\underline{C}H_2CH_2NH$), 69.0 (Ar-$O\underline{C}H_2$), 68.2 (C-4), 62.6 (C-6), 39.7 ($\underline{C}H_2NH$), – 2.02 ($CO\underline{C}H_2I$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 132.38 (s, Ar-$F_{\textit{trans}\text{-isomer}}$), – 131.96 (m, Ar-$F_{\textit{cis}\text{-isomer}}$) ppm;

IR (ATR): $\tilde{\nu}$ = 3313 (br), 2923 (m), 2237 (w), 2144 (w), 2002 (w), 1740 (m), 1649 (m), 1599 (m), 1546 (m), 1499 (s), 1423 (m), 1249 (s), 1219 (m), 1092 (s), 1063 (s), 998 (s), 926 (s), 835 (s), 679 (m), 619 (m), 519 (m), 462 (m), 423 (m) cm^{-1};

ESI-HRMS: m/z = 694.12639 $[M + H]^+$, (calc. 694.12674 for $C_{26}H_{33}FIN_3O_{10} + H^+$).

{4-[4'-(2-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}ethoxy)phenyl]azophenoxy}-*α*-D-mannopyranoside (H2)[438]



Compound **45** (100 mg, 0.12 mmol) was dissolved in methanol (5 mL) and sodium methoxide solution (2.5 µL, 5.4 M in methanol) was added. The reaction mixture was stirred at ambient conditions for 1 h, until TLC indicated full consumption of the starting material and was subsequently neutralised with Amberlite® IR 120. The resin was filtered off and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol/dH_20, 10:2:1) to afford compound **H2** as an orange solid.

Yield: 48.4 mg (0.07 mmol, 60 %);

TLC: R_f = 0 (cyclohexane/ethyl acetate/methanol, 5:5:1),
trans-isomer: 0.39 (cyclohexane/ethyl acetate/methanol, 20:2:1),
cis-isomer: 0.27 (cyclohexane/ethyl acetate/methanol, 20:2:1);

Melting point: 70.3 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 100.2 (c = 0.05, ethanol);

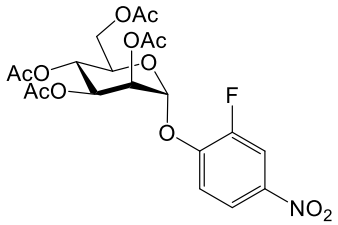
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.86 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.85 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.26 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.08 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.59 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.23 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.6 Hz, 2H, Ar-OCH_2), 4.04 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.5 Hz, 1H, H-2), 3.92 (dd, $^{3}J_{2,3}$ = 3.5 Hz, $^{3}J_{3,4}$ = 9.5 Hz, 1H, H-3), 3.89 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.6 Hz, 2H, Ar-$OCH_2C\underline{H}_2$), 3.79-3.72 (m, 5H, H-4, H-6, H-6', $OC\underline{H}_2CH_2O$), 3.68 (s, 2H, $NHCOC\underline{H}_2I$), 3.66-3.64 (m, 2H, $OCH_2C\underline{H}_2O$), 3.59 (ddd, 1H, $^{3}J_{5,4}$ = 9.7 Hz, $^{3}J_{5,6}$ = 5.1 Hz, $^{3}J_{5,6'}$ = 2.3 Hz, H-5), 3.55 (dd∼t, $^{3}J_{C\underline{H}2CH2NH,CH2NH}$ = 5.3 Hz, 2H, $C\underline{H}_2CH_2NH$), 3.36 (t, $^{3}J_{C\underline{H}2CH2NH,\ C\underline{H}2NH}$ = 5.3 Hz, 2H, $C\underline{H}_2NH$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 162.7 ($NH\underline{C}OCH_3$), 162.6 (Ar-C_4), 159.9 (Ar-$C_{4'}$), 149.3 (Ar-$C_{1'}$), 148.4 (Ar-C_1), 125.5 (2C, Ar-$C_{2/6}$), 125.2 (2C, Ar-$C_{2'/6'}$), 118 (2C, Ar-$C_{3'/5'}$), 116 (2C, Ar-$C_{3/5}$), 100.2 (C-1), 75.7 (C-5), 72.4 (C-3), 71.92 (C-2), 71.87 ($O\underline{C}H_2CH_2O$), 71.4 ($OCH_2\underline{C}H_2O$), 70.8 (Ar-$OCH_2O\underline{C}H_2$), 70.2 ($\underline{C}H_2CH_2NH$), 69 (Ar-$O\underline{C}H_2$), 68.3 (C-4), 62.7 (C-6), 41.1 ($\underline{C}H_2NH$), – 2.04 ($NHCO\underline{C}H_2I$) ppm;

IR (ATR): $\tilde{\nu}$ = 3330 (br), 2928(m), 1649 (m), 1597 (m), 1582 (m), 1497(m), 1231 (s), 1104 (s), 1005 (s), 975 (s), 841 (s), 799 (m), 682 (m) cm^{-1};

ESI-HRMS: *m/z* = 698.11753 $[M + Na]^+$, (calc. 698.11811 for $C_{26}H_{34}IN_3O_{10} + Na^+$).

2-Fluoro-4-nitrophenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (7)



O-(2,3,4,6-Tetra-*O*-acetyl-*α*-D-mannopyranosyl)trichloroacetimidate **5** (9.68 g; 19.65 mmol), 2-fluoro-4-nitrophenol **6** (3.13 g; 19.65 mmol) and a spatula tip of activated molecular sieves (3 Å) were dissolved in dry DCM (30 mL) and dry acetonitrile (12 mL) under a nitrogen atmosphere. After cooling to 0 °C, boron trifluoride diethyletherate (6.24 mL; 39.29 mmol) was added dropwise and the reaction mixture was stirred for 16 h at room temperature. DCM (100 mL) was added and the organic layer was sequentially washed with aqueous solutions of 1 N NaOH (50 mL), sat. $NaHCO_3$ (4 x 100 mL) and dH_2O (250 mL). The combined organic layers were dried over magnesium sulfate, filtered and concentrated under reduced pressure. The residue was purified by flash chromatography (cyclohexane/ethyl acetate, 2:1) to obtain compound **7** as a white solid.

Yield: 6.49 g (13.33 mmol, 68 %);

TLC: R_f = 0.33 (cyclohexane/ethyl acetate, 2:1);

Melting point: 144.2 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 88.3 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 8.07-8.03 (m, 2H, Ar-$H_{3/5}$), 7.37-7.34 (m, 1H, Ar-H_6), 5.65 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.55 (dd, $^{3}J_{3,4}$ = 9.8 Hz, $^{3}J_{3,2}$ = 3.55 Hz, 1H, H-3), 5.52 (dd, $^{3}J_{2,3}$ = 3.55 Hz, $^{3}J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.39 (dd~t, $^{3}J_{4,3}$ = 9.8 Hz, $^{3}J_{4,5}$ = 9.8 Hz, 1H, H-4), 4.27 (dd, $^{2}J_{6,6'}$ = 12.8 Hz, $^{3}J_{6,5}$ = 5.9 Hz, 1H, H-6), 4.11-4.07 (m, 2H, H-5, H-6'), 2.22 (s, 3H, $COCH_3$), 2.07 (s, 3H, $COCH_3$), 2.05 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$) ppm;

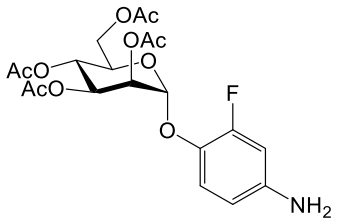
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.3 ($COCH_3$), 169.8 ($COCH_3$), 169.7 ($COCH_3$), 169.6 ($COCH_3$), 151.9 (d, $^{1}J_{Ar\text{-}C2,F}$ = 253 Hz, Ar-C_2), 148.8 (d, $^{2}J_{Ar\text{-}C1,F}$ = 10.8 Hz, Ar-C_1), 142.9 (d, $^{3}J_{Ar\text{-}C4,F}$ = 7 Hz, Ar-C_4), 120.5 (Ar-C_3) 117.0 (Ar-C_6), 113.1 (Ar-C_5), 96.8 (C-1), 70.1 (C-5), 68.8 (C-2), 68.3 (C-3), 65.5 (C-4), 61.9 (C-6), 20.8 ($COCH_3$), 20.63 ($COCH_3$), 20.62 ($COCH_3$), 20.61 ($COCH_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 127.57-127.64 (m, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3784 (w), 3362 (br), 2891 (w), 2317 (w), 2155 (w), 1747 (m), 1637 (m), 1529 (m), 1504 (w), 1247 (m), 1222 (m), 1212 (m), 1129 (m), 973 (m), 807 (m), 743 (m), 695 (m), 602 (m), 451 (m), 411 (m) cm^{-1};

ESI-HRMS: *m/z* = 487.11260 $[M]^+$, (calc. 487.11261 for $C_{20}H_{22}FNO_{12}$).

2-Fluoro-4-aminophenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (8)



To a solution of 2-fluoro-4-nitrophenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside **7** (8.2 g; 16.85 mmol) in dry MeOH (100 mL) and dry DCM (30 mL) a spatula tip of activated palladium catalyst (10 % on charcoal) was added. The reaction mixture was stirred under a hydrogen atmosphere for 16 h. The catalyst was removed by filtration over Celite® 545 and the solvent was removed under reduced pressure. The residue was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 1:1 under the addition of 1 % v/v trimethylamine) to yield compound **8** as a brown solid.

Yield: 7.17 g (15.68 mmol, 93 %);

TLC: R_f = 0.32 (cyclohexane/ethyl acetate, 1:1);

Melting point: 51.9 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 81.9 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 6.97 (t, $^{3}J_{Ar\text{-}H\text{-}3,F}$ = 8.8 Hz, 1H, Ar-H_3), 6.48 (dd, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 12.2 Hz, $^{4}J_{Ar\text{-}H6,F}$ = 2.7 Hz, 1H, Ar-H_6), 6.40-6.38 (m, 1H, Ar-H_5), 5.54 (dd, $^{3}J_{3,4}$ = 9.9 Hz, $^{3}J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.49 (dd, $^{3}J_{2,3}$ = 3.5 Hz, $^{3}J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.35 (dd~t, $^{3}J_{4,3}$ = 9.9 Hz, $^{3}J_{4,5}$ = 9.9 Hz, 1H, H-4), 5.29 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.32-4.26 (m, 2H, H-5, H-6), 4.13-4.11 (m, 1H, H-6'), 2.18 (s, 3H, $COCH_3$), 2.07 (s, 3H, $COCH_3$), 2.06 (s, 3H, $COCH_3$), 2.03 (s, 3H, $COCH_3$) ppm;

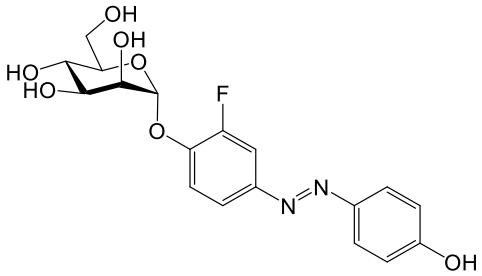
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.6 ($\underline{C}OCH_3$), 170.0 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.8 ($\underline{C}OCH_3$), 154.4 (d, $^1J_{Ar\text{-}C2,F}$ = 247 Hz, Ar-C_2), 143.4 (Ar-C_4), 135.6 (d, $^2J_{Ar\text{-}C1,F}$ = 12 Hz, Ar-C_1), 112.5 (Ar-C_3), 110.7 (Ar-C_5), 103.9 (Ar-C_6), 98.8 (C-1), 69.44 (C-2), 69.41 (C-5), 68.8 (C-3), 66.0 (C-4), 62.2 (C-6), 20.9 ($CO\underline{C}H_3$), 20.7 ($CO\underline{C}H_3$), 20.68 ($CO\underline{C}H_3$), 20.67 ($CO\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.8 (dd, $^3J_{Ar\text{-}F,Ar\text{-}H3}$ = 11.8 Hz, $^4J_{Ar\text{-}F,Ar\text{-}H6}$ = 9.2 Hz, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3470 (w), 3376 (w), 2927 (w), 2852 (w), 2367 (w), 2349 (w), 2318 (w), 1745 (s), 1634 (m), 1529 (w), 1515 (s), 1369 (m), 1213 (s), 1135 (m), 1086 (m), 1037 (m), 1028 (m), 978 (m), 837 (w), 750 (w), 602 (w), 450 (w), 411 (w) cm^{-1};

ESI-HRMS: m/z = 458.14508 $[M + H]^+$, (calc. 458.14570 for $C_{20}H_{24}FNO_{10} + H^+$).

4'-Hydroxyphenyl-4-(3-Fluoro)-azophenoxy α-D-mannopyranoside (10)



Procedure 1a:

2-Fluoro-4-aminophenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside **8** (1.25 g; 2.73 mmol) was dissolved in 1 N HCl_{aq} (10 mL) and cooled to 0 °C on ice. A solution of sodium nitrite (245 mg; 3.55 mmol) in dH_2O (2 mL) and methanol (15 mL) were added. Phenol (257 mg; 2.73 mmol) was dissolved in methanol (3 mL) and alkalised by addition of KOH (279 mg; 5.47 mmol). The phenolate solution was cooled on ice and added dropwise to the diazonium salt solution. The reaction mixture was stirred at ambient conditions for 16 h. If TCL indicated incomplete deacetylation, some more KOH was added and stirred under ambient conditions for another 2 h. The solvent was removed under reduced pressure and the residue was dissolved in a hot aqueous solution of 1 N NaOH and recrystallised under the addition of 1 N HCl_{aq}. The precipitate was washed with cold dH_2O and dried under reduced pressure. Compound **10** was obtained as brown crude product.

Yield: 0.88 g (crude product);

Procedure 1b:

2-Fluoro-4-aminophenyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside **8** (3.59 g; 12.79 mmol) was dissolved in 1 N HCl_{aq} (45 mL) and cooled to 0 °C on ice. A solution of sodium nitrite (1.14 g; 16.62 mmol) in dH_2O (10 mL) and methanol (67 mL) were added. Phenol (1.2 g; 12.79 mmol) was dissolved in methanol (14 mL) and alkalised by addition of KOH (1.31 g; 25.57 mmol). The phenolate solution was cooled on ice and added dropwise to the diazonium salt solution. The reaction mixture was stirred at ambient conditions for 16 h. If TCL indicated incomplete deacetylation, some more KOH was added and stirred under ambient conditions for another 2 h. The solvent was removed under reduced pressure and the residue was filtered over silica gel (ethyl acetate/methanol 9:1). Compound **10** was obtained as an orange crude product.

Yield: 2.79 g (crude product);

Procedure 2:

The crude product of compound **10** (procedures 1a and 1b; 2.8 g) was dissolved in pyridine (100 mL) and acetic anhydride (10 mL) was added. The reaction mixture was stirred at ambient temperature for 16 h and the solvent was removed under reduced pressure. The residue was dissolved in ethyl acetate (300 mL) and washed with 1 N HCl_{aq} (300 mL). The aqueous layer was extracted with DCM (2 x 200 mL) and the combined organic layers were dried over magnesium sulfate, filtered and concentrated under reduced pressure. Flash column chromatography on silica gel (cyclohexane/ethyl acetate, 2:1) yielded the peracetylated intermediate as an orange solid. The latter was dissolved in methanol (200 mL) and sodium methoxide solution (4.5 M in methanol; 100 µL) was added. The reaction mixture was stirred for 2 h under ambient conditions and was neutralised with Amberlite® IR 120. The resin was filtered off and the solvent was removed under reduced pressure to afford compound **10** in 53 % yield as an orange solid.

Yield: 2.26 g (5.73 mmol, 53 % over 2 steps);

TLC: R_f = 0.20 (ethyl acetate/methanol, 9:1);

Melting point: 206.9 °C (decomposition);

Rotational value: $[\alpha]_{20}^{D}$ = + 142 (c = 0.03 mmol · mL^{-1}, methanol);

^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.79 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 8.9 Hz, 2H, Ar-$H_{2'/6'}$), 7.69 (ddd, $^3J_{Ar\text{-}H6,\ Ar\text{-}H5}$= 8.7 Hz, $^4J_{Ar\text{-}H6,Ar\text{-}H2}$ = 2.3 Hz, $^5J_{Ar\text{-}H\text{-}6\text{-}F}$ = 1.3 Hz, 1H, Ar-H_6), 7.62 (dd, $^3J_{Ar\text{-}H2,F}$ = 12 Hz, $^4J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.3 Hz, 1H, Ar-H_2), 7.53 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 6.90 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 8.9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.09 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.94 (dd, $^3J_{3,4}$ = 9.4 Hz, $^3J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.79 (dd∼t, $^3J_{4,5}$ = 9.5 Hz, $^3J_{4,3}$ = 9.5 Hz, 1H, H-4), 3.78-3.71 (m, 2H, H-6, H-6'), 3.66 (ddd, $^3J_{5,4}$ = 9.5 Hz, $^3J_{5,6}$ = 5.6 Hz, $^3J_{5,6'}$ = 2.2 Hz, 1H, H-5) ppm;

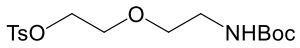
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 162.3 (Ar-$C_{4'}$), 154.8 (d, $^1J_{Ar\text{-}C3,F}$ = 247 Hz, Ar-C_3), 194.6 (d, $^3J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 147.31 (Ar-$C_{1'}$), 147.30 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 125.7 (2C, Ar-$C_{2'/6'}$), 122.6 (Ar-C_5), 119.3 (Ar-C_6), 116.8 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^2J_{Ar\text{-}C2,F}$ = 21 Hz, Ar-C_2), 101.3 (C-1), 76.1 (C-5), 72.3 (C-3), 71.8 (C-2), 68.2 (C-4), 62.7 (C-6) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 132.45 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3313 (br), 2921 (m), 2342 (w), 2296 (w), 2177 (w), 1599 (m), 1505 (s), a437 (m), 1273 (s), 1198 (m), 1273 (s), 1198 (m), 1069 (s), 1063 (s), 1045 (m), 1005 (s), 974 (s), 882 (m), 843 (s), 790 (m), 681 (m), 521 (m), 431 (w) cm^{-1};

ESI-HRMS: *m/z* = 395.12448 $[M + H]^+$, (calc. 395.125456 for $C_{18}H_{19}FN_2O_7 + H^+$); 417.10627 $[M + Na]^+$, (calc. 417.10685 for $C_{18}H_{19}FN_2O_7 + Na^+$).

_O_-Tosyl-{2-[2-(_tert_-Butoxycarbonylamino)ethoxy]}ethanol[438,446] (11)



4-Toluenesulfonyl chloride (2.96 g; 14.75 mmol) was added to a solution of commercially available 2-[2-(*tert*-butoxycarbonylamino)ethoxy]ethanol (2.30 g; 10.31 mmol) in pyridine (5 mL) and the reaction mixture was stirred for 20 h under ambient conditions. Deionised water (100 mL) was added and the aqueous layer was extracted with DCM (4 x 50 mL). The combined organic layers were washed with 1 N HCl_{aq}, dried over magnesium sulfate, the resin was filtered off and the filtrate was concentrated under reduced pressure. The residue was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 2:1) to yield compound **11** as colorless oil.

Yield: 2.17 g (6.03 mmol, 59 %);

TLC: R_f = 0.27 (cyclohexane/ethyl acetate, 2:1);

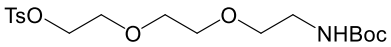
^{1}H-NMR ($CDCl_3$, 200 MHz, 298 K): δ = 7.81 (d, 2H, $^3J_{Ar\text{-}H,ortho,\ Ar\text{-}H,meta}$ = 8.3 Hz, Ar-$H_{(Tos)\text{-}S,ortho}$), 7.36 (d, 2H, $^3J_{Ar\text{-}H,ortho,\ Ar\text{-}H,meta}$ = 8.2 Hz, Ar-$H_{(Tos)\text{-}S,meta}$), 4.8 (br. s, 1H, NH), 4.17 (t, $^3J_{SOCH2,\ SOCH2C\underline{H}2}$ = 4.6 Hz, 2H, $SOCH_2$), 3.63 (t, $^3J_{SOCH2,\ SOCH2C\underline{H}2}$ = 4.8 Hz, 2H, $SOCH_2C\underline{H}_2$), 3.45 (t, $^3JC\underline{H}_2CH_2NH$, $C_{H2}NH$ = 5 Hz, 2H, $C\underline{H}_2CH_2NH$), 3.27-3.22 (m, 2H, $C\underline{H}_2NH$), 2.46 (s, 3H, Ar-CH_3) 1.45 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 155.9 ($\underline{C}O_2C(CH_3)_3$), 144.9 (Ar-$C_S$), 133.0 (Ar-$\underline{C}_{CH3}$), 129.8 (2C, Ar-$H_{(Tos)\text{-}S,ortho}$), 127.9 (2C, Ar-$H_{(Tos)\text{-}S,meta}$), 79.36 ($CO_2\underline{C}(CH_3)_3$), 70.4 ($O\underline{C}H_2CH_2NH$), 69.1 ($SOCH_2$), 68.4 ($SOCH_2CH_2$), 40.2 ($\underline{C}H_2NH$), 28.4 ($CO_2\underline{C}(CH_3)_3$), 21.6 (Ar-$CH_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3404 (br), 2976 (w), 2930 (w), 2252 (w), 1707 (m), 1509 (m), 1454 (m), 1356 (s), 1249 (m), 1173 (s), 1129 (s), 1017 (m), 919 (s), 815 (s), 775 (s), 662 (s) cm^{-1};

ESI-HRMS: *m/z* = 382.12891 $[M + Na]^+$, (calc. 382.12948 for $C_{16}H_{25}NO_6S + Na^+$).

_O_-Tosyl-(2-{2-[2-(_tert_-butyloxycarbonylamino)ethoxy]ethoxy})ethanol[438,448] (12)



Comercially available *O*-tosyl-{2-[2-(2-azidoethoxy)ethoxy]}ethanol (1.28 g; 3.89 mmol) and di-*tert*-butyl dicarbonate (1.27 g; 5.83 mmol) were dissolved in ethyl acetate (40 mL) under a nitrogen atmosphere. A spatula tip of activated palladium catalyst (10 % on charcoal) was added and the reaction mixture was stirred under a hydrogen atmosphere for 16 h. The catalyst was removed by filtration over Celite® 545 and the solvent was removed under reduced pressure. The residue was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 2:1 → 1:1) to yield compound **12** as a colorless oil.

Yield: 1.44 g (3.56 mmol, 92 %);

TLC: R_f = 0.16 (cyclohexane/ethyl acetate, 2:1);

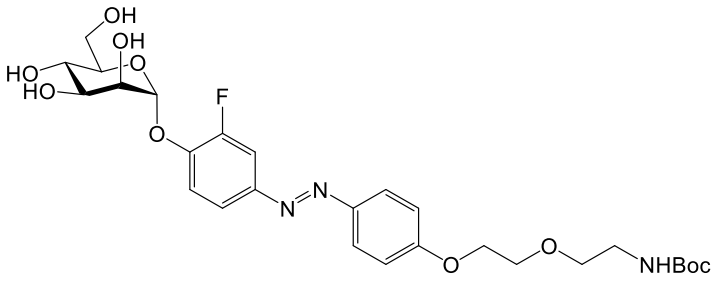
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.73 (d, 2H, $^3J_{Ar\text{-}H,ortho,Ar\text{-}H,meta}$ = 8.3 Hz, Ar-$H_{(Tos)\text{-}S,ortho}$), 7.28 (d, 2H, $^3J_{Ar\text{-}H,ortho,Ar\text{-}H,meta}$ = 8.2 Hz, Ar-$H_{(Tos)\text{-}S,meta}$), 4.87 (br. s, 1H, NH), 4.1 (t, $^3J_{SOC\underline{H}2,SOCH2C\underline{H}2O}$ = 4.8 Hz, 2H, $SOC\underline{H}_2$), 3.63 (t, $^3J_{SOC\underline{H}2CH2O,SOC\underline{H}2}$ = 4.9 Hz, 2H, $SOCH_2C\underline{H}_2O$), 3.56-3.47 (m, 4H, $OC\underline{H}_2CH_2O$, $OCH_2C\underline{H}_2O$), 3.43 (t, $^3J_{OC\underline{H}2CH2NH,C\underline{H}2NH}$ = 5.2 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.22 (t, $^3J_{C\underline{H}2NH,OC\underline{H}2CH2NH}$ = 5 Hz, 2H, $C\underline{H}_2NH$), 2.38 (s, 3H, Ar-CH_3), 1.37 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 155.9 ($\underline{C}O_2C(CH_3)_3$), 144.8 (Ar-$C_S$), 133.0 (Ar-$C_{CH3}$), 129.8 (2C, Ar-$C_{(Tos)\text{-}S\text{-}ortho}$), 128.0 (2C, Ar-$C_{(Tos)\text{-}S\text{-}meta}$), 79.23 ($CO_2\underline{C}(CH_3)_3$), 70.7 ($O\underline{C}H_2CH_2NH$), 70.3 ($OCH_2\underline{C}H_2O$), 70.2 ($O\underline{C}H_2CH_2O$), 69.2 ($SO\underline{C}H_2$), 68.7 ($SOCH_2\underline{C}H_2$), 40.3 ($\underline{C}H_2NH$), 28.4 ($CO_2C(\underline{C}H_3)_3$), 21.6 (Ar-$CH_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3360 (br), 2872 (w), 2250 (w), 1708 (s), 1512 (m), 1456 (m), 1363 (s), 1249 (m), 1174 (s), 1096 (s), 1016 (m), 919 (s), 816 (m), 775 (s), 663 (s) cm^{-1};

ESI-HRMS: m/z = 426.15493 $[M + Na]^+$, (calc. 426.15569 for $C_{18}H_{29}NO_7S + Na^+$).

{4-[3-(4'-{2-[2-(*tert*-Butoxycarbonylamino)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-α-D-mannopyranoside (13)



4'-Hydroxyphenyl-[4-(3-fluoro)azophenyl]-α-D-mannopyranoside **10** (2.26 g; 3.62 mmol), *O*-tosyl-{2-[2-(*tert*-butoxycarbonylamino)ethoxy]}ethanol **11** (1.30 g; 3.62 mmol) and potassium carbonate (1 g; 7.24 mmol) were dissolved in dry DMF (35 mL) under a nitrogen atmosphere and heated to 90 °C for 2 h under reflux. The mixture was subsequently stirred at room temperature for 16 h. The solvent was removed under reduced pressure and the residue was co-distilled with toluene (2 x 100 mL) and DCM (100 mL) in a rotary evaporator. The residue was filtered on silica gel (ethyl acetate/methanol, 9:1) to obtain **13** as an orange crude product.

Yield: 3.74 g (crude product);

TLC: R_f = 0.35 (ethyl acetate/methanol, 9:1);

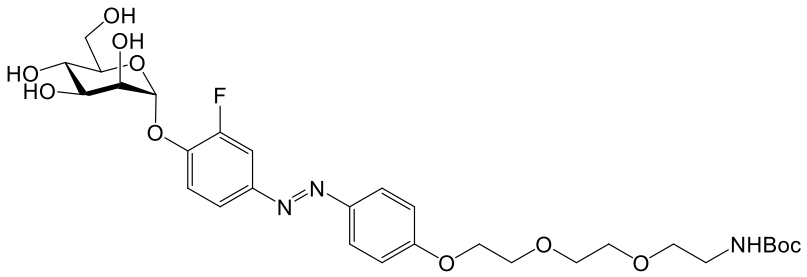
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.87 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^3J_{Ar\text{-}H6,\ Ar\text{-}H5}$ = 8.7 Hz, $^3J_{Ar\text{-}H3,Ar\text{-}H2}$ = 2.4 Hz, $^5J_{Ar\text{-}H6,F}$ = 1.2 Hz, 1H, Ar-H_6), 7.64 (dd, $^3J_{Ar\text{-}H2,F}$ = 11.9 Hz, $^4J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.3 Hz, 1H, Ar-H_2), 7.54 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.09 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, H-1), 4.59 (s, 2H, 2OH), 4.22 (dd~t, $^3J_{\ Ar\text{-}OC\underline{H}2,\ OCH2C\underline{H}2O}$ = 4.6 Hz, 2H, Ar-$OC\underline{H}_2$), 4.10 (dd, $^3J_{2,3}$ = 3.5 Hz, , $^3J_{2,1}$ = 1.9 Hz, 1H, H-2), 3.94 (dd, $^3J_{3,4}$ = 9.5 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 3.85 (dd~t, $^3J_{OCH2C\underline{H}2O,\ Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, OCH2C$\underline{H}_2$O), 3.80 (dd, $^2J_{6,6'}$ = 12 Hz, $^3J_{6,5}$ = 2.4 Hz, 1H, H-6), 3.77-3.71 (m, 2H, H-4, H-6'), 3.64 (ddd, $^3J_{5,4}$ = 9.7 Hz, $^3J_{5,6}$ = 5.4 Hz, $^3J_{5,6'}$ = 2.3 Hz, 1H, H-5), 3.59 (t, $^3J_{OC\underline{H}2CH2NH,\ C\underline{H}2NH}$ = 5.6 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.25 (t, $^3J_{C\underline{H}2NH,\ OC\underline{H}2CH2NH}$ = 5.6 Hz, 2H, $C\underline{H}_2NH$), 1.42 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 163.1 (Ar-$C_{4'}$), 158.5 ($\underline{C}O_2C(CH_3)_3$), 154.8 (d, $^1J_{Ar\text{-}C3,F}$ = 247 Hz, Ar-C_3), 149.6 ($^3J_{Ar\text{-}C1,F}$ = 5.5 Hz, Ar-C_1), 148.2 (Ar-$C_{1'}$), 147.6 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 125.8 (2C, Ar-$H_{2'/6'}$), 122.8 (Ar-C_6), 119.3 (Ar-C_5), 116.1 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 101.3 (C-1), 80.2 ($CO_2\underline{C}(CH_3)_3$), 76.1 (C-5), 72.3 (C-3), 71.79 (C-2), 71.3 ($O\underline{C}H_2CH_2NH$), 70.5 (OCH_2CH_2O), 69.1 (Ar-$O\underline{C}H_2$), 68.2 (C-4), 62.7 (C-6), 41.3 ($\underline{C}H_2NH$), 28.8 ($CO_2C(\underline{C}H_3)_3$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 132.26 (s, Ar-F) ppm;

ESI-HRMS: m/z = 582.24516 $[M + H]^+$, (calc. 582.24575 for $C_{27}H_{36}FN_3O_{10} + H^+$).

(4-{3-[4'-(2-{2-[2-(*tert*-Butoxycarbonylamino)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)-*α*-D-mannopyranoside (14)



4'-Hydroxyphenyl-[4-(3-fluoro)azophenyl]-*α*-D-mannopyranoside **10** (1.03 g; 2.62 mmol), *O*-tosyl-(2-{2-[2-(*tert*-butyloxycarbonylamino)ethoxy]ethoxy})ethanol **12** (1.06 g; 2.62 mmol) and potassium carbonate (0.72 g; 5.24 mmol) were dissolved in dry DMF (30 mL) under a nitrogen atmosphere and heated to 90 °C for 2 h under reflux. The mixture was subsequently stirred at room temperature for 16 h. The solvent was removed under reduced pressure and the residue was co-distilled with toluene (2 x 100 mL) and DCM (100 mL) in a rotary evaporator. The residue was filtered on silica gel (cyclohexane/ethylacetate/methanol, 5:5:1 → ethyl acetate/methanol, 9:1) to obtain **14** as an orange crude product.

Yield: 0.71 g (crude product);

TLC: R_f = 0.12 (cyclohexane/ethyl acetate/methanol, 5:5:1);

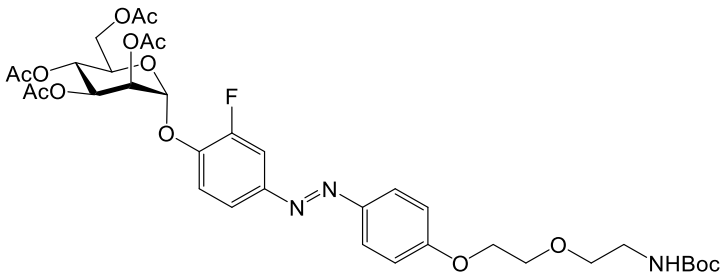
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^3J_{Ar\text{-}H6,\ Ar\text{-}H5}$ = 8.7 Hz, $^4J_{Ar\text{-}H6,Ar\text{-}H2}$ = 2.2 Hz, $^5J_{Ar\text{-}H6,F}$ = 1.1 Hz, 1H, Ar-H_6), 7.64 (dd, $^3J_{Ar\text{-}H2,F}$ = 11.9 Hz, $^4J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.3 Hz, 1H, Ar-H_2), 7.54 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.09 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.59 (s, 2H, 2OH), 4.24 (dd∼t, $^3J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.6 Hz, 2H, Ar-$OC\underline{H}_2$), 4.10 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.94 (dd, $^3J_{3,4}$ = 9.4 Hz, $^3J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.89 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.80 (dd, $^2J_{6,6'}$ = 11.6 Hz, $^3J_{6,5}$ = 3.4 Hz, 1H, H-6), 3.77-3.71 (m, 4H, H-4, H-6', $OC\underline{H}_2CH_2O$), 3.67-3.63 (m, 3H, H-5, $OCH_2C\underline{H}_2O$), 3.51 (t, $^3J_{OC\underline{H}2CH2NH,C\underline{H}2NH}$ = 5.6 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.22 (t, $^3J_{C\underline{H}2NH,OC\underline{H}2CH2NH}$ = 5.6 Hz, 2H, $C\underline{H}_2NH$), 1.42 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 163.1 (Ar-$C_{4'}$), 158.5 ($\underline{C}O_2C(CH_3)_3$), 154.7 (d, $^1J_{Ar\text{-}C3,F}$ = 248 Hz, Ar-C_3), 149.5 ($^3J_{Ar\text{-}C1,F}$ = 5.1 Hz, Ar-C_1), 148.2 (Ar-$C_{1'}$), 147.6 (d, $^2J_{Ar\text{-}C4,F}$ = 11 Hz, Ar-C_4), 125.8 (2C, Ar-$C_{2'/6'}$), 122.8 (Ar-C_6), 119.3 (Ar-C_5), 116.1 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 101.3 (C-1), 80.2 ($CO_2\underline{C}(CH_3)_3$), 76.1 (C-5), 72.3 (C-3), 71.84 ($O\underline{C}H_2CH_2O$), 71.79 (C-2), 71.3 ($OCH_2\underline{C}H_2O$), 71.2 ($O\underline{C}H_2CH_2NH$), 70.8 (Ar-$OCH_2\underline{C}H_2O$), 69.0 (Ar-$O\underline{C}H_2$), 68.2 (C-4), 62.7 (C-6), 41.3 ($\underline{C}H_2NH$), 28.8 ($CO_2C(\underline{C}H_3)_3$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 132.36 (s, Ar-F) ppm;

ESI-HRMS: *m/z* = 626.27114 $[M + H]^+$, (calc. 626.27196 for $C_{29}H_{40}FN_3O_{11} + H^+$).

(4-(3-{4'-{2-[2-(*tert*-Butoxycarbonylamino)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (15)



The crude product of compound **13** (4.44 g) was dissolved in pyridine (35 mL) and acetic anhydride (15 mL) was added. The reaction mixture was stirred under ambient conditions for 2 h and the solvent was removed under reduced pressure. The product was purified by flash chromatography on silica gel in darkness (DCM/diethyl ether, 3:1) to obtain title compound **15** as an orange foam.

Yield: 1.58 g (2.11 mmol, 58 % over 2 steps);

TLC: R_f = 0.33 (cyclohexane/ethyl acetate, 1:1),
trans-isomer: 0.56 (DCM/diethyl ether, 3:1),
cis-isomer: 0.43 (DCM/diethyl ether, 3:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 65.1 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.67 (m, 2H, Ar-$H_{2/6}$), 7.33 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.5 Hz, 1H, Ar-H_5), 7.04 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.39 (dd~t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.6 Hz, 1H, H-6), 4.24-4.19 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.11 (dd, $^2J_{6',6}$ = 12.2 Hz, $^3J_{6',5}$ = 2.1 Hz, 1H, H-6'), 3.86 (t, $^3J_{Ar\text{-}OCH2CH2O,Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, Ar-OCH$_2$C$\underline{H}_2$O), 3.63 (t, $^3J_{OC\underline{H}2OCH2NH,C\underline{H}2NH}$ = 5.1 Hz, 2H, OC$\underline{H}_2$CH$_2$NH), 3.36 (q, $^3J_{C\underline{H}2NH,OC\underline{H}2OCH2NH}$ = 5.1 Hz, 2H, C$\underline{H}_2$NH), 2.22 (s, 3H, COC$\underline{H}_3$), 2.08 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.04 (s, 3H, COC$\underline{H}_3$), 1.35 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

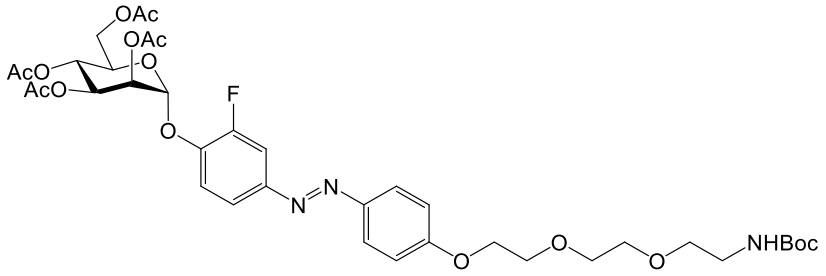
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.78 ($\underline{C}OCH_3$), 169.77 ($\underline{C}OCH_3$), 161.4 (Ar-$C_{4'}$), 156.0 ($\underline{C}O_2C(CH_3)_3$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 248 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.3 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (Ar-C_6), 118.2 (Ar-C_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 19.9 Hz, Ar-C_2), 97.2 (C-1), 79.4 ($CO_2\underline{C}(CH_3)_3$), 70.5 (O$\underline{C}H_2CH_2$NH), 69.8 (C-5), 69.3 (OCH$_2$CH$_2$O), 69.2 (C-2), 68.7 (C-3), 67.7 (Ar-O$\underline{C}$H$_2$), 65.8 (C-4), 62.1 (C-6), 40.3 ($\underline{C}$H$_2$NH), 29.7 ($CO_2C(\underline{C}H_3)_3$), 28.4 (3C, $CO_2C(\underline{C}H_3)_3$), 20.9 (CO$\underline{C}H_3$), 20.7 (CO$\underline{C}H_3$), 20.68 (2C, 2 x CO$\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.47 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 2249 (w), 2161 (w), 1741 (m), 1600 (w), 1503 (m), 1243 (s), 1138 (m), 1090 (w), 1037 (w), 626 (w), 592 (w), 464 (w), 447 (m) cm^{-1};

ESI-HRMS: *m/z* = 750.28720, $[M + H]^+$, (calc. 750.28801 for $C_{35}H_{44}FN_3O_{14} + H^+$).

(4-{3-[4'-(2-{2-[2-(tert-Butoxycarbonylamino)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (16)



The crude product of compound **14** (0.7 g) was dissolved in pyridine (15 mL) and acetic anhydride (4 mL) was added. The reaction mixture was stirred under ambient conditions for 3 h and the solvent was subsequently removed under reduced pressure. The residue was dissolved in ethyl acetate (200 mL) and washed with 1 N HCl_{aq} (200 mL) and dH_2O (200 mL). The aqueous layer was extracted with ethyl acetate (2 x 100 mL) and the combined organic layers were dried on magnesium sulfate and filtered. The product was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 1:1 ⟶ 1:2) to obtain compound **16** as a yellow foam.

Yield: 0.49 g (0.62 mmol, 24 % over 2 steps);

TLC: R_f = *trans*-isomer: 0.30 (cyclohexane/ethyl acetate ,1:1),
cis-isomer: 0.16 (cyclohexane/ethyl acetate, 1:1);

Rotational value: $[\alpha]_{20}^{D}$ = -1.4 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.69-7.66 (m, 2H, Ar-$H_{2/6}$), 7.32 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$= 8.5 Hz, 1H, Ar-$H_5$), 7.04 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 9.9 Hz, $^3J_{3,2}$ = 3.6 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 9.9 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.6 Hz, 1H, H-6), 4.25-4.19 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.11 (dd, $^2J_{6',6}$ = 12.3 Hz, $^3J_{6',5}$ = 2.4 Hz, 1H, H-6'), 3.90 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.7 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.75-3.72 (m, 2H, $OC\underline{H}_2CH_2O$), 3.67-3.65 (m, 2H, $OCH_2C\underline{H}_2O$), 3.56 (t, $^3J_{OC\underline{H}2OCH2NH,C\underline{H}2NH}$ = 5.1 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.33 (dd∼q, $^3J_{C\underline{H}2NH,OC\underline{H}2OCH2NH}$ = 5.1 Hz, 2H, $C\underline{H}_2NH$), 2.22 (s, 3H, $COC\underline{H}_3$), 2.08 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$), 1.45 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

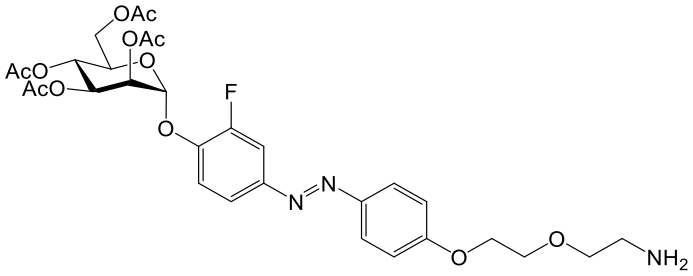
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.78 ($\underline{C}OCH_3$), 169.76 ($\underline{C}OCH_3$), 161.5 (Ar-$C_{4'}$), 155.9 ($\underline{C}O_2C(CH_3)_3$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 250 Hz, Ar-C_3), 149.0 (, $^3J_{Ar\text{-}C1,F}$ = 5.6 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.8 (2C, Ar-$C_{2'/6'}$), 121.4 (Ar-C_6), 118.2 (Ar-C_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 19.6 Hz, Ar-C_2), 97.2 (C-1), 79.9 ($CO_2\underline{C}(CH_3)_3$), 70.8 ($O\underline{C}H_2CH_2O$), 70.31 ($O\underline{C}H_2CH_2NH$), 70.28 ($OCH_2\underline{C}H_2O$), 69.8 (C-5), 69.6 ($O\underline{C}H_2CH_2O$), 69.2 (C-2), 68.7 (C-3), 67.7 (Ar-$OC\underline{H}_2$), 65.8 (C-4), 62.1 (C-6), 40.4 ($C\underline{H}_2NH$), 28.4, 26.9 (3C, $CO_2C(\underline{C}H_3)_3$), 20.8 ($CO\underline{C}H_3$), 20.68 ($CO\underline{C}H_3$), 20.67 ($CO\underline{C}H_3$), 20.66 ($CO\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.49 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3308 (w), 2977 (w), 2927 (w), 2872 /w), 2349 (w), 1715 (m), 1677 (s), 1595 (m), 1534 (m), 1504 (m), 1362 (m), 1288 (s), 1252 (s), 1174 (m), 1130 (s), 1111 (s), 1056 (m), 850 (m), 815 (m), 620 (m), 562 (w), 463 (m), 417 (w) cm^{-1};

ESI-HRMS: *m*/*z* = 794.31273, $[M + H]^+$, (calc. 794.31422 for $C_{37}H_{48}FN_3O_{15} + H^+$).

[4-(3-{4'-[2-(2-Aminoethoxy)ethoxy]phenyl}fluoro)azophenoxy]}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (17)



Procedure 1: *N*-Boc deprotection

Compound **15** (1.04 g; 1.39 mmol) was dissolved in DCM (50 mL) and trifluoroacetic acid (1.7 mL; 7.64 mmol) was added dropwise. The reaction mixture was stirred at ambient conditions for 3 h and subsequently washed with saturated $NaHCO_3$ solution (100 mL). The aqueous layer was extracted with DCM (3 x 100 mL) and the combined organic layers were dried over magnesium sulfate and filtered. The solvent was removed under reduced pressure and the crude TFA salt of compound **17** was obtained as an orange oil.

Yield: 0.91 g (crude product);

Procedure 2: Staudinger reduction

Compound **38** (544 mg, 0.81 mmol) and triphenylphosphane (423 mg, 1.61 mmol) were dissolved in THF (30 mL) and dH_2O (2 mL) under a nitrogen atmosphere. The reaction was stirred 48 h at room temperature and the solvent was removed under reduced pressure. The residue was roughly purified by column chromatography on silica gel (ethyl acetate/methanol, 10:1, under the addition of 1 % v/v diethyl amine), followed by size exclusion chromatography (LH-20, methanol). Crude compound **17** was obtained as an orange oil.

Yield: 460 mg (crude product);

TLC: R_f = 0.24 (ethyl acetate/methanol, 9:1), 0.42 (DCM/methanol, 9:1);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.65 (m, 2H, Ar-$H_{2/6}$), 7.32 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.5 Hz, 1H, Ar-H_5), 7.03 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.6 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.39 (dd∼t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.6 Hz, 1H, H-6), 4.24-4.19 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.11 (dd, $^2J_{6',6}$ = 12.2 Hz, $^3J_{6',5}$ = 2.1 Hz, 1H, H-6'), 3.87 (t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.5 Hz, 2H, Ar-OCH$_2$C$\underline{H}_2$O), 3.62 (dd∼t, $^3J_{OC\underline{H}2OCH2NH2,C\underline{H}2NH2}$ = 5.1 Hz, 2H, OC$\underline{H}_2$CH$_2$NH$_2$), 2.94 (t, $^3J_{C\underline{H}2NH2,OC\underline{H}2OCH2NH2,}$ = 5.1 Hz 2H, C$\underline{H}_2$NH$_2$), 2.22 (s, 3H, COC$\underline{H}_3$), 2.08 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.04 (s, 3H, COC$\underline{H}_3$) ppm;

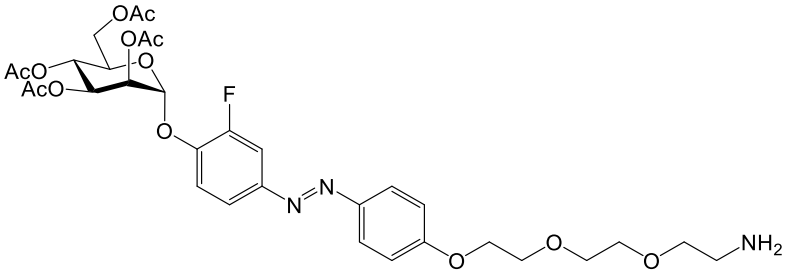
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (C$\underline{O}$CH$_3$), 169.9 (C$\underline{O}$CH$_3$), 169.78 (C$\underline{O}$CH$_3$), 169.75 (C$\underline{O}$CH$_3$), 161.5 (Ar-$C_{4'}$), 153.6 (d, $^1J_{Ar\text{-}C3,F}$ = 250 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.8 (2C, Ar-$C_{2'/6'}$), 121.4 (d, Ar-C_6, $^4J_{Ar\text{-}C6,F}$ = 3.2 Hz), 118.2 (Ar-C_6), 114.9 (2C, Ar$C_{3',5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-H_2), 97.2 (C-1), 70.1 (O$\underline{C}$H$_2$CH$_2$NH$_2$), 69.8 (C-5), 69.4 (OCH$_2\underline{C}$H$_2$O),

69.2 (C-2), 68.7 (C-3), 67.7 (Ar-O<u>C</u>H$_2$), 65.8 (C-4), 62.1 (C-6), 41.99 (<u>C</u>H$_2$NH$_2$), 23.3 (CO<u>C</u>H$_3$), 20.8 (CO<u>C</u>H$_3$), 20.69 (CO<u>C</u>H$_3$), 20.68 (CO<u>C</u>H$_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 76.30 (s, C$_2$H<u>F</u>$_3$O$_2$), – 130.49 (s, Ar-F) ppm;

ESI-HRMS: *m/z* = 650.23479, $[M + H]^+$, (calc. 650.23558 for $C_{30}H_{36}FN_3O_{12} + H^+$).

{4-[3-(4'-{2-[2-(2-Aminoethoxy)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (18)



Procedure 1: *N*-Boc deprotection

Compound **16** (0.66 g; 0.84 mmol) was dissolved in DCM (15 mL) and trifluoroacetic acid (1.3 mL; 5.84 mmol) was added dropwise. The reaction mixture was stirred at ambient conditions for 3 h and subsequently washed with saturated $NaHCO_3$ solution (100 mL) and Brine (200 mL). The aqueous layer was extracted with DCM (2 x 150 mL) and the combined organic layers were dried over magnesium sulfate and filtered. The solvent was removed under reduced pressure and the crude TFA salt of compound **18** was obtained as an orange oil.

Yield: 0.56 g (crude product);

Procedure 2: Staudinger reduction

Compound **39** (350 mg, 0.49 mmol) and triphenylphosphane (191 mg, 0.73 mmol) were dissolved in THF (15 mL) and dH_2O (1 mL) under a nitrogen atmosphere. The reaction was stirred 20 h at room temperature. The solvent was removed under reduced pressure. Crude compound **18** was obtained as an orange oil.

Yield: 519 mg (crude product);

TLC: R_f = 0.21 (ethyl acetate/methanol, 9:1);

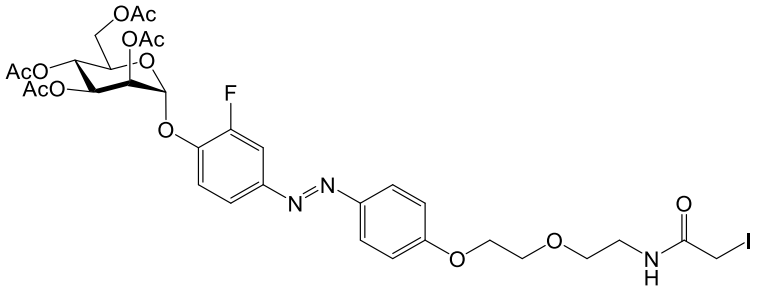
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-H$_{2'/6'}$), 7.70-7.66 (m, 2H, Ar-H$_{2/6}$), 7.32 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.4 Hz, 1H, Ar-H$_5$), 7.03 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-H$_{3'/5'}$), 5.59 (d, $^3J_{1,2}$ = 1.9 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.9 Hz, 1H, H-2), 5.38 (dd~t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.29 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.4 Hz, 1H, H-6), 4.24-4.19 (m, 3H, H-5, Ar-OC<u>H</u>$_2$), 4.11 (dd, $^2J_{6',6}$ = 12.2 Hz, $^3J_{6',5}$ = 2.1 Hz, 1H, H-6'), 3.89 (t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.5 Hz, 2H, Ar-OCH$_2$C<u>H</u>$_2$O), 3.75-3.71 (m, 2H, OC<u>H</u>$_2$CH$_2$O), 3.70-3.67 (m, 4H, OCH$_2$C<u>H</u>$_2$O, OC<u>H</u>$_2$CH$_2$NH$_2$), 3.06 (t, 2H, C<u>H</u>$_2$NH2), 2.21 (s, 3H, COC<u>H</u>$_3$), 2.08 (s, 3H, COC<u>H</u>$_3$), 2.05 (s, 3H, COC<u>H</u>$_3$), 2.04 (s, 3H, COC<u>H</u>$_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.78 ($\underline{C}OCH_3$), 169.75 ($\underline{C}OCH_3$), 161.5 (Ar-$C_{4'}$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 250 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 145 (d, $^2J_{Ar\text{-}C4,F}$ = 11.9 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (d, Ar-C_6, $^4J_{Ar\text{-}C6,F}$ = 3.2 Hz), 118.2 (Ar-C_6), 114.9 (2C, $ArC_{3',5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-H_2), 97.2 (C-1), 70.6 ($OCH_2\underline{C}H_2O$), 70.1 ($O\underline{C}H_2CH_2NH_2$), 69.8 (C-5), 69.5 ($O\underline{C}H_2CH_2O$), 69.2 (C-2), 68.7 (C-3), 67.6 (Ar-$O\underline{C}H_2$), 65.8 (C-4), 62.1 (C-6), 55 ($\underline{C}H_2NH_2$), 21.9 ($CO\underline{C}H_3$), 20.9 ($CO\underline{C}H_3$), 20.7 ($CO\underline{C}H_3$), 20.68 ($CO\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = -76.25 (s, $C_2H\underline{F}_3O_2$), -128.45 (s, Ar-F) ppm;

ESI-HRMS: *m*/*z* = 694.26040, $[M + Na]^+$, (calc. 694.26053 for $C_{34}H_{40}FN_3O_{14} + Na^+$).

{4-[3-(4'-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (19)



Crude compound **17** (450 mg; ~ 0.69 mmol) was dissolved in dry DCM (25 mL) under a nitrogen atmosphere and 2-iodoacetic anhydride (295 mg; 0.83 mmol) was added. The reaction mixture was stirred 16 h in darkness at room temperature. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (cyclohexane/ethyl acetate/methanol, 5:5:1), whereas the column was wrapped in aluminium foil in order to prevent *cis*/*trans*-isomerisation during purification. The compound **19** was obtained as an orange foam.

Yield: 0.32 g (0.39 mmol, ~ 57 %);

TLC: R_f = *trans*-isomer: 0.39 (cyclohexane/ethyl acetate/methanol, 5:5:1),
cis-isomer: 0.32 (cyclohexane/ethyl acetate/methanol, 5:5:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 47.5 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.90 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.71-7.66 (m, 2H, Ar-$H_{3/5}$), 7.33 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.5 Hz, 1H, Ar-H_5), 7.05 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.9 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.9 Hz, 1H, H-2), 5.39 (dd~t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.5 Hz, 1H, H-6), 4.25-4.19 (m, 3H, H-5, Ar-O$\underline{C}H_2$), 4.11 (dd, $^2J_{6',6}$ = 12.2 Hz, $^3J_{6',5}$ = 2.2 Hz, 1H, H-6'), 3.89 (t, $^3J_{OCH2C\underline{H}2O,\ Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, $OCH_2C\underline{H}_2O$), 3.67 (s, 2H, $COC\underline{H}_2I$), 3.66 (t, $^3J_{OC\underline{H}2OCH2NH,\ C\underline{H}2NH}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.51 (q, $^3J_{C\underline{H}2NH,\ OC\underline{H}2OCH2NH}$ = 5.3 Hz, 2H, $C\underline{H}_2NH$), 2.22 (s, 3H, $COC\underline{H}_3$), 2.08 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$) ppm;

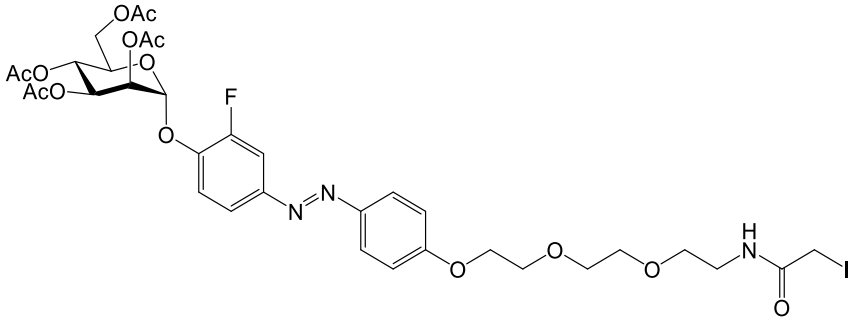
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.78 ($\underline{C}OCH_3$), 169.76 ($\underline{C}OCH_3$), 167.0 ($NH\underline{C}{=}OCH_2I$), 161.4 (Ar-$C_{4'}$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 250 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.4 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 145.0 (d, $^2J_{Ar\text{-}C4,F}$ = 12.1 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (Ar-C_6), 118.2 (Ar-C_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 19.8 Hz, Ar-C_2), 97.2 (C-1), 69.8 (C-5), 69.6 ($O\underline{C}H_2CH_2NH$), 69.5 ($OCH_2\underline{C}H_2O$), 69.2 (C-2), 68.7 (C-3), 67.6 (Ar-$O\underline{C}H_2$), 65.8 (C-4), 62.1 (C-6), 40.1 ($\underline{C}H_2NH$), 20.8 ($CO\underline{C}H_3$), 20.69 ($CO\underline{C}H_3$), 20.66 (2C, 2 x $CO\underline{C}H_3$), – 0.74 ($CO\underline{C}H_2I$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.44 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3368 (br), 2925 (m), 2835 (m), 2384 (w), 2342 (w), 2298 (w), 2178 (w), 2041 (w) 1750 (m), 1655 (m), 1600 (m), 1501 (m), 1455 (w), 1501 (m), 1371 (m), 1248 (s), 1132 (m), 1022 (s), 673 (m), 627 (m), 528 (w), 479 (w) 449 (w) cm^{-1};

ESI-HRMS: m/z = 818.14108 $[M + H]^+$, (calc. 818.14038 for $C_{32}H_{37}FIN_3O_{13} + H^+$).

(4-{3-[4'-(2-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (20)



Crude compound **18** (100 mg; ~ 0.13 mmol) was dissolved in dry DCM (5 mL) under a nitrogen atmosphere and 2-iodoacetic anhydride (56.8 mg; 0.16 mmol) was added. The reaction mixture was stirred 16 h in darkness at room temperature. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (cyclohexane/ethyl acetate/methanol, 10:10:1), whereas the column was wrapped in aluminium foil in order to prevent *cis*/*trans*-isomerisation during purification. The compound **20** was obtained as an orange solid.

Yield: 51.4 mg (0.059 mmol, ~ 47 %);

TLC: R_f = *trans*-isomer: 0.38 (cyclohexane/ethyl acetate/methanol, 10:10:1), *cis*-isomer: 0.29 (cyclohexane/ethyl acetate/methanol, 10:10:1);

Melting point: 67.1 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 55.3 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.71-7.66 (m, 2H, Ar-$H_{2/6}$), 7.33 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.3 Hz, 1H, Ar-H_5), 7.03 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^{3}J_{3,4}$ = 9.9 Hz, $^{3}J_{3,2}$ = 3.6 Hz, 1H, H-3), 5.55 (dd, $^{3}J_{2,3}$ = 3.5 Hz, $^{3}J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.39 (dd∼t, $^{3}J_{4,5}$ = 10 Hz, $^{3}J_{4,3}$ = 9.9 Hz, 1H, H-4), 4.29 (dd, $^{2}J_{6,6'}$ = 12.2 Hz, $^{3}J_{6,5}$ = 5.5 Hz, 1H, H-6), 4.25-4.19 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.11 (dd, $^{2}J_{6',6}$ = 12.2 Hz, $^{3}J_{6',5}$ = 2.2 Hz, 1H, H-6'), 3.92 (dd∼t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,\ Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77-3.73 (m, 2H, $OC\underline{H}_2CH_2O$), 3.72-3.68 (m, 2H, $OCH_2C\underline{H}_2O$), 3.67 (s, 2H, $COC\underline{H}_2I$), 3.63-3.57 (m, 2H, $OC\underline{H}_2CH_2NH$), 3.54-3.46 (m, 2H, $C\underline{H}_2NH$), 2.22 (s, 3H, $COC\underline{H}_3$), 2.08 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$) ppm;

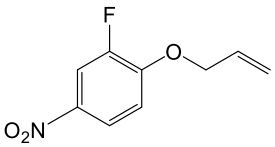
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.78 ($\underline{C}OCH_3$), 169.76 ($\underline{C}OCH_3$), 167.0 ($NH\underline{C}{=}OCH_2I$), 161.3 (Ar-$C_{4'}$), 153.5 (d, $^{1}J_{Ar\text{-}C3,F}$ = 250 Hz, Ar-C_3), 149.0 (d, $^{3}J_{Ar\text{-}C1,F}$ = 5.1 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 145.0 (d, $^{2}J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (d, $^{4}J_{Ar\text{-}C6,F}$ = 3 Hz Ar-C_6), 118.2 (Ar-H_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^{2}J_{Ar\text{-}C2,F}$ = 19.6 Hz, Ar-H_2), 97.2 (C-1), 70.9 ($O\underline{C}H_2CH_2O$), 70.4 ($OCH_2\underline{C}H_2O$), 69.8 (C-5), 69.6 ($OCH_2\underline{C}H_2O$), 69.4 ($O\underline{C}H_2CH_2NH$), 69.2 (C-2), 68.7 (C-3), 67.8 (Ar-$O\underline{C}H_2$), 65.8 (C-4), 62.1 (C-6), 40.2 ($\underline{C}H_2NH$), 20.9 ($CO\underline{C}H_3$), 20.70 ($CO\underline{C}H_3$), 20.68 (2C, 2 x $CO\underline{C}H_3$), – 0.63 ($CO\underline{C}H_2I$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.46 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3361 (br), 3071 (w), 2929 (w), 2359 (w), 2326 (w), 2145 (w), 1748 (s), 1655 (m), 1599 (m), 1500 (m), 1245 (s), 1226 (s), 1134 (m), 1094 (m), 1055 (m), 1029 (m) 978 (m), 841 (m), 627 (w), 512 (w), 441 (w) cm^{-1};

ESI-HRMS: m/z = 862.16815 $[M + H]^+$, (calc. 862.16900 for $C_{34}H_{41}FIN_3O_{14} + H^+$).

1-Allyloxy-2-fluoro-4-nitrophenol[449] (22)



2-Fluoro-4-nitrophenol (**6**, 3.14 g, 20 mmol) and potassium carbonate (2.78 g, 20 mmol) were dissolved in acetonitrile (350 mL) under a nitrogen atmosphere. Allyl bromide (1.67 mL, 20 mmol) was added dropwise and the reaction mixture was heated to 80 °C under reflux for 3 h and subsequently stirred at room temperature overnight. After filtration over Celite® 545, the solvent was removed under reduced pressure. The residue was dissolved in ethyl acetate (300 mL) and washed with 1 N NaOH solution, saturated $NaHCO_3$ solution and dH_2O (300 mL each). The organic was dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. A purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 10:1) yielded the title compound **22** as a yellow oil.

Yield: 3.69 g (18.7 mmol, 93 %);

TLC: R_f = 0.38 (cyclohexane/ethyl acetate, 10:1);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 8.04 (dd, $^{3}J_{Ar\text{-}H5,\ Ar\text{-}H6}$ = 9.1 Hz, $^{4}J_{Ar\text{-}H5,Ar\text{-}H3}$= 2.7 Hz 1H, Ar-$H_5$), 7.99 (dd, $^{3}J_{Ar\text{-}H3,F}$ = 10.6 Hz, $^{4}J_{Ar\text{-}H3,Ar\text{-}H5}$= 2.7 Hz, 1H, Ar-$H_3$), 7.06 (t, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 9.1 Hz, 1H, Ar-H_6) 6.1-6.02 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.47 (dd, $^{2}J_{CH\underline{H}',C\underline{H}H'}$ = 1.2 Hz,$^{3}J_{C\underline{H}=CHH',CH=CH\underline{H}'}$ = 17 Hz, 1H, CH=CH$\underline{H}$'), 5.39 (dd, $^{2}J_{C\underline{H}H',CH\underline{H}'}$ = 1.2 Hz, $^{3}J_{CH=C\underline{H}H',C\underline{H}=CHH'}$ = 10.6 Hz, 1H, CH=C$\underline{H}$H'), 4.73 (d, $^{3}J_{OC\underline{H}2CH,OCH2C\underline{H}}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH$) ppm;

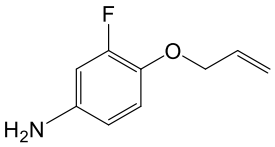
^{3}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 152.3 (d, $^{2}J_{Ar\text{-}C1,F}$ = 10.5 Hz, Ar-C_1), 152.3 (d, $^{1}J_{Ar\text{-}C2,F}$ = 247 Hz, Ar-C_2), 140.9 (d, $^{3}J_{Ar\text{-}C4,F}$ = 7 Hz Ar-C_4), 131.3 ($OCH_2\underline{C}H=CH_2$), 120.78 (d, $^{4}J_{Ar\text{-}C5,F}$ = 3.5 Hz, Ar-H_5), 119.3 ($CH_2=\underline{C}HH'$), 113.4 (d, $^{3}J_{Ar\text{-}C6,F}$ = 2 Hz, Ar-C_6), 112.4 (d, $^{2}J_{Ar\text{-}C3,F}$ = 23 Hz, Ar-C_3), 70.3 ($O\underline{C}H_2CH$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.48 (s, Ar-F) ppm;

EI-MS: *m/z* = 197.05 $[M]^+$; (calc. 197.04882 for $C_9H_8FNO_3$).

Further analytical data corresponds to the literature.

1-Allyloxy-2-fluoro-4-anilin (24)



Procedure 1[450]

1-Allyloxy-2-fluoro-4-nitrophenol **22** (1.79 g, 9.08 mmol) was dissolved in dH_2O (60 mL) and iron powder (7.3 g, 0.131 mol) and glacial acetic acid (300 µL) were added. The reaction mixture was heated to 100 °C under reflux for 30 min. When the mixture was cooled to room temperature, the iron powder was filtered off and the filtrate was extracted with ethyl acetate (4 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography on silica gel (cyclohexane/ethyl acetate 7:3) to obtain compound **24** as a brown oil.

Yield: 998.5 mg (5.97 mmol, 66 %);

Procedure 2[452]

1-Allyloxy-2-fluoro-4-nitrophenol **22** (1.79 g, 9 mmol) and $Sn_2Cl \cdot 2\ H_2O$ (10.2 g, 45.4 mmol) were dissolved in dry ethanol (60 mL) under a nitrogen atmosphere and refluxed at 80 °C for 20 min. Subsequently, cold dH_2O (200 mL) and saturated solution of $NaHCO_3$ (150 mL) were added and the aqueous phase was extracted with DCM and ethyl acetate (1 L each). The combined organic layers dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The crude product **24** was purified by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 7:3).

Yield: 741 mg (4.43 mmol, 49 %);

TLC: R_f = 0.20 (cyclohexane/ethyl acetate, 5:1);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 6.80 (t $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 9 Hz, Ar-H_6), 6.46 (dd, $^3J_{Ar\text{-}H3\text{-}F}$ = 12.6 Hz, $^4J_{Ar\text{-}3,Ar\text{-}H5}$ = 2.7 Hz, Ar-H_3) 6.35 (ddd, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 9 Hz, $^4J_{Ar\text{-}5,Ar\text{-}H3}$ = 2.7 Hz, $^5J_{Ar\text{-}H5,F}$ = 1.4 Hz, Ar-H_5), 6.98-6.0 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.37 (dd, $^2J_{CH\underline{H}',C\underline{H}H'}$ = 1.5 Hz,$^3J_{C\underline{H}=CHH',\ CH=CH\underline{H}'}$ = 17.2 Hz, 1H, CH=CH$\underline{H}$'), 5.25 (dd, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.5 Hz, $^3J_{CH=C\underline{H}H',OCH2C\underline{H}=CH2}$ = 10.5 Hz, 1H, CH=C$\underline{H}$H'), 4.5 (d, $^3J_{OC\underline{H}2CH,OCH2C\underline{H}}$ = 5.6 Hz, 2H, $OC\underline{H}_2CH$), 3.5 (br. s, 2H NH_2) ppm;

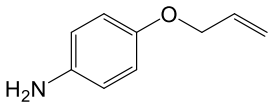
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 153.9 (d, $^1J_{Ar\text{-}C2,F}$ = 244 Hz, Ar-C_2), 141.5 (d, $^3J_{Ar\text{-}C4,F}$ = 8.9 Hz, Ar-C_4), 138.9 (d, $^2J_{Ar\text{-}C1,F}$ = 11.3 Hz, Ar-C_1), 133.5 ($OCH_2\underline{C}H=CH_2$), 118.47 (d, $^3J_{Ar\text{-}C6,F}$ = 3.1 Hz, Ar-C_6), 117.9 ($CH_2=\underline{C}HH'$), 110.3 (d, $^4J_{Ar\text{-}C5,F}$ = 3.2 Hz, Ar-H_5), 104.1 (d, $^2J_{Ar\text{-}C3,F}$ = 21.6 Hz, Ar-C_3), 71.6 ($O\underline{C}H_2CH$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 132.72 (s, Ar-F) ppm;

ESI-HRMS: *m/z* = 168.08168 $[M + H]^+$, (calc. 168.08192 for $C_9H_{10}FNO + H^+$).

Further analytical data corresponds to the literature.

1-Allyloxy-4-anilin[450,451] (25)



Commercially available 4-nitrophenol (**21**, 3 g, 21.7 mmol) and potassium carbonate (3 g, 21.7 mmol) were dissolved in acetonitrile (230 mL) under a nitrogen atmosphere. Allyl bromide (1.87 mL, 21.7 mmol) was added dropwise and the reaction mixture was heated to 80 °C under reflux for 3 h and subsequently was allowed to cool down to room temperature After filtration over Celite® 545, the solvent was removed under reduced pressure. The residue was dissolved in ethyl acetate (200 mL) and washed with 1 N NaOH solution, sat. $NaHCO_3$ solution and dH_2O (200 mL each). The organic was dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The resulting 1-allyloxy-4-nitrophenol **23** was dissolved in dH_2O (133 mL) and iron powder (16.7 g, 299 mmol) and glacial acetic acid (660 µL) were added. The reaction mixture was heated to 85 °C under reflux for 30 min and subsequently was allowed to cool to room temperature. The reaction mixture was neutralised with sat. $NaHCO_3$ solution (10 mL), the iron powder was filtered off and the filtrate was extracted with ethyl acetate (3 x 150 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 7:3) to obtain compound **25** as a brown oil.

Yield: 1.4 g (9.62 mmol, % 44 over 2 steps);

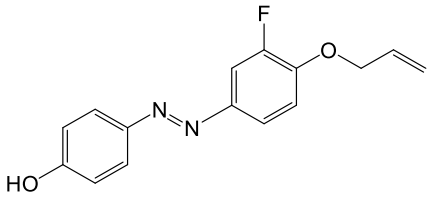
TLC: R_f = 0.25 (cyclohexane/ethyl acetate, 3:1);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 6.76 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 8.8 Hz, 2H, Ar-$H_{3/5}$), 6.63 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 8.8 Hz, 2H, Ar-$H_{2/6}$), 6.08-5.99 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.38 (dd, $^2J_{CH\underline{H}',C\underline{H}H'}$ = 1.6 Hz,$^3J_{C\underline{H}=CHH',CH=CH\underline{H}'}$ = 17.3 Hz, 1H, CH=CH$\underline{H}$'), 5.25 (dd, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.6 Hz, $^3J_{CH=C\underline{H}H',OCH2C\underline{H}=CH2}$ = 10.5 Hz, 1H, CH=C$\underline{H}$H'), 4.45 (d, $^3J_{OC\underline{H}2CH,OCH2C\underline{H}}$ = 5.4 Hz, 2H, $OC\underline{H}_2CH$), 3.28 (br. s, 2H NH_2) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 151.8 (Ar-C_1), 140.1 (Ar-C_4), 133.8 ($OCH_2\underline{C}H=CH_2$), 117.3 ($CH_2=\underline{C}HH'$), 116.3 (Ar-$C_{2/6}$), 115.9 (Ar-$C_{3/5}$), 69.6 ($O\underline{C}H_2CH$) ppm;

ESI-HRMS: *m/z* = 150.09118 $[M + H]^+$, (calc. 150.09134 for $C_9H_{11}NO + H^+$).

4'-{[4-(Allyloxy)-3-fluorophenyl]diazenyl}phenol (26)



1-Allyloxy-2-fluoro-4-anilin **24** (3.1 g; 18.5 mmol) was dissolved in dH_2O (57 mL), acidified with conc. HCl_{aq} (2.45 mL) and cooled to 0 °C on ice. A solution of sodium nitrite (1.27 g; 18.5 mmol) in dH_2O (57 mL) and methanol (57 mL) were added and the mixture was stirred 1 h at 10 °C. Phenol (1.74 g; 18.5 mmol) was dissolved in dH_2O (57 mL) and alkalised by addition of NaOH (0.74 g; 18.5 mmol) and $NaCO_3$ (1.96g, 18.5 mmol). The phenolate solution was cooled on ice and added dropwise to the diazonium salt solution. The reaction mixture was stirred at 10 °C for 1 h, ethyl acetate (300 mL) was added and subsequently washed with dH_2O (300 mL) The aqueous layer was extracted with ethyl acetate (4 x 250 mL) and the combined organic layers were dried over magnesium sulfate and filtered. The solvent was removed under reduced pressure and the residue was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 5:1). The product **26** was obtained as a brown solid.

Yield: 4.37 g (16.05 mmol, 87 %);

TLC: R_f = 0.43 (toluene/ethyl acetate, 10:1), 0.34 (cyclohexane/ethyl acetate, 5:1);

Melting point: 118.6 °C;

^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.77 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 8.8 Hz, 2H, Ar-$H_{2'/6'}$),7.67 (d, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ =8.6 Hz, 1H, Ar-H_5) 7.59 (dd, $^3J_{Ar\text{-}H2,F}$ = 12.3 Hz, $^4J_{Ar\text{-}H2,Ar\text{-}H6}$ = 1.6 Hz, 1H, Ar-H_2), 7.21 (t, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.6 Hz, 1H, Ar-H_6) 6.89 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 8.8 Hz, 2H, Ar-$H_{3'/5'}$), 6.13-6.07 (m, 1H, $OCH_2C\underline{H}{=}CH_2$), 5.46 (dd, $^2J_{CH\underline{H}',CH\underline{H}'}$ = 1.2 Hz, $^3J_{C\underline{H}{=}CHH',CH{=}CH\underline{H}'}$ = 17.2 Hz, 1H, CH=CH$\underline{H}$'), 5.31 (dd, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.2 Hz, $^3J_{CH{=}C\underline{H}H',OCH2C\underline{H}{=}CH2}$ = 10.2 Hz, 1H, CH=C$\underline{H}$H'), 4.69 (d, $^3J_{OC\underline{H}2CH,OCH2C\underline{H}}$ = 4.9 Hz, 2H, O$C\underline{H}_2$CH) ppm;

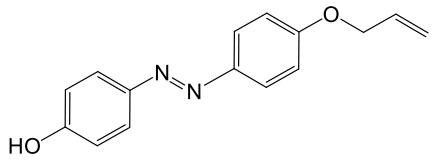
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 162.0 (Ar-$C_{4'}$), 154.2 (d, $^1J_{Ar\text{-}C3,F}$ = 247.4 Hz, Ar-C_3), 149.8 (d, $^2J_{Ar\text{-}C4,F}$ = 11.6 Hz, Ar-C_4), 148.3 (d, $^3J_{Ar\text{-}C1,F}$ = 5.1 Hz, Ar-C_1), 147.3 (Ar-$C_{1'}$), 134.2 ($OCH_2\underline{C}H{=}CH_2$), 125.8 (2C, Ar-$C_{2'/6'}$), 123.0 (d, $^3J_{Ar\text{-}C5,F}$ = 2.7 Hz, Ar-C_5), 118.37 ($CH_2{=}\underline{C}HH'$), 116.8 (2C, Ar-$C_{3'/5'}$), 115.7 (Ar-C_6), 108.24 (d, $^2J_{Ar\text{-}C2,F}$ = 19.6 Hz, Ar-C_2), 71.1 ($O\underline{C}H_2CH$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 133.56 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3320 (br), 3084 (m), 3027 (m), 2985 (m), 2602 (w), 2359 (w), 2344 (w), 1595 (s), 1503 (s), 1425 (m), 1326 (m), 1242 (s), 1211 (s), 1140 (m), 1097 (s), 1004 (s), 93 (s), 928 (s), 875 (s), 836 (s), 808 (s), 748 (m), 619 (s), 522 /s), 481 (m), 417 (m) cm^{-1};

ESI-HRMS: m/z = 273.10352 $[M + H]^+$, (calc. 273.10338 for $C_{15}H_{13}FN_2O_2 + H^+$).

4'-[4-(Allyloxy)phenyl]azophenol[683] (27)



1-Allyloxy-4-anilin **25** (2.35 g, 16.15 mmol) was dissolved in dH_2O (48 mL), acidified with conc. HCl_{aq} (1.27 mL) and cooled to 0 °C on ice. A solution of sodium nitrite (1.11 g; 16.15 mmol) in dH_2O (48 mL) and methanol (50 mL) were added and the mixture was stirred 1 h at 10 °C. Phenol (1.52 g; 16.15 mmol) was dissolved in dH_2O (248 mL) and alkalised by addition of NaOH (645 mg; 16.15 mmol) and $NaCO_3$ (1.71 g, 16.15 mmol). The phenolate solution was cooled on ice and added dropwise to the diazonium salt solution. The reaction mixture was stirred at 10 °C for 1 h and 12 h under ambient conditions. The organic solvent was removed under reduced pressure and the remaining aqueous layer was extracted with ethyl acetate (4 x 250 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. A purification by flash chromatography on silica gel (cyclohexane/ethyl acetate, 4:1). The product **27** was obtained as a yellow solid.

Yield: 2.43 g (9.56 mmol, 58 %), Lit. 86 %;[683]

TLC: R_f = 0.25 (cyclohexane/ethyl acetate, 3:1);

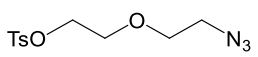
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88-7.80 (m, 4H, Ar-$H_{2/6}$, Ar-$H_{2'/6'}$), 7.01 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 8.9 Hz, 2H, Ar-$H_{3/5}$), 6.92 (d, 2H, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 8.8 Hz, 2H, Ar-$H_{3'/5'}$), 6.11-6.04 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.44 (dd, $^2J_{CH\underline{H}',C\underline{H}H'}$ = 1.3 Hz, $^3J_{C\underline{H}=CHH',CH=CH\underline{H}'}$ = 17.2 Hz, 1H, CH=CH$\underline{H}$'), 5.32 (dd, $^2J_{C\underline{H}H',C\underline{H}H'}$ = 1.3 Hz, $^3J_{CH=C\underline{H}H',OCH2C\underline{H}=CH2}$= 10.5 Hz, 1H, CH=C$\underline{H}$H'), 4.61 (d, $^3J_{OC\underline{H}2CH,OCH2C\underline{H}}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 160.6 (Ar-$C_{4'}$), 157.8 (Ar-C_4), 147.2 (Ar-C_1), 147.0 (Ar-$C_{1'}$), 132.8 ($OCH_2\underline{C}H=CH_2$), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 118.1 (CH=$\underline{C}$HH'), 115.8 (2C, Ar-$C_{3'/5'}$), 114.9 (2C, Ar-$C_{3/5}$), 69.0 ($O\underline{C}H_2CH$) ppm;

ESI-HRMS: m/z = 255.11256 $[M + H]^+$, (calc. 255.10553 for $C_{15}H_{14}N_2O_2 + H^+$).

Further analytical data corresponds to the literature.

***O*-Tosyl-[2-(2-Azidoethoxy)ethanol][454] (28)**



Commercially acquired diethylene glycol bis(*p*-toluenesulfonate) (5 g, 12.06 mmol) was dissolved in dry DMF (100 mL) under a nitrogen atmosphere and sodium azide (785 mg; 12.06 mmol) was added. The reaction mixture was stirred at room temperature for 20 h. DCM (100 mL) was added and the precipitating solid was removed by filtration over Celite® 545. The solvent was removed from the filtrate under reduced pressure. A further purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 3:1) yielded compound **28** as a colourless oil.

Yield: 2.3 g (8.08 mmol, 67 %), Lit.: 62 %;[454]

TLC: R_f = 0.27 (cyclohexane/ethyl acetate, 4:1);

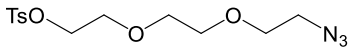
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.8 (d, 2H, $^{3}J_{\text{Ar-H,}ortho\text{, Ar-H,}meta}$ = 8.3 Hz, Ar-$H_{(Tos)\text{-S,}ortho}$), 7.35 (d, 2H, $^{3}J_{\text{Ar-H,}ortho\text{, Ar-H,}meta}$ = 8 Hz, Ar-$H_{(Tos)\text{-S,}meta}$), 4.17 (dd~t, $^{3}J_{SO\underline{H}2,SOCH2C\underline{H}2}$ = 4.7 Hz, 2H, $SO\underline{CH}_2$), 3.70 (t, $^{3}J_{SO\underline{CH}2,SOCH2C\underline{H}2}$ = 4.7 Hz, 2H, $SOCH_2C\underline{H}_2$), 3.6 (t, $^{3}J_{C\underline{H}2CH2N3,C\underline{H}2N3}$ = 5.1 Hz, 2H, $C\underline{H}_2CH_2N_3$), 3.32 (d, $^{3}J_{C\underline{H}2N3,C\underline{H}2CH2N3}$ = 5.1 Hz, 2H, $C\underline{H}_2N_3$), 2.45 (s, 3H,Ar-CH_3) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 144.94 (Ar-C_S), 132.8 (Ar-$\underline{C}_{CH3}$), 129.9 (2C, Ar-$H_{(Tos)\text{-S,}ortho}$), 127.9 (2C, Ar-$H_{(Tos)\text{-S,}meta}$), 70.1 ($O\underline{C}H_2CH_2N_3$), 69.2 ($SOCH_2$), 68.6 ($SOCH_2CH_2$), 50.6 ($\underline{C}H_2N_3$), 21.6 (Ar-CH_3) pm;

ESI-HRMS: m/z = 286.08520 $[M + H]^+$, (calc. 286.08560 for $C_{11}H_{15}N_3O_4S + H^+$).

Further analytical data corresponds to the literature.

O-Tosyl-{2-[2-(2-azidoethoxy)ethoxy]}ethanol[438,447] (29)



Commercially available triethylene glycol di(*p*-toluenesulfonate) (5 g; 10.30 mmol) was dissolved in dry DMF (100 mL) under a nitrogen atmosphere and sodium azide (669.5 mg; 10.30 mmol) was added. The reaction mixture was stirred at room temperature for 16 h. DCM (200 mL) was added and the precipitating solid was removed by filtration. The filtrate was concentrated under reduced pressure and purified by column chromatography on silica gel (cyclohexane/ethyl acetate 3:1) to obtain compound **29** as a colorless oil.

Yield: 1.75 g (5.31 mmol, 52 %), Lit.: 43 %;[447]

TLC: R_f = 0.23 (cyclohexane, ethyl acetate, 3:1), 0.19 (ethyl acetate/ethanol, 3:1);

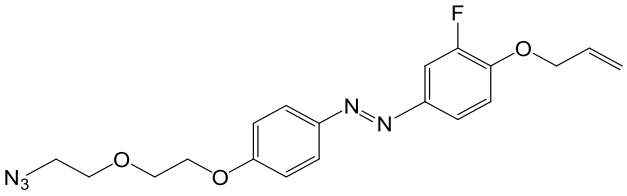
^{1}H-NMR ($CDCl_3$, 200 MHz, 298 K): δ = 7.81 (d, 2H, $^{3}J_{\text{Ar-H,}ortho\text{,Ar-H,}meta}$ = 8.3 Hz, Ar-$H_{(Tos)\text{-S-}ortho}$), 7.35 (d, 2H, $^{3}J_{\text{Ar-H,}ortho\text{,Ar-H,}meta}$ = 8.5 Hz, Ar-$H_{(Tos)\text{-S,}meta}$), 4.17 (t, $^{3}J_{SO\underline{H}2,SOCH2C\underline{H}2O}$ = 4.7 Hz, 2H, $SO\underline{CH}_2$), 3.7 (t, $^{3}J_{SOCH2,SOCH2CH2O}$ = 4.5 Hz, 2H, $SO\underline{CH}_2CH_2$), 3.65 (t, $^{3}J_{C\underline{H}2CH2N3,CH2C\underline{H}2N3}$ = 5 Hz, 2H, $C\underline{H}_2CH_2N_3$), 3.61 (s, 4H, $OC\underline{H}_2CH_2O$, $OCH_2C\underline{H}_2O$), 3.36 (t, $^{3}J_{CH2C\underline{H}2N3,C\underline{H}2CH2N3}$ = 5 Hz, 2H, $C\underline{H}_2N_3$), 2.45 (s, 3H, Ar-CH_3) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 146.4 (Ar-C_S), 134.4 (Ar-C_{CH3}), 131.0 (2C, Ar-$C_{(Tos)\text{-S,}ortho}$), 129.0 (2C, Ar-$H_{,meta}$), 71.6 ($CH_2\underline{C}H_2O$), 71.43 ($\underline{C}H_2CH_2O$), 71.1 ($SO\underline{C}H_2$), 70.8 ($C\underline{H}_2CH_2N_3$), 69.7 ($SOCH_2\underline{C}H_2$), 51.7 ($\underline{C}H_2N_3$), 21.6 (Ar-CH_3) ppm;

IR (ATR): $\tilde{\nu}$ = 2870 (m), 2101 (m), 1598 (m), 1451 (m), 1353 (s), 1291 (m), 1174 (s), 1120 (s), 1096 (s), 1013 (m), 917 (s), 815 (s), 773 (s), 662 (s) cm^{-1};

ESI-HRMS: m/z = 352.09317 $[M + Na]^+$, (calc. 352.09376 for $C_{13}H_{19}N_3O_5S + Na^+$).

[4-(Allyloxy)-3-fluorophenyl]-{4'-[2-(2-azidoethoxy)ethoxy]phenyl}diazene (30)



Compound **26** (0.43 g, 1.57 mmol), compound **28** (0.58 g, 2.04 mmol), and potassium carbonate (0.47 g, 3.4 mmol) were dissolved in dry DMF (40 mL) under a nitrogen atmosphere. The reaction mixture was refluxed at 90 °C for 2 h and stirred overnight at room temperature. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (toluene/ethyl acetate 10:1), whereas the column was wrapped in aluminium foil in order to prevent *cis*/*trans*-isomerisation during purification. Compound **30** was obtained as an orange solid.

Yield: 0.55 g (1.42 mmol, 91 %);

TLC: R_f = *trans*-isomer: 0.45 (toluene/ethyl acetate, 10:1),
cis-isomer: 0.13 (toluene/ethyl acetate, 10:1);

Melting point: 85.0 °C;

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.87 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.66 (m, 2H, Ar-H_2, Ar-H_5), 7.06 (t, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.9 Hz, 1H, Ar-H_6), 7.03 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$) 6.12-6.06 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.47 (dd, $^2J_{CH\underline{H}',C\underline{H}H'}$ = 1.5 Hz, $^3J_{C\underline{H}=CHH',CH=CH\underline{H}'}$ =17.3 Hz, CH=C$H\underline{H}$'), 5.35 (d, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.3 Hz, $^3J_{CH=C\underline{H}H'}$ = 10.5 Hz, CH=C$\underline{H}$H'), 4.69 (d, $^3J_{OC\underline{H}2CH=CH2,OCH2C\underline{H}=CH2}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH=CH_2$), 4.23 (dd∼t, $^3J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.6 Hz, 2H, Ar-$OC\underline{H}_2CH_2$), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77 (t, $^3J_{OC\underline{H}2CH2N3,C\underline{H}2N3}$ = 5 Hz, 2H, $OC\underline{H}_2CH_2N_3$), 3.43 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5 Hz, 2H, $C\underline{H}_2N_3$) ppm;

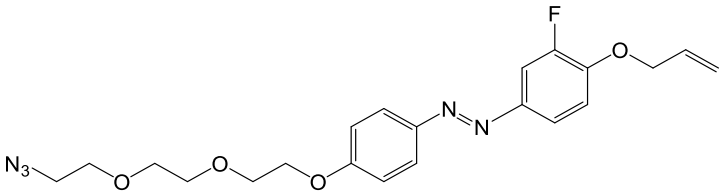
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 161 (Ar-$C_{4'}$), 152.9 (d, $^1J_{Ar\text{-}C3,F}$ = 247.3 Hz, Ar-C_3), 148.7 (d, $^2J_{Ar\text{-}C4,F}$ = 11.5 Hz, Ar-C_4), 147 (d, $^3J_{Ar\text{-}C1,F}$ = 5.3 Hz, Ar-C_1), 146.9 (Ar-$C_{1'}$), 132.4 ($OCH_2\underline{C}H=CH_2$), 124.6 (2C, Ar-$C_{2'/6'}$), 122 (d, $^3J_{Ar\text{-}H5,F}$ = 2.6 Hz, Ar-H_5), 118.5 (CH=$\underline{C}$HH'), 114.8 (2C, Ar-$C_{3'/5'}$), 114.2 (Ar-C_6), 107.9 (d, $^2J_{Ar\text{-}C2,F}$ = 19.5 Hz, Ar-C_2), 70.3 ($O\underline{C}H_2CH_2N_3$), 70.1 ($O\underline{C}H_2CH=CH_2$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 67.7 (Ar-$O\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 133.12 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3336 (br), 2919 (m), 2891 (m), 2358 (w), 2326 (w), 2080 (m), 1742 (w), 1588 (s), 1500 (s), 1424 (m), 1351 (m) 1250 (s), 1216 (s), 1060 (s), 1093 (s), 1125 (s), 981 (s), 940 (s), 882 (s), 841 (s), 818 (s), 797 (m), 619 (m), 552 (m), 521 (m), 502 (m) 465 (m) cm^{-1};

ESI-HRMS: *m*/*z* = 386.16162 $[M + H]^+$, (calc. 386.16229 for $C_{19}H_{20}FN_5O_3 + H^+$).

[4-(Allyloxy)-3-fluorophenyl]-(4'-{2-[2-(2-azidoethoxy)ethoxy]ethoxy}phenyl)diazene (31)



Compound **26** (1.01 g, 3.72 mmol), compound **29** (1.57 g, 4.75 mmol), and potassium carbonate (1.03 g, 7.4 mmol) were dissolved in dry DMF (100 mL) under a nitrogen atmosphere. The reaction mixture was refluxed at 90 °C for 2 h. The solvent was removed under reduced pressure and the residue was coevaporated with toluene (2 x 150 mL). The crude product was purified by flash chromatography on silica gel (ethyl acetate/cyclohexane, 2:1), whereas the column was wrapped in aluminium foil in order to prevent *cis*/*trans*-isomerisation during purification. Compound **31** was obtained as a yellow solid.

Yield: 1.42 g (3.3 mmol, 89 %);

TLC: R_f = *trans*-isomer: 0.45 (cyclohexane/ethyl acetate, 2:1),
cis-isomer: 0.31 (cyclohexane/ethyl acetate, 2:1);

Melting point: 81.2 °C;

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.86 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.65 (m, 2H, Ar-H_2, Ar-H_5), 7.06 (t, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.3 Hz, 1H, Ar-H_6) 7.02 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 6.13-6.05 (m, 1H, $OCH_2\underline{C}H{=}CH_2$), 5.47 (dd, $^2J_{CH\underline{H}',C\underline{H}H'}$ = 1.5 Hz, $^3J_{C\underline{H}{=}CHH',CH{=}CH\underline{H}'}$ =17.3 Hz, CH=CH$\underline{H}$'), 5.35 (d, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.3 Hz, $^3J_{CH{=}C\underline{H}H'}$ = 10.5 Hz, CH=C$\underline{H}$H'), 4.69 (d, $^3J_{OC\underline{H}2CH{=}CH2,OCH2C\underline{H}{=}CH2}$ = 5.4 Hz, 2H, $OC\underline{H}_2CH{=}CH_2$), 4.21 (dd∼t, $^3J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2O}$ = 4.9 Hz, 2H, Ar-$OC\underline{H}_2CH_2$), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.9 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77-3.74 (m, 2H, $OC\underline{H}_2CH_2O$), 3.72-3.68 (m, 4H, $OCH_2C\underline{H}_2O$, $OC\underline{H}_2CH_2N_3$), 3.39 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5 Hz, $C\underline{H}_2N_3$) ppm;

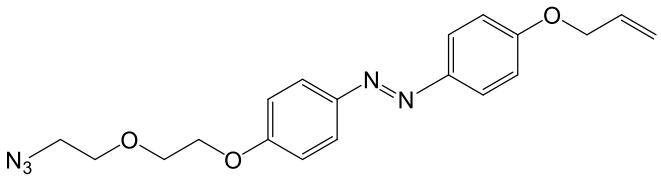
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 161 (Ar-$C_{4'}$), 153.0 (d, $^1J_{Ar\text{-}C3,F}$ = 247.3 Hz, Ar-C_3), 148.6 (d, $^2J_{Ar\text{-}C4,F}$ = 11.5 Hz, Ar-C_4), 147 (d, $^3J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 146.9 (Ar-$C_{1'}$), 132.4 ($OCH_2\underline{C}H{=}CH_2$), 124.6 (2C, Ar-$C_{2'/6'}$), 122 (d, $^3J_{Ar\text{-}H5,F}$ = 2.7 Hz, Ar-H_5), 118.5 ($O\underline{C}H_2CH{=}CH_2$), 114.9 (2C, Ar-$C_{3'/5'}$), 114.2 (Ar-$C_6$), 107.9 (d, $^2J_{Ar\text{-}C2,F}$ = 19.6 Hz, Ar-C_2), 70.9 ($O\underline{C}H_2CH_2O$), 70.8 ($OCH_2\underline{C}H_2O$), 70.1 (2C, $O\underline{C}H_2CH{=}CH_2$, $O\underline{C}H_2CH_2N_3$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 67.7 (Ar-$O\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 133.13 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 2920 (m), 2870 (m), 2326 (w), 2102 (s), 1729 (w), 1595 (s), 1501 (s), 1425 (m) 1283 (s), 1244 (s), 1139 (s), 1102 (s), 1060 (m), 989 (s), 942 (s), 883 (m), 812 (s), 798 (s), 619 (m), 530 (m), 519 (m), 468 (m) 417 (m) cm^{-1};

ESI-HRMS: *m*/*z* = 430.18792 $[M + H]^+$, (calc. 430.18851 for $C_{21}H_{24}FN_5O_4 + H^+$).

[4-(Allyloxy)phenyl]-{4'-[2-(2-azidoethoxy)ethoxy]phenyl}diazene (32)



Compound **27** (300 mg, 1.2 mmol), *O*-tosyl-[2-(2-azidoethoxy)ethanol] **28** (337 mg, 1.2 mmol) and potassium carbonate (326 mg, 2.4 mmol) were dissolved in dry DMF (3 mL) under a nitrogen atmosphere. The reaction mixture was refluxed at 90 °C for 2 h and stirred overnight at room temperature. The solvent was removed under reduced pressure, the residue was washed with dH_2O (100 mL) and the aqueous layer was extracted with ethyl acetate (3 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. A purification by flash chromatography (cyclohexane/ethyl acetate, 2:1) yielded compound **32** as an orange foam.

Yield: 0.44 g (1.2 mmol, quant.);

TLC: R_f = 0.22 (cyclohexane/ethyl acetate, 2:1);

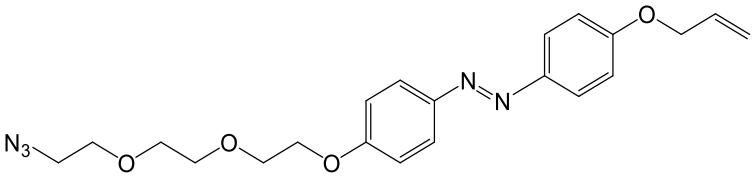
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89-7.85 (m, 4H, Ar-$H_{2/6}$, Ar-$H_{2'/6'}$), 7.04-6.99 (m, 4H, Ar-$H_{3/5}$, Ar-$H_{3'/5'}$), 6.13-6.03 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.45 (dd, $^2J_{CHH',CHH'}$ = 1.5 Hz, $^3J_{C\underline{H}=CHH',CH=CH\underline{H}'}$ =17.2 Hz, CH=CH$\underline{H}$'), 5.32 (d, $^2J_{C\underline{H}H',CH\underline{H}'}$ = 1.5 Hz, $^3J_{CH=C\underline{H}H'}$ = 10.5 Hz, CH=C$\underline{H}$H'), 4.62 (d, $^3J_{OC\underline{H}2CH=CH2,OCH2C\underline{H}=CH2}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH=CH_2$), 4.22 (dd~t, $^3J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.8 Hz, 2H, Ar-$OC\underline{H}_2CH_2$), 3.91 (dd~t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77 (dd~t, $^3J_{OC\underline{H}2CH2N3,C\underline{H}2N3}$ = 5.1 Hz, 2H, $OC\underline{H}_2CH_2N_3$), 3.43 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5.1 Hz, 2H, $C\underline{H}_2N_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 160.7 (Ar-$C_{4'}$), 160.6 (Ar-C_4), 147.2 (Ar-C_1), 147.1 (Ar-$C_{1'}$), 132.8 ($OCH_2\underline{C}H=CH_2$), 124.3 (4C, Ar-$C_{2/6}$, Ar-$C_{2'/6'}$), 118 (CH=$\underline{C}$HH'), 114.9 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 70.3 ($O\underline{C}H_2CH_2N_3$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 69.0 ($O\underline{C}H_2CH=CH_2$), 67.7 (Ar-$O\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 2926 (w), 2886 (w), 2082 (m), 1650 (w), 1593 (s), 1497 (s), 1450 (m), 1352 (m), 1248 (s), 1232 (s), 1147 (m), 1126 (s), 1111 (m), 1060 (m), 1010 (m) 938 (m), 842 (s), 768 (m), 640 (w), 580 (w), 466 (w) cm^{-1};

ESI-HRMS: *m/z* = 368.16389 $[M + H]^+$, (calc. 368.17172 for $C_{19}H_{21}N_5O_3 + H^+$).

[4-(Allyloxy)phenyl]-(4'-{2-[2-(2-azidoethoxy)ethoxy]ethoxy}phenyl)diazene (33)



Compound **27** (1.05 g, 4.13 mmol), *O*-tosyl-{2-[2-(2-azidoethoxy)ethoxy]}ethanol **29** (1.36 g, 4.13 mmol), and potassium carbonate (2.1 g, 15.2 mmol) were dissolved in dry DMF (70 mL) under a nitrogen atmosphere. The reaction mixture was refluxed at 90 °C for 2 h and stirred overnight at room temperature.

The solvent was removed under reduced pressure, the residue was washed with dH_2O (200 mL) and the aqueous layer was extracted with ethyl acetate (3 x 200 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. A purification by flash chromatography (cyclohexane/ethyl acetate, 2:1) yielded compound **33** as an orange foam.

Yield: 1.49 g (3.62 mmol, 88 %);

TLC: R_f = 0.13 (cyclohexane/ethyl acetate, 3:1);

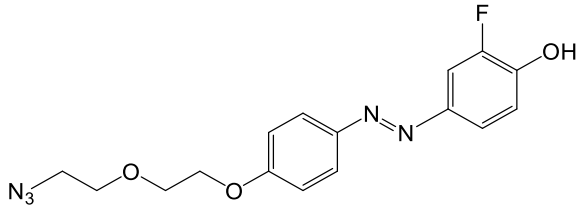
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88-7.84 (m, 4H, Ar-$H_{2/6}$, Ar-$H_{2'/6'}$), 7.03-6.99 (m, 4H, Ar-$H_{3/5}$, Ar-$H_{3'/5'}$), 6.13-6.03 (m, 1H, $OCH_2C\underline{H}=CH_2$), 5.44 (dd, $^2J_{C\underline{H}H',C\underline{H}H'}$ = 1.5 Hz, $^3J_{C\underline{H}=CHH',CH=C\underline{H}H'}$ =17.3 Hz, CH=C$\underline{H}$H'), 5.32 (d, $^2J_{C\underline{H}H',C\underline{H}H'}$ = 1.5 Hz, $^3J_{CH=C\underline{H}H'}$ = 10.5 Hz, CH=C$\underline{H}$H'), 4.61 (d, $^3J_{OC\underline{H}2CH=CH2,OCH2C\underline{H}=CH2}$ = 5.3 Hz, 2H, $OC\underline{H}_2CH=CH_2$), 4.21 (dd∼t, $^3J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2O}$ = 4.9 Hz, 2H, Ar-$OC\underline{H}_2CH_2$), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.9 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77-3.74 (m, 2H, $OC\underline{H}_2CH_2O$), 3.72-3.67 (m, 4H, $OCH_2C\underline{H}_2O$, $OC\underline{H}_2CH_2N_3$), 3.39 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 4.8 Hz, 2H, $C\underline{H}_2N_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 160.8 (Ar-$C_{4'}$), 160.6 (Ar-C_4), 147.2 (Ar-C_1), 147.1 (Ar-$C_{1'}$), 132.8 ($OCH_2\underline{C}H=CH_2$), 124.3 (4C, Ar-$C_{2/6}$, Ar-$C_{2'/6'}$), 118 (CH=$\underline{C}$HH'), 114.9 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 70.9 ($O\underline{C}H_2CH_2O$), 70.8 ($OCH_2\underline{C}H_2O$), 70.1 ($O\underline{C}H_2CH_2N_3$), 69.8 (Ar-$OCH_2\underline{C}H_2O$), 69.0 ($O\underline{C}H_2CH=CH_2$), 67.7 (Ar-$O\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 2926 (w), 2873 (w), 2102 (m), 1650 (w), 1593 (s), 1498 (s), 1452 (m), 1357 (w), 1298 (m), 1246 (s), 1233 (s), 1147 (s), 1105 (s), 1061 (m), 1011 (m), 933 (s), 912 (m), 651 (w), 555 (w), 454 (w) cm^{-1};

ESI-HRMS: m/z = 412.19748 $[M + H]^+$, (calc. 412.19793 for $C_{21}H_{25}N_5O_4 + H^+$).

4-(3-{4'-[2-(2-Azidoethoxy)ethoxy]phenyl}fluoro)azophenol (34)



Compound **30** (451 mg, 1.4 mmol) and tetrakis(triphenylphosphine)palladium(0) (16.1 mg, 10.4 µmol) were dissolved in dry methanol (25 mL) under a nitrogen atmosphere. Potassium carbonate (582 mg, 4.2 mmol) was added and the reaction mixture was stirred 5 h at room temperature. The solvent was removed under reduced pressure, the residue was dissolved in 1 N HCl_{aq} (100 mL) and the aqueous layer was extracted with ethyl acetate (3 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 4:1 → 3:1) gave compound **34** as an orange solid.

Yield: 456 mg (1.32 mmol, 94 %);

TLC: R_f = 0.29 (cyclohexane/ethyl acetate, 4:1);

Melting point: 110.0 °C;

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.86 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.64 (m, 2H, Ar-H_2, Ar-H_5), 7.12 (t, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.4 Hz, 1H, Ar-H_6), 7.02 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.55 (d, $^{4}J_{Ar\text{-}OH,F}$ = 3.7 Hz, 1H, Ar-OH), 4.23 (dd~t, $^{3}J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.8 Hz, 2H, Ar-O$C\underline{H}_2CH_2$), 3.91 (dd~t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77 (t, $^{3}J_{OC\underline{H}2CH2N3,C\underline{H}2N3}$ = 5.1 Hz, 2H, O$C\underline{H}_2CH_2N_3$), 3.44 (t, $^{3}J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5.1 Hz, 2H, $C\underline{H}_2N_3$) ppm;

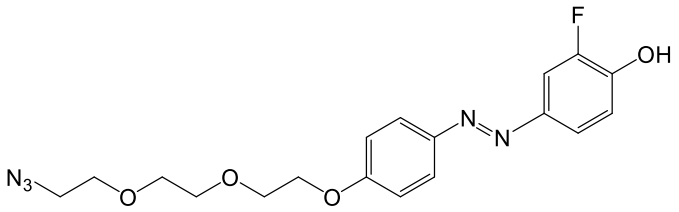
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 161.0 (Ar-$C_{4'}$), 152.1 (d, $^{1}J_{Ar\text{-}C3,F}$ = 241 Hz, Ar-C_3), 146.9 (d, $^{3}J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 146.87 (Ar-$C_{1'}$), 145.7 (d, $^{2}J_{Ar\text{-}C4,F}$ = 15.4 Hz, Ar-C_4), 124.6 (2C, Ar-$C_{2'/6'}$), 122.3 (d, $^{3}J_{Ar\text{-}H5,F}$ = 2.4 Hz, Ar-H_5), 116.9 (Ar-C_6), 114.9 (2C, Ar-$C_{3'/5'}$), 107.6 (d, $^{2}J_{Ar\text{-}C2,F}$ = 19.3 Hz, Ar-C_2), 70.3 (O$\underline{C}H_2CH_2N_3$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 67.7 (Ar-O$\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 139.9 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3254 (br), 2926 (m), 2873 (m), 2111 (s), 1600 (s), 1581 (s), 1497 (s), 1450 (m), 1371 (m), 1305 (m), 1242 (s), 1128 (s), 1111(s), 1062 (s), 926 (m), 847 (m), 788 (m), 745 (m), 618 (m), 566 (m), 509 (m), 468 (w) 413 (w) cm^{-1};

ESI-HRMS: *m/z* = 346.13049 $[M + H]^+$, (calc. 346.13099 for $C_{16}H_{16}FN_5O_3 + H^+$).

4-[3-(4'-{2-[2-(2-Azidoethoxy)ethoxy]ethoxy}phenyl)fluoro]azophenol (35)



Compound **31** (1.4 g, 3.28 mmol) and tetrakis(triphenylphosphine)palladium(0) (419 mg, 362.4 µmol) were dissolved in dry methanol (60 mL) under a nitrogen atmosphere. Potassium carbonate (1.36 g, 9.84 mmol) was added and the reaction mixture was stirred 5 h at room temperature. The solvent was removed under reduced pressure, the residue was dissolved in 1 N HCl_{aq} (200 mL) and the aqueous layer was extracted with ethyl acetate (3 x 200 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 3:1) gave compound **35** as an orange-brown solid.

Yield: 1.12 g (2.87 mmol, 88 %);

TLC: R_f = 0.19 (cyclohexane/ethyl acetate, 3:1);

Melting point: 77.2 °C;

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.84 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',\ Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.67-7.62 (m, 2H, Ar-H_2, Ar-H_5), 7.10 (t, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.6 Hz, 1H, Ar-H_6), 6.99 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',\ Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.72 (d, $^{4}J_{Ar\text{-}OH,F}$ = 3.9 Hz, 1H, Ar-OH), 4.21 (dd~t, $^{3}J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2O}$ = 4.8 Hz, 2H, Ar-O$C\underline{H}_2CH_2$), 3.91 (dd~t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.78-3.75 (m, 2H, O$C\underline{H}_2CH_2O$), 3.73-3.68 (m, 4H, $OCH_2C\underline{H}_2O$, O$C\underline{H}_2CH_2N_3$), 3.96 (t, $^{3}J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 4.9 Hz, 2H, $C\underline{H}_2N_3$) ppm;

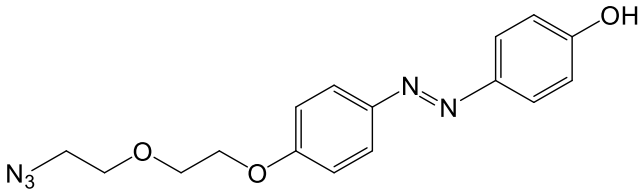
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 161.3 (Ar-$C_{4'}$), 151.4 (d, $^{1}J_{Ar\text{-}C3,F}$ = 238 Hz, Ar-C_3), 146.9 (d, $^{3}J_{Ar\text{-}C1,F}$ = 4.9 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 145.7 (d, $^{2}J_{Ar\text{-}C4,F}$ = 15.2 Hz, Ar-C_4), 124.5 (2C, Ar-$C_{2'/6'}$), 122.2 (d, $^{3}J_{Ar\text{-}H5,F}$ = 2.8 Hz, Ar-H_5), 116.9 (d, $^{4}J_{Ar\text{-}C6,F}$ = 2.3 Hz, Ar-C_6), 114.8 (2C, Ar-$C_{3'/5'}$), 107.7 (d, $^{2}J_{Ar\text{-}C2,F}$ = 19.3 Hz, Ar-C_2), 70.9 (O$\underline{C}H_2CH_2O$), 70.8 ($OCH_2\underline{C}H_2O$), 70.1 (O$\underline{C}H_2CH_2N_3$), 69.8 (Ar-$OCH_2\underline{C}H_2O$), 67.7 (Ar-O$\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 139.67 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3251 (br), 2885 (w), 2385 (w), 2349 (w), 2108 (m), 1599 (s), 1502 (s), 1296 (m), 1231 (s), 1133 (m), 939 (w), 879 (m), 618 (w), 451 (m), 412 (w) cm^{-1};

ESI-HRMS: *m*/*z* = 390.15663 $[M + H]^+$, (calc. 390.15721 for $C_{18}H_{20}FN_5O_4 + H^+$).

4-{4'-[2-(2-Azidoethoxy)ethoxy]phenyl}azophenol (36)



Compound **32** (513 mg, 1.4 mmol) and tetrakis(triphenylphosphine)palladium(0) (16.1 mg, 10.4 µmol) were dissolved in dry methanol (27 mL) under a nitrogen atmosphere. Potassium carbonate (578 mg, 4.18 mmol) was added and the reaction mixture was stirred 20 h at room temperature. The reaction mixture was filtered, and the solvent was removed from the filtrate under reduced pressure. The residue was dissolved in 1 N HCl_{aq} (350 mL) and the aqueous layer was extracted with ethyl acetate (5 x 300 mL). The combined organic layers were washed with Brine (200 mL), dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 2:1) gave compound **36** as an orange foam.

Yield: 246 mg (0.75 mmol, 54 %);

TLC: R_f = 0.33 (cyclohexane/ethyl acetate, 3:1);

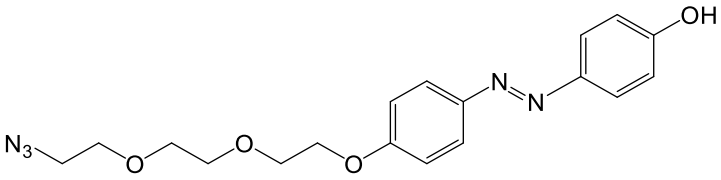
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89-7.80 (m, 4H, Ar-$H_{2/6}$, Ar-$H_{2'/6'}$), 7.04-6.99 (m, 4H, Ar-$H_{3/5}$, Ar-$H_{3'/5'}$), 5.37 (s, 1H, Ar-OH), 4.22 (dd~t, $^{3}J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.8 Hz, 2H, Ar-O$C\underline{H}_2CH_2$), 3.91 (dd~t, $^{3}J_{Ar\text{-}CH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.77 (dd~t, $^{3}J_{OC\underline{H}2CH2N3,C\underline{H}2N3}$ = 5.3 Hz, 2H, O$C\underline{H}_2CH_2N_3$), 3.44 (t, $^{3}J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5.3 Hz, 2H, $C\underline{H}_2N_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 160.7 (Ar-$C_{4'}$), 157.8 (Ar-C_4), 147.2 (Ar-C_1), 147.15 (Ar-$C_{1'}$), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 114.9 (2C, Ar-$C_{3/5}$), 114.8 (2C, Ar-$C_{3'/5'}$), 70.3 (O$\underline{C}H_2CH_2N_3$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 67.7 (Ar-O$\underline{C}H_2CH_2$), 50.7 ($\underline{C}H_2N_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3098 (br), 2024 (m), 2869 (m), 2573 (w), 2249 (w), 2097 (s), 1599 (s), 1583 (s), 1500 (s)m 1436 (s), 1325 (w), 1273 (m), 1249 (s), 1226 (s), 1133 (s), 1115 (s), 1058 (s), 929 (m), 839 (s), 766 (m), 634 (m), 552 (s), 533 (m), 511 (m), 470 (w), 417 (m) cm^{-1};

ESI-HRMS: *m*/*z* = 328.13997 $[M + H]^+$, (calc. 328.14042 for $C_{16}H_{17}N_5O_3 + H^+$).

4-(4'-{2-[2-(2-Azidoethoxy)ethoxy]ethoxy}phenyl)azophenol (37)



Compound **33** (1.49 g, 3.6 mmol) and tetrakis(triphenylphosphine)palladium(0) (1641 mg, 36 µmol) were dissolved in dry methanol (70 mL) under a nitrogen atmosphere. Potassium carbonate (1.5 g, 10.87 mmol) was added and the reaction mixture was stirred 48 h at room temperature. The solvent was removed under reduced pressure and the residue was dissolved in 1 N HCl_{aq} (150 mL). The aqueous layer was extracted with ethyl acetate (6 x 200 mL). The combined organic layers were neutralised with sat. $NaHCO_3$ solution (150 mL), dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The crude product was purified by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 2:1) and compound **37** was obtained as an orange oil.

Yield: 766 mg (2.06 mmol, 57 %);

TLC: R_f = 0.16 (cyclohexane/ethyl acetate, 2:1);

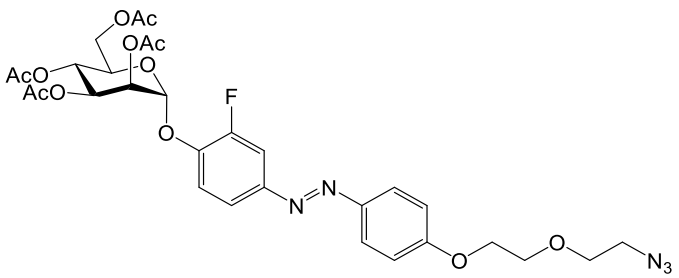
^{1}H-NMR ($CDCl_3$, 600 MHz, 298 K): δ = 7.88-7.85 (m, 4H, Ar-$H_{2/6}$, Ar-$H_{2'/6'}$), 6.97-6.88 (m, 4H, Ar-$H_{3/5}$, Ar-$H_{3'/5'}$), 6.00 (s, 1H, Ar-OH), 4.17 (dd~t, $^3J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2O}$ = 4.8 Hz, 2H, Ar-OC$\underline{H}_2$$CH_2$), 3.91 (dd~t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.8 Hz, 2H, Ar-OCH_2C$\underline{H}_2$O), 3.79-3.76 (m, 2H, OC$\underline{H}_2$$CH_2$O), 3.74-3.67 (m, 2H, OCH_2C$\underline{H}_2$O), 3.69 (dd~t, $^3J_{OC\underline{H}2CH2N3,C\underline{H}2N3}$ = 5.2 Hz, 2H, OC$\underline{H}_2$$CH_2N_3$), 3.39 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 5.2 Hz, 2H, C$\underline{H}_2$$N_3$) ppm;

^{13}C-NMR ($CDCl_3$, 151 MHz, 298 K): δ = 160.7 (Ar-$C_{4'}$), 158.0 (Ar-C_4), 147.01 (Ar-C_1), 146.98 (Ar-$C_{1'}$), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 115.8 (2C, Ar-$C_{3'/5'}$), 114.7 (2C, Ar-$C_{3/5}$), 70.9 (O$\underline{C}$$H_2CH_2O$), 70.7 ($OCH_2\underline{C}H_2O$), 70.1 (O$\underline{C}$$H_2CH_2N_3$), 69.8 (Ar-$OCH_2\underline{C}H_2O$), 67.6 (Ar-O$\underline{C}$$H_2CH_2$), 50.6 ($\underline{C}H_2N_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3232 (br), 2925 (m), 2873 (m), 2360 (w), 2265 (w), 2097 (m), 1745 (m), 1585 (s), 1584 (s), 1499 (s), 1459 (m), 1370 (m), 1244 (s), 1219 (s), 1142 (s), 1104 (s), 1057 (s), 932 (m), 841 (s), 764 (m), 644 (m), 549 (s), 529 (m), 504 (m), 645 (w), 447 (w) 408 (w) cm^{-1};

ESI-HRMS: m/z = 372.16624 $[M + H]^+$, (calc. 372.16663 for $C_{18}H_{21}N_5O_4 + H^+$).

[4-(3-{4'-[2-(2-Azidoethoxy)ethoxy]phenyl}fluoro)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (38)



The mannosyl donor, *O*-(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl)trichloroacetimidate **5** (602 mg, 1.55 mmol), compound **34**, (446 mg; 1.29 mmol), and a spatula tip of activated molecular sieves (3 Å) were dissolved in dry DCM (10 mL) and dry acetonitrile (7 mL) under a nitrogen atmosphere.

After cooling to 0 °C on ice, boron trifluoride diethyletherate (327 µL; 2.58 mmol) was added dropwise and the reaction mixture was stirred for 16 h at room temperature. Ethyl acetate (100 mL) was added and the organic layer was washed with sat. $NaHCO_3$ solution (4 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by flash chromatography (cyclohexane/ethyl acetate, 2:1 → 1:1) to obtain compound **38** as an orange foam.

Yield: 545 mg (0.81 mmol, 63 %);

TLC: R_f = 0.25 (cyclohexane/ethyl acetate, 2:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 64.8 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.66 (m, 2H, Ar-$H_{3/5}$), 7.32 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.5 Hz, 1H, Ar-H_5), 7.04 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.6 Hz, 1H, H-6), 4.25-4.20 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.13-4.09 (m, 3H, H-6', Ar-$OCH_2C\underline{H}_2O$), 3.77 (t, $^3J_{O\underline{CH}2OCH2N3,C\underline{H}2N3}$ = 5.1 Hz, 2H, $OC\underline{H}_2CH_2N_3$), 3.44 (dd∼t, $^3J_{C\underline{H}2N3,O\underline{CH}2OCH2N3}$ = 5.1 Hz, 2H, $C\underline{H}_2N_3$), 2.22 (s, 3H, $COC\underline{H}_3$), 2.08 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$) ppm;

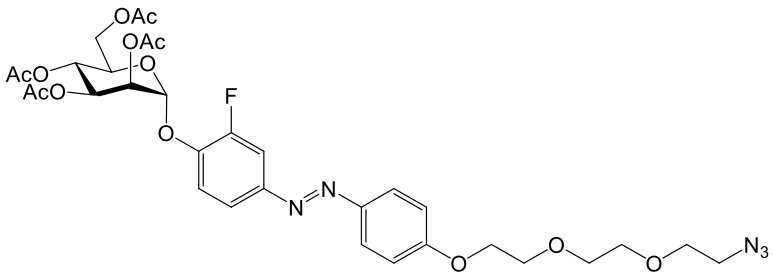
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.8 (2C, 2 x $\underline{C}OCH_3$), 161.4 (Ar-$C_{4'}$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 249 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.4 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.8 (2C, Ar-$C_{2'/6'}$), 121.3 (Ar-C_6), 118.2 (Ar-C_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 97.2 (C-1), 70.3 ($O\underline{C}H_2CH_2N_3$), 69.8 (C-5), 69.6 (Ar-$OCH_2\underline{C}H_2O$), 69.2 (C-2), 68.7 (C-3), 67.6 (Ar-$O\underline{C}H_2$), 65.8 (C-4), 62.1 (C-6), 50.7 ($\underline{C}H_2N_3$), 20.8 ($CO\underline{C}H_3$), 20.69 ($CO\underline{C}H_3$), 20.66 (2C, 2 x $CO\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.49 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3368 (br), 2926 (w), 2370 (w), 2342 (w), 2106 (m), 1748 (s), 1599 (m), 1500 (s), 1368 (m), 1244 (s), 1220 (s), 1131 (s), 1055 (m), 1016 (m), 977 (m), 838 (m), 672 (w), 463 (w) cm^{-1};

ESI-HRMS: m/z = 676.22662 $[M + H]^+$, (calc. 676.22608 for $C_{30}H_{34}FN_5O_{12} + H^+$).

{4-[3-(4'-{2-[2-(2-Azidoethoxy)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (39)



O-(2,3,4,6-Tetra-*O*-acetyl-*α*-D-mannopyranosyl)trichloroacetimidate **5** (1.52 g, 3.09 mmol), compound **35** (1.08 g; 2.77 mmol) and a spatula tip of activated molecular sieves (3 Å) were dissolved in dry DCM (20 mL) and dry acetonitrile (15 mL) under a nitrogen atmosphere. After cooling to 0 °C, boron trifluoride diethyletherate (703 µL; 5.55 mmol) was added dropwise and the reaction mixture was stirred for 16 h at room temperature.

Ethyl acetate (150 mL) was added and the organic layer was washed with sat. $NaHCO_3$ solution (3 x 150 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by flash chromatography on silica gel (cyclohexane/ethyl acetate, 2:1 → 1:1) to obtain the title compound **39** as an orange foam.

Yield: 1.86 g (2.59 mmol, 84 %);

TLC: R_f = 0.18 (cyclohexane/ethyl acetate, 2:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 63.6 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.66 (m, 2H, Ar-$H_{3/5}$), 7.32 (t, $^3J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.5 Hz, 1H, Ar-H_5), 7.03 (d, $^3J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 5.6 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{4,5}$ = 10 Hz, $^3J_{4,3}$ = 10 Hz, 1H, H-4), 4.29 (dd, $^2J_{6,6'}$ = 12.2 Hz, $^3J_{6,5}$ = 5.5 Hz, 1H, H-6), 4.24-4.19 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.11 (dd, $^2J_{6',6}$ = 12.2 Hz, $^3J_{6',5}$ = 2.2 Hz, 1H, H-6'), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OCH2}$ = 4.7 Hz, 2H, Ar-OCH$_2$C$\underline{H}_2$O), 3.78-3.75 (m, 2H, OC$\underline{H}_2$CH$_2$O), 3.72-3.68 (m, 4H, OCH$_2$C$\underline{H}_2$O, OC$\underline{H}_2$CH$_2$N$_3$), 3.39 (dd∼t, $^3J_{C\underline{H}2N3,OC\underline{H}2OCH2N3}$ = 5 Hz, 2H, C$\underline{H}_2$N$_3$), 2.21 (s, 3H, COC$\underline{H}_3$), 2.08 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.04 (s, 3H, COC$\underline{H}_3$) ppm;

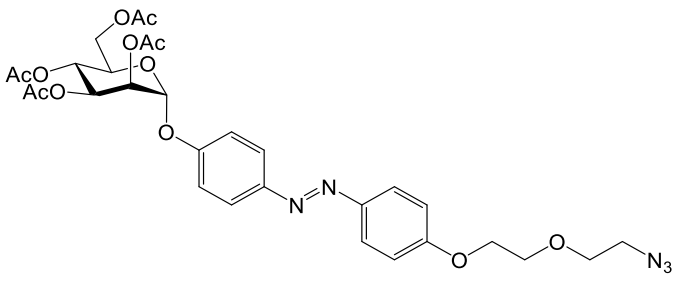
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (C$\underline{O}$CH$_3$), 169.9 (C$\underline{O}$CH$_3$), 169.8 (2C, 2 x C$\underline{O}$CH$_3$), 161.5 (Ar-$C_{4'}$), 153.5 (d, $^1J_{Ar\text{-}C3,F}$ = 251 Hz, Ar-C_3), 149.0 (d, $^3J_{Ar\text{-}C1,F}$ = 5.41Hz, Ar-C_1), 146.7 (Ar-$C_{1'}$), 144.9 (d, $^2J_{Ar\text{-}C4,F}$ = 12 Hz, Ar-C_4), 124.8 (2C, Ar-$C_{2'/6'}$), 121.3 (d, $^4J_{Ar\text{-}C6,F}$ = 2.8 Hz, Ar-C_6), 118.2 (Ar-C_5), 114.9 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^2J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 97.2 (C-1), 70.9 (O$\underline{C}$H$_2$CH$_2$O), 70.8 (OCH$_2\underline{C}$H$_2$O), 70.1 (O$\underline{C}$H$_2$CH$_2$N$_3$), 69.8 (C-5), 69.7 (Ar-OCH$_2\underline{C}$H$_2$O), 69.2 (C-2), 68.7 (C-3), 67.7 (Ar-O$\underline{C}$H$_2$), 65.8 (C-4), 62.1 (C-6), 50.7 ($\underline{C}$H$_2$N$_3$), 20.8 (CO$\underline{C}$H$_3$), 20.69 (CO$\underline{C}$H$_3$), 20.66 (2C, 2 x CO$\underline{C}$H$_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.50 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 2926 (br), 2383 (w), 2358 (w), 2326 (w), 2301 (w), 2102 (m), 1749 (s), 1599 (m), 1500 (s), 1455 (w), 1369 (m), 1244 (s), 1223 (s), 1133 (s), 1055 (s), 1032 (m), 978 (m), 841 (w), 600 (w), 625 (w), 406 (w) cm^{-1};

ESI-HRMS: *m/z* = 720.25115 [M + H]$^+$, (calc. 720.25229 for $C_{32}H_{38}FN_5O_{13}$ + H$^+$).

(4-{4'-[2-(2-Azidoethoxy)ethoxy]phenyl}azophenoxy)-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (40)



O-(2,3,4,6-Tetra-*O*-acetyl-*α*-D-mannopyranosyl)trichloroacetimidate **5** (1.29 g, 2.36 mmol), compound **36** (774 mg; 2.37 mmol) and a spatula tip of activated molecular sieves (3 Å) were dissolved in dry DCM (17 mL) and dry acetonitrile (17 mL) under a nitrogen atmosphere. After cooling to 0 °C on ice, boron trifluoride diethyletherate (600 µL; 4.73 mmol) was added dropwise.

The reaction mixture was stirred for 18 h at room temperature. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (cyclohexane/ethyl acetate, 1:1). Compound **40** was obtained as an orange foam.

Yield: 1.35 g (2.05 mmol, 78 %);

TLC: R_f = 0.15 (cyclohexane/ethyl acetate, 2:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 84.4 (c = 0.02 mmol · mL^{-1}, methanol);

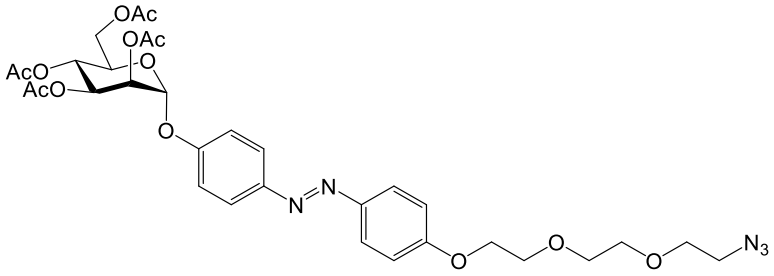
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.21 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.03 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.61 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.58 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.6 Hz, 1H, H-3), 5.48 (dd, $^3J_{2,3}$ = 3.6 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{4,3}$ = 10 Hz, $^3J_{4,5}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^2J_{6,6'}$ = 12.6 Hz, $^3J_{6,5}$ = 5.8 Hz, 1H, H-6), 4.23 (dd∼t, $^3J_{Ar\text{-}OCH2,OCH2CH2O}$ = 4.7 Hz, 2H, Ar-OCH_2CH_2), 4.13-4.06 (m, 2H, H-5, H-6'), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2CH2O,Ar\text{-}OCH2}$ = 4.7 Hz, 2H, Ar-OCH_2CH_2O), 3.77 (dd∼t, $^3J_{OCH2CH2N3,CH2N3}$ = 5.2 Hz, 2H, O$CH_2CH_2N_3$), 3.43 (t, $^3J_{CH2N3,OCH2CH2N3}$ = 5.2 Hz, 2H, CH_2N_3) 2.22 (s, 3H, COCH_3), 2.06 (s, 3H, COCH_3), 2.05 (s, 3H, COCH_3), 2.03 (s, 3H, COCH_3) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (COCH_3), 169.95 (COCH_3), 169.92 (COCH_3), 169.7 (COCH_3), 161 (Ar-$C_{4'}$), 157.2 (Ar-C_4), 148.5 (Ar-C_1), 147.1 (Ar-$C_{1'}$), 124.6 (2C, Ar-$C_{2/6}$), 124.2 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.6 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70.3 (O$CH_2CH_2N_3$), 69.7 (Ar-OCH_2CH_2O), 69.4 (C-5), 69.3 (C-2), 68.8 (C-3), 67.7 (Ar-OCH_2CH_2), 65.9 (C-4), 62.1 (C-6), 50.7 (CH_2N_3), 20.8 (COCH_3), 20.86 (COCH_3), 20.68 (COCH_3), 20.68 (2C, 2 x COCH_3) ppm;

IR (ATR): $\tilde{\nu}$ = 2930 (w), 2370 (w), 2326 (w), 2252 (w), 2107 (w), 2040 (w), 1750 (s), 1598 (m), 1583 (w), 1499 (m), 1369 (m), 1223 (s), 1129 (m), 1033 (m), 982 (w), 846 (w), 750 (w), 552 (w), 466 (w), 405 (w) cm^{-1};

ESI-HRMS: m/z = 658.23470 $[M + H]^+$, (calc. 658.23550 for $C_{30}H_{35}N_5O_{12}$ + H^+).

[4-(4'-{2-[2-(2-Azidoethoxy)ethoxy]ethoxy}phenyl)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (41)



The mannosyl donor *O*-(2,3,4,6-Tetra-*O*-acetyl-*α*-D-mannopyranosyl)trichloroacetimidate **5** (1.29 g, 2.36 mmol), azobenzene derivative **37** (752 mg; 2.02 mmol) and a spatula tip of activated molecular sieves (3 Å) were dissolved in dry DCM (17 mL) and dry acetonitrile (17 mL) under a nitrogen atmosphere. After cooling to 0 °C, boron trifluoride diethyletherate (513 µL; 4.05 mmol) was added dropwise and the reaction mixture was stirred for 18 h at room temperature. The solvent was removed under reduced pressure and the residue was roughly purified by flash chromatography on silica gel (cyclohexane/ethyl acetate 1:1 → 1:2).

Since a large part of the obtained fractions still contained the glycosyl donor, the crude product was dissolved in a mixture of pyridine/acetic anhydride 4:1 (20 mL) and stirred 16 h at room temperature. The solvent was removed under reduced pressure and the residue was subjected to a second purification by column chromatography (cyclohexane/ethyl acetate, 1:1). Compound **41** was obtained as an orange oil.

Yield: 1.20 g (1.71 mmol, 65 %);

TLC: R_f = *trans*-isomer: 0.42 (cyclohexane/ethyl acetate, 1:1),
cis-isomer: 0.25 (cyclohexane/ethyl acetate, 1:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 79.6 (c = 0.007 mmol · mL^{-1}, methanol);

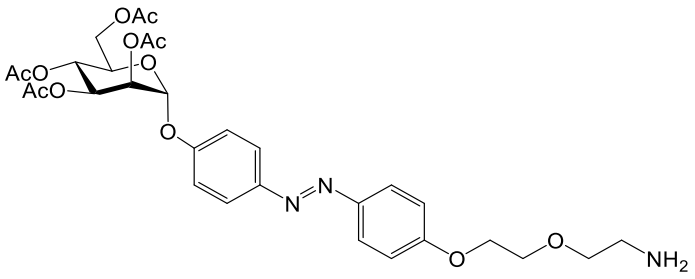
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.21 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.03 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.61 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.58 (dd, $^3J_{3,4}$ = 10.1 Hz, $^3J_{3,2}$ = 3.6 Hz, 1H, H-3), 5.48 (dd, $^3J_{2,3}$ = 3.6 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{4,3}$ = 10.1 Hz, $^3J_{4,5}$ = 10.1 Hz, 1H, H-4), 4.28 (dd, $^2J_{6,6'}$ = 12.5 Hz, $^3J_{6,5}$ = 5.8 Hz, 1H, H-6),4.22 (dd∼t, $^3J_{Ar\text{-}OCH2,OCH2CH2O}$ = 4.9 Hz, 2H, Ar-OC$\underline{H}_2$,OCH2C$\underline{H}_2$O = 4.9 Hz, 2H, Ar-OC$\underline{H}_2CH_2$), 4.13-4.07 (m, 2H, H-5, H-6'), 3.91 (dd∼t, $^3J_{Ar\text{-}OCH2CH2O,Ar\text{-}OCH2}$ = 4.9 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.78-3.74 (m, 2H, $OC\underline{H}_2CH_2O$), 3.72-3.68 (m, 4H, $OCH_2C\underline{H}_2O$, $OC\underline{H}_2CH_2N_3$), 3.39 (t, $^3J_{C\underline{H}2N3,OC\underline{H}2CH2N3}$ = 4.9 Hz, 2H, $C\underline{H}_2N_3$) 2.22 (s, 3H, $COC\underline{H}_3$), 2.06 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.03 (s, 3H, $COC\underline{H}_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.95 ($\underline{C}OCH_3$), 169.92 ($\underline{C}OCH_3$), 169.7 ($\underline{C}OCH_3$), 161.1 (Ar-$C_{4'}$), 157.1 (Ar-C_4), 148.5 (Ar-C_1), 147.0 (Ar-$C_{1'}$), 124.5 (2C, Ar-$C_{2/6}$), 124.2 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70.9 ($O\underline{C}H_2CH_2O$), 70.8 ($OCH_2\underline{C}H_2O$), 70.1 ($O\underline{C}H_2CH_2N_3$), 69.7 (Ar-$OCH_2\underline{C}H_2O$), 69.4 (C-5), 69.3 (C-2), 68.8 (C-3), 67.7 (Ar-$O\underline{C}H_2CH_2$), 65.9 (C-4), 62.1 (C-6), 50.7 ($\underline{C}H_2N_3$), 20.9 ($CO\underline{C}H_3$), 20.86 ($CO\underline{C}H_3$), 20.68 (2C, 2 x $CO\underline{C}H_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 2936 (w), 2342 (w), 2308 (w), 2105 (m), 1747 (s), 1598 (m), 1583 (m), 1498 (m), 1454 (w), 1368 (m), 1216 (s), 1128 (s), 1032 (s), 981 (m), 9201 (w), 690 (w), 638 (w), 552 (w), 434 (w) cm^{-1};

ESI-HRMS: *m*/*z* = 702.26116 [M + H]$^+$, (calc. 702.26171 for $C_{32}H_{39}N_5O_{13}$ + H$^+$).

(4-{4'-[2-(2-Aminoethoxy)ethoxy]phenyl}azophenoxy)-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside[438] (42)



Compound **40** (180 mg, 0.28 mmol) and triphenylphosphane (144 mg, 0.55 mmol) were dissolved in THF (30 mL) and dH_2O (2 mL) under a nitrogen atmosphere. The reaction was stirred 20 h at room temperature. The solvent was removed under reduced pressure and the residue was roughly purified by column chromatography on silica gel (ethyl acetate/methanol, 10:1, under the addition of 1 % v/v diethyl amine). Crude compound **42** was obtained as an orange oil.

Yield: 179 mg (crude product);

TLC: R_f = 0 (cyclohexane/ethyl acetate, 1:1);

Rotational value: $[\alpha]_{20}^{D}$ = -+ 129.5 (c = 0.06, methanol);

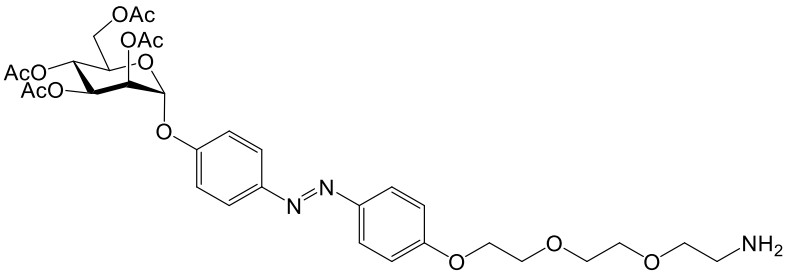
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 8.9 Hz, 2H, Ar-$H_{2/6}$), 7.86 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.20 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 6.99 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 5.6 (d, $^3J_1$, 2 = 1.7 Hz, 1H, H-1), 5.58 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{3,4}$ = 10 Hz, 1H, H-3), 5.48 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{3,4}$ = $^3J_{4,5}$ = 10. 1 Hz, 1H, H-4) 5.32 (br. s, 2H, N$\underline{H}_2$), 4.29 (dd, $^3J_{5,6}$ = 5.5 Hz, $^2J_{6,6'}$ = 12.2 Hz, 1H, H-6), 4.19 (dd∼t, $^3J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2}$ = 4.3 Hz, 2H, Ar-OC$\underline{H}_2$), 4.13-4.06 (m, 2H, H-5 ,H-6'), 3.89 (dd∼t, $^3J_{Ar\text{-}OCH2C\underline{H}2,Ar\text{-}OC\underline{H}2}$ = 4.2 Hz, 2H, Ar-$OCH_2C\underline{H}_2$), 3.8 (dd∼t, $^3J_{C\underline{H}2CH2NH2,C\underline{H}2NH2}$ = 4.7 Hz, 2H, $C\underline{H}_2CH_2NH_2$), 3.27 (m, 2H, $C\underline{H}_2NH_2$) 2.22 (s, 3H, $COC\underline{H}_3$), 2.06 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.03 (s, 3H, $COC\underline{H}_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.8 ($\underline{C}OCH_3$), 170.09 ($\underline{C}OCH_3$), 170.08 ($\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 161.5 (Ar-C_4), 157.3 (Ar-$C_{4'}$), 148.4 (Ar-$C_{1'}$), 147.3 (Ar-C_1), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2/6'}$), 116.8 (2C, Ar-$C_{3/5}$), 114.8 (2C, Ar-$C_{3'/5'}$), 95.7 (C-1), 69.7 (Ar-$OCH_2\underline{C}H_2$), 69.4 (C-5), 69.3 (C-2), 68.9 (C-3), 67.46 (Ar-O$\underline{C}H_2$), 67.38 ($\underline{C}H_2CH_2NH_2$), 65.92 (C-4), 62.16 (C-6), 40.08 ($\underline{C}H_2NH_2$), 20.9 ($CO\underline{C}H_3$), 20.8 ($CO\underline{C}H_3$), 20.7 (2C, 2 x $CO\underline{C}H_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 2934 (br), 2326 (w), 2050 (w), 1744 (m), 1677 (m), 1597 (m), 1497 (m), 1369 (m), 1214 (s), 1125 (s), 1031 (s), 979 (s), 839 (s), 798 (m), 721 (m) cm^{-1};

EI-MS: m/z = 631.25 $[M + H]^+$, (calc. 631.24 for $C_{30}H_{37}N_3O_{12} + H^+$).

[4-(4'-{2-[2-(2-Aminoethoxy)ethoxy]ethoxy}phenyl)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside[438] (43)



Compound **41** (556 mg, 0.79 mmol) and triphenylphosphane (208 mg, 0.79 mmol) were dissolved in THF (10 mL) and dH_2O (2 mL) under a nitrogen atmosphere. The reaction was stirred 3 h at room temperature. The solvent was removed under reduced pressure and the residue was roughly purified by column chromatography on silica gel (cyclohexane/ethyl acetate 1:1 → ethyl acetate, under the addition of 1 % v/v diethyl amine). The residue was coevaporated with toluene (2 x 100 mL) and DCM (100 mL). Crude compound **43** was obtained as an orange oil.

Yield: 576 mg (crude product);

TLC: R_f = 0 (cyclohexane/ethyl acetate, 1:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 100.0 (c = 0.05, methanol);

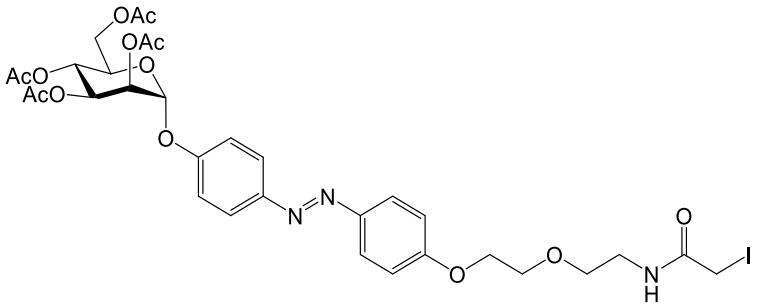
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.87 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9 Hz, 2H, Ar-$H_{2/6}$), 7.86 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 8.9 Hz, 2H, Ar-$H_{2'/6'}$), 7.20 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 7.0 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 5.6 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, H-1), 5.58 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{3,4}$ = 10 Hz, 1H, H-3), 5.48 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, H-2), 5.38 (dd~t, $^3J_{3,4}$ = $^3J_{4,5}$ = 10. 1 Hz, 1H, H-4), 4.29 (dd, $^3J_{5,6}$ = 5.6 Hz, $^2J_{6,6'}$ = 12.2 Hz, 1H, H-6), 4.21 (dd~t, $^3J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2}$ = 4.6 Hz, 2H, Ar-O$C\underline{H}_2$), 4.12-4.07 (m, 2H, H-5 ,H-6'), 3.87 (dd~t, $^3J_{Ar\text{-}OCH2C\underline{H}2,\ Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, Ar-$OCH_2C\underline{H}_2$), 3.75-3.67 (m, 6H, $OC\underline{H}_2CH_2O$, $OCH_2C\underline{H}_2O$, $C\underline{H}_2CH_2NH_2$), 3.15 (m, 2H, $C\underline{H}_2NH_2$) 2.21 (s, 3H, $COC\underline{H}_3$), 2.06 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$), 2.03 (s, 3H, $COC\underline{H}_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 169.9 (2 C, 2 x $\underline{C}OCH_3$), 169.8 ($\underline{C}OCH_3$), 160.6 (Ar-$C_{4'}$), 157.2 (Ar-C_4), 148.4 (Ar-C_1), 147.3 (Ar-$C_{1'}$), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70.5 ($O\underline{C}H_2CH_2O$), 69.9 ($OCH_2\underline{C}H_2O$), 69.5 ($O\underline{C}H_2CH_2NH_2$), 69.4 (Ar-$OCH_2\underline{C}H_2O$), 69.3 (C-5), 68.8 (C-2), 67.4 (C-3), 66.4 (Ar-$O\underline{C}H_2CH_2$), 65.9 (C-4), 62.1 (C-6), 39.8 ($\underline{C}H_2NH_2$), 20.9 ($CO\underline{C}H_3$), 20.68 ($CO\underline{C}H_3$), 20.66 (2C, 2 x $CO\underline{C}H_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3083 (br), 2886 (w), 2362 (w), 2000 (w), 1745 (s), 1598 (m), 1498 (m), 1370 (s), 1213 (s), 1029 (s), 980 (s), 842 (m), 798 (m), 722(m), 705 (m) cm^{-1};

EI-MS: *m/z* = 675.30 $[M + H]^+$, (calc 675.26 for $C_{32}H_{41}N_3O_{13} + H^+$).

[4-(4'-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}phenyl)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside[438] (44)



Crude compound **42** (179 mg, ~ 0.25 mmol) was dissolved in dry DCM (10 mL) under a nitrogen atmosphere and 2-iodoacetic anhydride (89.4 mg; 0.25 mmol) was added. The reaction mixture was stirred 16 h in darkness at room temperature. The crude compound was washed with dH_20 (50 mL) and the aqueous layer was extracted with ethyl acetate (2 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by flash chromatography (cyclohexane/ethyl acetate, 2:3 → 4:1) to obtain **44** as an orange solid.

Yield: 99.4 mg (0.12 mmol, 49 % over 2 steps);

TLC: R_f = 0.3 (cyclohexane/ethyl acetate, 2:3),
0.17 (cyclohexane/ethyl acetate, 1:1);

Melting point: 84.6 °C;

Rotational value: $[\alpha]_{20}^{D}$ = -+ 76.9 (c = 0.1, methanol);

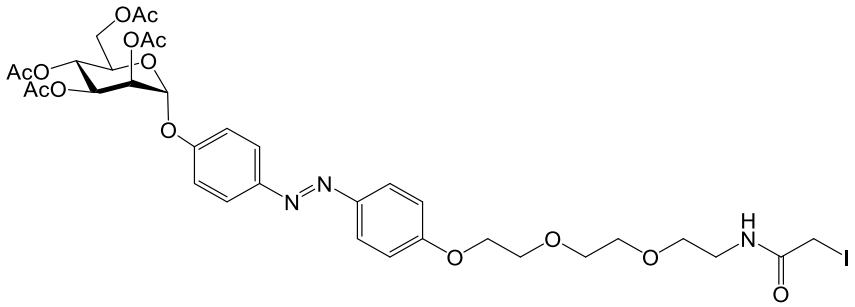
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.21 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.04 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$),6.41 (br. s, 1H, NH), 5.61 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.58 (dd, $^{3}J_{2,3}$ = 3.6 Hz, $^{3}J_{3,4}$ = 10.3 Hz, 1H, H-3), 5.48 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.5 Hz, 1H, H-2), 5.39 (dd~t, $^{3}J_{3,4}$ = 10.3 Hz, $^{3}J_{4,5}$ = 10.3 Hz, 1H, H-4), 4.3 (dd, $^{3}J_{5,6}$ = 5.5 Hz, $^{2}J_{6,6'}$ = 12.3 Hz, 1H, H-6), 4.23 (t, $^{3}J_{OC\underline{H}2,OCH2C\underline{H}2}$ = 4.6 Hz, 2H, Ar-OC$\underline{H}_2$), 4.13-4.08 (m, 2H, H-5, H-6'), 3.88 (t, $^{3}J_{Ar\text{-}OC\underline{H}2,Ar\text{-}OCH2C\underline{H}2}$ = 4.6 Hz, 2H, Ar-OCH_2C$\underline{H}_2$), 3.66 (s, 2H, COC$\underline{H}_2$I), 3.65 (t, $^{3}J_{C\underline{H}2CH2NH,CH2NH}$ = 5.2 Hz, 2H, C$\underline{H}_2$CH_2NH), 3.51 (dd~t, $^{3}J_{C\underline{H}2CH2NH,C\underline{H}2NH}$ = 5.2 Hz, 2H, C$\underline{H}_2$NH), 2.22 (s, 3H, COC$\underline{H}_3$), 2.06.9 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.04 (s, 3H, COC$\underline{H}_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (C$\underline{O}$CH_3), 169.98 (C$\underline{O}$CH_3), 169.95 (C$\underline{O}$CH_3), 169.7 (C$\underline{O}$CH_3), 167 (NHC$\underline{O}$CH_2I), 161 (Ar-C_4), 157.2 (Ar-$C_{4'}$), 148.5 (Ar-$C_{1'}$), 147.2 (Ar-C_1), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.9 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 69.6 (C$\underline{H}_2$CH_2NH), 69.5 (Ar-OCH_2C$\underline{H}_2$), 69.4 (C-5), 69.3 (C-2), 68.8 (C-3), 67.6 (Ar-O$\underline{C}H_2$), 65.9 (C-4), 62.1 (C-6), 40.1 (C$\underline{H}_2$NH), 20.9 (CO$\underline{C}H_3$), 20.7 (2C, 2 x CO$\underline{C}H_3$), 19.1 (CO$\underline{C}H_3$), – 0.7 (CO$\underline{C}H_2$I) ppm;

IR (ATR): $\tilde{\nu}$ = 3296 (br), 2928 (br), 2316 (w), 2079 (w), 1744 (s), 1597 (m), 1497 (m), 1367 (m), 1211 (s), 1125 (s), 1032 (s), 979 (s), 842 (s), 694 (w) cm^{-1};

ESI-HRMS: *m/z* = 800.15051 $[M + H]^+$, (calc. 800.15221 for $C_{32}H_{38}IN_3O_{13} + H^+$).

{4-[4'-(2-{2-[2-(2-Iodoacetamido)ethoxy]ethoxy}ethoxy)phenyl]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside[438] (45)



Crude compound **43** (161 mg, ~ 0.22 mmol) was dissolved in dry DCM (7 mL) under a nitrogen atmosphere and 2-iodoacetic anhydride (93.1 mg; 0.26 mmol) was added. The reaction mixture was stirred 16 h in darkness at room temperature. The solvent was removed under reduced pressure and the residue was purified by flash chromatography (cyclohexane/ethyl acetate/methanol, 5:5:1) to obtain **45** as an orange solid.

Yield: 101 mg (0.12 mmol, 55 % over 2 steps);

TLC: R_f = 0.21 (cyclohexane/ethyl acetate, 1:2),
0.3 (cyclohexane/ethyl acetate/methanol, 5:5:1);

Melting point: 87.9 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 53.4 (c = 0.06, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.88 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.20 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 7.02 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 6.56 (br. s, 1H, NH), 5.61 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.58 (dd, $^{3}J_{2,3}$ = 3.6 Hz, $^{3}J_{3,4}$ = 10.2 Hz, 1H, H-3), 5.48 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.5 Hz, 1H, H-2), 5.38 (dd∼t, $^{3}J_{3,4}$ = 10.2 Hz, $^{3}J_{4,5}$ = 10.2 Hz, 1H, H-4), 4.29 (dd, $^{3}J_{5,6}$ = 5.5 Hz, $^{2}J_{6,6'}$ = 12.3 Hz, 1H, H-6), 4.24 (dd∼t, $^{3}J_{OCH2,OCH2CH2}$ = 4.8 Hz, 2H, Ar-OCH_2), 4.15-4.06 (m, 2H, H-5, H-6'), 3.91 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.8 Hz, 2H, Ar-OCH$_2$CH_2), 3.77-3.74 (m, 2H, OCH_2CH$_2$O), 3.71-3.67 (m, 2H, OCH$_2$CH_2O), 3.66 (s, 2H, COCH_2I), 3.59 (dd∼t, $^{3}J_{CH2CH2NH,CH2NH}$ = 4.8 Hz, 2H, CH_2CH$_2$NH), 3.48 (dd∼t, $^{3}J_{CH2CH2NH,CH2NH}$ = 4.8 Hz, 2H, CH_2NH), 2.21 (s, 3H, COCH_3), 2.06 (s, 3H, COCH_3), 2.05 (s, 3H, COCH_3), 2.03 (s, 3H, COCH_3) ppm;

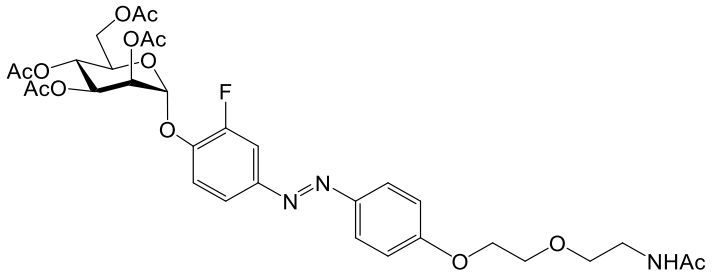
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (COCH_3), 169.96 (COCH_3), 169.93 (COCH_3), 169.7 (COCH_3), 167 (NHCOCH_2I), 160.9 (Ar-C_4), 157.2 (Ar-$C_{4'}$), 148.5 (Ar-$C_{1'}$), 147.1 (Ar-C_1), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 116.8 (2C, Ar-$C_{3'/5'}$), 114.9 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70.9 (OCH_2CH$_2$O), 70.4 (OCH$_2$CH_2O), 69.7 (Ar-OCH$_2$CH_2), 69.42 (CH_2CH$_2$NH), 69.40 (C-5), 69.3 (C-2), 68.8 (C-3), 67.8 (Ar-OCH_2), 65.9 (C-4), 62.1 (C-6), 40.2 (CH_2NH), 20.9 (COCH_3), 20.69 (COCH_3), 20.68 (2C, 2 x COCH_3), – 0.65 (COCH_2I) ppm;

IR (ATR): $\tilde{\nu}$ = 3299 (w), 2928 (w), 2366 (w), 1743 (s), 1597 (m), 1498 (m), 1367 (m), 1211 (s);

1229 (s), 979 (s), 843 (m), 789 (w), 730 (w) cm^{-1};

ESI-HRMS: m/z = 844.17682 $[M + H]^+$, (calc. 844.17842 for $C_{34}H_{42}IN_3O_{14} + H^+$).

4-[3-(4'-{2-[2-(*N*-Acetylaminoethoxy)ethoxy]phenyl}fluoro)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (46)



Crude compound **17** (260 mg, ∼ 0.4 mmol) was dissolved in pyridine (20 mL) and acetic anhydride (5 mL) was added. The reaction mixture was stirred 16 h under ambient conditions. 1 N HCl_{aq} was added (150 mL) and the aqueous layer was extracted with ethyl acetate (2 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by column chromatography on silica gel (ethyl acetate → ethyl acetate/ methanol, 10:1) gave compound **46** as a viscous, orange oil

Yield: 156 mg (0.23 mmol, ∼ 56 %);

TLC: R_f = 0.22 (ethyl acetate), 0.5 (ethyl acetate/methanol, 10:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 68 (c = 0.01 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.89 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.71-7.66 (m, 2H, Ar-$H_{3/5}$), 7.33 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 9 Hz, 1H, Ar-H_5), 7.04 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 5.90 (br. s, 1H, NH), 5.60 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.59 (dd, $^{3}J_{3,4}$ = 10 Hz, $^{3}J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^{3}J_{2,3}$ = 3.5 Hz, $^{3}J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.39 (dd∼t, $^{3}J_{4,5}$ = 10 Hz, $^{3}J_{4,3}$ = 10 Hz, 1H, H-4), 4.29 (dd, $^{2}J_{6,6'}$ = 12.2 Hz, $^{3}J_{6,5}$ = 5.5 Hz, 1H, H-6), 4.24-4.20 (m, 3H, H-5, Ar-OC$\underline{H}_2$), 4.13-4.09 (m, 1H, H-6'), 3.87 (dd∼t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2,Ar\text{-}OC\underline{H}2}$ = 4.7 Hz, 2H, Ar-$OCH_2C\underline{H}_2$), 3.65 (t, $^{3}J_{OC\underline{H}2OCH2NH,C\underline{H}2NH}$ = 5.1 Hz, 2H, $OC\underline{H}_2CH_2NH$), 3.44 (dd∼q, $^{3}J_{C\underline{H}2NH,OC\underline{H}2OCH2NH}$ = 5.1 Hz, 2H, $C\underline{H}_2NH$), 2.22 (s, 3H, $COC\underline{H}_3$), 2.08 (s, 3H, $COC\underline{H}_3$), 2.05 (s, 3H, $COC\underline{H}_3$), 2.04 (s, 3H, $COC\underline{H}_3$), 1.96 (s, 3H, $NHCOC\underline{H}_3$) ppm;

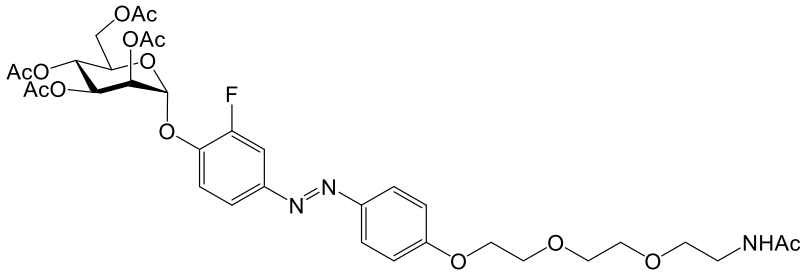
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 ($\underline{C}OCH_3$), 170.2 ($NH\underline{C}OCH_3$), 169.9 ($\underline{C}OCH_3$), 169.8 (2C, 2 x $\underline{C}OCH_3$), 161.3 (Ar-$C_{4'}$), 153.4 (d, $^{1}J_{Ar\text{-}C3,F}$ = 249.6 Hz, Ar-C_3), 149.0 (d, $^{3}J_{Ar\text{-}C1,F}$ = 5 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^{2}J_{Ar\text{-}C4,F}$ = 11.8 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (Ar-C_6), 118.2 (Ar-C_5), 114.8 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^{2}J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 97.2 (C-1), 70.1 ($O\underline{C}H_2CH_2NH$), 69.8 (C-5), 69.4 (Ar-$OCH_2\underline{C}H_2O$), 69.2 (C-2), 68.7 (C-3), 67.6 (Ar-$O\underline{C}H_2$), 65.8 (C-4), 62.1 (C-6), 39.2 ($\underline{C}H_2NH$), 23.3 ($NHCO\underline{C}H_3$), 21 ($CO\underline{C}H_3$), 20.8 ($CO\underline{C}H_3$), 20.68 ($CO\underline{C}H_3$), 20.66 ($CO\underline{C}H_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.45 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3391 (br), 2926 (m), 2366(w), 2308 (w), 2168 (w), 1746 (s), 1655 (m), 1599 (m), 1500 (s), 1426 (w), 1369 (m), 1242 (s), 1221 (s), 1131 (s), 1089 (m), 1054 (m), 1026 (s), 977 (m), 839 (m), 749 (w), 625 (m), 600 (m), 523 (w), 447 (w) cm^{-1};

ESI-HRMS: *m/z* = 692.24575 $[M + H]^+$, (calc. 692.24614 for $C_{30}H_{36}FN_3O_{12} + H^+$).

(4-{3-[4'-(2-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (47)



Crude compound **18** (187.3 mg, ∼ 0.18 mmol) was dissolved in pyridine (20 mL) and acetic anhydride (5 mL) was added. The reaction mixture was stirred 16 h under ambient conditions. 1 N HCl_{aq} was added (150 mL) and the aqueous layer was extracted with ethyl acetate (2 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by column chromatography on silica gel (ethyl acetate → ethyl acetate/methanol, 10:1) gave compound **47** as a viscous, orange oil

Yield: 132 mg (0.14 mmol, ∼ 79 %);

TLC: R_f = 0 (cyclohexane/ethyl acetate, 1:1), 0.25 (ethyl acetate /methanol, 10:1);

Rotational value: $[\alpha]_{20}^{D}$ =+ 64.5 (c = 0.005 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 600 MHz, 298 K): δ = 7.89 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.70-7.65 (m, 2H, Ar-$H_{3/5}$), 7.32 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.8 Hz, 1H, Ar-H_5), 7.02 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 6.07 (br. s, 1H, NH), 5.60 (d, $^{3}J_{1,2}$ = 1.7 Hz, 1H, H-1), 5.59 (dd, $^{3}J_{3,4}$ = 10.1 Hz, $^{3}J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.55 (dd, $^{3}J_{2,3}$ = 3.5 Hz, $^{3}J_{2,1}$ = 1.7 Hz, 1H, H-2), 5.39 (dd∼t, $^{3}J_{4,5}$ = 10.1 Hz, $^{3}J_{4,3}$ = 10.1 Hz, 1H, H-4), 4.30 (dd, $^{2}J_{6,6'}$ = 12.2 Hz, $^{3}J_{6,5}$ = 5.6 Hz, 1H, H-6), 4.24 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.6 Hz, 2H, Ar-OCH_2), 4.22 (ddd , $^{3}J_{5,4}$ = 10.1 Hz, $^{3}J_{5,6}$ = 5.6 Hz, $^{3}J_{5,6'}$= 2.1 Hz, 1H, H-5), 4.13-4.09 (m, 1H, H-6'), 3.91 (dd∼t, $^{3}J_{Ar\text{-}OCH2CH2,Ar\text{-}OCH2}$ = 4.6 Hz, 2H, Ar-OCH_2CH_2), 3.76-3.73 (m, 2H, OCH_2CH_2O), 3.69-3.66 (m, 2H, OCH_2CH_2O), 3.58 (dd∼t, $^{3}J_{OCH2OCH2NH,CH2NH}$ = 5.2 Hz, 2H, OCH_2CH_2NH), 3.44 (dd∼q, $^{3}J_{CH2NH,OCH2OCH2NH}$ = 5.2 Hz, 2H, CH_2NH), 2.22 (s, 3H, $COCH_3$), 2.08 (s, 3H, $COCH_3$), 2.05 (s, 3H, $COCH_3$), 2.04 (s, 3H, $COCH_3$), 1.96 (s, 3H, $NHCOCH_3$) ppm;

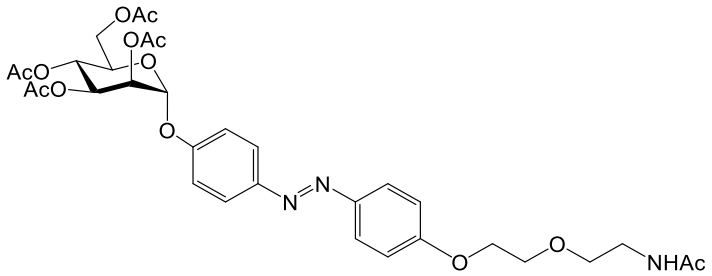
^{13}C-NMR ($CDCl_3$, 151 MHz, 298 K): δ = 170.5 ($COCH_3$), 170.2 ($NHCOCH_3$), 169.9 ($COCH_3$), 169.8 ($COCH_3$), 196.75 ($COCH_3$), 161.3 (Ar-$C_{4'}$), 153.5 (d, $^{1}J_{Ar\text{-}C3,F}$ = 249.6 Hz, Ar-C_3), 149.0 (d, $^{3}J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 146.8 (Ar-$C_{1'}$), 144.9 (d, $^{2}J_{Ar\text{-}C4,F}$ = 11.9 Hz, Ar-C_4), 124.9 (2C, Ar-$C_{2'/6'}$), 121.4 (d, $^{4}J_{Ar\text{-}C6,F}$ = 2.6 Hz, Ar-C_6), 118.2 (Ar-C_5), 114.8 (2C, Ar-$C_{3'/5'}$), 108.7 (d, $^{2}J_{Ar\text{-}C2,F}$ = 19.5 Hz, Ar-C_2), 97.2 (C-1), 70.8 (OCH_2CH_2O), 70.2 (OCH_2CH_2O), 69.9 (OCH_2CH_2NH), 69.8 (C-5), 69.6 (Ar-OCH_2CH_2O), 69.2 (C-2), 68.7 (C-3), 67.7 (Ar-OCH_2), 65.8 (C-4), 62.1 (C-6), 39.2 (CH_2NH), 23.2 ($NHCOCH_3$), 20.9 ($COCH_3$), 20.7 ($COCH_3$), 20.66 (2C, 2 x $COCH_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 130.47 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3299 (br), 2924 (m), 2370 (w), 2185 (w), 1968 (w), 1749 (s), 1655 (m), 1599 (m), 1500 (m), 1369 (m), 1244 (s), 1223 (s), 1133 (s), 1091 (m), 1055 (m), 1027 (m), 977 (m), 841 (m), 601 (w), 528 (m), 467 (m), 448 (m) cm^{-1};

ESI-HRMS: *m/z* = 736.27039 $[M + H]^+$, (calc. 736.27236 for $C_{34}H_{42}FN_3O_{14} + H^+$).

[4-(4'-{2-[2-(*N*-Acetylamino)ethoethoxy}phenyl)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (48)



Crude amine **42** (182 mg, ∼ 0.27 mmol) was dissolved in pyridine (15 mL) and acetic anhydride (5mL) was added. The reaction mixture was stirred 18 h under ambient conditions. The solvent was removed under reduced pressure, the residue was dissolved in ethyl acetate (100 mL) and washed with 1 N HCl_{aq} (100 mL). The aqueous layer was extracted with ethyl acetate (100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by column chromatography on silica gel (ethyl acetate → ethyl acetate/methanol, 10:1) gave the peracetylated compound **48** as a viscous, orange oil

Yield: 133 mg (0.2 mmol, ∼ 73 %);

TLC: R_f = 0.15 (ethyl acetate), 0.52 (ethyl acetate/methanol, 10:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 69.8 (c = 0.015 mmol · mL^{-1}, methanol);

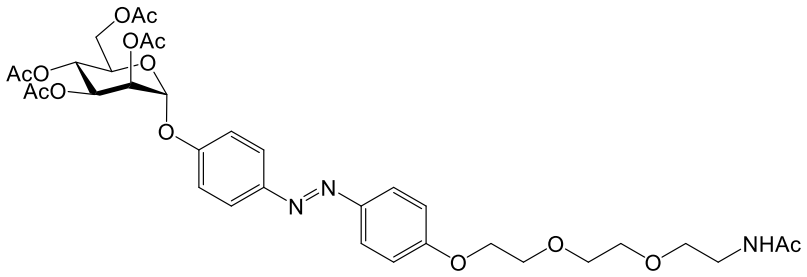
^{1}H-NMR ($CDCl_3$, 600 MHz, 298 K): δ = 7.89 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 8.9 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 8.9 Hz, 2H, Ar-$H_{2'/6'}$), 7.21 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 8.9 Hz, 2H, Ar-$H_{3'/5'}$), 7.03 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 8.9 Hz, 2H, Ar-$H_{3/5}$), 5.94 (br. s, 1H, NH), 5.61 (d, $^{3}J_{1,2}$ = 1.7 Hz, 1H, H-1), 5.58 (dd, $^{3}J_{2,3}$ = 3.6 Hz, $^{3}J_{3,4}$ = 10.1 Hz, 1H, H-3), 5.48 (dd, $^{3}J_{1,2}$ = 1.7 Hz, $^{3}J_{2,3}$ = 3.6 Hz, 1H, H-2), 5.38 (dd∼t, $^{3}J_{3,4}$ = 10.1 Hz, $^{3}J_{4,5}$ = 10.1 Hz, 1H, H-4), 4.30 (dd, $^{3}J_{5,6}$ = 5.6 Hz, $^{2}J_{6,6'}$ = 12.5 Hz, 1H, H-6), 4.22 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH_2), 4.14-4.07 (m, 2H, H-5, H-6'), 3.87 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH_2CH_2), 3.65 (dd∼t, $^{3}J_{CH2CH2NH,CH2NH}$ = 5 Hz, 2H, CH_2CH_2NH), 3.49 (dd∼q, $^{3}J_{CH2CH2NH,CH2NH}$ = 5 Hz, 2H, CH_2NH), 2.22 (s, 3H, $COCH_3$), 2.06 (s, 3H, $COCH_3$), 2.05 (s, 3H, $COCH_3$), 2.04 (s, 3H, $COCH_3$), 1.97 (s, 2H, $NHCOCH_3$) ppm;

^{13}C-NMR ($CDCl_3$, 151 MHz, 298 K): δ = 170.5 ($COCH_3$), 170.3 ($NHCOCH_3$), 169.98 ($COCH_3$), 169.96 ($COCH_3$), 169.7 ($COCH_3$), 160.9 (Ar-C_4), 157.2 (Ar-$C_{4'}$), 148.4 (Ar-$C_{1'}$), 147.1 (Ar-C_1), 124.6 (2C, Ar-$C_{2/6}$), 124.3 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70 (CH_2CH_2NH), 69.42 (Ar-OCH_2OCH_2), 69.37 (C-5), 69.3 (C-2), 68.8 (C-3), 67.5 (Ar-OCH_2), 65.9 (C-4), 62.1 (C-6), 39.3 (CH_2NH), 23.3 ($NHCOCH_3$), 20.9 ($COCH_3$), 20.7 ($COCH_3$), 20.68 (2C, 2 x $COCH_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3299 (br), 2919 (m), 2850 (m), 2370 (w), 2343 (w), 2076 (m), 1743 (s), 1658 (m), 1598 (m), 1583 (m), 1498 (m), 1369 (m), 1218 (s), 1127 (s), 1932 (s), 981 (m), 844 (m), 719 (w), 602 (w), 552 (m). 494 (w), 468 (w) cm^{-1};

ESI-HRMS: *m/z* = 674.25421 $[M + H]^+$, (calc. 674.25556 for $C_{32}H_{39}N_3O_{13} + H^+$).

{4-[4'-(2-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}ethoxy)phenyl]azophenoxy}-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (49)



Crude amine **43** (182 mg, ∼ 0.27 mmol) was dissolved in pyridine (20 mL) and acetic anhydride (5mL) was added. The reaction mixture was stirred 18 h under ambient conditions. The solvent was removed under reduced pressure, the residue was dissolved in ethyl acetate (100 mL) and washed with 1 N HCl_{aq} (100 mL). The aqueous layer was extracted with ethyl acetate (2 x 100 mL). The combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by column chromatography on silica gel (ethyl acetate → ethyl acetate/ methanol, 10:1) gave peracetylated compound **49** as a viscous, brown oil

Yield: 120 mg (0.17 mmol, ∼ 62 %);

TLC: R_f = 0.05 (ethyl acetate), 0.42 (ethyl acetate/methanol, 10:1);

Rotational value: $[\alpha]_{20}^{D}$ = + 74.9 (c = 0.014 mmol · mL^{-1}, methanol);

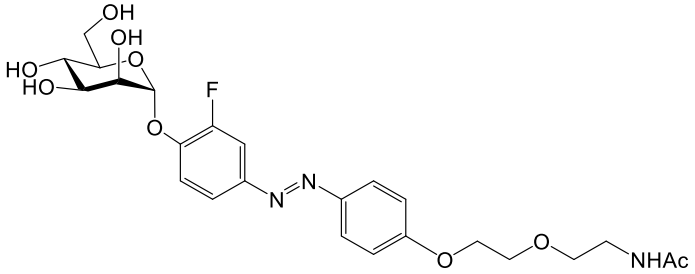
^{1}H-NMR ($CDCl_3$, 600 MHz, 298 K): δ = 7.88 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9 Hz, 2H, Ar-$H_{2/6}$), 7.87 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.21 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 7.02 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 6.11 (br. s, 1H, NH), 5.61 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, H-1), 5.58 (dd, $^3J_{2,3}$ = 3.6 Hz, $^3J_{3,4}$ = 10 Hz, 1H, H-3), 5.48 (dd, $^3J_{1,2}$ = 1.7 Hz, $^3J_{2,3}$ = 3.6 Hz, 1H, H-2), 5.38 (dd∼t, $^3J_{3,4}$ = 10 Hz, $^3J_{4,5}$ = 10 Hz, 1H, H-4), 4.30 (dd, $^3J_{5,6}$ = 5.6 Hz, $^2J_{6,6'}$ = 12.5 Hz, 1H, H-6), 4.23 (t, $^3J_{OCH2,OCH2CH2}$ = 4.8 Hz, 2H, Ar-OCH_2), 4.14-4.07 (m, 2H, H-5, H-6'), 3.91 (t, $^3J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.8 Hz, 2H, Ar-OCH_2CH_2), 3.75-3.73 (m, 2H, OCH_2CH_2O), 3.68-3.66 (m, 2H, OCH_2CH_2O), 3.58 (dd∼t, $^3J_{CH2CH2NH,CH2NH}$ = 5.3 Hz, 2H, CH_2CH_2NH), 3.49 (dd∼q, $^3J_{CH2CH2NH,\ CH2NH}$ = 5.3 Hz, 2H, CH_2NH), 2.22 (s, 3H, $COCH_3$), 2.06 (s, 3H, $COCH_3$), 2.05 (s, 3H, $COCH_3$), 2.04 (s, 3H, $COCH_3$), 1.96 (s, 2H, $NHCOCH_3$) ppm;

^{13}C-NMR ($CDCl_3$, 151 MHz, 298 K): δ = 170.5 ($COCH_3$), 170.3 ($NHCOCH_3$), 169.98 ($COCH_3$), 169.95 ($COCH_3$), 169.7 ($COCH_3$), 160.9 (Ar-C_4), 157.2 (Ar-$C_{4'}$), 148.4 (Ar-$C_{1'}$), 147.1 (Ar-C_1), 124.6 (2C, Ar-$C_{2/6}$), 124.2 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3'/5'}$), 114.8 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 70.8 (OCH_2CH_2O), 70.2 (OCH_2CH_2O), 69.9 (CH_2CH_2NH), 69.6 (Ar-OCH_2OCH_2), 69.4 (C-5), 69.3 (C-2), 68.8 (C-3), 67.7 (Ar-OCH_2), 65.9 (C-4), 62.1 (C-6), 39.3 (CH_2NH), 23.2 ($NHCOCH_3$), 20.9 ($COCH_3$), 20.7 ($COCH_3$), 20.68 (2C, 2 x $COCH_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3328 (br), 2924 (m), 2370 (w), 2299 (w), 2186 (w), 1742 (s), 1653 (m), 1597 (m), 1497 (m), 1368 (m), 1215 (s), 1125 (s), 1031 (s), 978 (m), 843 (m), 686 (m), 600 (m), 551 (m), 495 (m), 469 (w), 439 (w) cm^{-1};

ESI-HRMS: m/z = 718.28040 $[M + H]^+$, (calc. 718.28178 for $C_{34}H_{43}N_3O_{14} + H^+$).

{4-[3-(4'-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}phenyl)fluoro]azophenoxy}-*α*-D-mannopyranoside (51)



Compound **46** (126 mg, 0.18 mmol) was dissolved in methanol (10 mL) and a solution of sodium methanolate (20 µL, 5.4 M) was added. The reaction was stirred for 2 h under ambient conditions until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed from the filtrate under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol, 10:1 → ethyl acetate/d$H_2$0, 20:2:1), followed by size exclusion chromatography (LH-20, methanol) to afford title compound **51** as a yellow solid.

Yield: 89.9 mg (0.17 mmol, 94 %);

TLC: R_f = 0.08 (ethyl acetate/methanol, 10:1),
trans-isomer: 0.29 (ethyl acetate/methanol/d$H_2$0, 20:2:1),
cis-isomer: 0.2 (ethyl acetate/methanol/d$H_2$0, 20:2:1);

Melting point: 169.7 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 111.5 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.88 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.7 Hz, $^{4}J_{Ar\text{-}H6,Ar\text{-}H2}$ = 2.4 Hz, $^{5}J_{Ar\text{-}H6\text{-}F}$ = 1.2 Hz, 1H, Ar-H_6), 7.64 (dd, $^{3}J_{Ar\text{-}H2,F}$ = 12 Hz, $^{4}J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.4 Hz, 1H, Ar-H_2), 7.54 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.09 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^{3}J_{1,2}$ = 1.7 Hz, 1H, H-1), 4.23 (dd~t, $^{3}J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.5 Hz, 2H, Ar-OC$\underline{H}_2$), 4.09 (dd, $^{3}J_{2,3}$ = 3.4 Hz, $^{3}J_{2,1}$ = 1.7 Hz, 1H, H-2), 3.94 (dd, $^{3}J_{3,4}$ = 9.4 Hz, $^{3}J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.85 (dd~t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.5 Hz, 2H, Ar-$OCH_2C\underline{H}_2O$), 3.80 (dd, $^{2}J_{6,6'}$ = 11.8 Hz, $^{3}J_{6,5}$ = 2.4 Hz, 1H, H-6), 3.78-3.71 (m, 2H, H-4, H-6'), 3.65 (ddd, $^{3}J_{5,4}$ = 9.8 Hz, $^{3}J_{5,6}$ = 5.4 Hz, $^{3}J_{5,6'}$ = 2.3 Hz, 1H, H-5), 3.62 (t, $^{3}J_{OC\underline{H}2CH2NH,C\underline{H}2NH}$ = 5.6 Hz, 2H, OC$\underline{H}_2CH_2$NH), 3.39 (t, $^{3}J_{C\underline{H}2NH,OC\underline{H}2CH2NH}$ = 5.6 Hz, 2H, C$\underline{H}_2$NH), 1.93 (s, 3H, NHCOC$\underline{H}_3$) ppm;

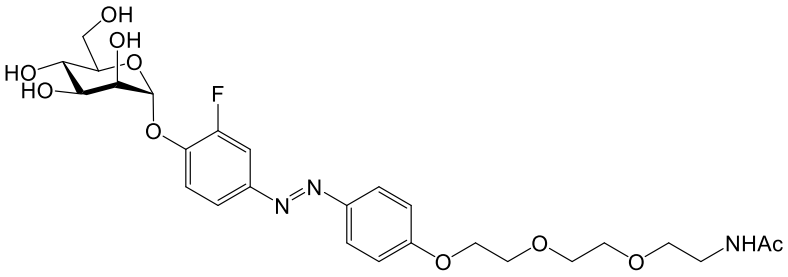
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 173.5 (NHC$\underline{O}$CH$_3$), 163.1 (Ar-$C_{4'}$), 154.7 (d, $^{1}J_{Ar\text{-}C3,F}$ = 248 Hz, Ar-C_3), 149.5 ($^{3}J_{Ar\text{-}C1,F}$ = 5.2 Hz, Ar-C_1), 148.1 (Ar-$C_{1'}$), 145.7 (d, $^{2}J_{Ar\text{-}C4,F}$ = 11.9 Hz, Ar-C_4), 125.8 (2C, Ar-$H_{2'/6'}$), 122.8 (Ar-C_6), 119.2 (Ar-C_5), 116 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^{2}J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 101.2 (C-1), 76.1 (C-5), 72.3 (C-3), 71.7 (C-2), 70.8 (O$\underline{C}H_2CH_2$NH), 70.5 (Ar-OCH$_2\underline{C}$H$_2$O), 69.1 (Ar-O$\underline{C}$H$_2$), 68.2 (C-4) 62.7 (C-6), 40.5 ($\underline{C}H_2$NH), 22.5 (NHCO$\underline{C}$H$_3$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 133.81 (s, Ar-$F_{\textit{trans}\text{-isomer}}$), – 133.03 (m, Ar-$F_{\textit{cis}\text{-isomer}}$) ppm;

IR (ATR): $\tilde{\nu}$ = 3533 (m), 3319 (br), 2918 (w), 1383 (w), 2325 (w), 2299 (w), 1646 (m), 1602 (m), 1583 (m), 1504 (m), 1425 (w), 1321 (m), 1270 (m), 1244(s), 1130 (m), 1096 (s), 1002 (s), 973 (m), 844 (m), 749 (m), 637 (w), 619 (w), 503 (w), 468 (w), 401 (w) cm^{-1};

ESI-HRMS: *m/z* = 524.20313 $[M + H]^+$, (calc. 524.20388 for $C_{24}H_30FN_3O_9 + H^+$).

(4-{3-[4'-(2-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}ethoxy)phenyl]fluoro}azophenoxy)-*α*-D-mannopyranoside (52)



Compound **47** (98.5 mg, 0.13 mmol) was dissolved in methanol (10 mL) and a solution of sodium methanolate (20 µL, 5.4 M) was added. The reaction was stirred for 2 h under ambient conditions until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed from the filtrate under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol 10:1 → ethyl acetate/d$H_2$0, 20:2:1) to afford compound **52** as a yellow solid.

Yield: 72 mg (0.127 mmol, 95 %);

TLC: R_f =0.05 (ethyl acetate/methanol, 10:1),
trans-isomer: 0.23 (ethyl acetate/methanol/d$H_2$0 20:2:1),
cis-isomer: 0.13 (ethyl acetate/methanol/d$H_2$0, 20:2:1);

Melting point: 121.0 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 106.2 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.87 (d, $^{3}J_{Ar\text{-}H2'/Ar\text{-}H6',Ar\text{-}H3'/Ar\text{-}H5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.71 (ddd, $^{3}J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.7 Hz, $^{4}J_{Ar\text{-}H6,Ar\text{-}H2}$ = 2.4 Hz, $^{5}J_{Ar\text{-}H6\text{-}F}$ = 1.1 Hz, 1H, Ar-H_6), 7.63 (dd, $^{3}J_{Ar\text{-}H2,F}$ = 11.9 Hz, $^{4}J_{Ar\text{-}H2,Ar\text{-}H6}$ = 2.4 Hz, 1H, Ar-H_2), 7.53 (t, $^{3}J_{Ar\text{-}H5,Ar\text{-}H6}$ = 8.7 Hz, 1H, Ar-H_5), 7.08 (d, $^{3}J_{Ar\text{-}H3'/Ar\text{-}H5',Ar\text{-}H2'/Ar\text{-}H6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 5.61 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.23 (dd~t, $^{3}J_{Ar\text{-}OC\underline{H}2,OCH2C\underline{H}2O}$ = 4.6 Hz, 2H, Ar-OC$\underline{H}_2$), 4.01 (dd, $^{3}J_{2,3}$ = 3.4 Hz, $^{3}J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.94 (dd, $^{3}J_{3,4}$ = 9.5 Hz, $^{3}J_{3,2}$ = 3.4 Hz, 1H, H-3), 3.89 (dd~t, $^{3}J_{Ar\text{-}OCH2C\underline{H}2O,Ar\text{-}OC\underline{H}2}$ = 4.6 Hz, 2H, Ar-OCH$_2$C$\underline{H}_2$O), 3.79 (dd, $^{2}J_{6,6'}$ = 11.9 Hz, $^{3}J_{6,5}$ = 2.3 Hz, 1H, H-6), 3.78-3.69 (m, 4H, H-4, H-6', OC$\underline{H}_2$CH$_2$O), 3.68-3.63 (m, 3H, H-5, OCH$_2$C$\underline{H}_2$O), 3.54 (t, $^{3}J_{OC\underline{H}2CH2NH,C\underline{H}2NH}$ = 5.4 Hz, 2H, OC$\underline{H}_2$CH$_2$NH), 3.31 (t, $^{3}J_{C\underline{H}2NH,OC\underline{H}2CH2NH}$ = 5.4 Hz, 2H, C$\underline{H}_2$NH), 1.92 (s, 3H, NHCOC$\underline{H}_3$) ppm;

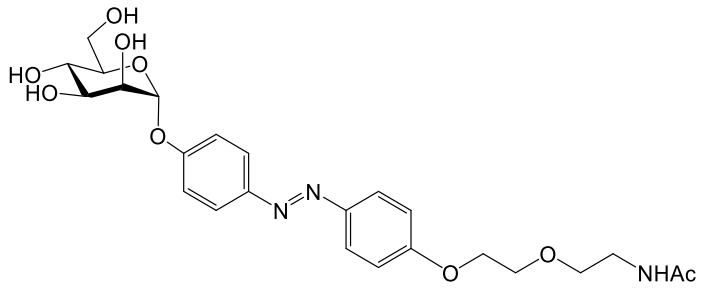
^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 173.4 (NH$\underline{C}$OCH$_3$), 163 (Ar-$C_{4'}$), 154.7 (d, $^{1}J_{Ar\text{-}C3,F}$ = 247.8 Hz, Ar-C_3), 149.5 ($^{3}J_{Ar\text{-}C1,F}$ = 5.4 Hz, Ar-C_1), 148.1 (Ar-$C_{1'}$), 147.5 (d, $^{2}J_{Ar\text{-}C4,F}$ = 11.9 Hz, Ar-C_4), 125.8 (2C, Ar-$H_{2'/6'}$), 122.8 (d, $^{4}J_{Ar\text{-}C6,F}$ = 2.8 Hz, Ar-C_6), 119.3 (Ar-C_5), 116 (2C, Ar-$C_{3'/5'}$), 108.8 (d, $^{2}J_{Ar\text{-}C2,F}$ = 20 Hz, Ar-C_2), 101.3 (C-1), 76.1 (C-5), 72.3 (C-3), 71.8 (O$\underline{C}$H$_2$CH$_2$O), 71.77 (C-2), 71.3 (OCH$_2\underline{C}$H$_2$O); 70.8 (Ar-OCH$_2\underline{C}$H$_2$O), 70.6 (O$\underline{C}$H$_2$CH$_2$NH), 69 (Ar-O$\underline{C}$H$_2$), 68.2 (C-4) 62.7 (C-6), 40.5 ($\underline{C}$H$_2$NH), 22.6 (NHCO$\underline{C}$H$_3$) ppm;

^{19}F-NMR ($CDCl_3$, 471 MHz, 298 K): δ = – 133.84 (s, Ar-$F_{\textit{trans}\text{-isomer}}$), – 133.03 (m, Ar-$F_{\textit{cis}\text{-isomer}}$) ppm;

IR (ATR): $\tilde{\nu}$ = 3335 (br), 2936 (m), 2371 (w), 2342 (m), 2326 (w), 2160 (w), 2104 (w), 2042 (w), 1639 (m), 1601 (m), 1501 (s), 1501 (s), 145 (w), 1258 (s), 1130 (s), 1099 (s), 1058 (m), 1002 (s), 975 (m), 839 (m), 603 (w), 499 (w), 464 (m), 419 (m) cm^{-1};

ESI-HRMS: *m*/*z* = 568.22937 [M + H]$^+$, (calc. 568.23010 for $C_{26}H_{34}FN_3O_{10}$ + H$^+$).

[4-(4'-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}phenyl)azophenoxy]-*α*-D-mannopyranoside (53)



Peracetylated azobenzene mannoside derivative **48** (120 mg, 0.18 mmol) was dissolved in methanol (10 mL) and a solution of sodium methanolate (20 µL, 5.4 M) was added. The reaction was stirred for 2 h under ambient conditions until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed from the filtrate under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol, 10:1 → ethyl acetate/d$H_2$0, 20:2:1) to afford deptrotected compound **53** as a yellow solid.

Yield: 84.6 mg (0.17 mmol, 94 %);

TLC: R_f = *trans*-isomer: 0.27 (ethyl acetate/methanol/dH_2O, 20:2:1),
cis-isomer: 0.13 (ethyl acetate/methanol/dH_2O, 20:2:1);

Melting point: 178.9 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 116.2 (c = 0.02 mmol · mL^{-1}, methanol);

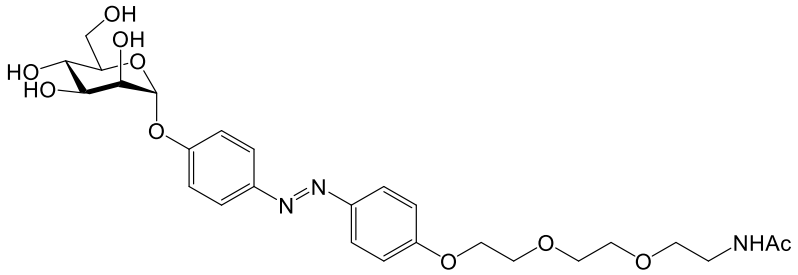
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.86 (d, $^{3}J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.85 (d, $^{3}J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.26 (d, $^{3}J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.08 (d, $^{3}J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.59 (d, $^{3}J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.22 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH_2), 4.04 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.4 Hz, 1H, H-2), 3.92 (dd, $^{3}J_{2,3}$ = 3.4 Hz, $^{3}J_{3,4}$ = 9.4 Hz, 1H, H-3), 3.85 (dd∼t, $^{3}J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH$_2$CH_2), 3.80-3.70 (m, 3H, H-4, H-6, H-6'), 3.62 (dd∼t, $^{3}J_{CH2CH2NH,CH2NH}$ = 5.5 Hz, 2H, CH_2CH$_2$NH), 3.59 (ddd, 1H, $^{3}J_{5,4}$ = 9.8 Hz, $^{3}J_{5,6}$ = 5.3 Hz, $^{3}J_{5,6'}$ = 2.4 Hz, H-5), 3.38 (t, $^{3}J_{CH2CH2NH,CH2NH}$ = 5.5 Hz, 2H, CH_2NH), 1.93 (s, 2H, NHCOCH_3) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 162.7 (NHCOCH$_3$), 159.9 (Ar-C_4), 149.3 (Ar-$C_{4'}$), 148.4 (Ar-$C_{1'}$), 147.1 (Ar-C_1), 125.5 (2C, Ar-$C_{2/6}$), 125.2 (2C, Ar-$C_{2'/6'}$), 118 (2C, Ar-$C_{3'/5'}$), 116 (2C, Ar-$C_{3/5}$), 100.2 (C-1), 75.7 (C-5), 72.4 (C-3), 71.9 (C-2). 70.8 (CH$_2$CH$_2$NH), 70.5 (Ar-OCH$_2$OCH$_2$), 69.1 (Ar-OCH$_2$), 68.3 (C-4), 62.7 (C-6), 40.5 (CH$_2$NH), 22.5 (NHCOCH$_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3302 (br), 2919 (m), 2385 (w), 2358 (w), 2293 (w), 2151 (w), 1640 (m), 1581 (s), 1498 (m), 1378 (m), 1309 (m), 1233 (s), 1125 (s), 1107 (s), 1047 (s), 604 (s), 841 (s), 749 (w), 662 (w), 547 (m), 435 (w), 419 (w) cm^{-1};

ESI-HRMS: *m/z* = 506.21274 [M + H]$^{+}$, (calc. 506.21331 for $C_{24}H_{31}N_3O_9$ + H^{+}).

{4-[4'-(2-{2-[2-(*N*-Acetylamino)ethoxy]ethoxy}ethoxy)phenyl]azophenoxy}-*α*-D-mannopyranoside (54)



Peracetylated azobenzene mannoside derivative **49** (105 mg, 0.15 mmol) was dissolved in methanol (10 mL) and a solution of sodium methanolate (20 µL, 5.4 M) was added. The reaction was stirred for 2 h under ambient conditions until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed from the filtrate under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol, 10:1 → ethyl acetate/d$H_2$0, 20:2:1) to afford deprotected compound **54** as a yellow solid.

Yield: 74.8 mg (0.14 mmol, 93 %);

TLC: R_f = *trans*-isomer: 0.22 (ethyl acetate/methanol/dH_2O, 20:2:1),
cis-isomer: 0.12 (ethyl acetate/methanol/dH_2O, 20:2:1);

Melting point: 136.9 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 107.6 (c = 0.009 mmol · mL^{-1}, methanol);

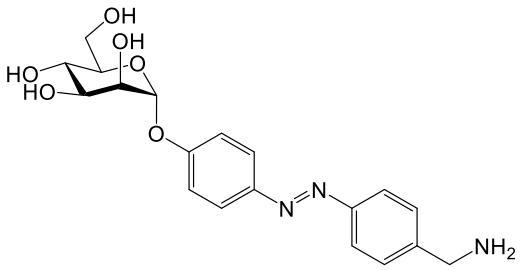
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.86 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.84 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9.1 Hz, 2H, Ar-$H_{2'/6'}$), 7.26 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9.1 Hz, 2H, Ar-$H_{3'/5'}$), 7.08 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.59 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.22 (dd∼t, $^3J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH_2), 4.04 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, H-2), 3.92 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{3,4}$ = 9.5 Hz, 1H, H-3), 3.88 (dd∼t, $^3J_{Ar\text{-}OCH2,Ar\text{-}OCH2CH2}$ = 4.5 Hz, 2H, Ar-OCH_2CH_2), 3.80-3.71 (m, 5H, H-4, H-6, H-6', OCH_2CH_2O), 3.66-3.63 (m, 2H, OCH_2CH_2O), 3.60 (ddd, 1H, $^3J_{5,4}$ = 9.7 Hz, $^3J_{5,6}$ = 5.1 Hz, $^3J_{5,6'}$ = 2.3 Hz, H-5), 3.55 (dd∼t, $^3J_{CH2CH2NH,CH2NH}$ = 5.6 Hz, 2H, CH_2CH_2NH), 3.45 (t, $^3J_{CH2CH2NH,\ CH2NH}$ = 5.6 Hz, 2H, CH_2NH), 1.92 (s, 2H, $NHCOCH_3$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 163.4 ($NHCOCH_3$), 162.6 (Ar-C_4), 159.9 (Ar-$C_{4'}$), 149.3 (Ar-$C_{1'}$), 148.4 (Ar-C_1), 125.5 (2C, Ar-$C_{2/6}$), 125.2 (2C, Ar-$C_{2'/6'}$), 118 (2C, Ar-$C_{3'/5'}$), 116 (2C, Ar-$C_{3/5}$), 100.1 (C-1), 75.7 (C-5), 72.4 (C-3), 71.9 (C-2), 71.8 (OCH_2CH_2O), 71.3 (OCH_2CH_2O), 70.8 (Ar-OCH_2OCH_2), 70.6 (CH_2CH_2NH), 69 (Ar-OCH_2), 68.3 (C-4), 62.7 (C-6), 40.5 (CH_2NH), 22.6 ($NHCOCH_3$) ppm;

IR (ATR): $\tilde{\nu}$ = 3339 (br), 2919 (m), 2382 (w), 2342 (w), 2249 (w), 2012(w), 1630 (m), 1597 (s), 1581 (m), 1497 (m), 154 (m), 1300 (m), 1234 (s), 1122 (m), 1104 (m), 1009 (m), 978 (m), 841 (m), 672 (m), 558 (m), 411 (m) cm^{-1};

ESI-HRMS: *m/z* = 550.23877 $[M + H]^+$, (calc. 550.23952 for $C_{26}H_{35}N_3O_{10} + H^+$).

4-{4'-[(Aminomethyl)phenyl]azophenoxy}-*α*-D-mannopyranoside (56)



The deacetylated azobenzene mannoside derivative **67** (1.16 g, 2.37 mmol) was dissolved in DCM (30 mL) and trifluoroacetic acid (2.79 mL, 36.32 mmol) was added dropwise. The reaction mixture was stirred 16 h under ambient conditions. Since TLC indicated an incomplete conversion, probably due to a low solubility of the acid in DCM, methanol (20 mL) and additional trifluoroacetic acid (2.79 mL, 36.32 mmol) was added. The reaction was stirred 2.5 h at room temperature, until TLC indicated full consumption of the starting material. The product was coevaporated with toluene (2 x 100 mL) and DCM (100 mL). The TFA salt of compound **56** was obtained as an orange-brown solid.

Yield: 1.19 g (2.37 mmol, 99 %);

TLC: R_f = 0 (cyclohexane/ethyl acetate, 2:1);

Melting point: 203.5 °C (decomposition);

Rotational value: $[\alpha]_{20}^{D}$ = + 133.8 (c = 0.03 mmol · mL^{-1}, methanol);

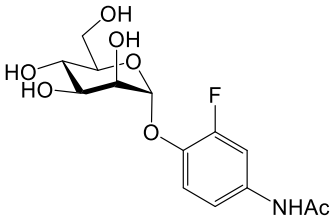
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.94 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.92 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9 Hz, 2H, Ar-$H_{2/6}$), 7.62 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 7.29 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 5.61 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.21 (s, 2H, Ar-CH_2NH_2), 4.04 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, H-2), 3.92 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{3,4}$ = 9.4 Hz, 1H, H-3), 3.79-3.71 (m, 3H, H-4, H-6, H-6'), 3.58 (ddd, $^3J_{5,4}$ = 9.8 Hz, $^3J_{5,6}$ = 5.4 Hz, $^3J_{5,6'}$ = 2.5 Hz, 1H, H-5) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 160.8 (Ar-C_1), 154.4 (Ar-$C_{1'}$), 149.1 (Ar-C_4), 136.8 (Ar-$C_{4'}$), 131 (2C, Ar-$C_{3'/5'}$), 125.8 (2C, Ar-$C_{2/6}$), 124.2 (2C, Ar-$C_{2'/6'}$), 118.1 (2C, Ar-$C_{3/5}$), 100.1 (C-1), 75.8 (C-5), 72.4 (C-3), 71.9 (C-2), 68.3 (C-4), 62.7 (C-6), 44 (Ar-$\underline{C}H_2NH_2$) ppm;

IR (ATR): $\tilde{\nu}$ = 3379 (br), 2937 (m), 2370 (w), 2348 (w), 2326 (w), 2299 (w), 1676 (s), 1599 (m), 1498 (m), 1425 (w), 1382 (w), 1235 (s), 1293 (s), 1139 (s), 1014 (s), 843 (m), 724 (m), 687 (w), 561 (w), 427 (w) cm^{-1};

ESI-HRMS: m/z = 390.16507 $[M + H]^+$, (calc. 389.158687 for $C_{19}H_{23}N_3O_6 + H^+$).

2-Fluoro-4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside (57)



2-Fluoro-4-(*N*-acetylamino)phenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside **68** (190 mg, 0.38 mmol) was dissolved in methanol (15 mL) and a solution of sodium methanolate (20 µL, 5.4 M) was added. The reaction was stirred for 1 h under ambient conditions until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed from the filtrate under reduced pressure. The residue was purified by flash column chromatography on silica gel (ethyl acetate/methanol/dH_2O, 20:2:1). The deprotected 2-fluoro-4-(*N*-acetylamino)phenyl mannoside **57** was obtained as a white solid.

Yield: 124 mg (0.37 mmol, 98 %);

TLC: R_f = 0 (cyclohexane/ethyl acetate, 1:4),
0.28 (ethyl acetate/methanol/dH_2O, 20:2:1);

Melting point: 118.2 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 134.6 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.55 (dd, $^3J_{Ar\text{-}H3,F}$ = 13.1 Hz, $^4J_{Ar\text{-}H3,Ar\text{-}H5}$ = 2.5 Hz, 1H, Ar-H_3), 7.28 (t, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.9 Hz, 1H, Ar-H_6), 7.14 (ddd, $^3J_{Ar\text{-}H6,Ar\text{-}H5}$ = 8.9 Hz, $^4J_{Ar\text{-}H5,Ar\text{-}H3}$ = 2.5 Hz, $^5J_{Ar\text{-}H5,F}$ = 1.5 Hz, 1H, Ar-H_5), 5.39 (d, $^3J_{1.2}$ = 1.8 Hz, 1H, H-1), 4.05 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.89 (dd, $^3J_{3,2}$ = 3.4 Hz, $^3J_{3,4}$ = 8.8 Hz, 1H, H-3), 3.79-3.69 (m, 4H, H-4, H-5, H-6, H-6'), 2.10 (s, 3H, NHCOC$\underline{H}_3$) ppm;

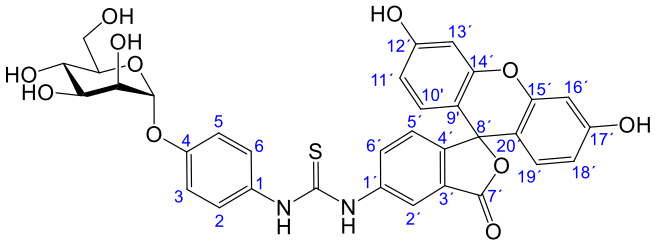
^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 171.6 (NH$\underline{C}OCH_3$), 154.4 (d, ^{1} $_{Ar\text{-}C2,F}$ = 243.3 Hz, Ar-C2), 141.6 (d, $^2J_{Ar\text{-}C1,F}$ = 11.4 Hz, Ar-C_1), 135.8 (d, $^3J_{Ar\text{-}C4,F}$ = 9.7 Hz, Ar-C_4), 120.8 (Ar-C_6), 116.8 (d, $^4J_{Ar\text{-}5,F}$ = 3.3 Hz, Ar-C_5), 109.8 (d, $^2J_{Ar\text{-}C3,F}$ = 23.6 Hz, Ar-C_3), 102.1 (C-1),75.9 (C-5), 73.4 (C-3), 71.9 (C-2), 68.3 (C-4), 62.7 (C-6), 23.7 (NHCO$\underline{C}H_3$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 131.97 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3305 (s, br), 2932 (w), 2851 (w), 2837 (w), 2365 (w), 2297 (w), 2012 (w), 1843 (w), 1662 (m), 1609 (m), 1550 (m), 1511 (s), 1426 (m), 1374 (m), 1274 (m), 1212 (m), 1111 (m), 1007 (s), 972 (s), 827 (m), 751 (m), 60 (m), 589 (m), 463 (m), 439 (m) cm^{-1};

ESI-HRMS: *m/z* = 349.14001 $[M + NH_4]^+$, (calc. 349.14056 for $C_{14}H_{18}FNO_7 + NH_4^+$); 376.10443 $[M + HCOO]^-$, (calc. 349.14056 for $C_{14}H_{18}FNO_7 + HCOO^-$).

Fluorescein-5-thiourea-*N*-(*p*-phenyoxy)-*α*-D-mannopyranoside (59)



4-Aminophenyl *α*-D-mannopyranoside **64** (10 mg, 36.8 µmol) and fluorescein-5-isothiocyanate (FITC isomer I) (14.3 mg, 36.8 µmol) were dissolved in carbonate buffer I (2 mL, Na_2CO_3, 1.59 g/L, $NaHCO_3$, 2.52 g/L, pH 9.5, Table 5.7-1) and stirred 16 h under ambient conditions. The aqueous phase was washed with ethyl acetate (50 mL) and the aqueous phase was concentrated to a final volume of 1 mL under reduced pressure. The crude product was purified by column chromatography on a C_{18}-reversed phase silica cartidge (dH_2O → acetonitrile). The carboxylic form of the fluoresceinyl-labeled phenyl mannoside **59** eluted first from the column and was obtained as a yellow solid. The subsequently eluted lacton form was afforded as a red solid.

Yield: total: 17.3 mg (26.19 µmol, 71 %);
lactone: 8.1 mg (12.2 µmol, 33 %),
carboxylic acid: 9.2 mg (13.9 µmol, 38 %);

TLC: R_f = 0.45 (ethyl acetate/methanol/dH_2O, 20:2:1);

Melting point: 313.8 °C (decomposition);

Rotational value: $[\alpha]_{20}^{D}$ = + 84.7 (c = ,0.0002 mmol · mL^{-1}, methanol);

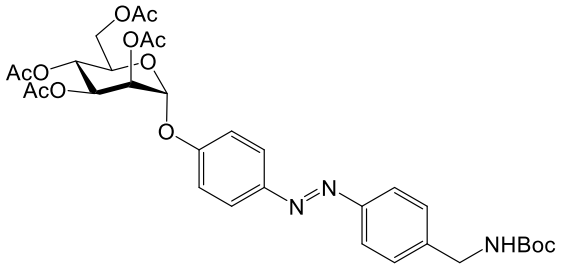
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 8.11 (d, $^4J_{Ar\text{-}H2',Ar\text{-}H6'}$= 1.6 Hz, 1H, Ar-$H_{2'}$), 7.80 (dd, $^3J_{Ar\text{-}H6',Ar\text{-}H5'}$ = 8.2 Hz, $^4J_{Ar\text{-}H6',Ar\text{-}H2'}$= 1.6 Hz, 1H, Ar-$H_{6'}$), 7.36 (d, $^3J_{Ar\text{-}H5/3,Ar\text{-}H2/6}$ = 9.2 Hz, 2H, Ar-$H_{5/3}$), 7.18-7.14 (m, 3H, Ar-$H_{5'}$, Ar-$H_{2/6}$), 6.76 (d, $^3J_{Ar\text{-}H10'/19',Ar\text{-}H11'/18'}$ = 8.7, 2H, Ar-$H_{10'/19'}$), 6.69 (d, $^4J_{Ar\text{-}H13'/16',Ar\text{-}H11'/18'}$ = 2.3 Hz, 2H, Ar-$H_{13'/16'}$), 6.58 (dd, $^3J_{Ar\text{-}H11'/18',Ar\text{-}H10'/19'}$ = 8.7 Hz, $^4J_{Ar\text{-}H11'/18',Ar\text{-}H13'/16'}$ = 2.3Hz, 2H, Ar-$H_{11'/18'}$), 5.50 (d, $^3J_{1.2}$ = 1.7 Hz, 1H, H-1), 4.03 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 3.92 (dd, $^3J_{3,2}$ = 3.4 Hz, $^3J_{3,4}$ = 9.4 Hz, 1H, H-3), 3.979-3.70 (m, 3H, H-4, H-6, H-6'), 3.62 (ddd, $^3J_{5,4}$ = 9.7 Hz, $^3J_{5,6}$ = 5.1 Hz, $^3J_{5,6'}$ = 2.5 Hz, 1H, H-5) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 182.5 (NHCSNH), 171.7 ($C_{7'}$), 156.1 (Ar-C_4), 155 (2C, Ar-$C_{9'/20'}$), 142.5 (Ar-$C_{4'}$), 134.2 (Ar-C_1), 132.2 (Ar-$C_{6'}$), 131.1 (Ar-$C_{1'}$), 130.8 (2C, Ar-$C_{10'/19'}$), 127.9 (4C, Ar-$C_{14'/15'}$, Ar-$C_{3/5}$), 126.4 (Ar-$C_{5'}$), 122.6 (Ar-$C_{3'}$), 122.2 (Ar-$C_{2'}$), 118.4 (2C, Ar-$C_{2/6}$), 115.1 (2C, Ar-$C_{11'/18'}$), 115.0 (Ar-$C_{8'}$), 112.3 (2C, Ar-$H_{12'/17'}$), 103.7 (2C, Ar-$C_{13'/16'}$), 100.3 (C-1), 175.4 (C-5), 72.4 (C-3), 72 (C-2), 68.3 (C-4), 62.6 (C-6) ppm;

IR (ATR): $\tilde{\nu}$ = 3351 (br), 2925 (m), 2359 (w), 2167 (w), 2123 (w), 1997 (w), 1653 (m), 1576 (s), 1507 (s), 14666 (s), 1389 (m), 1327 (m), 1212 (s), 1111 (m), 1050 (m), 850 (w), 750 (m), 602 (m), 480 (w), 459 (w) cm^{-1};

ESI-HRMS: *m/z* = 661.14759 $[M + H]^+$, (calc. 661.14866 for $C_{33}H_{28}N_2O_{11}S + H^+$).

[4-(4'-{[(*tert*-Butoxycarbonyl)amino]methyl}phenyl)azophenoxy]-2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside[458] (66)



tert-Butyl-(4-aminobenzyl)carbamate **65** (2.21 g, 9.94 mmol) was dissolved in DCM (85 mL) under a nitrogen atmosphere and a solution of potassium monopersulfate triple salt (Oxone®, Sigma-Aldrich, 9.14 g, 14.92 mmol) in dH_2O (85 mL) was added. The reaction mixture was stirred vigorously at room temperature for 16 h. The phases were separated and the aqueous layer was extracted with DCM (2 x 100 mL). The combined organic layers were sequentially washed with 1 N HCl_{aq} sat. solution of $NaHCO_3$ and dH_2O (100 mL each), dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The resulting intermediate, *tert*-butyl-(4-nitrosobenzyl)carbamate, was obtained as a green solid and was directly dissolved in a mixture of DMSO and glacial acetic acid (50 mL each). 4-Aminophenyl *α*-D-mannopyranoside **64** (1.41 g, 3.21 mmol) was added and the reaction mixture was stirred under ambient conditions for 17 h. Subsequently, dH_2O (350 mL) was added and the crude product was filtered off over Celite® 545. The precipitate was washed from the filter material with ethyl acetate (250 mL). The organic layer was dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. Purification by flash column chromatography on silica gel (cyclohexane/ethyl acetate, 2:1) afforded the title compound **66** as an orange solid.

Yield: 1.79 g (2.72 mmol, 85 %);

TLC: R_f = 0.15 (cyclohexane/ethyl acetate, 2:1);

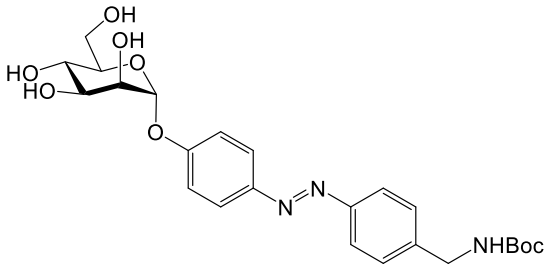
^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.91 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9.1 Hz, 2H, Ar-$H_{2/6}$), 7.86 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 8.5 Hz, 2H, Ar-$H_{2'/6'}$), 7.42 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 8.5 Hz, 2H, Ar-$H_{3'/5'}$), 7.22 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9.1 Hz, 2H, Ar-$H_{3/5}$), 5.62 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, H-1), 5.58 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{3,4}$ = 10 Hz, 1H, H-3), 5.48 (dd, $^3J_{1,2}$ = 1.7 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, H-2), 5.39 (dd~t, $^3J_{4,5}$ = 9.5 Hz, $^3J_{4,3}$ = 9.5 Hz, 1H, H-4), 4.92 (br. s, 1H, NH), 4.40 (d, $^3J_{Ar\text{-}CH2NH,NH}$ = 5.2 Hz, 2H, Ar-C$\underline{H}_2$NH), 4.32-4.27 (m, 1H, H-6), 4.12-4.07 (m, 2H, H-5, H-6'), 2.22 (s, 3H, COC$\underline{H}_3$), 2.06 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.04 (s, 3H, COC$\underline{H}_3$), 1.48 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR ($CDCl_3$, 126 MHz, 298 K): δ = 170.5 (2C, 2x $\underline{C}OCH_3$), 169.96 ($\underline{C}OCH_3$), 169.94 ($\underline{C}O_2C(CH_3)_3$), 169.7 ($\underline{C}OCH_3$), 157.6 (Ar-$C_{4'}$), 151.9 (Ar-$C_1$), 148.4.2 (Ar-$C_{1'}$), 141.8 (Ar-$C_4$), 128.0 (2C, Ar-$C_{3'/5'}$), 124.6 (2C, Ar-$C_{2/6}$), 123.0 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3/5}$), 95.7 (C-1), 79.9 ($CO_2\underline{C}(CH_3)_3$), 69.4 (C-5), 69.3 (C-3), 68.8 (C-2), 65.8 (C-4), 62.1 (C-6), 44.4 (Ar-$\underline{C}H_2NH_2$), 28.4 ($CO_2C(\underline{C}H_3)_3$), 20.9 ($CO\underline{C}H_3$), 20.7 (3C, 3 x $CO\underline{C}H_3$) ppm;

ESI-HRMS: *m/z* = 658.26077 $[M + H]^+$, (calc. 658.26065 for $C_{32}H_{39}N_3O_{12} + H^+$).

Further analytical data corresponds to the literature.

(4-{4'-[(*tert*-Butoxycarbonylamino)methyl]phenyl}azophenoxy)-*a*-D-mannopyranoside[459] (67)



The acetyl-protected azobenzene mannoside derivative **66** (1.75 g, 2.66 mmol) was dissolved in methanol (25 mL) and sodium (3.5 mg, 0.15 mmol) was added. The reaction mixture was stirred for 1.6 h at room temperature, until TLC indicated full consumption of the starting material. The mixture was neutralised with Amberlite® IR 120, the resin was filtered off and the solvent was removed under reduced pressure. The deacetylated compound **67** was obtained as an orange solid.

Yield: 1.21 g (2.48 mmol, 93 %);

TLC: R_f =0 (cyclohexane/ethyl acetate, 2:1);

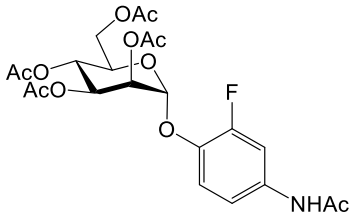
^{1}H-NMR (CD_3OD, 500 MHz, 298 K): δ = 7.89 (d, $^3J_{Ar\text{-}H2/6,Ar\text{-}H3/5}$ = 9 Hz, 2H, Ar-$H_{2/6}$), 7.83 (d, $^3J_{Ar\text{-}H2'/6',Ar\text{-}H3'/5'}$ = 9 Hz, 2H, Ar-$H_{2'/6'}$), 7.43 (d, $^3J_{Ar\text{-}H3'/5',Ar\text{-}H2'/6'}$ = 9 Hz, 2H, Ar-$H_{3'/5'}$), 7.28 (d, $^3J_{Ar\text{-}H3/5,Ar\text{-}H2/6}$ = 9 Hz, 2H, Ar-$H_{3/5}$), 5.60 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 4.31 (s, 2H, Ar-C$\underline{H}_2$$NH_2$), 4.04 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.4 Hz, 1H, H-2), 3.92 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{3,4}$ = 9.4 Hz, 1H, H-3), 3.79-3.69 (m, 3H, H-4, H-6, H-6'), 3.59 (ddd $^3J_{5,4}$ = 9.6 Hz, $^3J_{5,6}$ = 5.3 Hz, $^3J_{5,6'}$ = 2.4 Hz, 1H, H-5), 1.47 (s, 9H, $CO_2C(C\underline{H}_3)_3$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 160.4 ($\underline{C}O_2C(CH_3)_3$), 153.2 (Ar-$C_{4'}$),149.3 (2C, Ar-$C_1$, Ar-$C_{1'}$), 141.8 (Ar-$C_4$), 128.0 (2C, Ar-$C_{3'/5'}$), 124.6 (2C, Ar-$C_{2/6}$), 122.9 (2C, Ar-$C_{2'/6'}$), 116.7 (2C, Ar-$C_{3/5}$), 100.1 (C-1), 80.4 ($CO_2\underline{C}(CH_3)_3$), 75.7 (C-5), 72.4 (C-3), 71.9 (C-2), 68.3 (C-4), 62.7 (C-6), 44.8 (Ar-$\underline{C}H_2NH_2$), 28.8 ($CO_2C(\underline{C}H_3)_3$) ppm;

ESI-HRMS: m/z = 490.21750 $[M + H]^+$, (calc. 490.21839 for $C_{24}H_{31}N_3O_8 + H^+$).

Further analytical data corresponds to the literature.

2-Fluoro-4-(*N*-acetylamino)phenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside (68)



2-Fluoro-4-aminophenyl 2,3,4,6-tetra-*O*-acetyl-*α*-D-mannopyranoside **8** (229 mg, 0.5 mmol) was dissolved in pyridine (10 mL) and acetic anhydride (2.5 mL) was added. The reaction mixture was stirred 18 h under ambient conditions. Ethyl acetate (50 mL) was added and the organic layer was washed with dH_2O (100 mL) and 1 N HCl_{aq} (100 mL). The aqueous layer was extracted with ethyl acetate (100 mL) and the combined organic layers were dried over magnesium sulfate, filtered and the solvent was removed under reduced pressure. The residue was purified by column chromatography on silica gel (cyclohexane/ethyl acetate, 1:3) to afford the peracetylated 2-fluoro-4-aminophenyl mannoside **68** as a white solid.

Yield: 235 mg (0.47 mmol, 94 %);

TLC: R_f = 0.3 (cyclohexane/ethyl acetate, 1:4);

Melting point: 136.7 °C;

Rotational value: $[\alpha]_{20}^{D}$ = + 73.5 (c = 0.02 mmol · mL^{-1}, methanol);

^{1}H-NMR ($CDCl_3$, 500 MHz, 298 K): δ = 7.54-7.50 (m, 1H, Ar-H_3), 6.48 (s, 1H, Ar-H_5), 6.13-6.06 (m, 1H, Ar-H_6), 5.55 (dd, $^3J_{3,4}$ = 10 Hz, $^3J_{3,2}$ = 3.5 Hz, 1H, H-3), 5.51 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{2,1}$ = 1.8 Hz, 1H, H-2), 5.41 (d, $^3J_{1,2}$ = 1.8 Hz, 1H, H-1), 5.36 (dd∼t, $^3J_{4,3}$ = 10 Hz, $^3J_{4,5}$ = 10 Hz, 1H, H-4), 4.29-4.23 (m, 2H, H-5, H-6), 4.12-4.08 (m, 1H, H-6'), 2.19 (s, 3H, COC$\underline{H}_3$), 2.17 (s, 3H, NHCOC$\underline{H}_3$) 2.07 (s, 3H, COC$\underline{H}_3$), 2.05 (s, 3H, COC$\underline{H}_3$), 2.03 (s, 3H, COC$\underline{H}_3$) ppm;

^{13}C-NMR (CD_3OD, 126 MHz, 298 K): δ = 172.3 ($\underline{C}OCH_3$), 171.63 ($\underline{C}OCH_3$), 171.55 (2C, 2x $\underline{C}OCH_3$) 171.5 (NH$\underline{C}OCH_3$), 154.6 (d, $^1J_{Ar\text{-}C2,F}$ = 244 Hz, Ar-C_2), 140.9 (d, $^2J_{Ar\text{-}C1,F}$ = 10.6 Hz, Ar-C_1), 136.7 (Ar-C_4), 121.1 (Ar-C_5), 116.8 (Ar-C_6), 109.8 (d, $^2J_{Ar\text{-}C3,F}$ = 23.7 Hz,-Ar-C_3), 99.1 (C-1), 71.1 (C-5), 70.5 (C-2), 70.3 (C-3), 67 (C-4), 63.4 (C-6), 30.8 (NHCO$\underline{C}H_3$), 23.8 (CO$\underline{C}H_3$), 20.6 (3C, 3 x CO$\underline{C}H_3$) ppm;

^{19}F-NMR (CD_3OD, 471 MHz, 298 K): δ = – 131.2 (s, Ar-F) ppm;

IR (ATR): $\tilde{\nu}$ = 3749 (w), 3359 (w), 2957 (w), 2926 (w), 2852 (w), 2385 (w), 2297 (w), 2193 (w), 1747 (s), 1674 (m), 1608 (m), 1541 (m), 1211 (s), 1115 (m), 1027 (m), 977 (m), 818 (m), 690 (m), 532 (w), 447 (w) cm^{-1};

ESI-HRMS: *m/z* = 522.13743 $[M + Na]^+$, (calc. 522.13821 for $C_{22}H_{26}FNO_{11} + Na^+$).

5.4 ^{1}H and ^{13}C NMR spectra of synthesised compounds

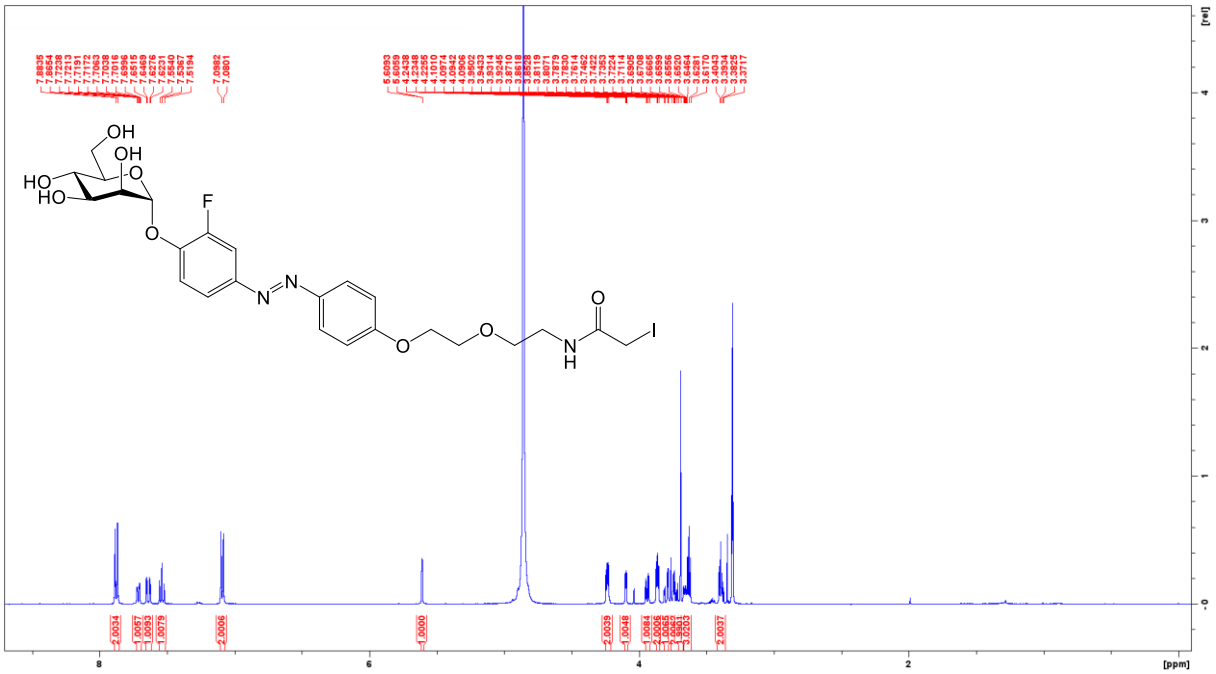


Figure 5.4-1: ^{1}H NMR spectrum of compound **F1** (500 MHz, CD_3OD, 298 K).

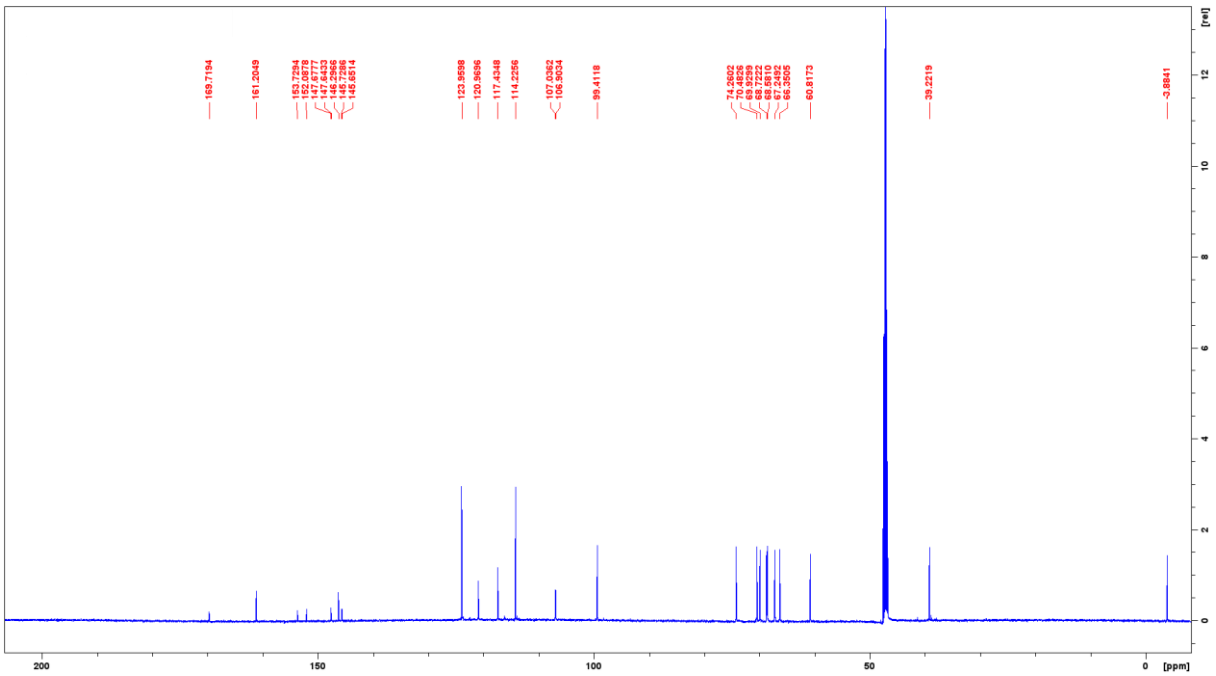


Figure 5.4-2: ^{13}C NMR spectrum of compound **F1** (126 MHz, CD_3OD, 298 K).

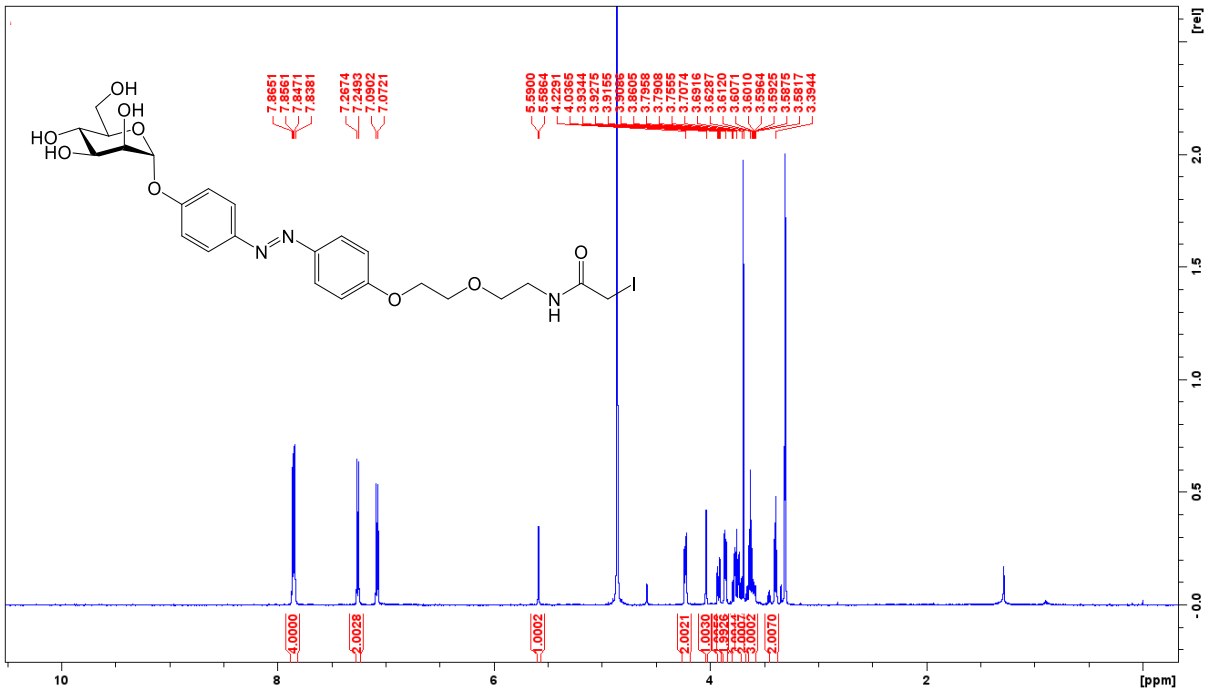


Figure 5.4-3: ^{1}H NMR spectrum of compound **H1** (500 MHz, CD_3OD, 298 K).

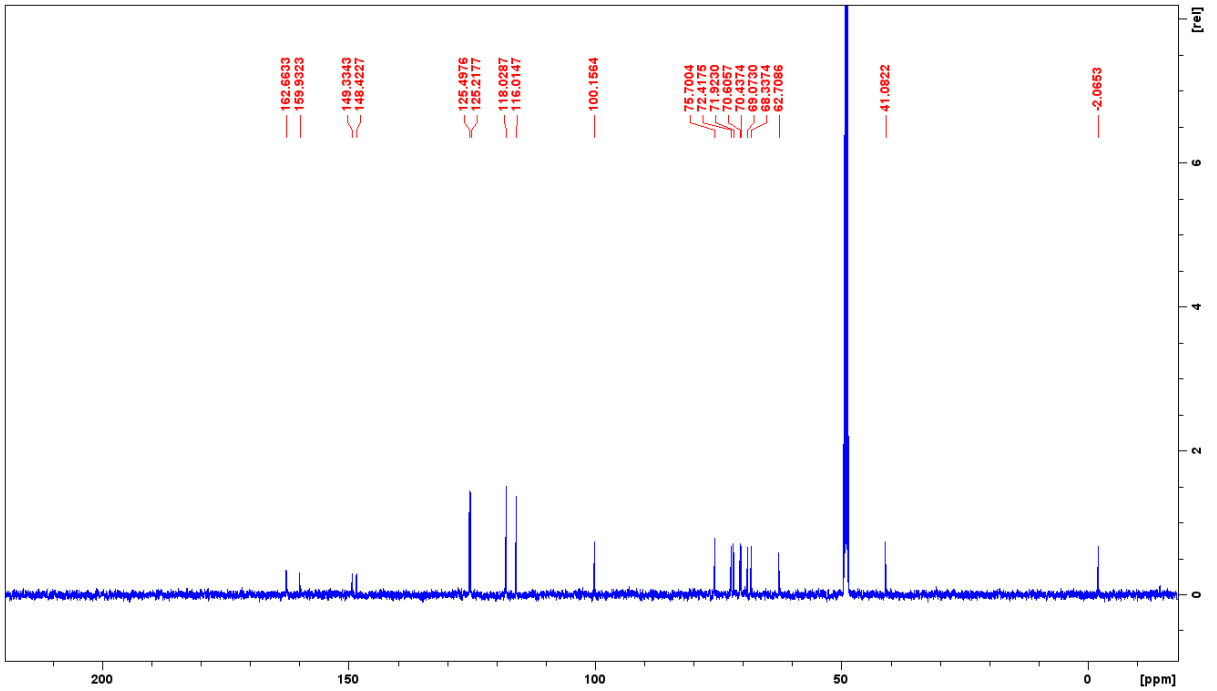


Figure 5.4-4: ^{13}C NMR spectrum of compound **H1** (126 MHz, CD_3OD, 298 K).

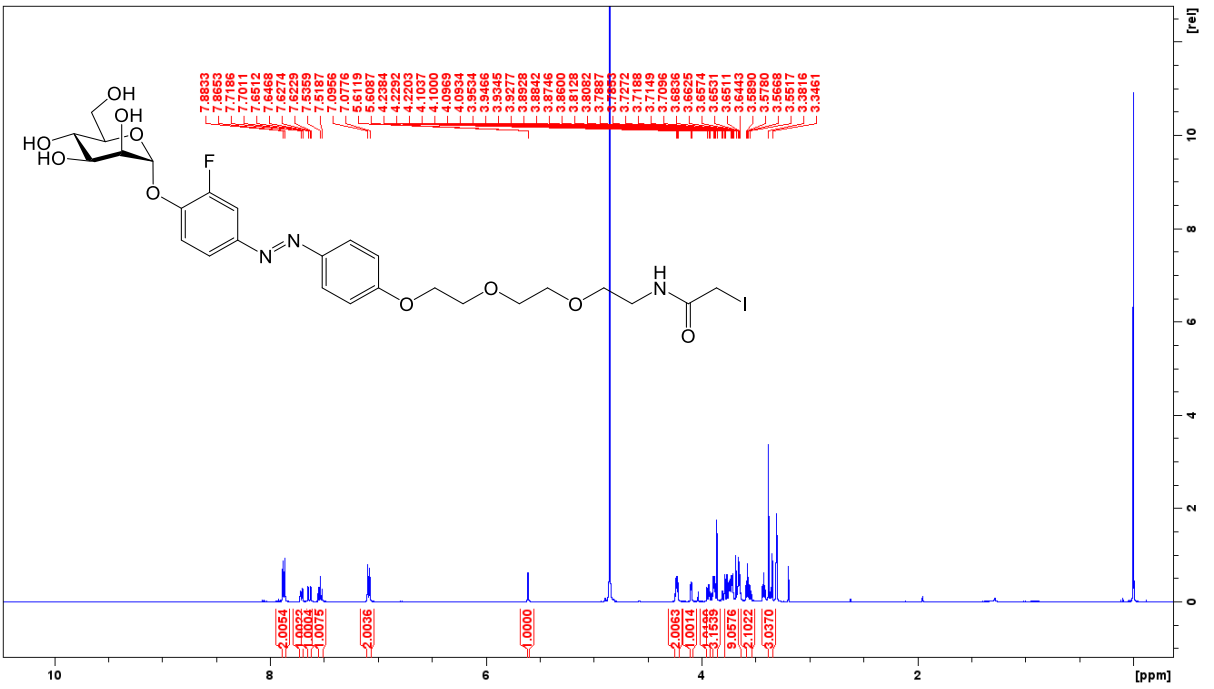


Figure 5.4-5: ^{1}H NMR spectrum of compound **F2** (500 MHz, CD_3OD, 298 K).

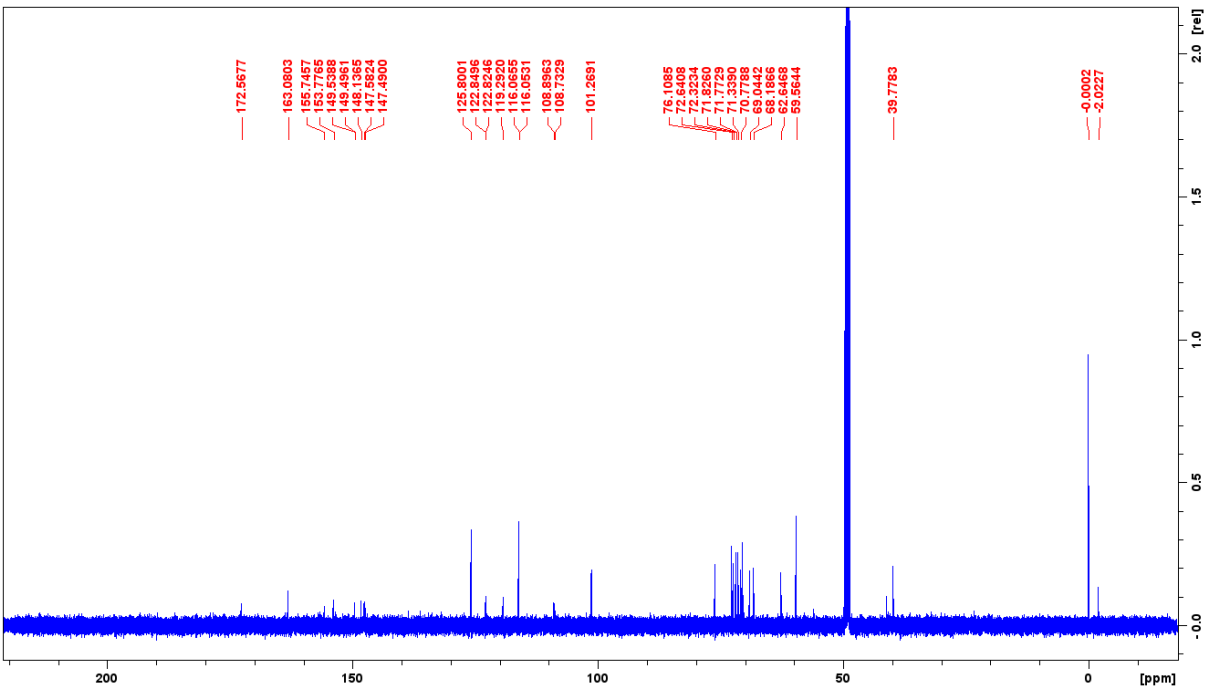


Figure 5.4-6: ^{13}C NMR spectrum of compound **F2** (126 MHz, CD_3OD, 298 K).

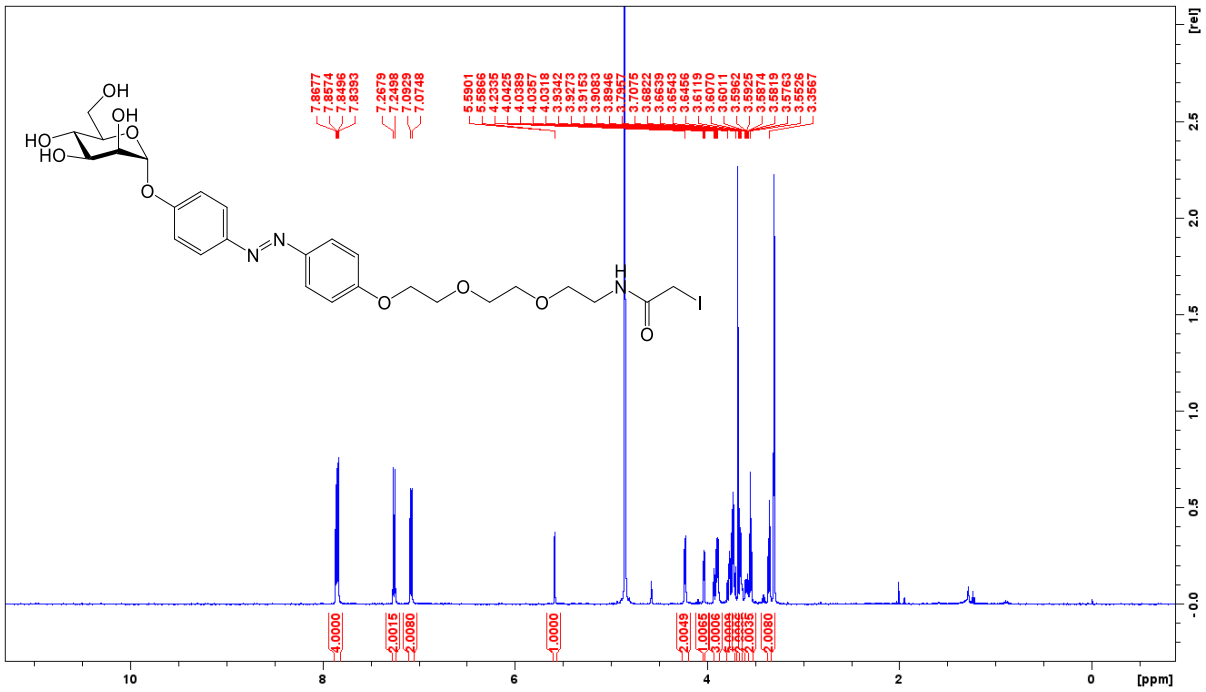


Figure 5.4-7: ^{1}H NMR spectrum of compound **H2** (500 MHz, CD_3OD, 298 K).

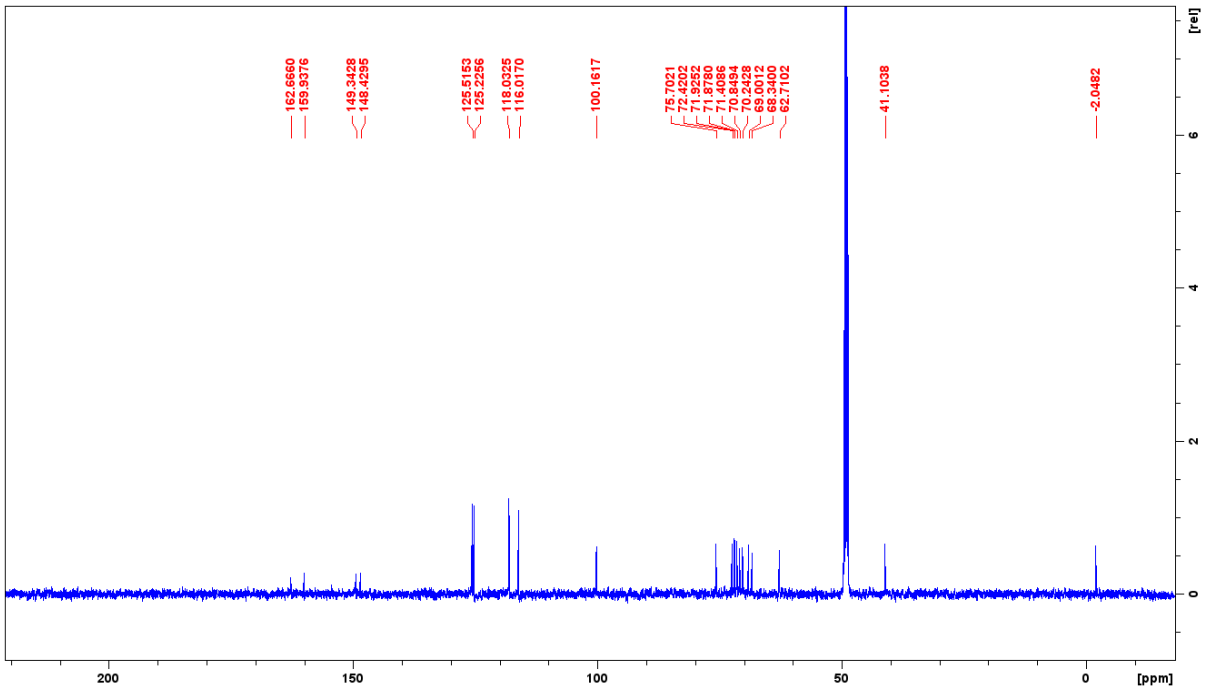


Figure 5.4-8: ^{13}C NMR spectrum of compound **H2** (126 MHz, CD_3OD, 298 K).

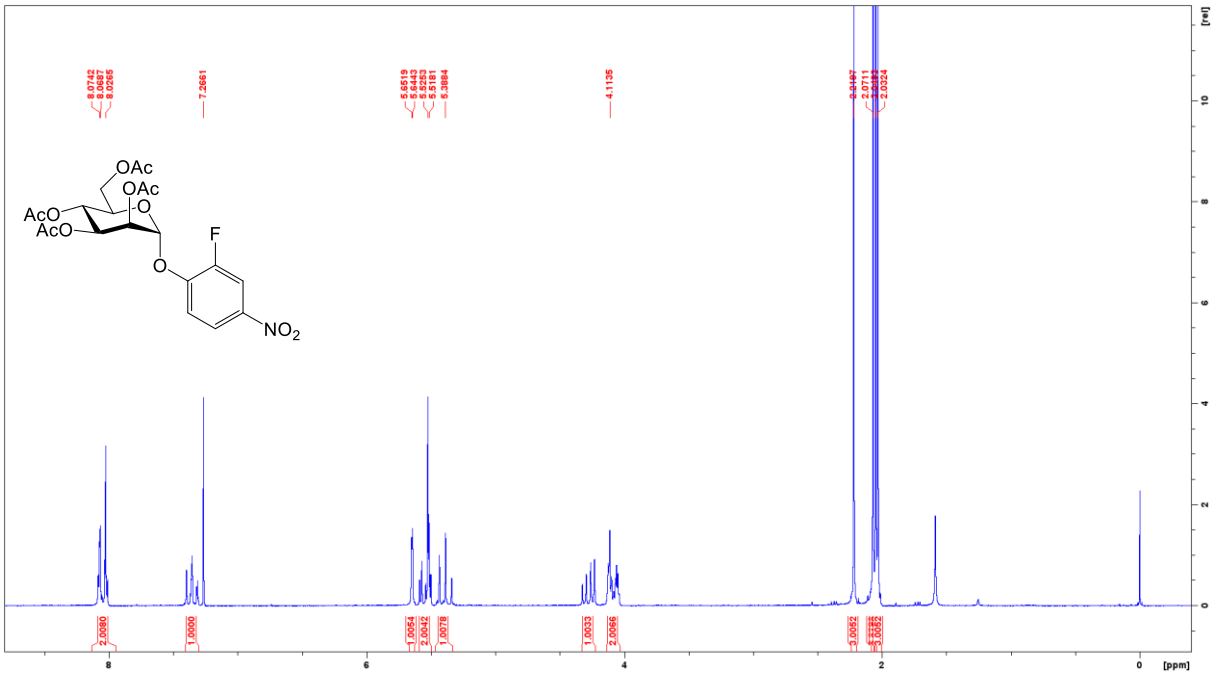


Figure 5.4-9: ^{1}H NMR spectrum of compound **7** (200 MHz, $CDCl_3$, 298 K).

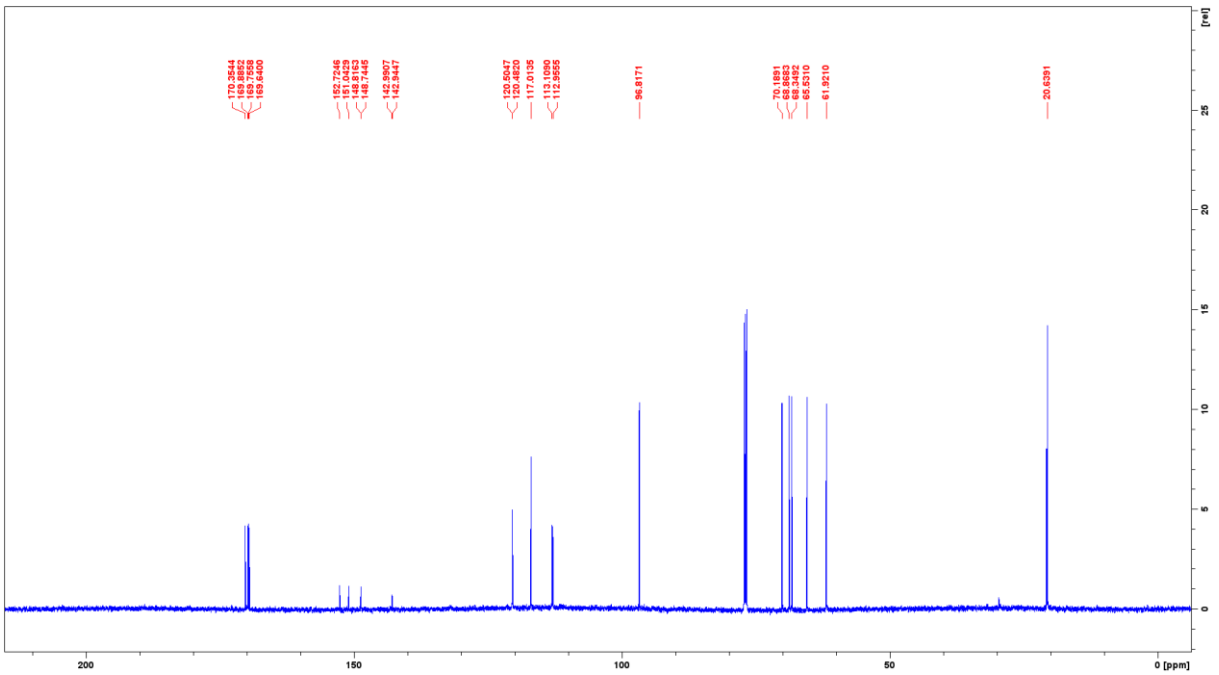


Figure 5.4-10: ^{13}C NMR spectrum of compound **7** (126 MHz, $CDCl_3$, 298 K).

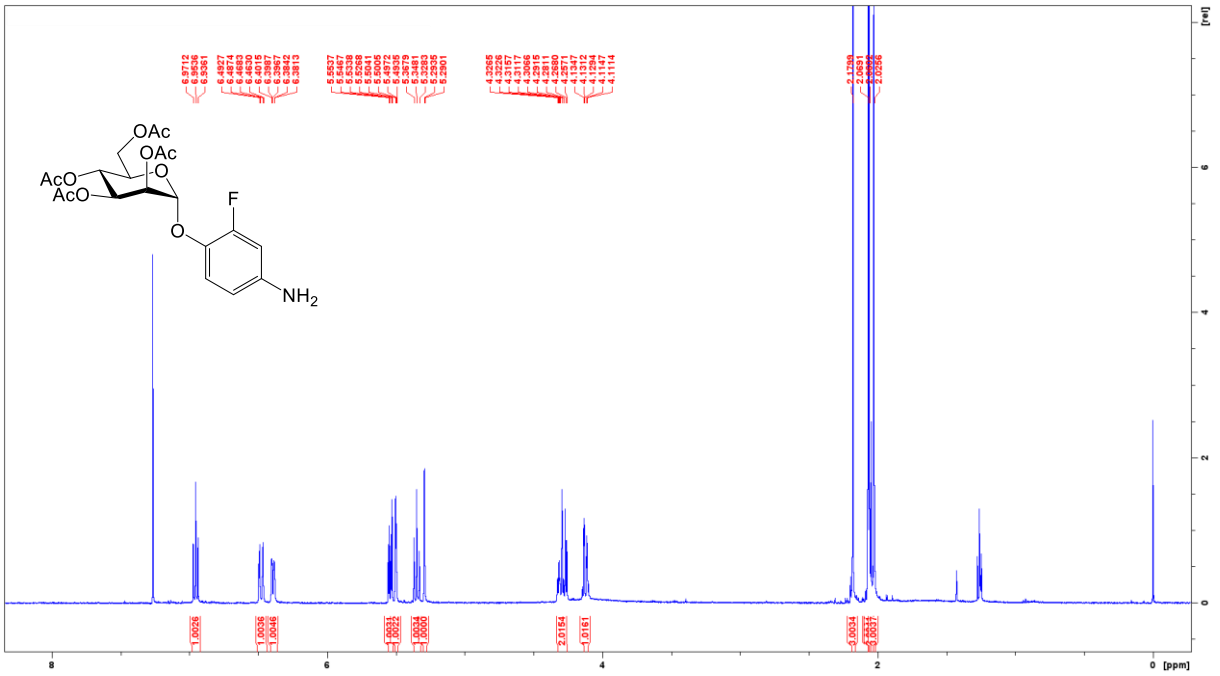


Figure 5.4-11: ^{1}H NMR spectrum of compound **8** (500 MHz, $CDCl_3$, 298 K).

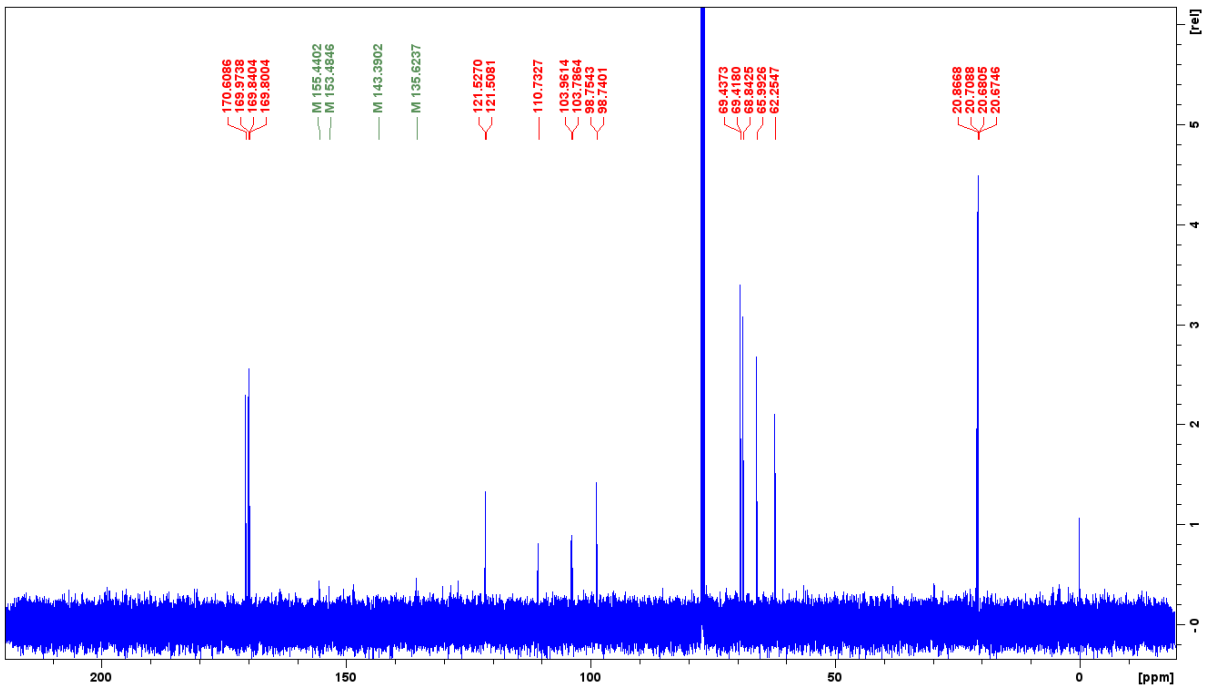


Figure 5.4-12: ^{13}C NMR spectrum of compound **8** (126 MHz, $CDCl_3$, 298 K).

The manually amended carbon signal assignments (green) were allocated using the respective HMBC spectrum of compound **8** (Figure 5.4-13).

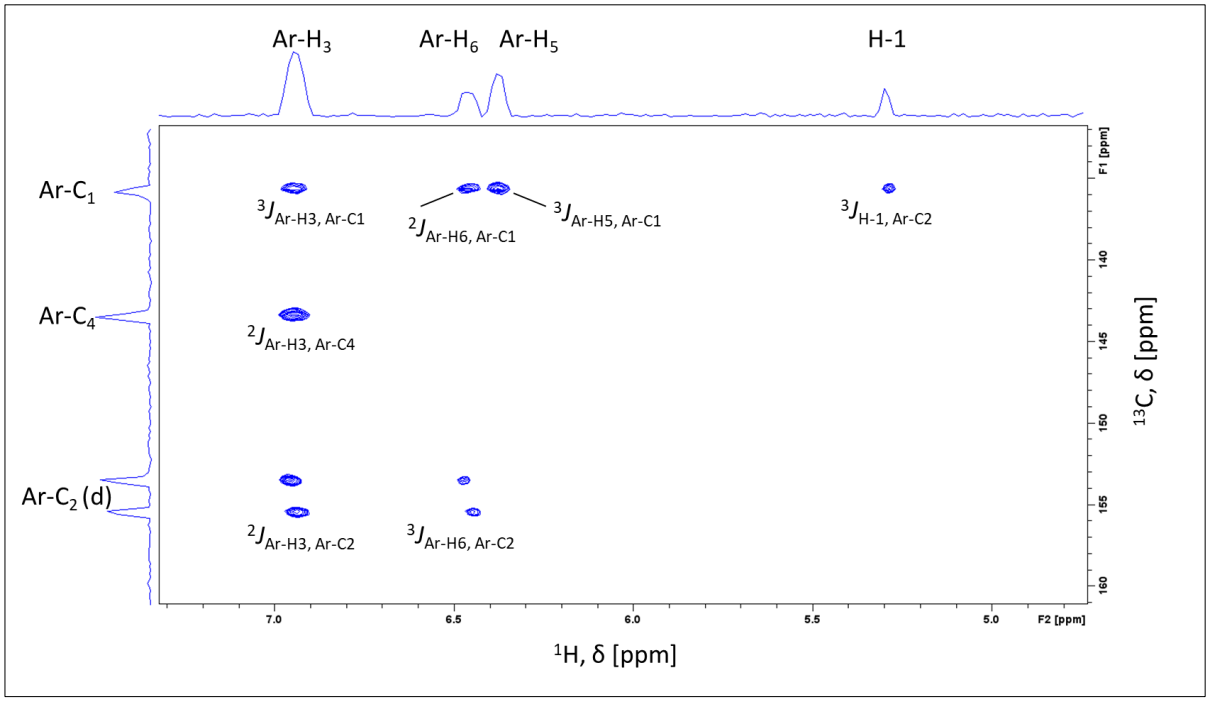


Figure 5.4-13: Image section from the HMBC spectrum of compound **8** (500 MHz, $CDCl_3$, 298 K).

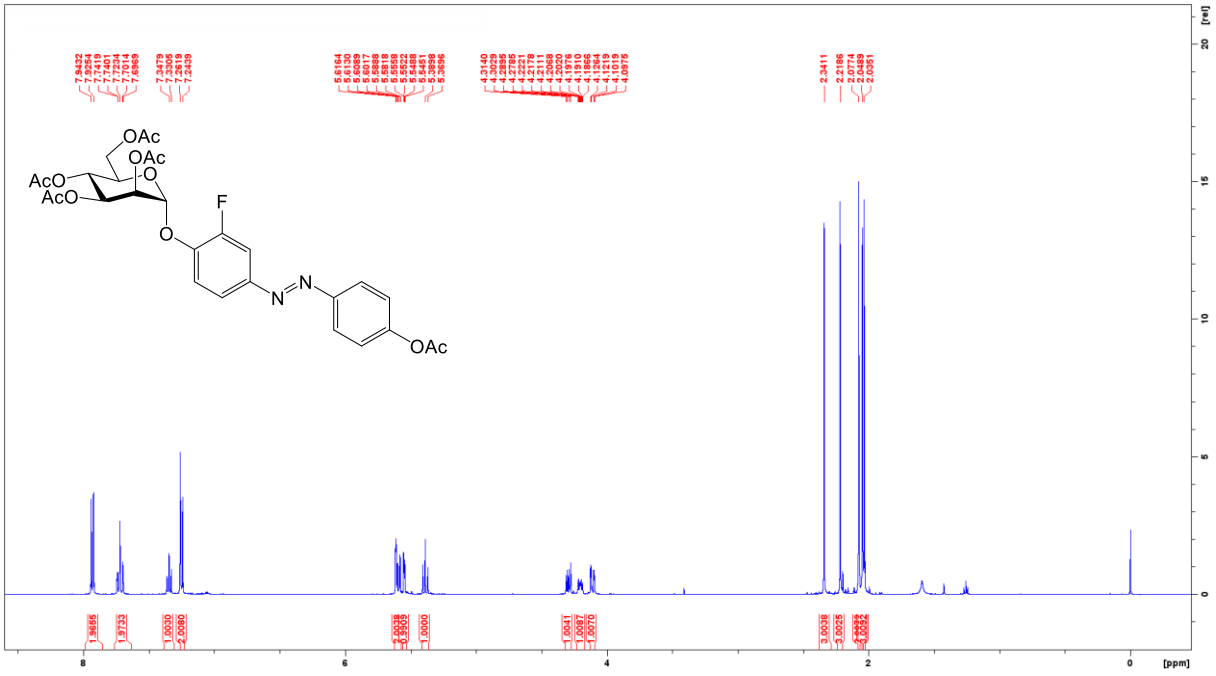


Figure 5.4-14: ^{1}H NMR spectrum of compound **9** (500 MHz, $CDCl_3$, 298 K).

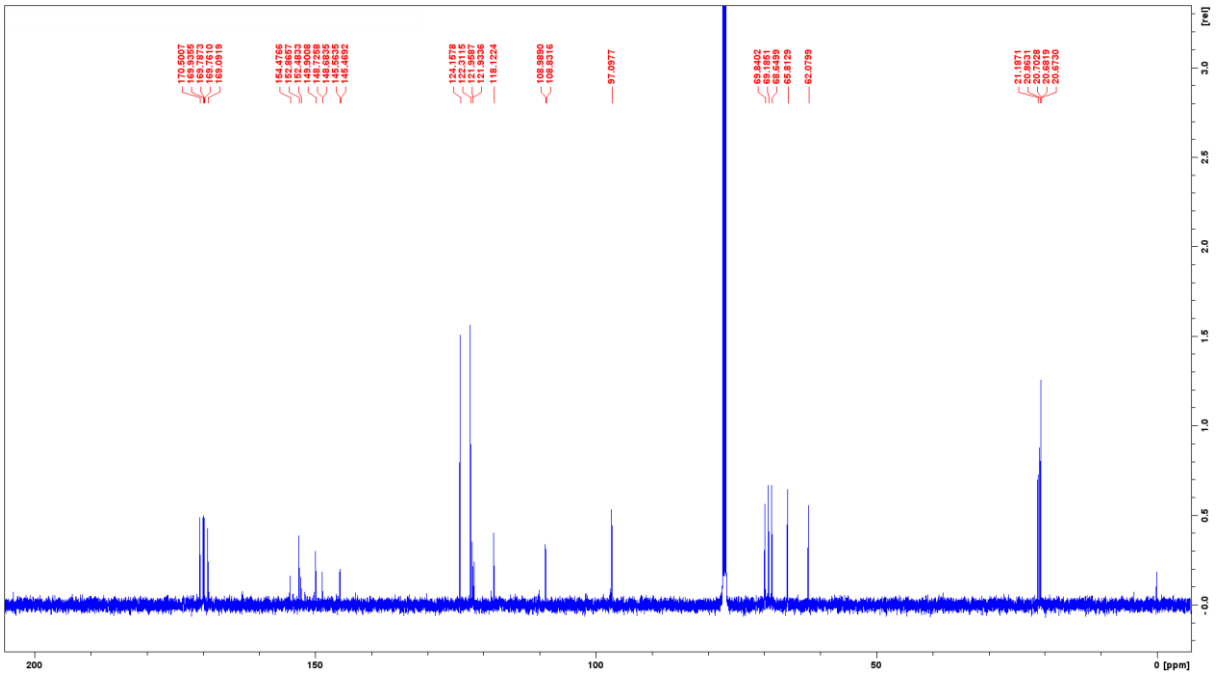


Figure 5.4-15: ^{13}C NMR spectrum of compound **9** (126 MHz, $CDCl_3$, 298 K).

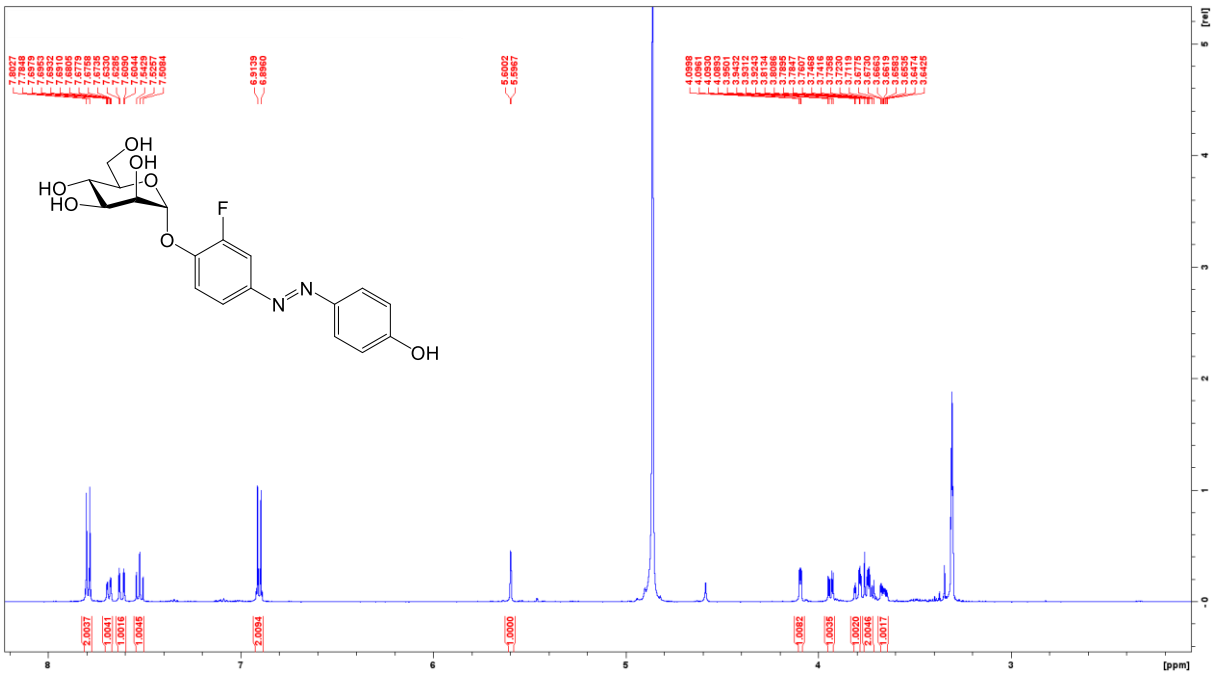


Figure 5.4-16: ^{1}H NMR spectrum of compound **10** (500 MHz, CD_3OD, 298 K).

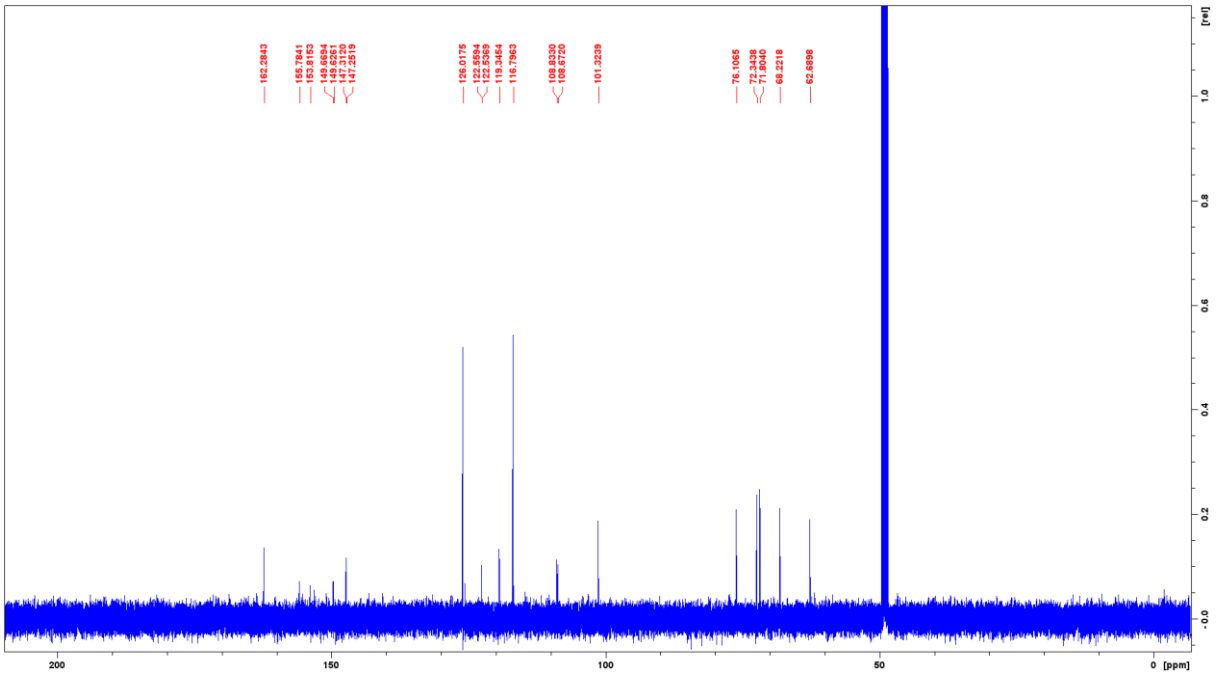


Figure 5.4-17: ^{13}C NMR spectrum of compound **10** (126 MHz, CD_3OD, 298 K).

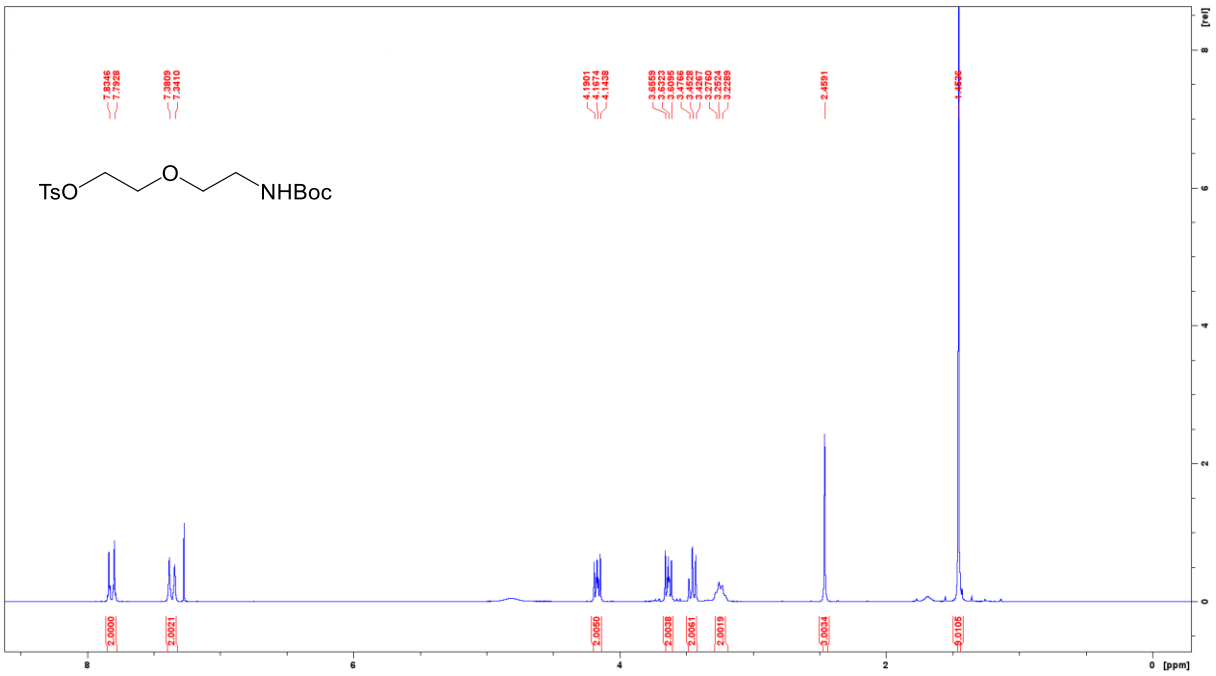


Figure 5.4-18: ^{1}H NMR spectrum of compound **11** (200 MHz, $CDCl_3$, 298 K).

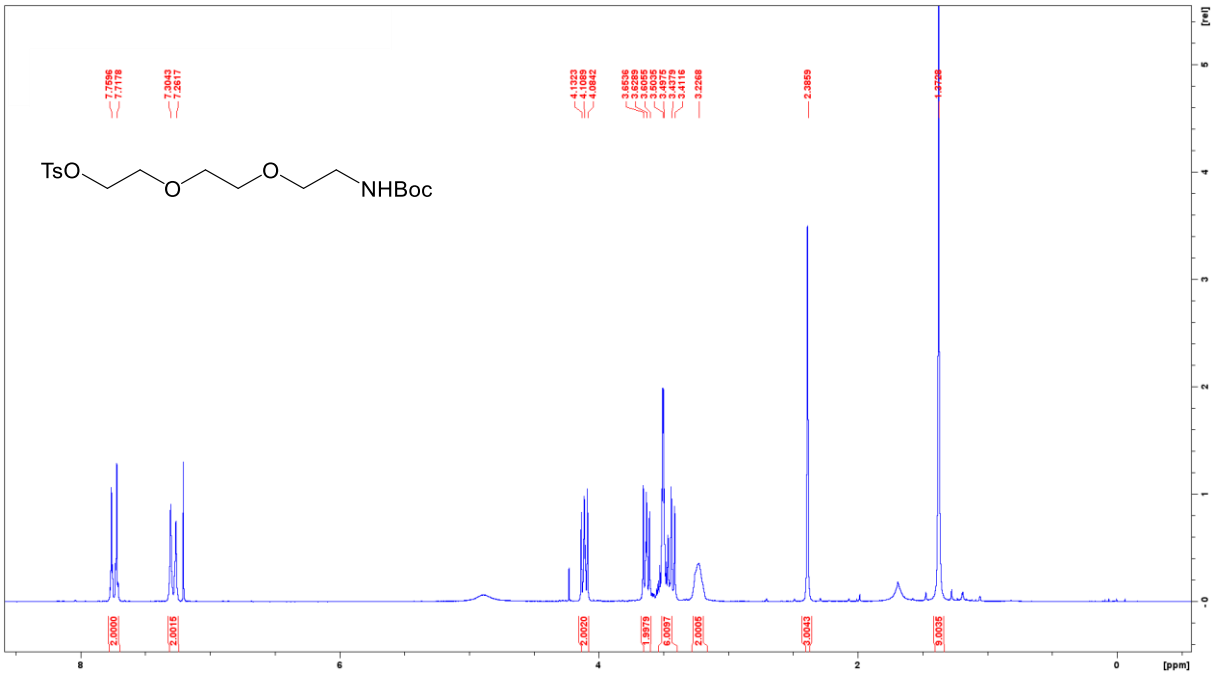


Figure 5.4-19: ^{1}H NMR spectrum of compound **12** (200 MHz, $CDCl_3$, 298 K).

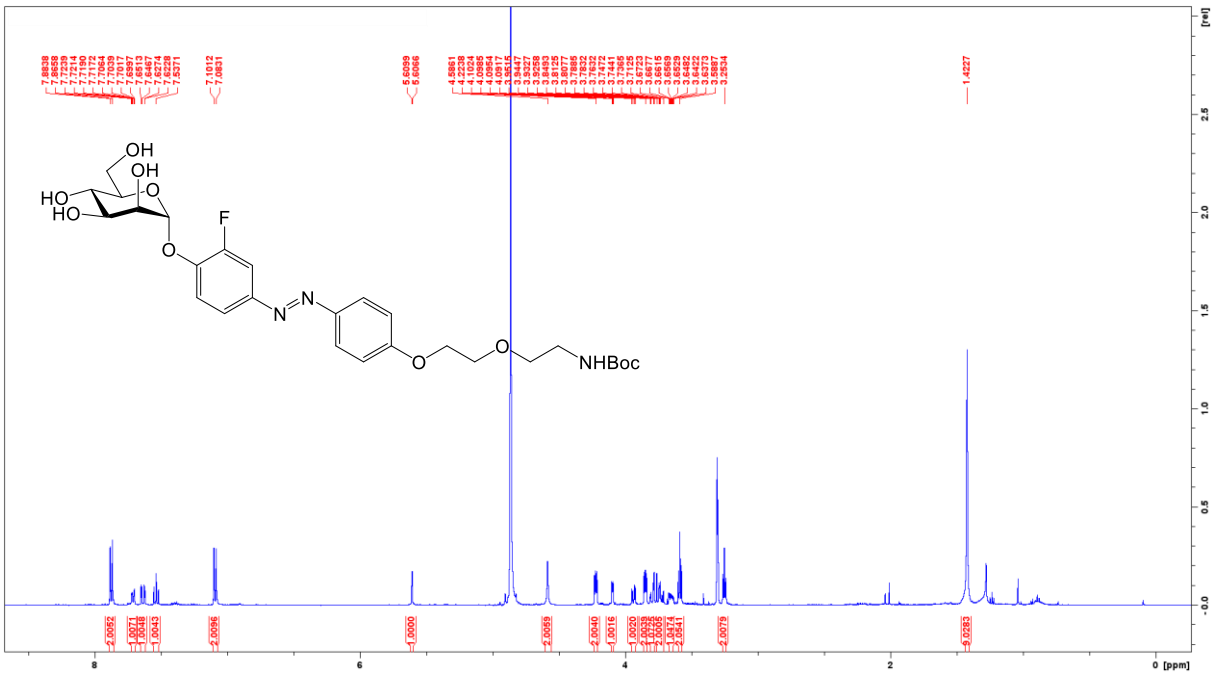


Figure 5.4-20: ^{1}H NMR spectrum of compound **13** (500 MHz, CD_3OD, 298 K).

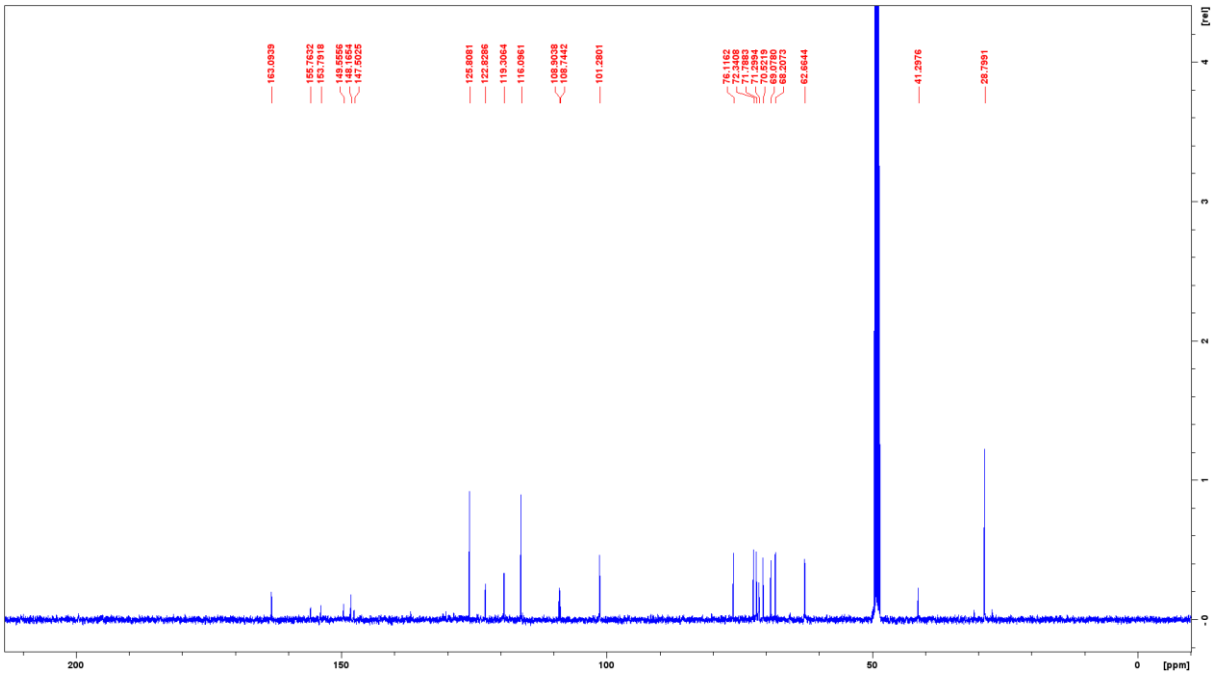


Figure 5.4-21: ^{13}C NMR spectrum of compound **13** (126 MHz, CD_3OD, 298 K).

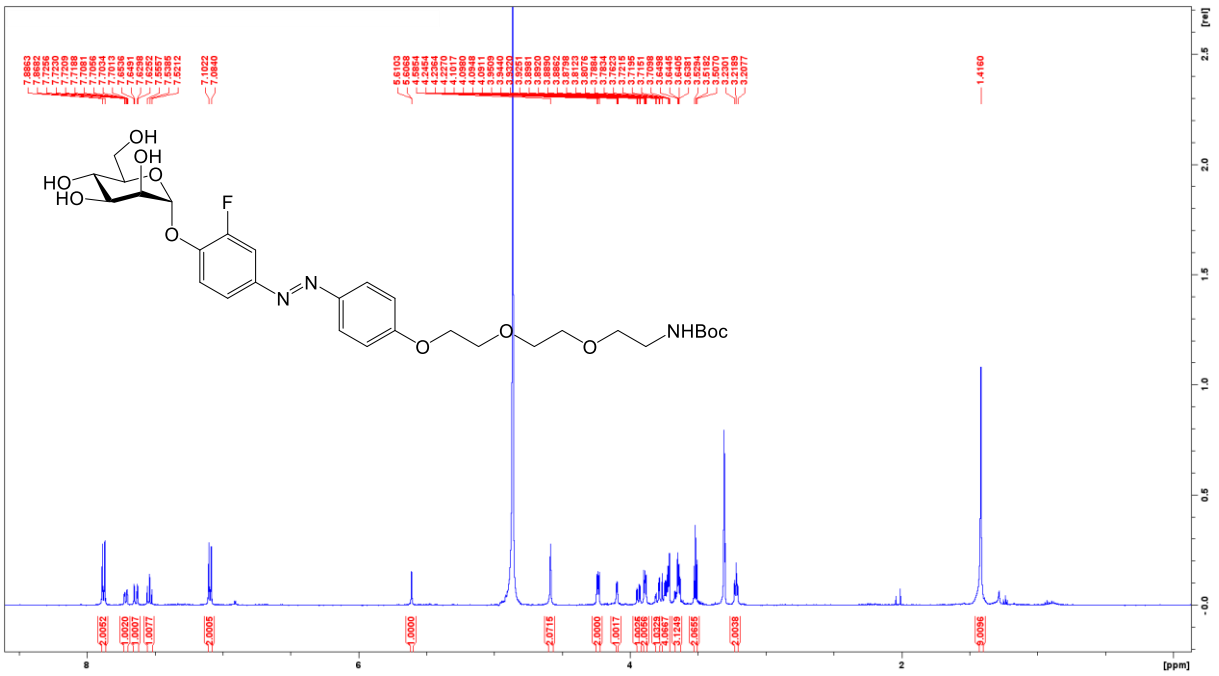


Figure 5.4-22: ^{1}H NMR spectrum of compound **14** (500 MHz, CD_3OD, 298 K)

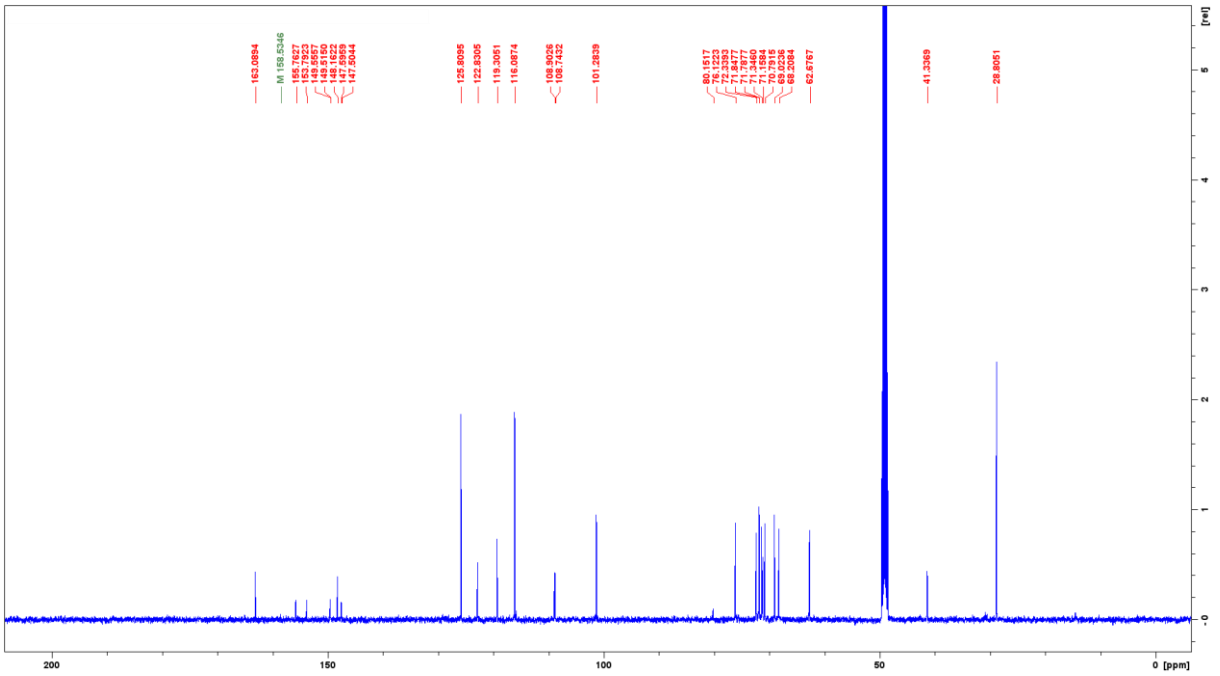


Figure 5.4-23: ^{13}C NMR spectrum of compound **14** (126 MHz, CD_3OD, 298 K).

The manually amended carbon signal assignment at 158.5 ppm (green) was allocated using the respective HMBC spectrum of compound **14** (Figure 5.4-24).

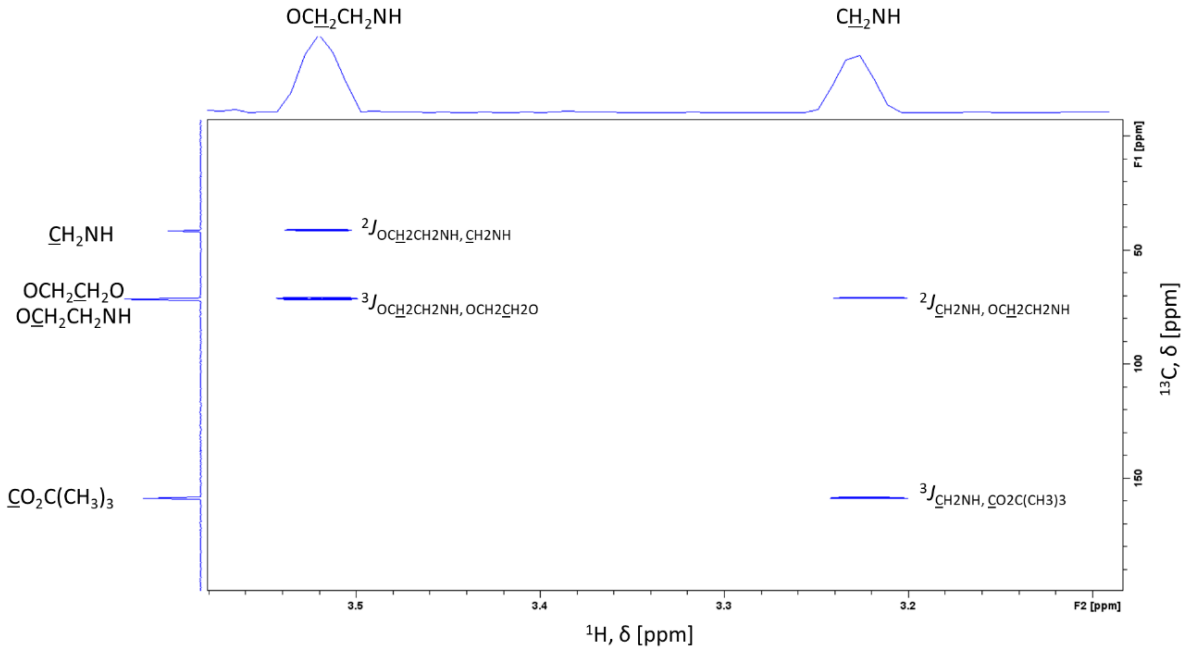


Figure 5.4-24: Image section (^{1}H: 3.0 – 3.58 ppm) from the HMBC spectrum of compound **14** (500 MHz, $CDCl_3$, 298 K).

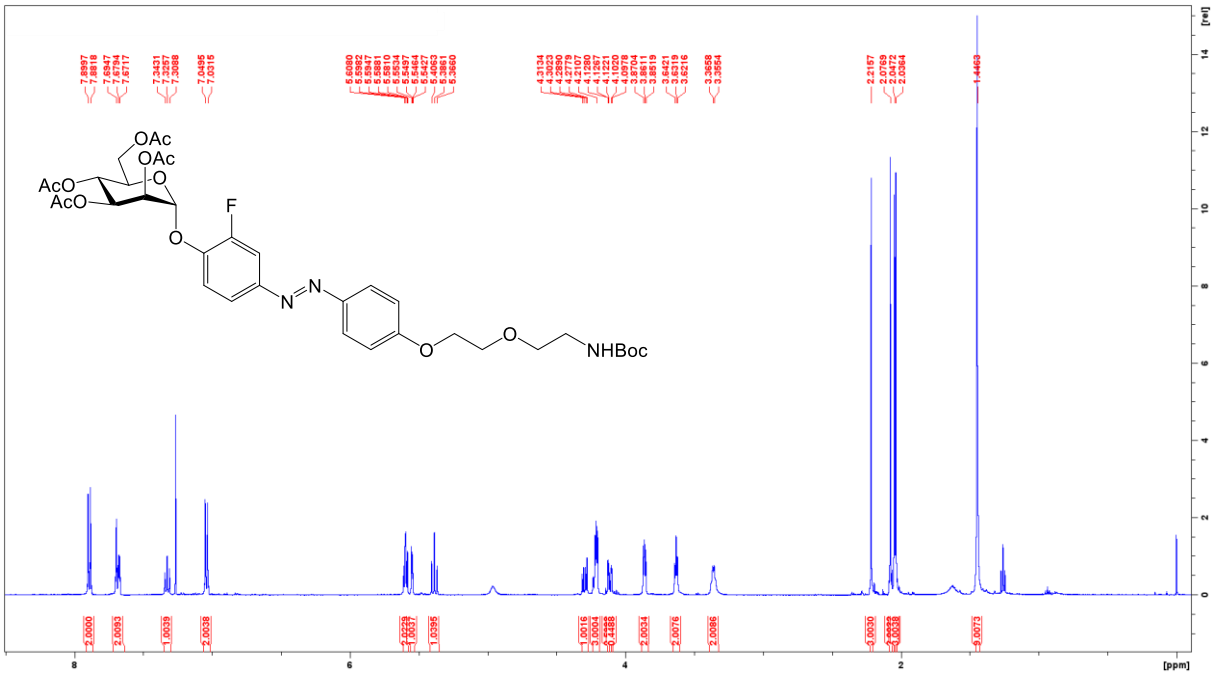


Figure 5.4-25: ^{1}H NMR spectrum of compound **15** (500 MHz, $CDCl_3$, 298 K).

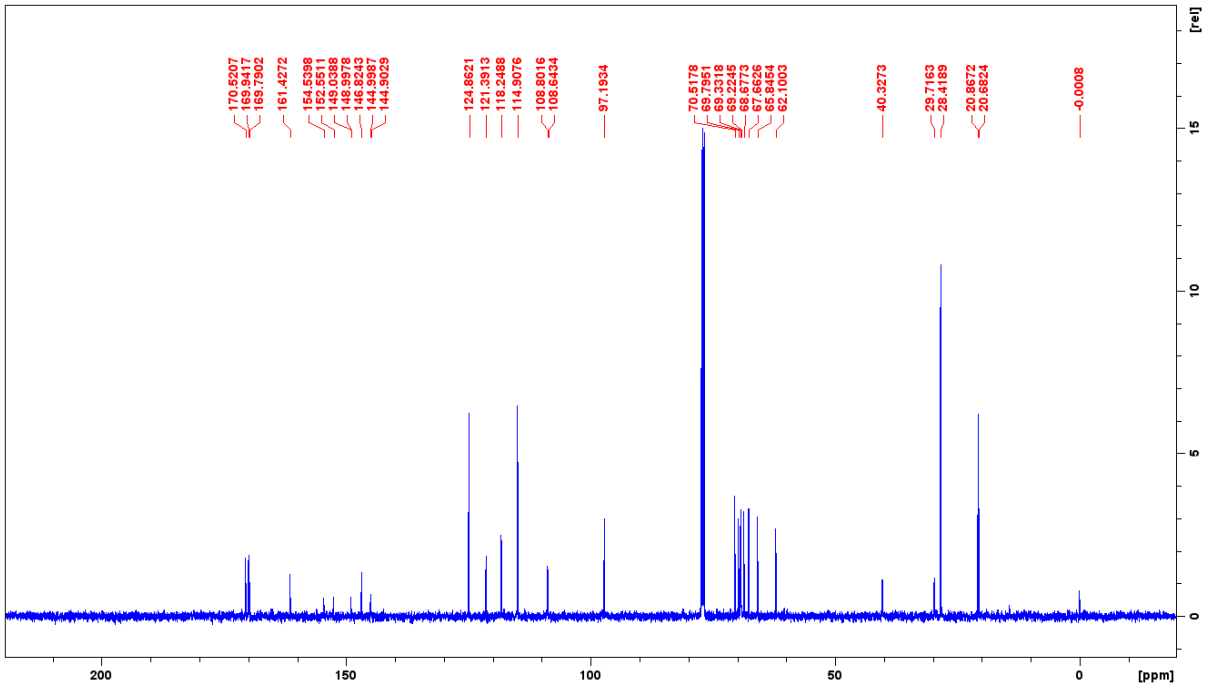


Figure 5.4-26: ^{13}C NMR spectrum of compound **15** (126 MHz, $CDCl_3$, 298 K).

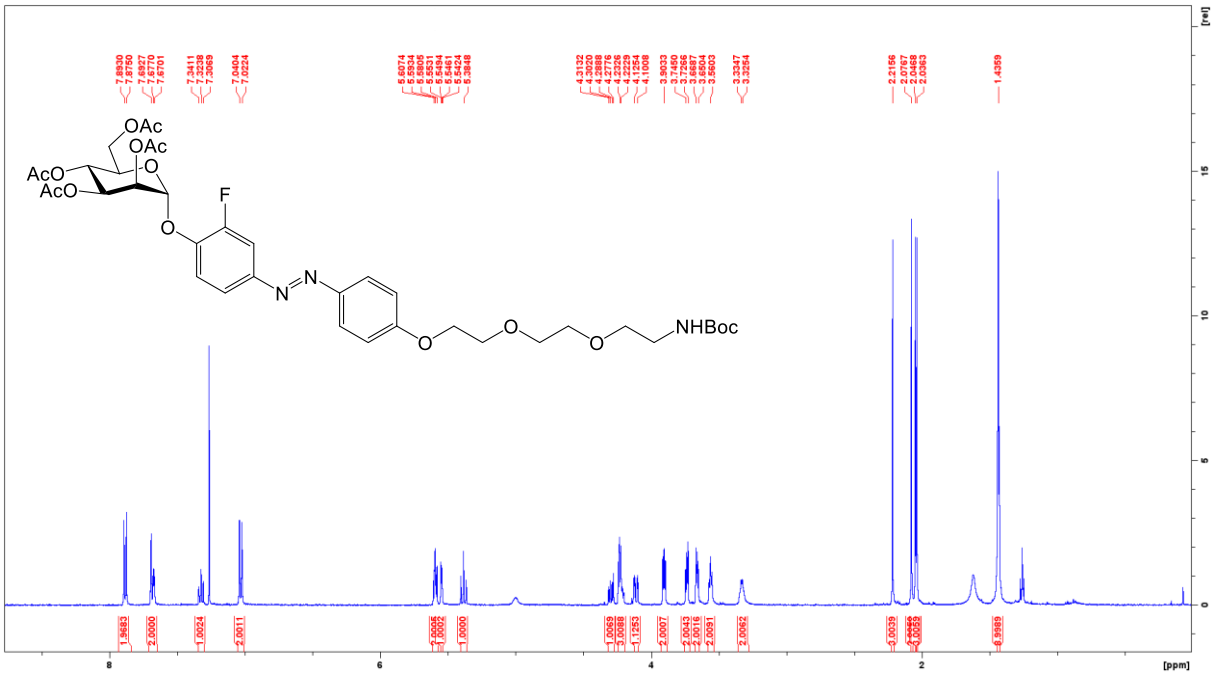


Figure 5.4-27: ^{1}H NMR spectrum of compound **16** (500 MHz, $CDCl_3$, 298 K).

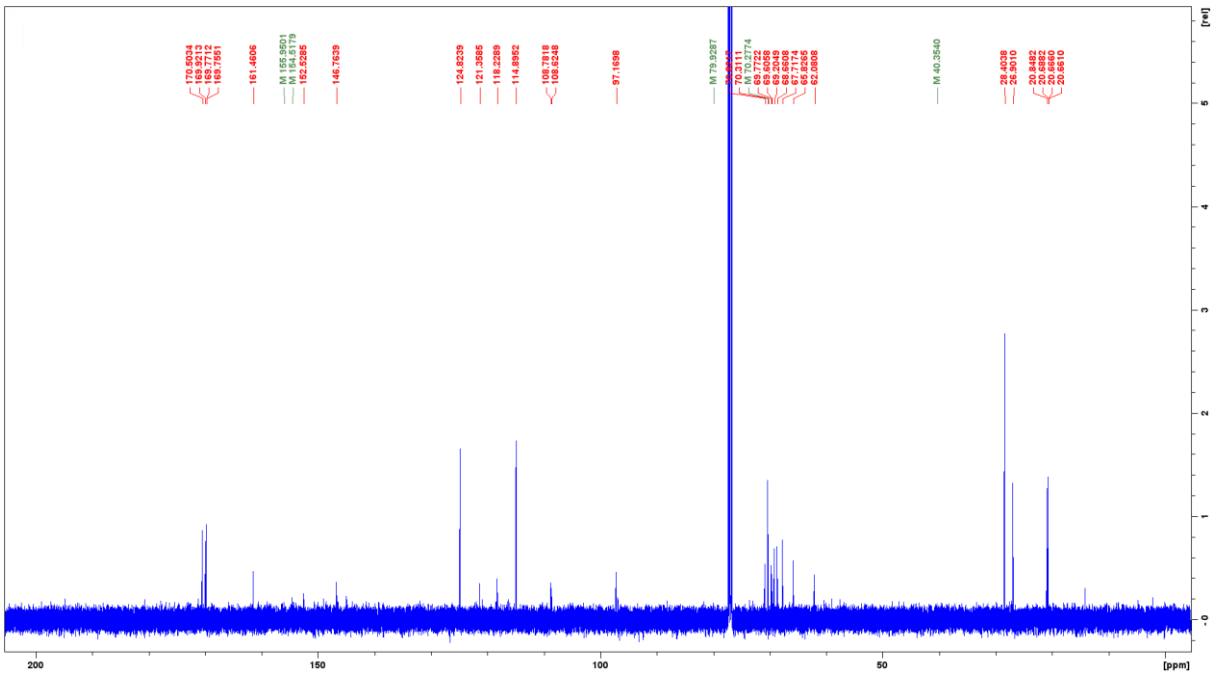


Figure 5.4-28: ^{13}C NMR spectrum of compound **16** (126 MHz, $CDCl_3$, 298 K).

The manually amended carbon signal assignments at 155.95 and 154.5 ppm (green) were allocated using the respective HMBC spectrum of compound **16** (Figure 5.4-29). The labelled signals at 79.9, 70.28 and 40.4 ppm were assigned according to the HSQC spectrum.

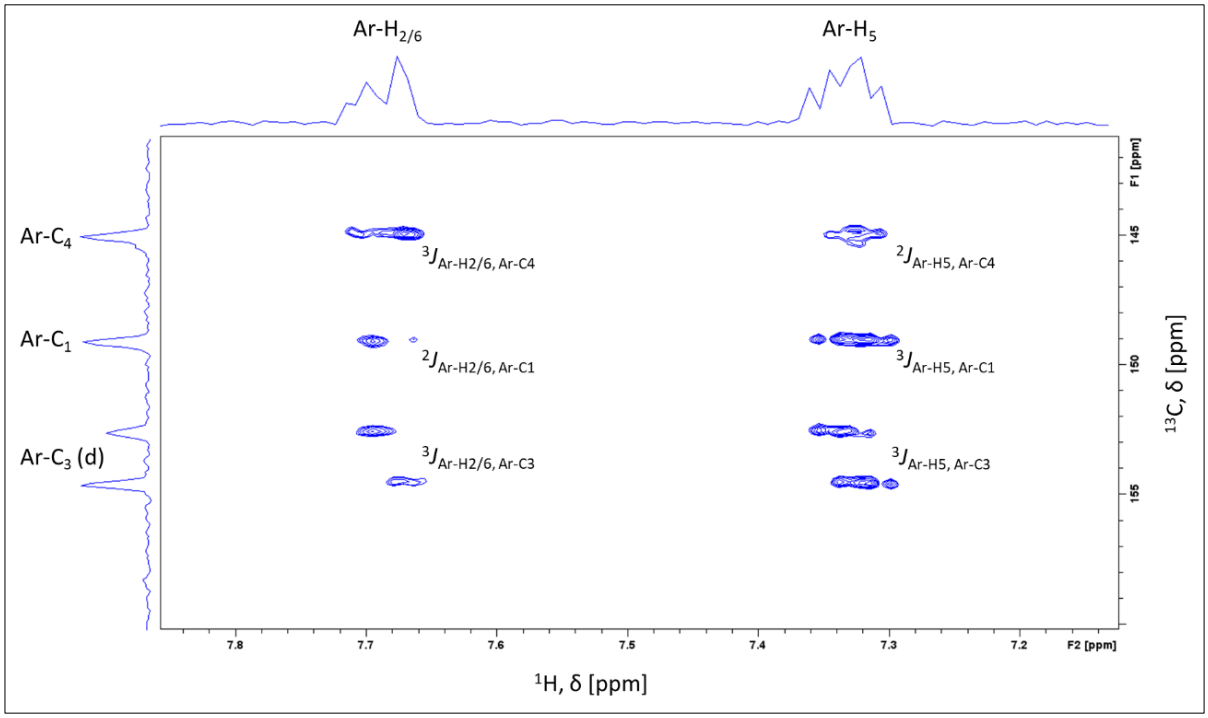


Figure 5.4-29: Image section (^{1}H: 7.8 – 7.2 ppm) from the HMBC spectrum of compound **16** (500 MHz, $CDCl_3$, 298 K).

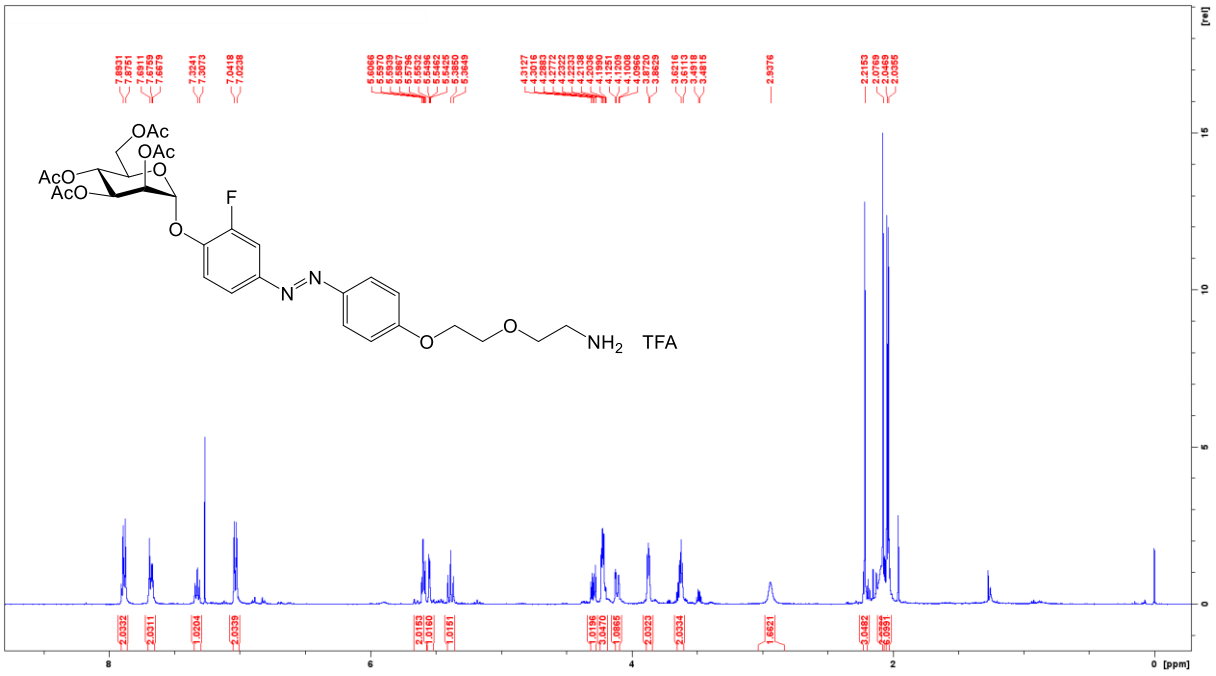


Figure 5.4-30: ^{1}H NMR spectrum of crude compound **17** (500 MHz, $CDCl_3$, 298 K).

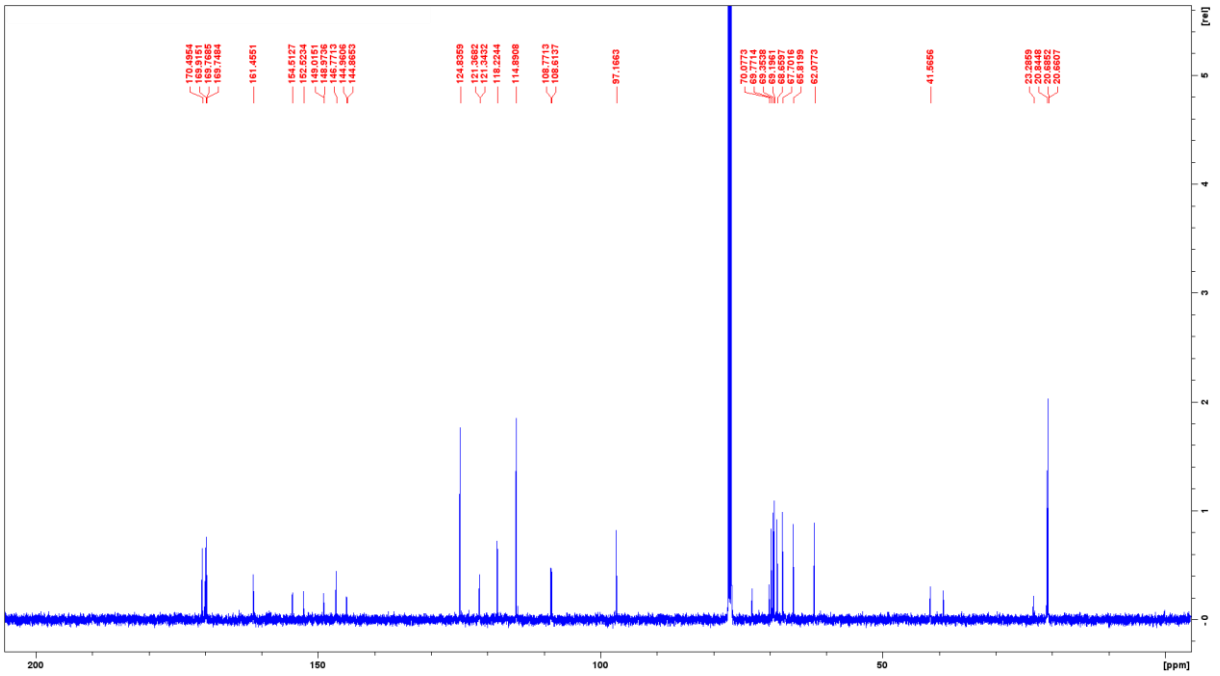


Figure 5.4-31: ^{13}C NMR spectrum of crude compound **17** (126 MHz, $CDCl_3$, 298 K).

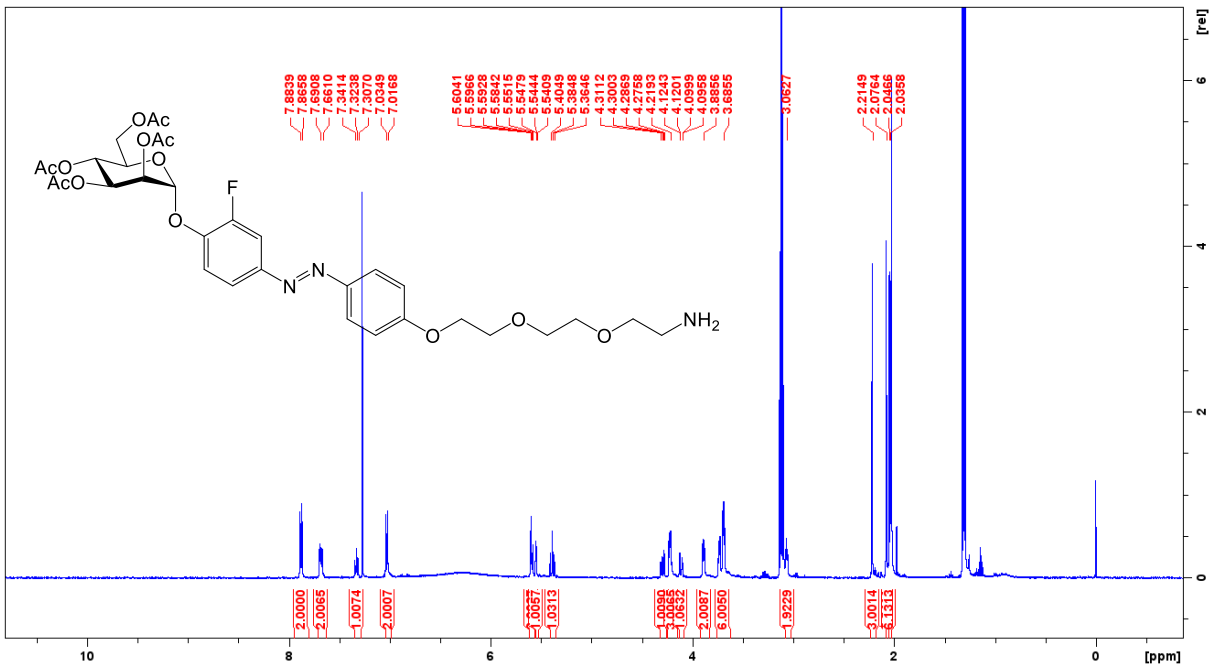


Figure 5.4-32: ^{1}H NMR spectrum of crude compound **18** (500 MHz, $CDCl_3$, 298 K).

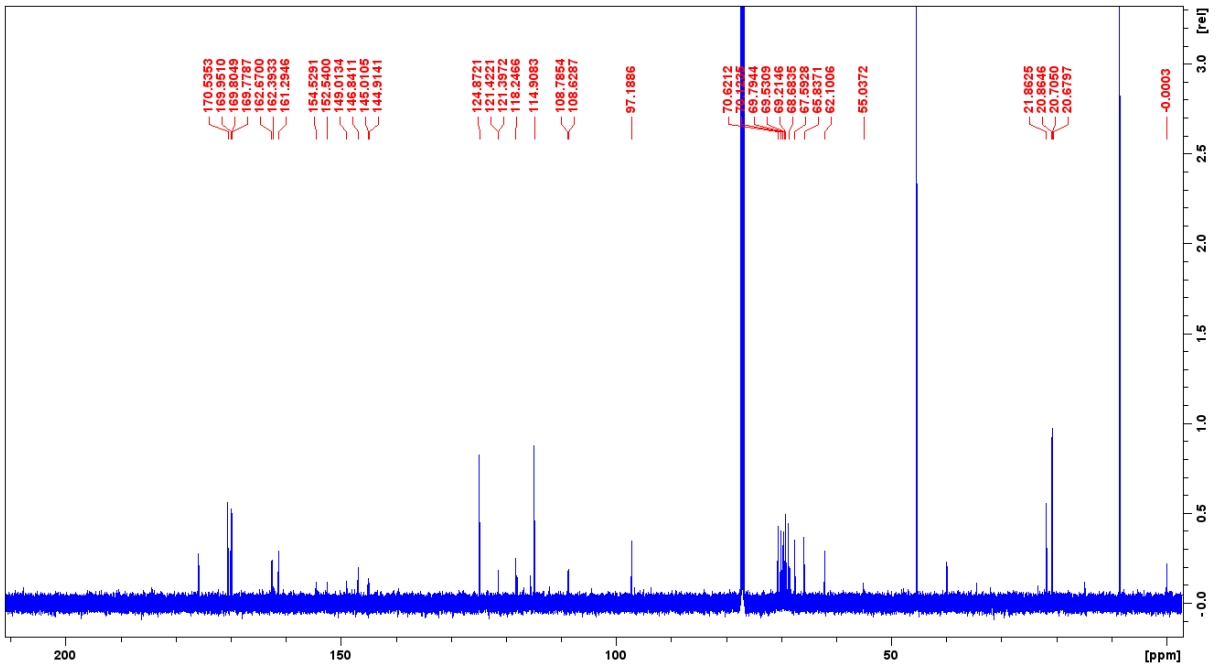


Figure 5.4-33: ^{13}C NMR spectrum of crude compound **18** (126 MHz, $CDCl_3$, 298 K).

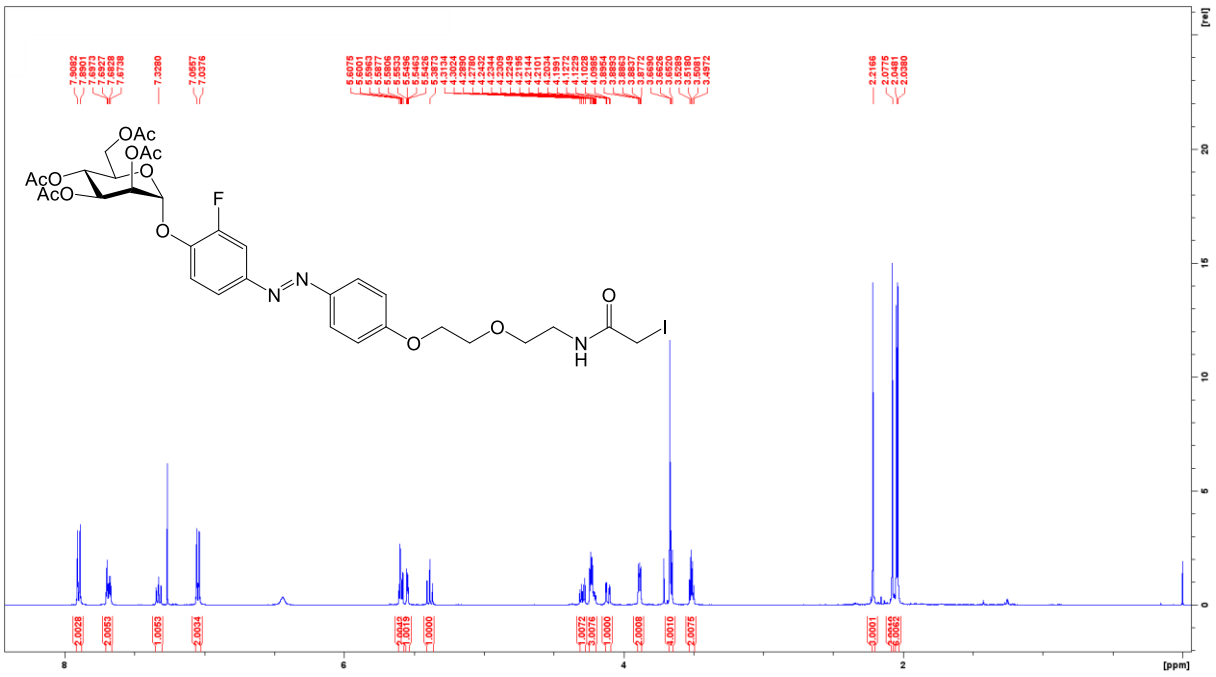


Figure 5.4-34: ^{1}H NMR spectrum of compound **19** (500 MHz, $CDCl_3$, 298 K).

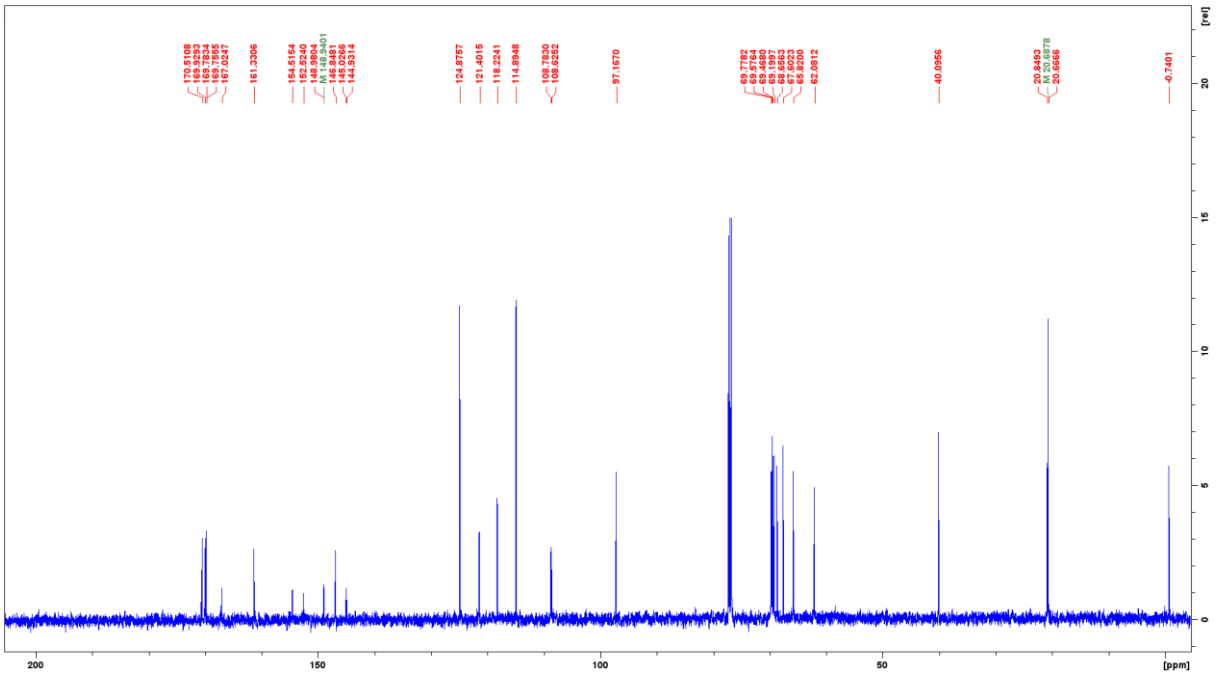


Figure 5.4-35: ^{13}C NMR spectrum of compound **19** (126 MHz, $CDCl_3$, 298 K).
The second signal of the doublet at 149.9 ppm (green) has been manually amended.

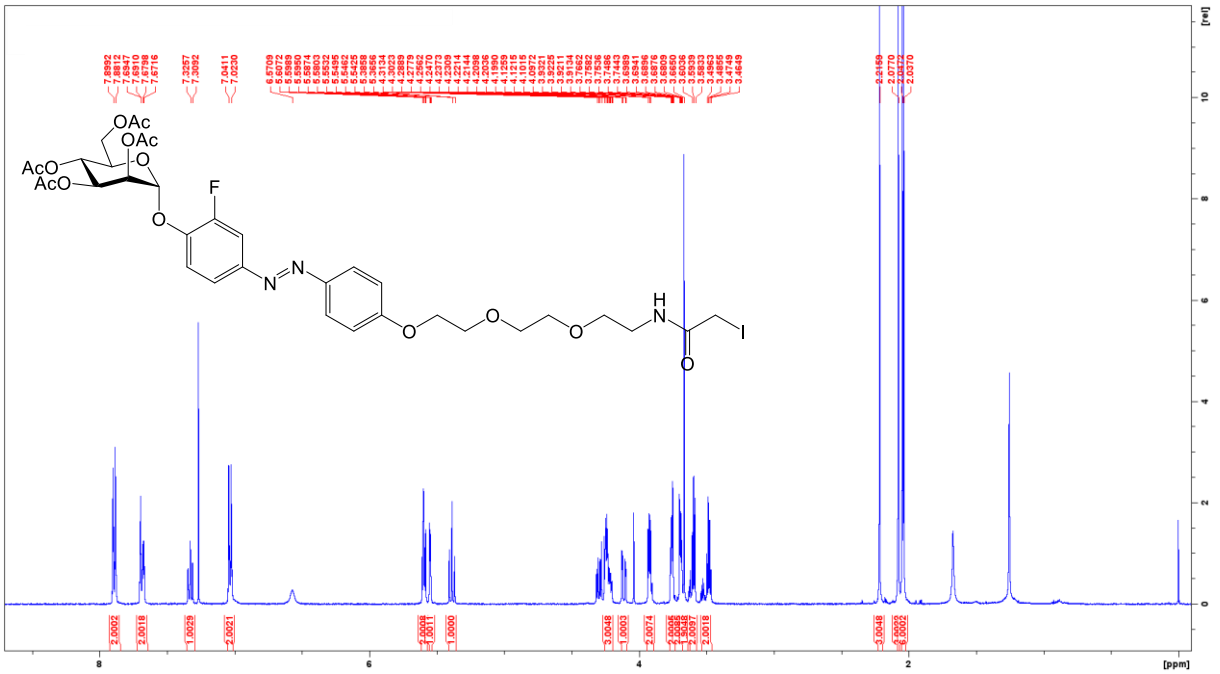


Figure 5.4-36: ^{1}H NMR spectrum of compound **20** (500 MHz, $CDCl_3$, 298 K).

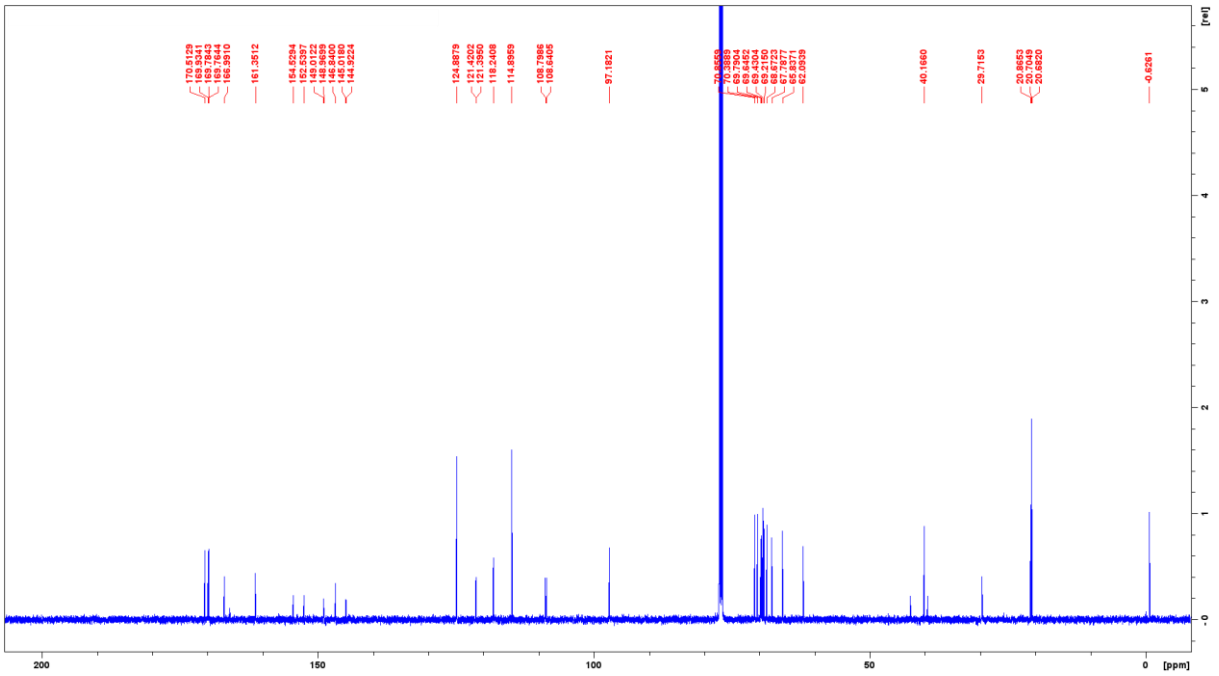


Figure 5.4-37: ^{13}C NMR spectrum of compound **20** (126 MHz, $CDCl_3$, 298 K).

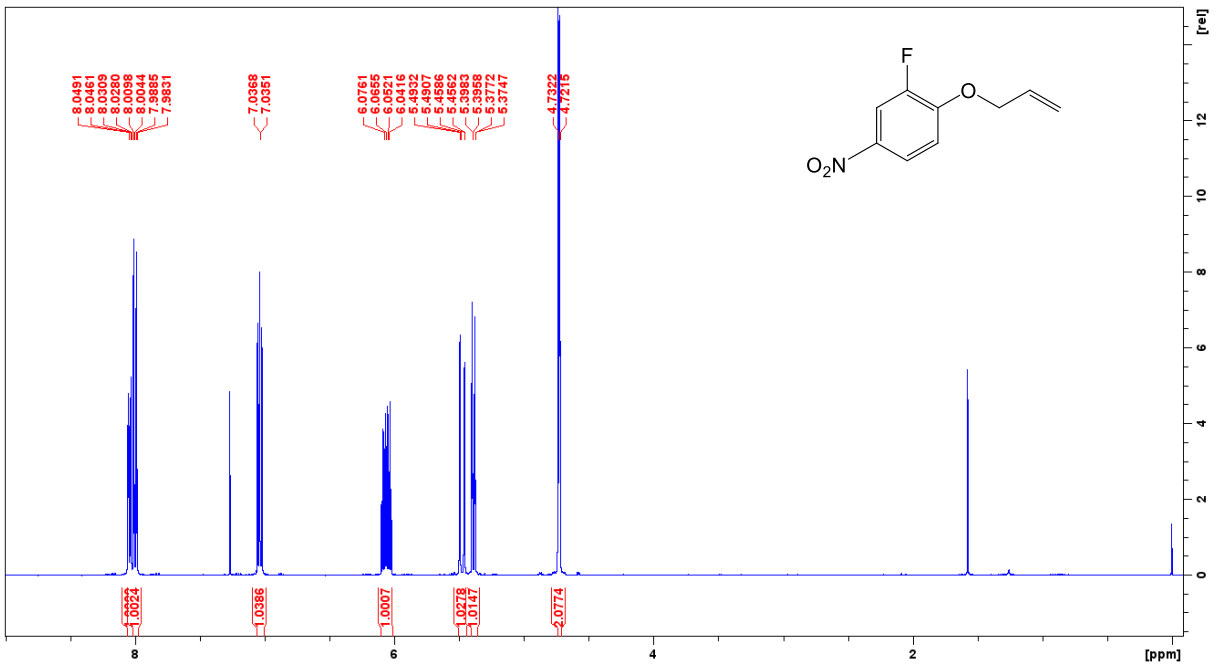


Figure 5.4-38: ^{1}H NMR spectrum of compound **22** (500 MHz, $CDCl_3$, 298 K).

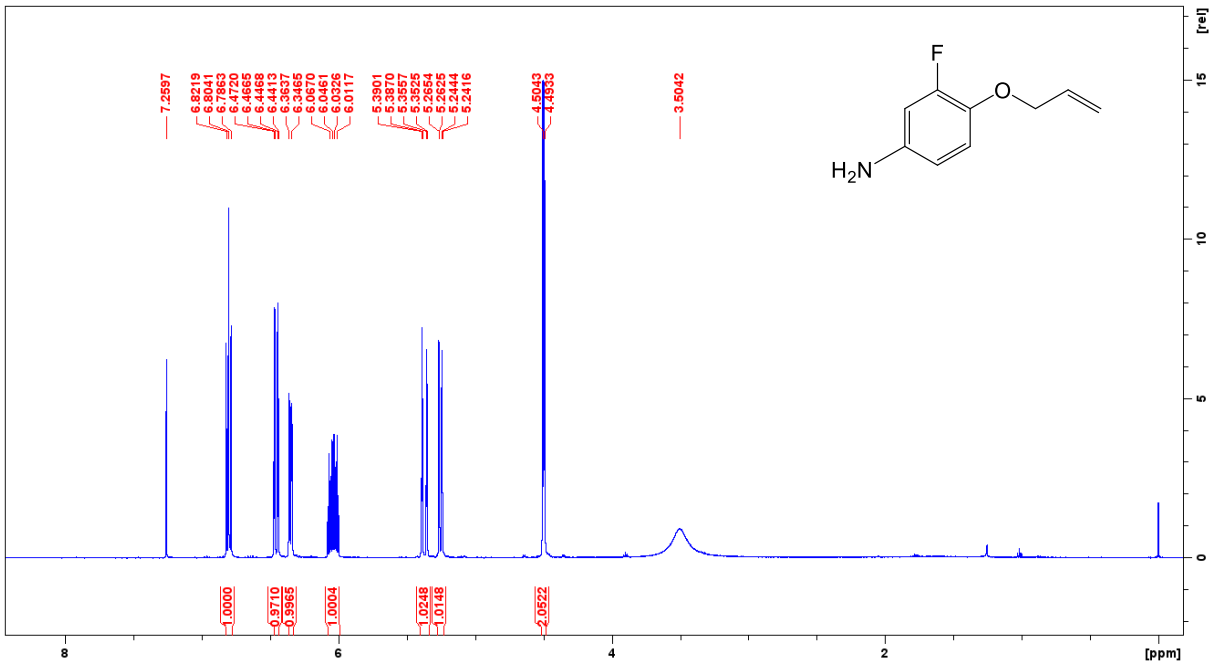


Figure 5.4-39: ^{1}H NMR spectrum of compound **24** (500 MHz, $CDCl_3$, 298 K).

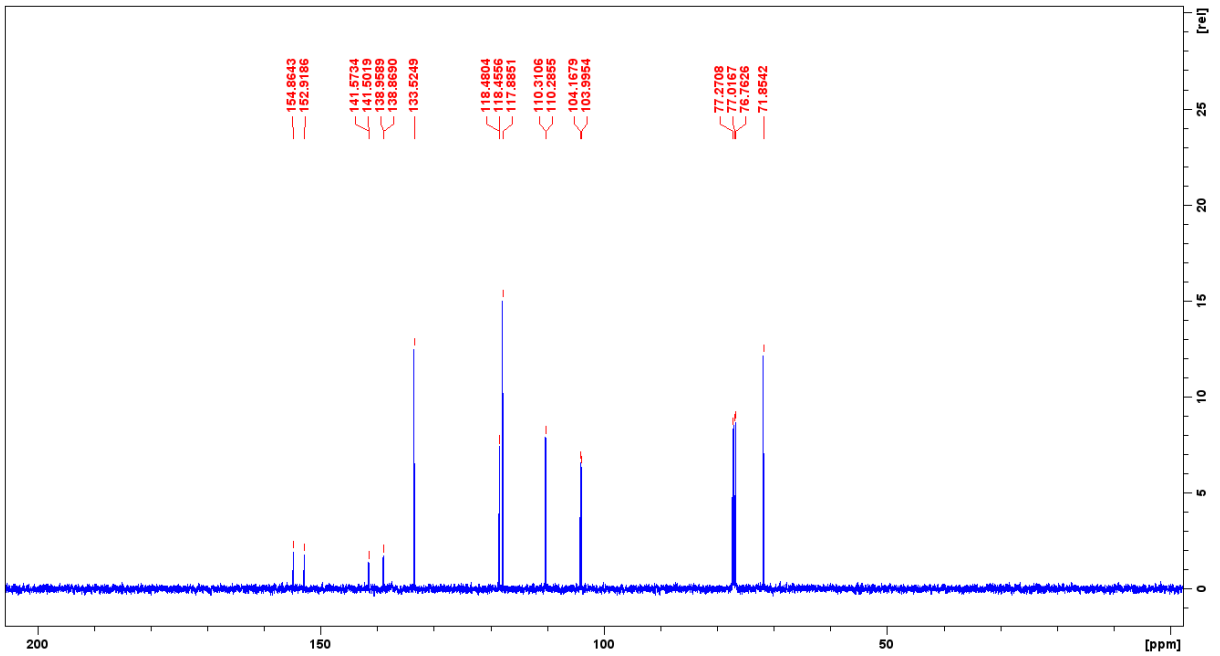


Figure 5.4-40: ^{13}C NMR spectrum of compound **24** (126 MHz, $CDCl_3$, 298 K).

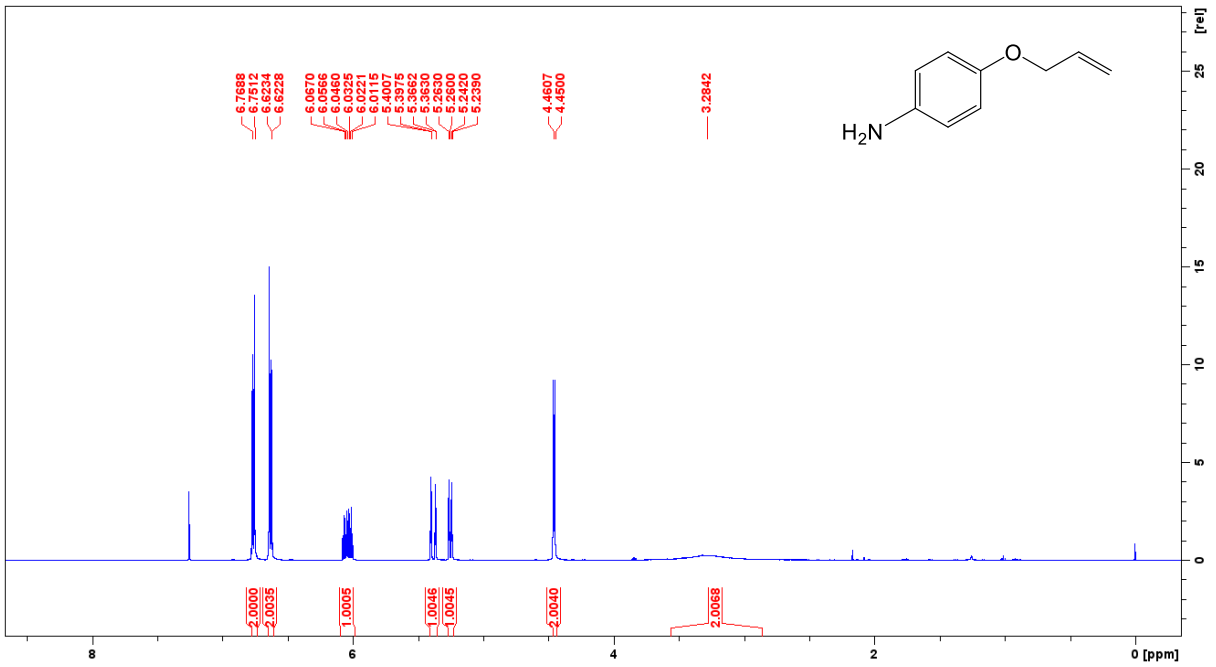


Figure 5.4-41: ^{1}H NMR spectrum of compound **25** (500 MHz, $CDCl_3$, 298 K).

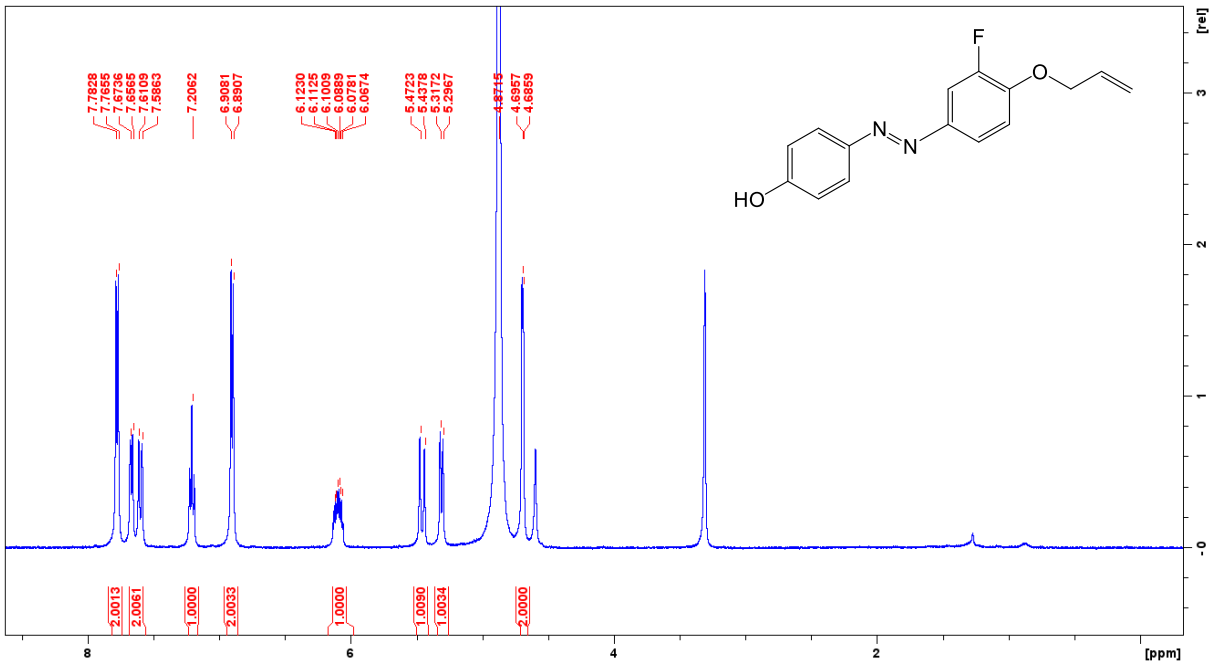


Figure 5.4-42: ^{1}H NMR spectrum of compound **26** (500 MHz, CD_3OD, 298 K).

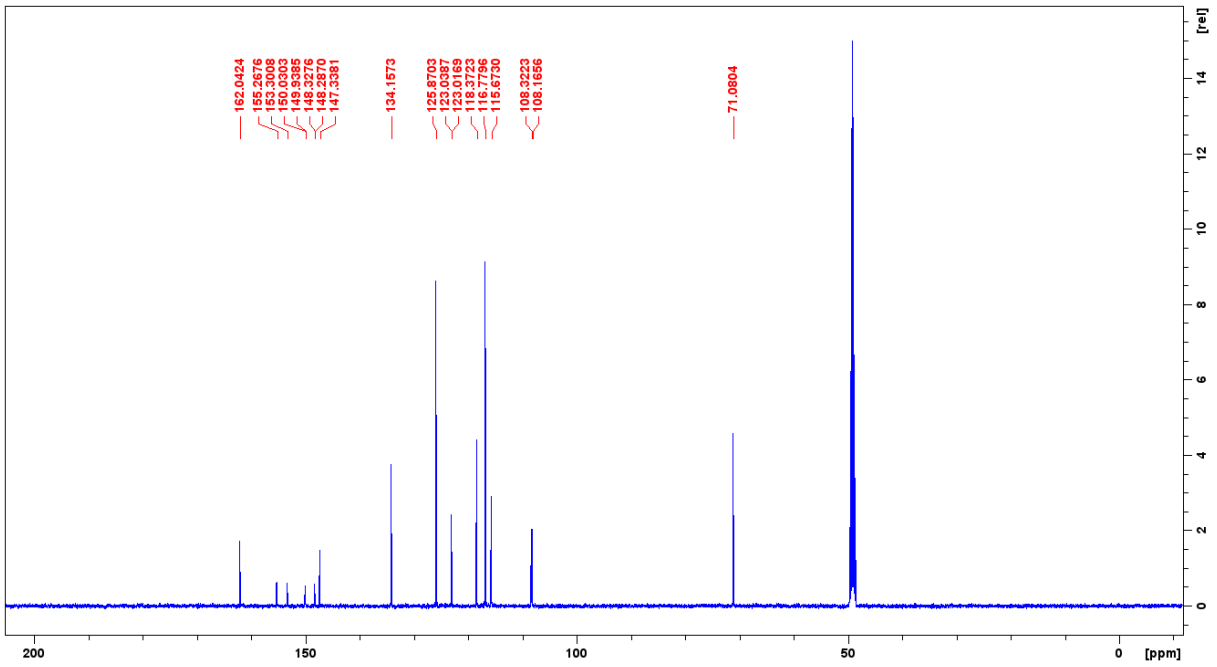


Figure 5.4-43: ^{13}C NMR spectrum of compound **26** (126 MHz, CD_3OD, 298 K).

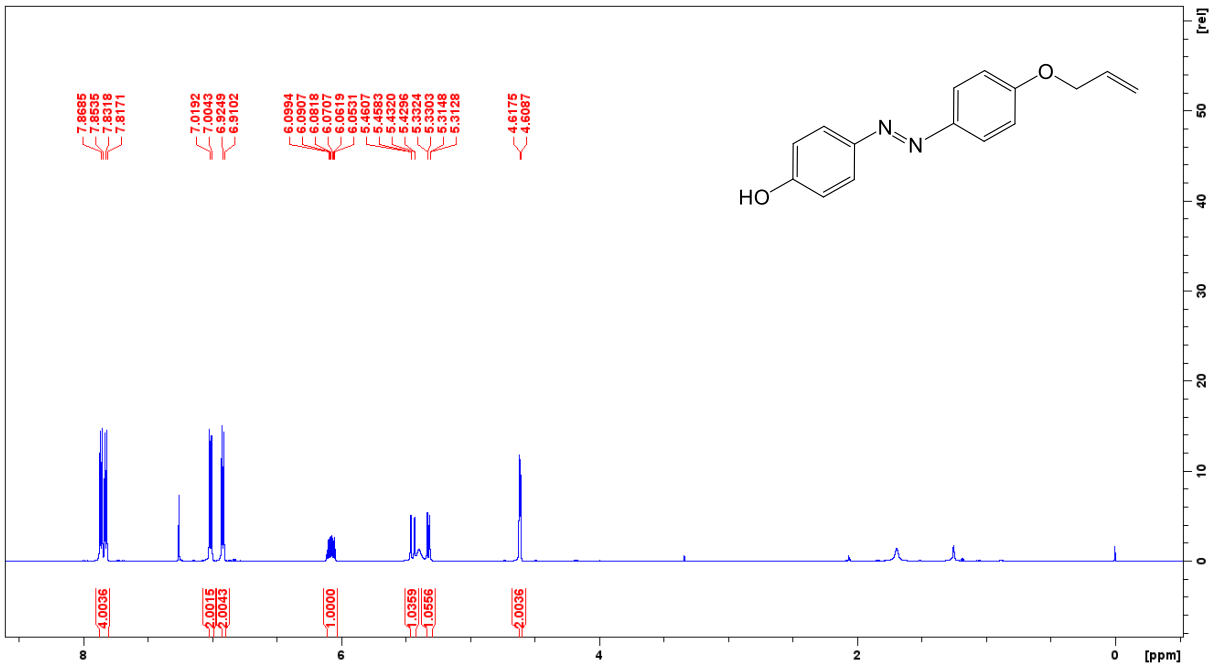


Figure 5.4-44: ^{1}H NMR spectrum of compound **27** (500 MHz, $CDCl_3$, 298 K).

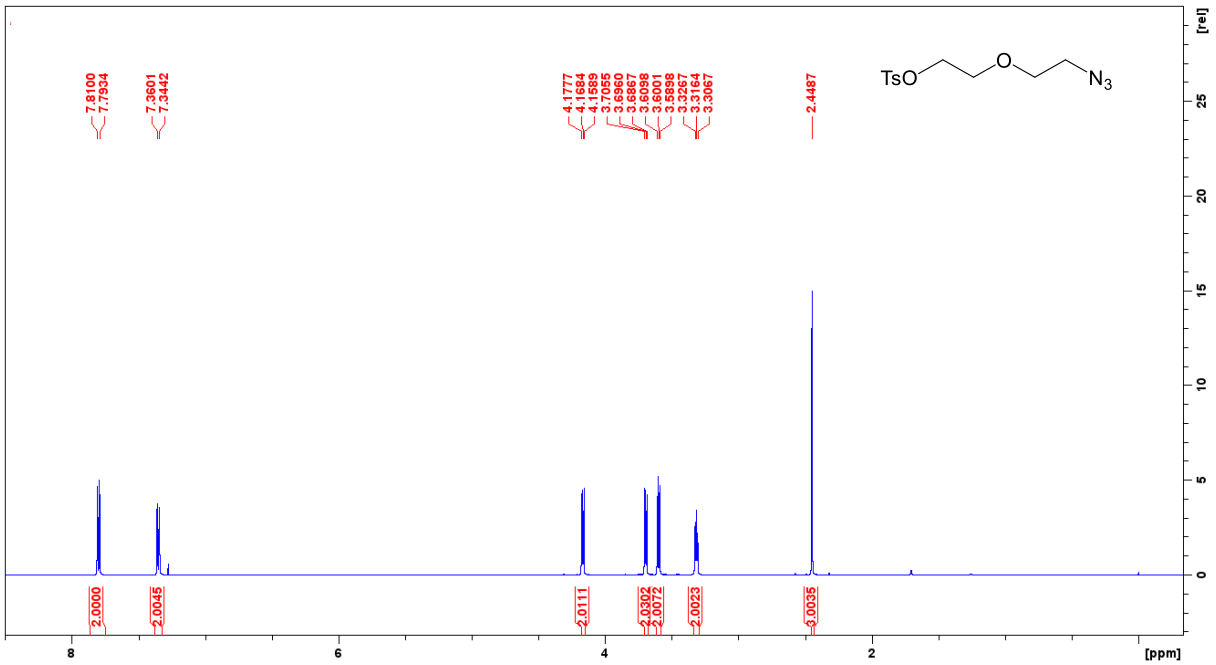


Figure 5.4-45: ^{1}H NMR spectrum of compound **28** (500 MHz, $CDCl_3$, 298 K).

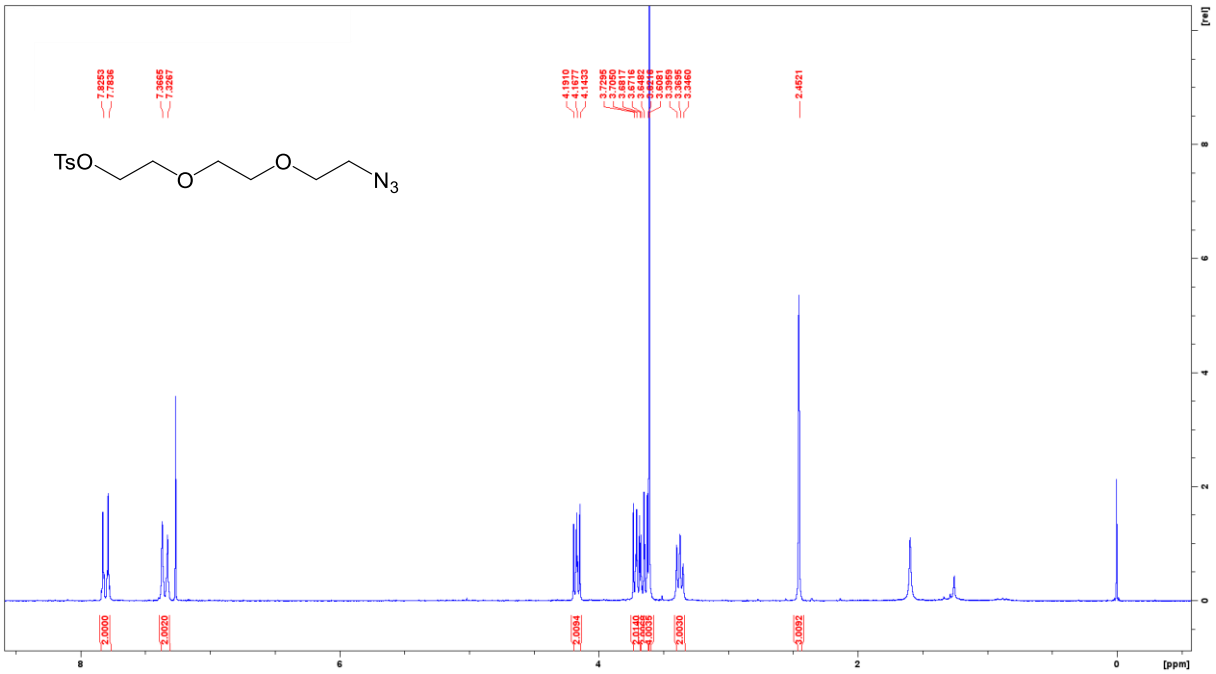


Figure 5.4-46: ^{1}H NMR spectrum of compound **29** (200 MHz, $CDCl_3$, 298 K).

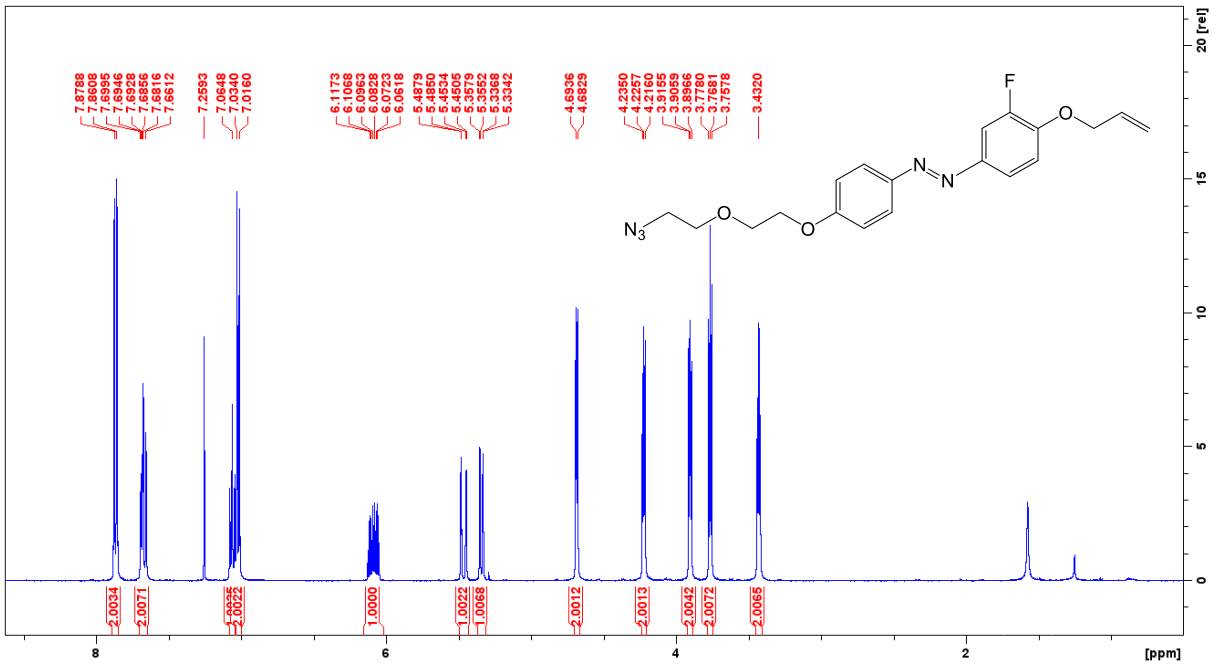


Figure 5.4-47: ^{1}H NMR spectrum of compound **30** (500 MHz, $CDCl_3$, 298 K).

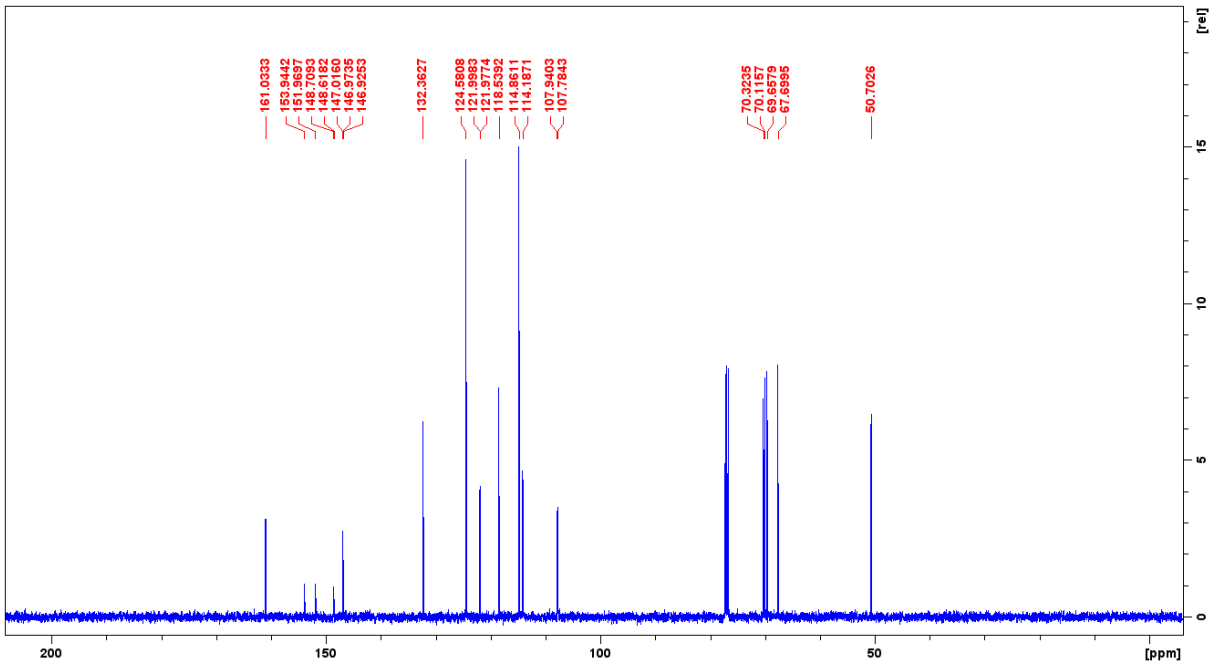


Figure 5.4-48: ^{13}C NMR spectrum of compound **30** (126 MHz, $CDCl_3$, 298 K).

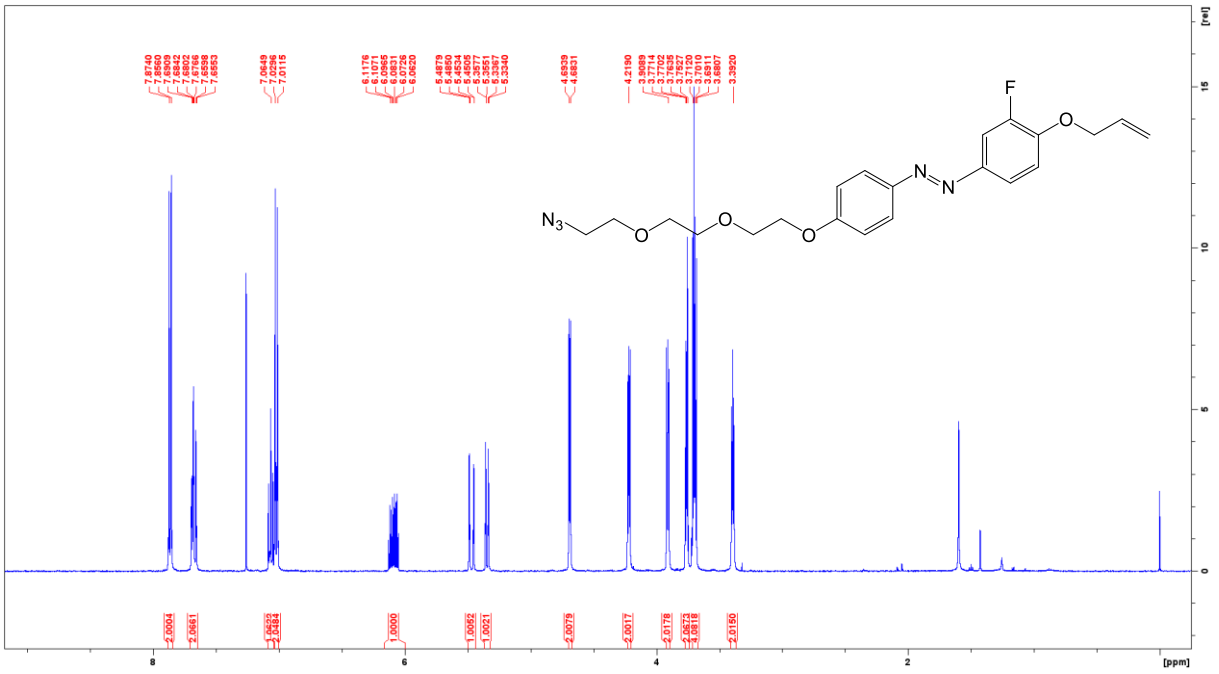


Figure 5.4-49: ^{1}H NMR spectrum of compound **31** (500 MHz, $CDCl_3$, 298 K).

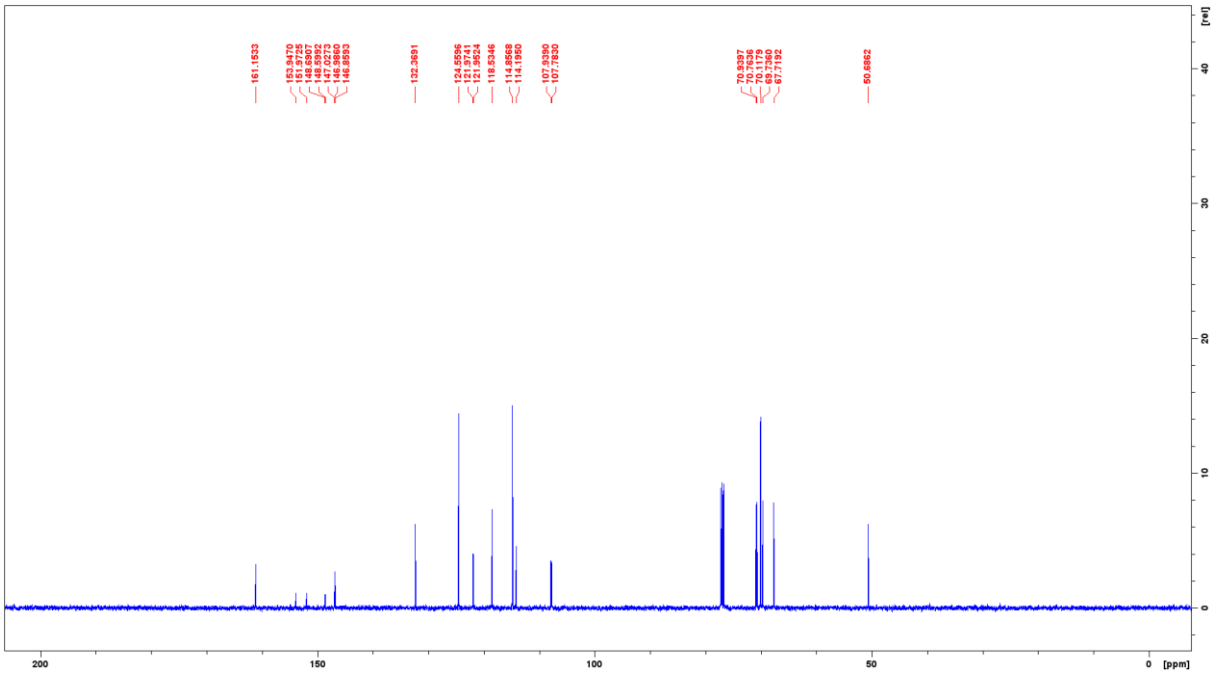


Figure 5.4-50: ^{13}C NMR spectrum of compound **31** (126 MHz, $CDCl_3$, 298 K).

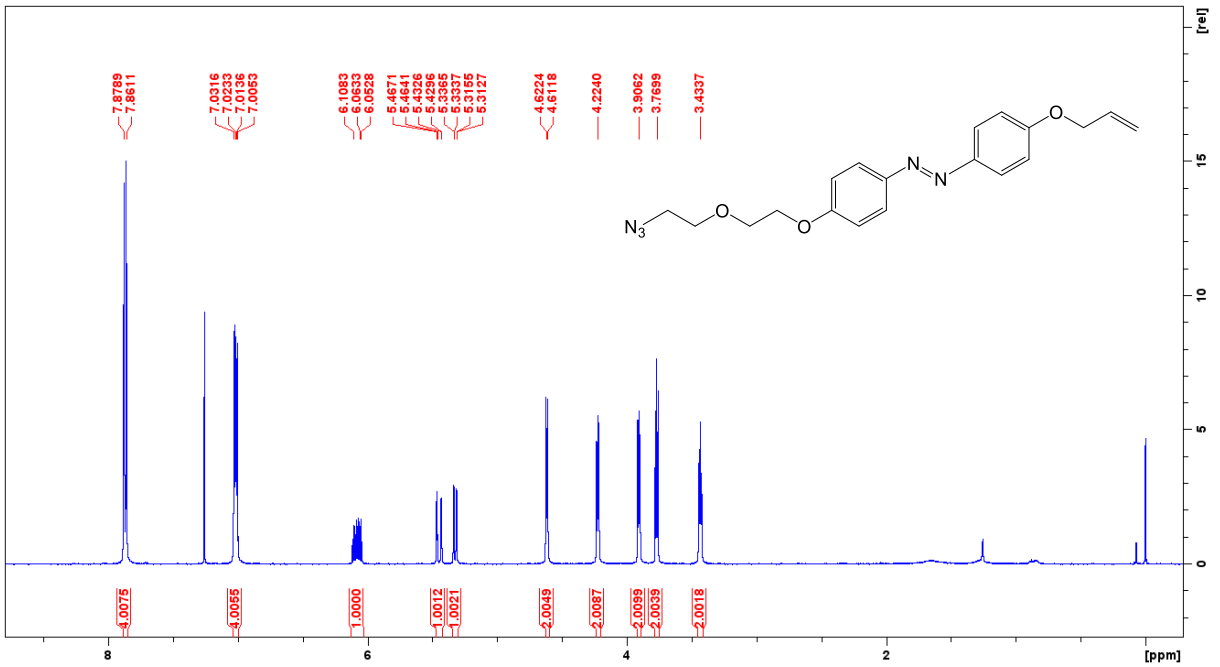


Figure 5.4-51: ^{1}H NMR spectrum of compound **32** (500 MHz, $CDCl_3$, 298 K).

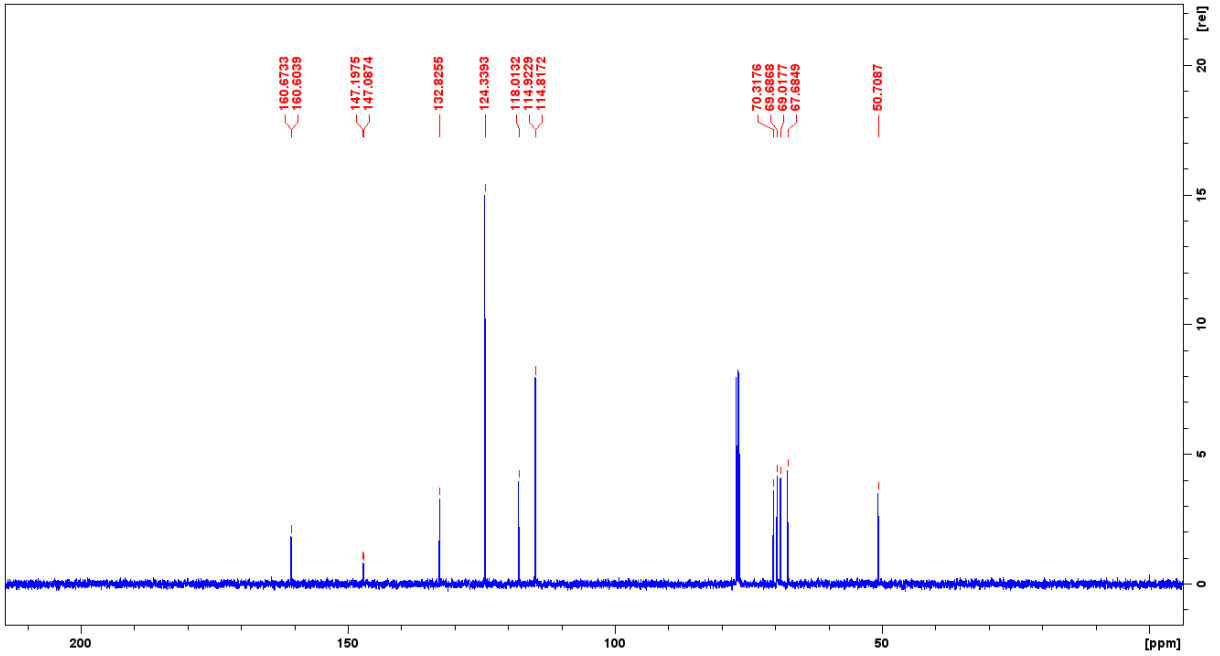


Figure 5.4-52: ^{13}C NMR spectrum of compound **32** (126 MHz, $CDCl_3$, 298 K).

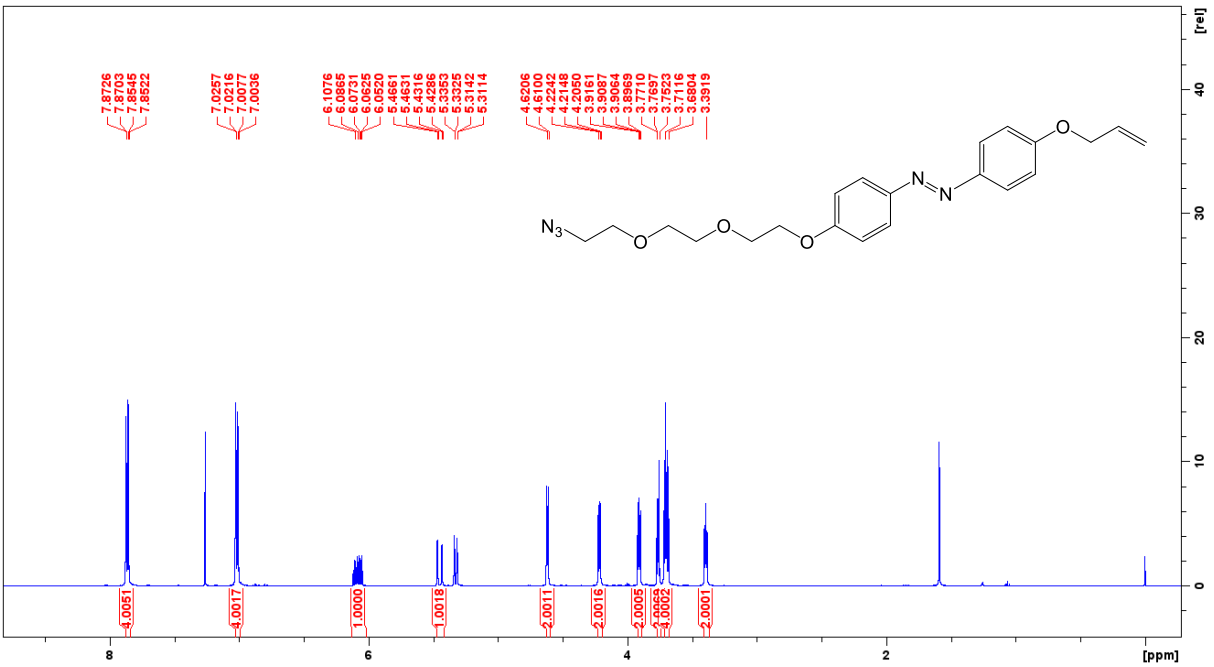


Figure 5.4-53: ^{1}H NMR spectrum of compound **33** (500 MHz, $CDCl_3$, 298 K).

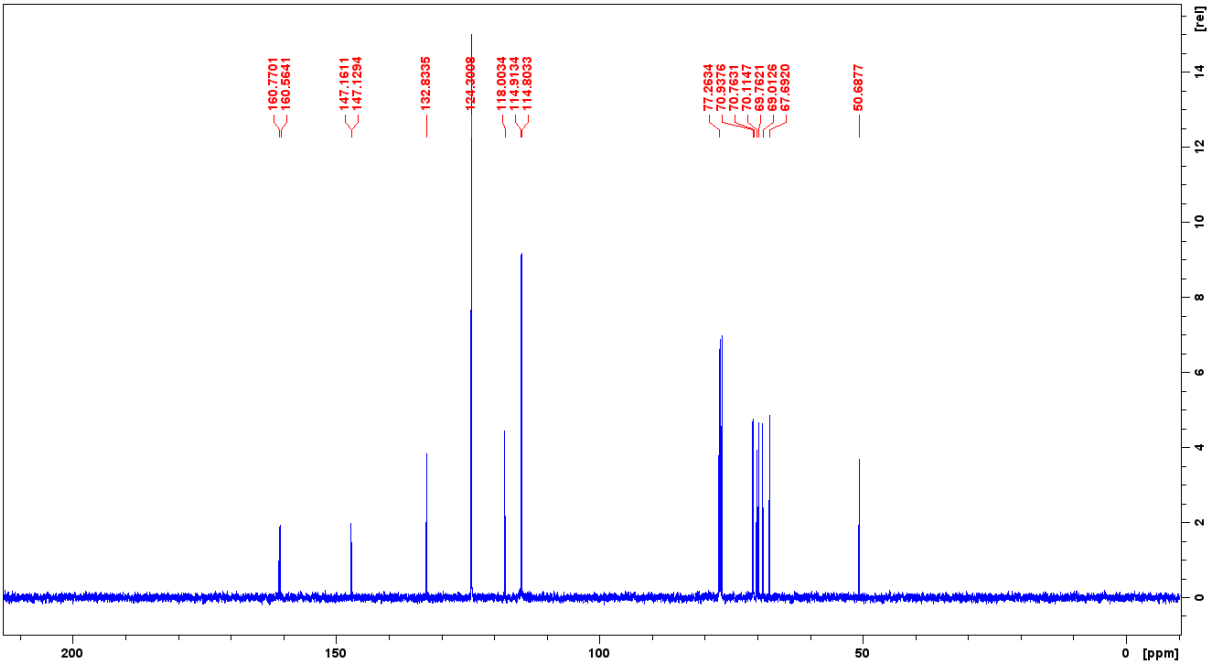


Figure 5.4-54: ^{13}C NMR spectrum of compound **33** (126 MHz, $CDCl_3$, 298 K).

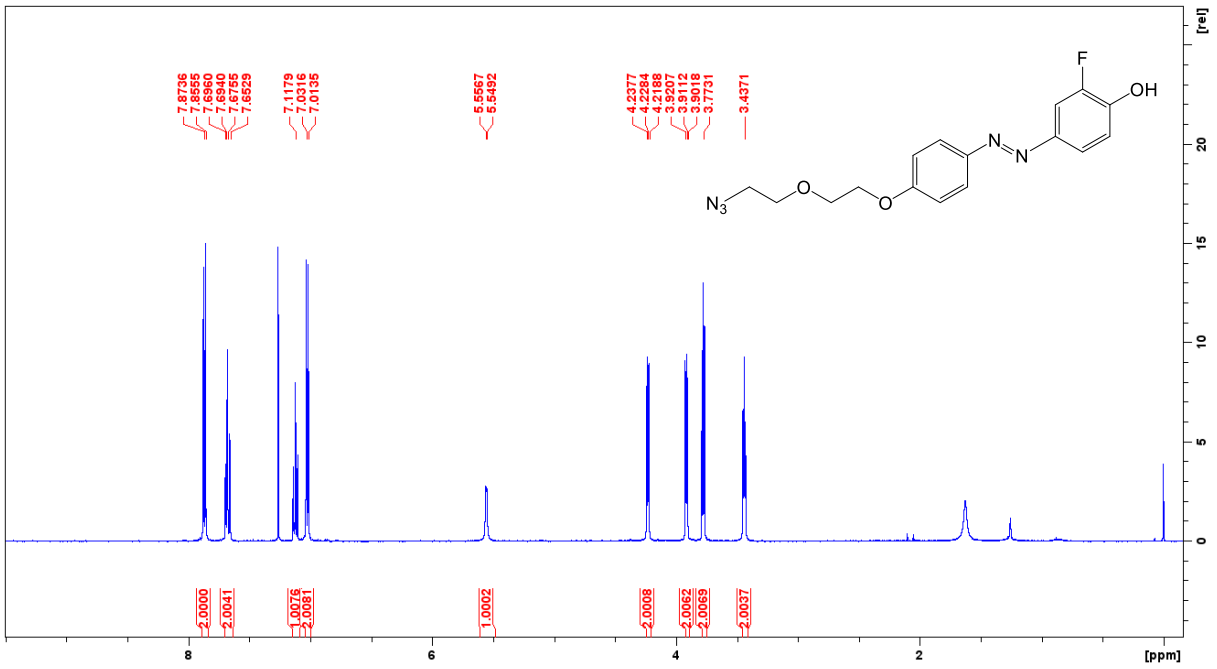


Figure 5.4-55: ^{1}H NMR spectrum of compound **34** (500 MHz, $CDCl_3$, 298 K).

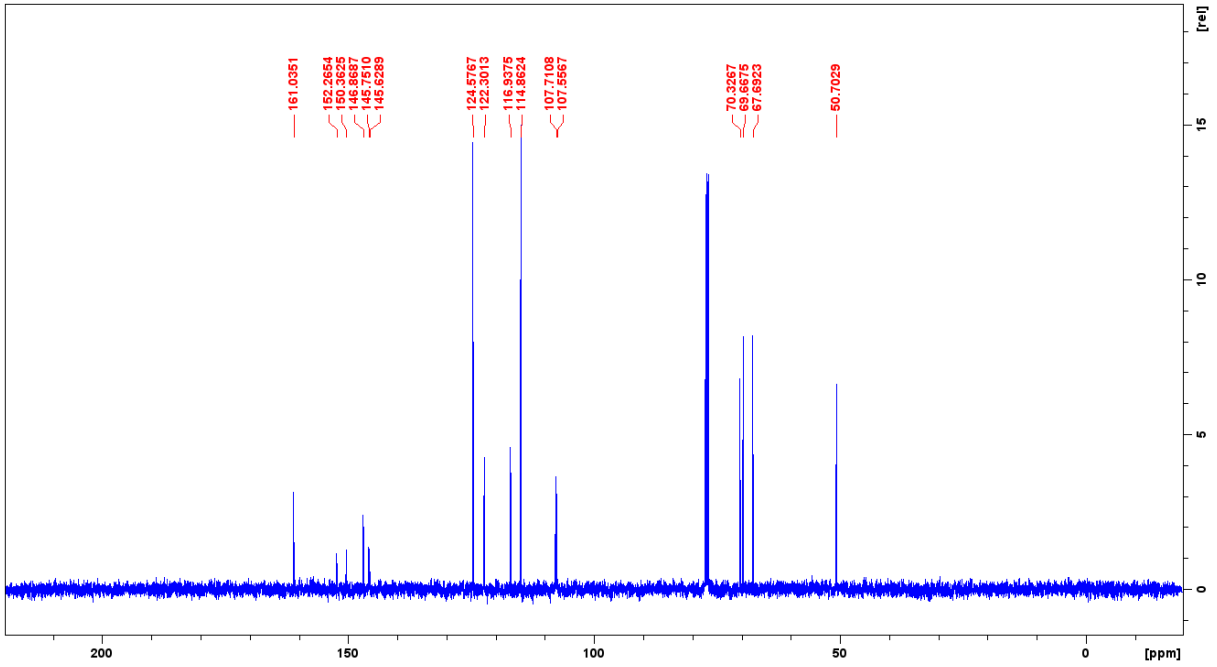


Figure 5.4-56: ^{13}C NMR spectrum of compound **34** (126 MHz, $CDCl_3$, 298 K).

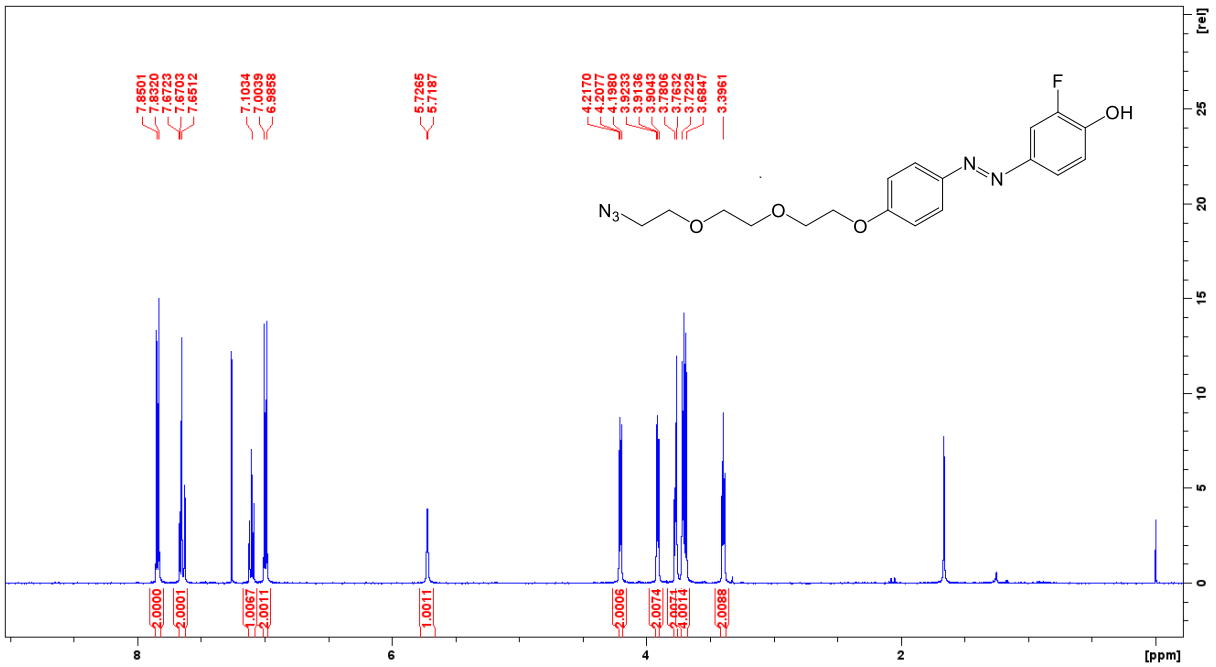


Figure 5.4-57: ^{1}H NMR spectrum of compound **35** (500 MHz, $CDCl_3$, 298 K).

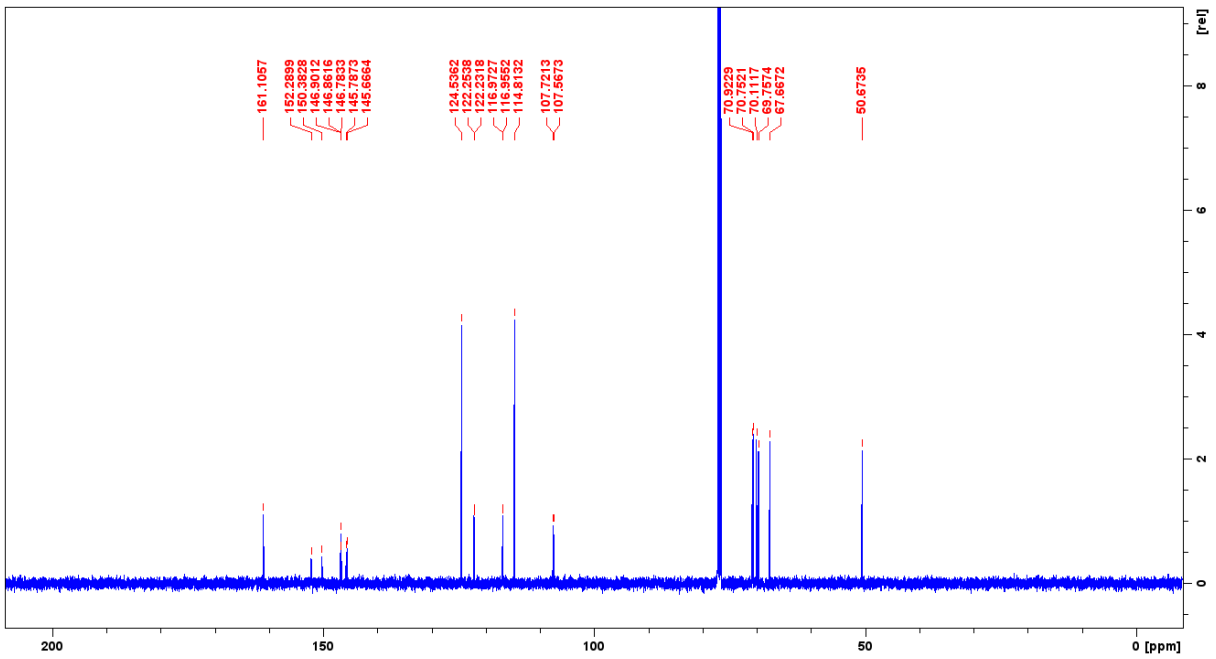


Figure 5.4-58: ^{13}C NMR spectrum of compound **35** (126 MHz, $CDCl_3$, 298 K).

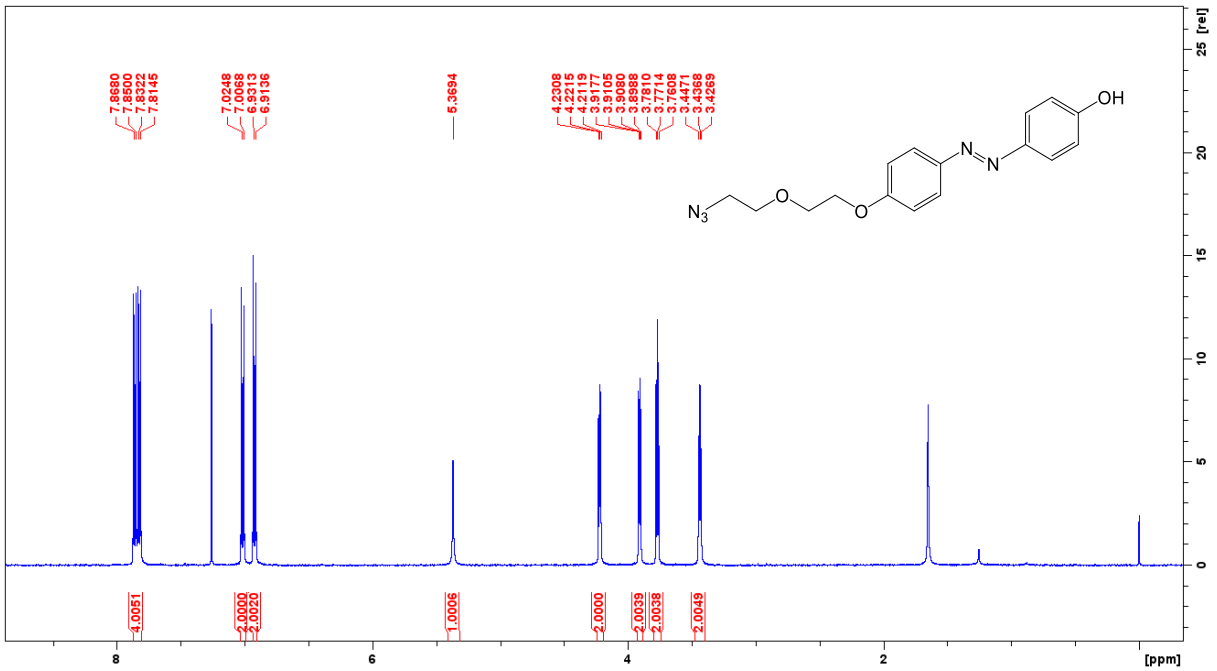


Figure 5.4-59: ^{1}H NMR spectrum of compound **36** (500 MHz, $CDCl_3$, 298 K).

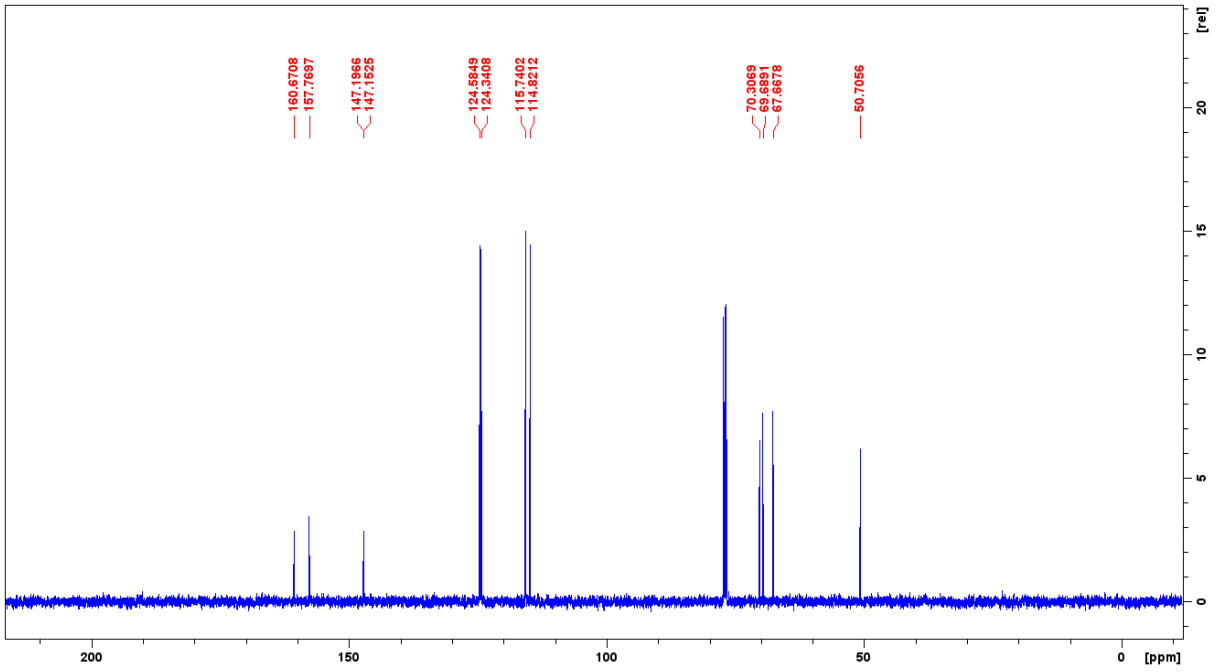


Figure 5.4-60: ^{13}C NMR spectrum of compound **36** (126 MHz, $CDCl_3$, 298 K).

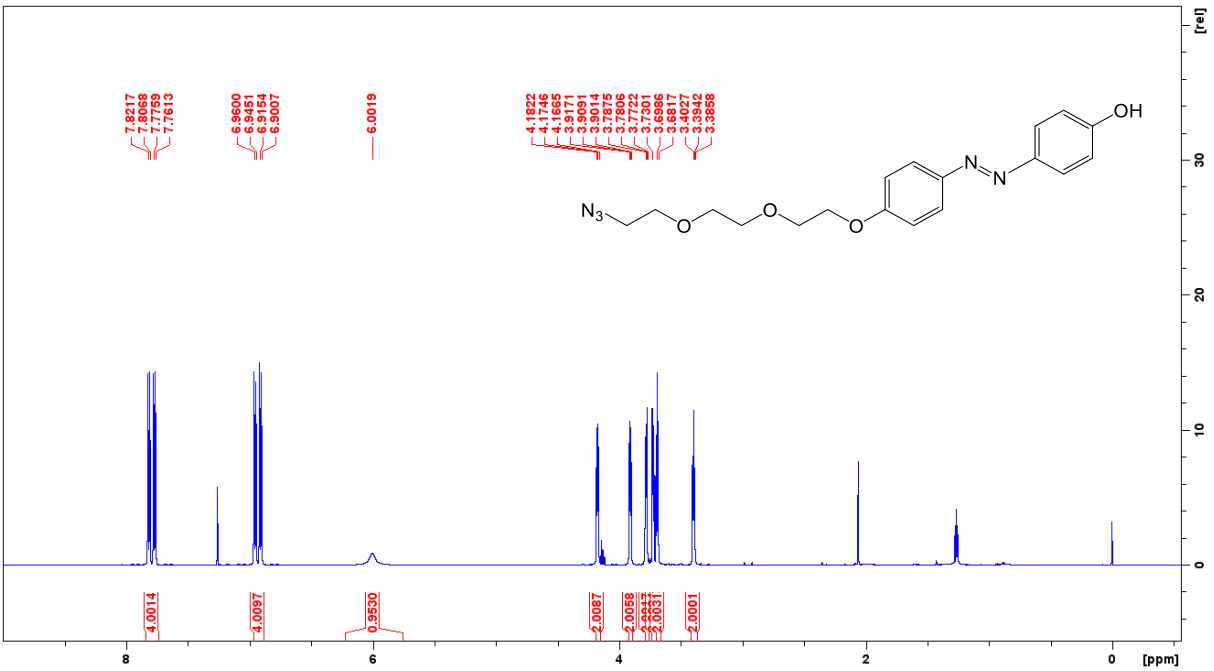


Figure 5.4-61: ^{1}H NMR spectrum of compound **37** (600 MHz, $CDCl_3$, 298 K).

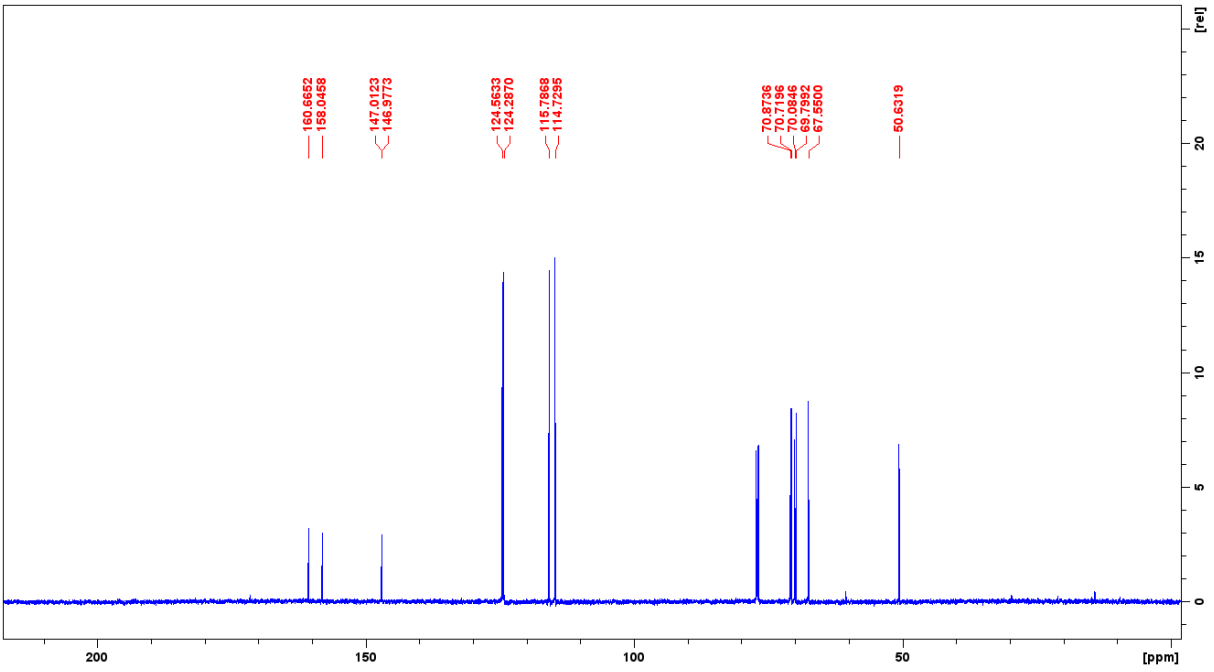


Figure 5.4-62: ^{13}C NMR spectrum of compound **37** (151 MHz, $CDCl_3$, 298 K).

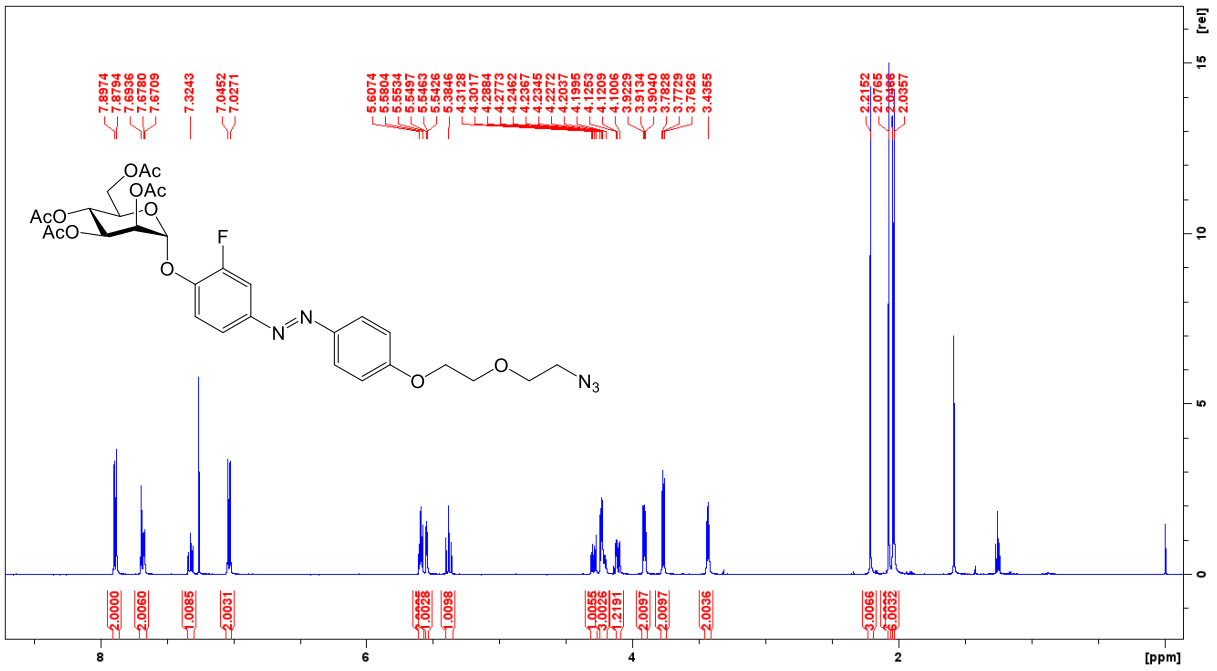


Figure 5.4-63: ^{1}H NMR spectrum of compound **38** (500 MHz, $CDCl_3$, 298 K).

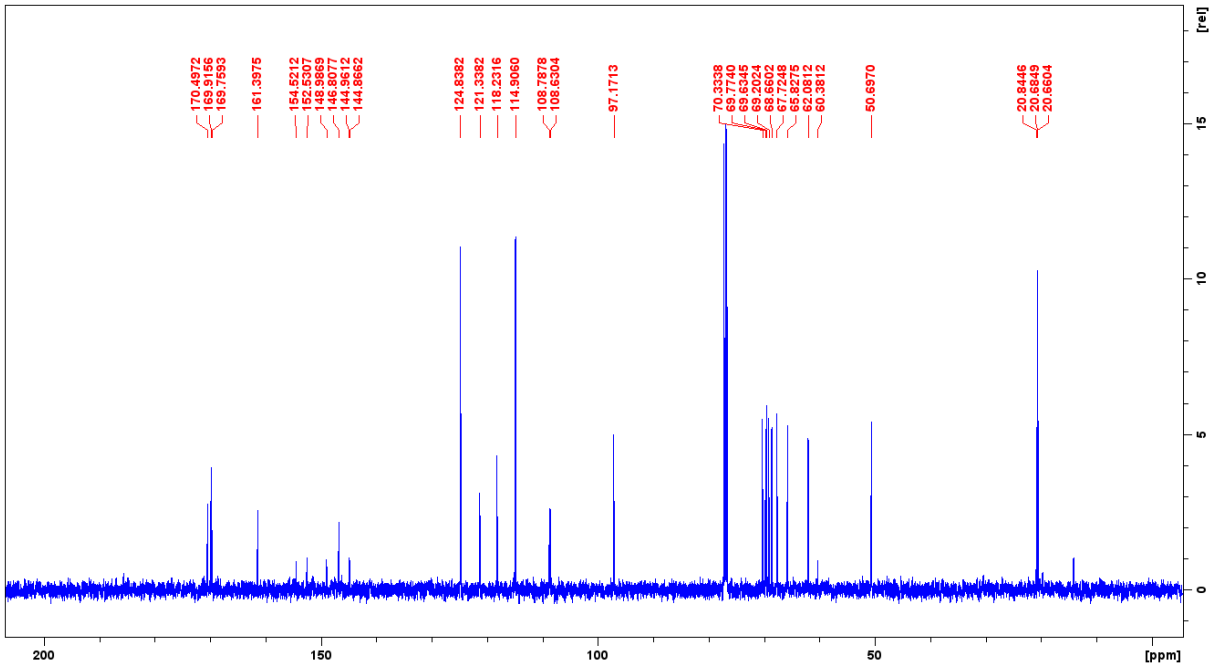


Figure 5.4-64: ^{13}C NMR spectrum of compound **38** (126 MHz, $CDCl_3$, 298 K).

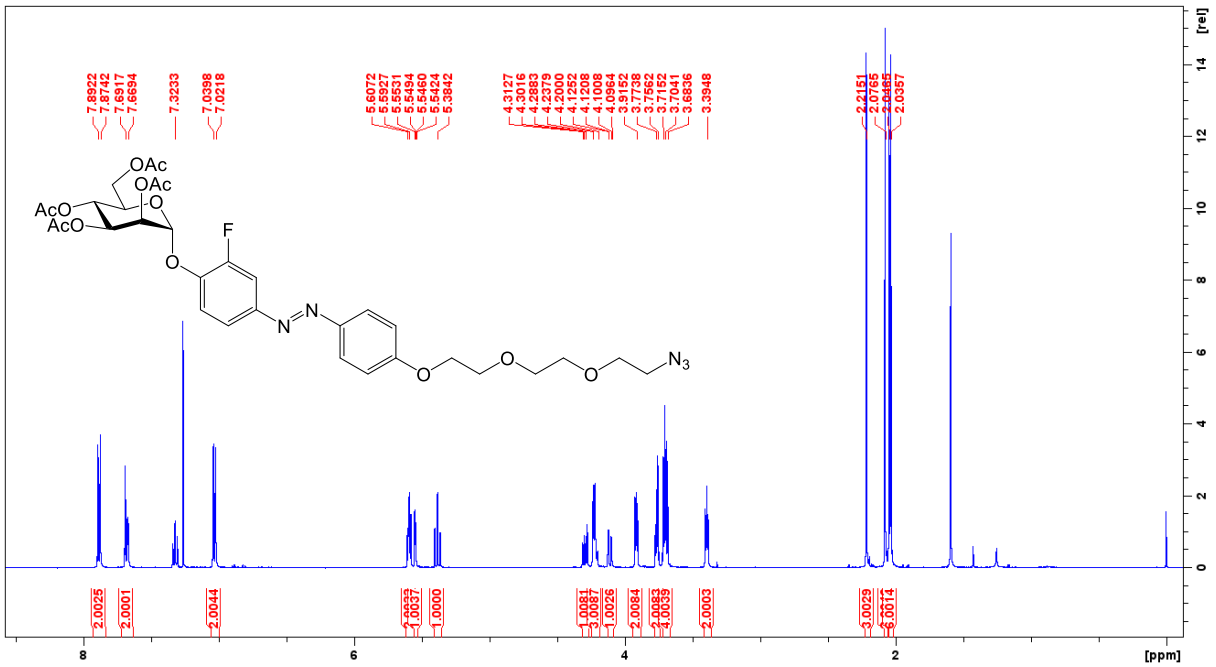


Figure 5.4-65: ^{1}H NMR spectrum of compound **39** (500 MHz, $CDCl_3$, 298 K).

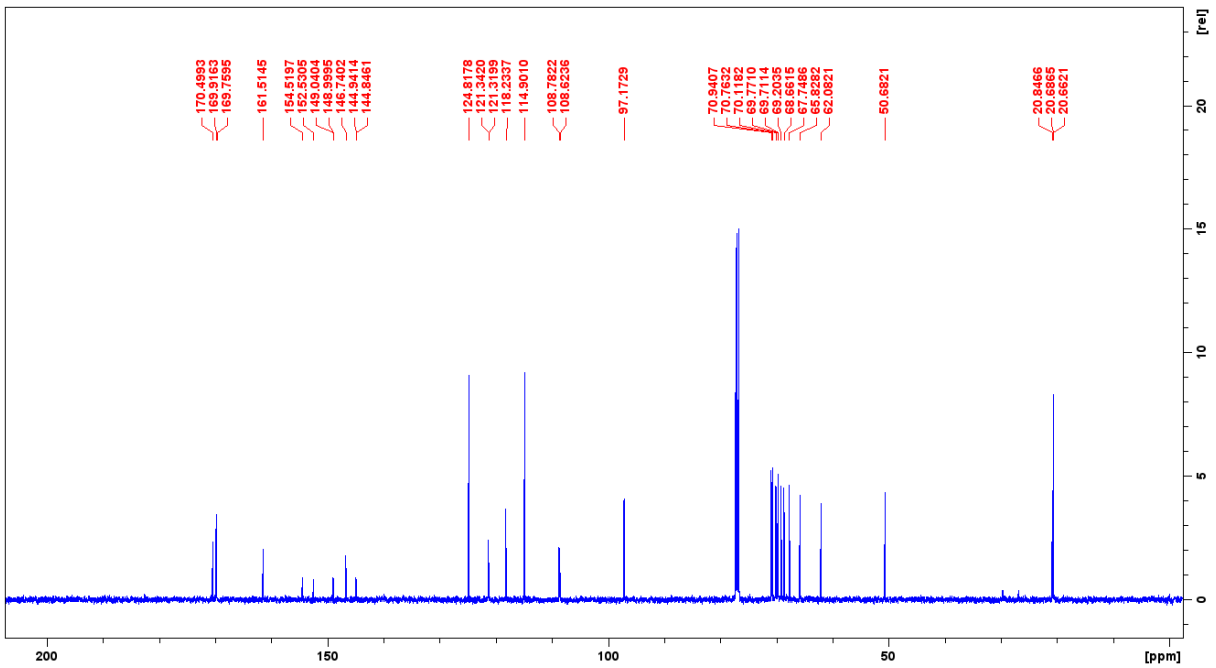


Figure 5.4-66: ^{13}C NMR spectrum of compound **39** (126 MHz, $CDCl_3$, 298 K).

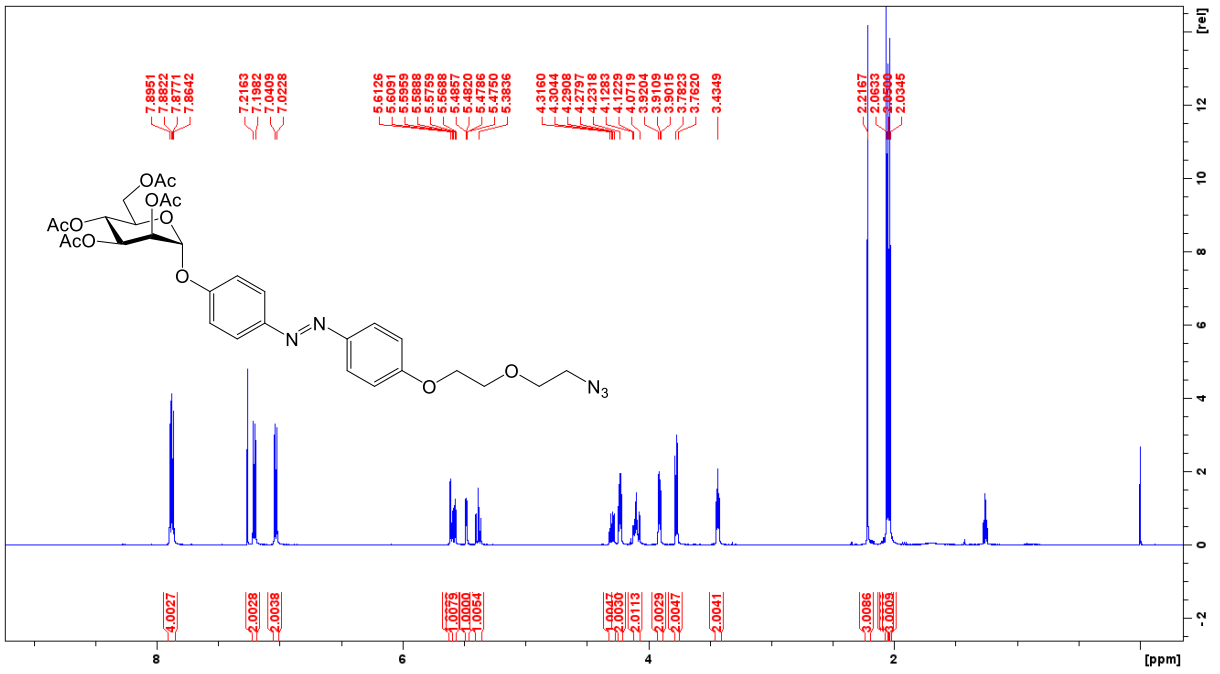


Figure 5.4-67: ^{1}H NMR spectrum of compound **40** (500 MHz, $CDCl_3$, 298 K).

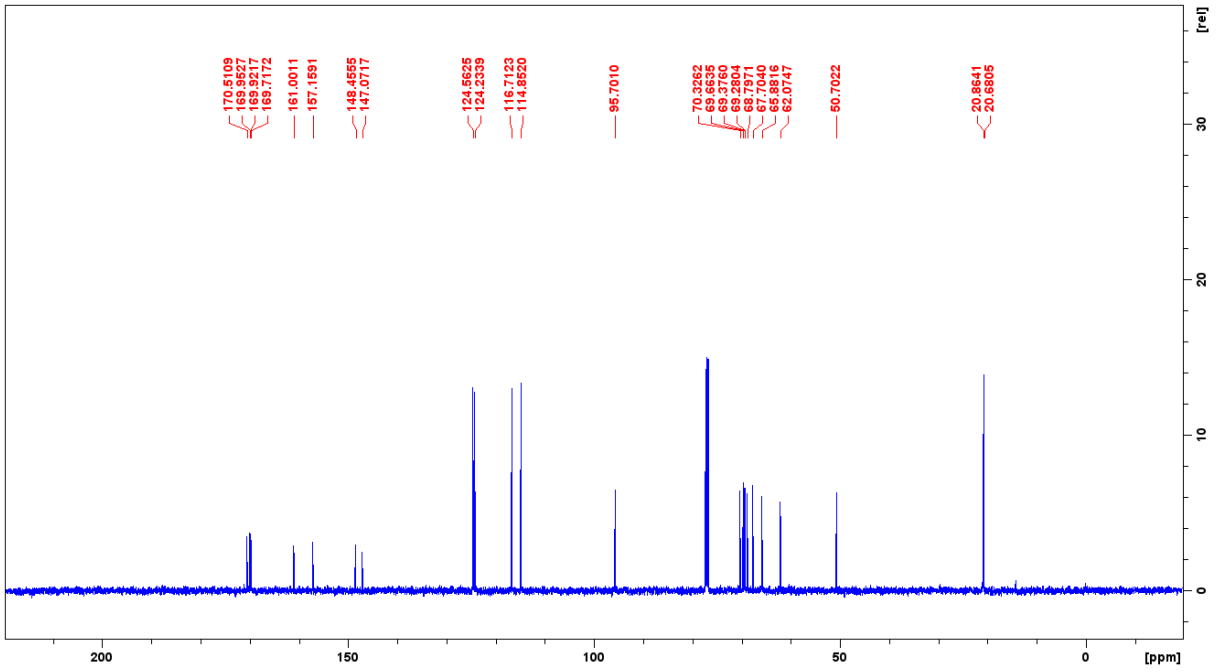


Figure 5.4-68: ^{13}C NMR spectrum of compound **40** (126 MHz, $CDCl_3$, 298 K).

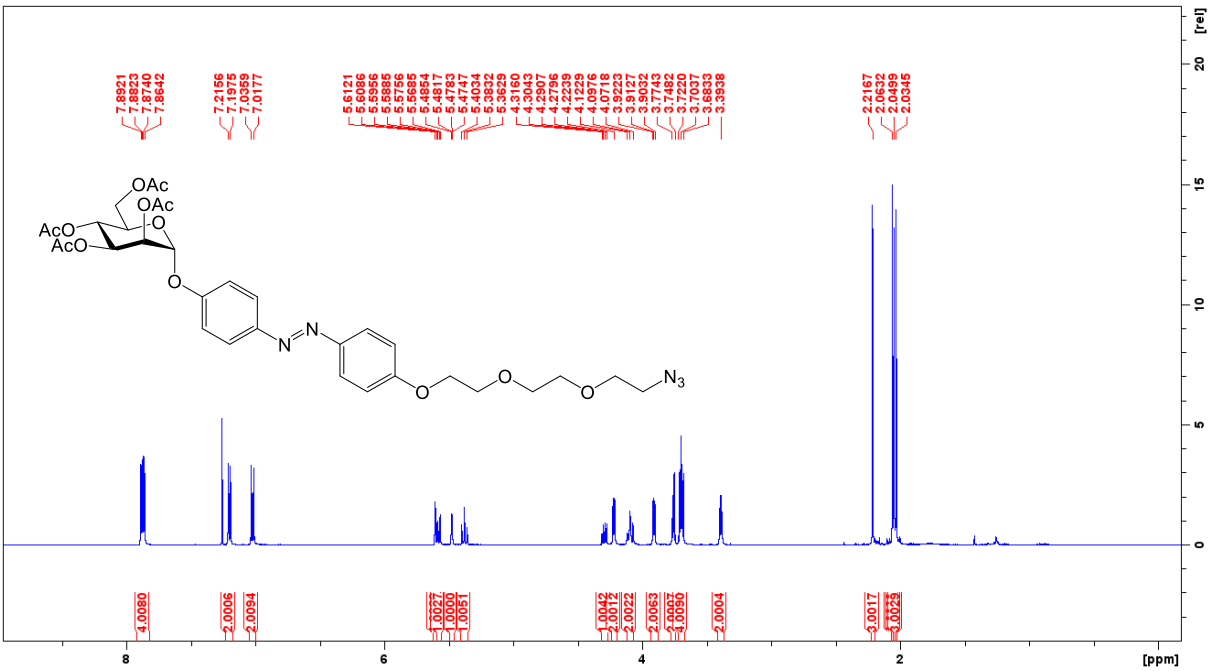


Figure 5.4-69: ^{1}H NMR spectrum of compound **41** (500 MHz, $CDCl_3$, 298 K).

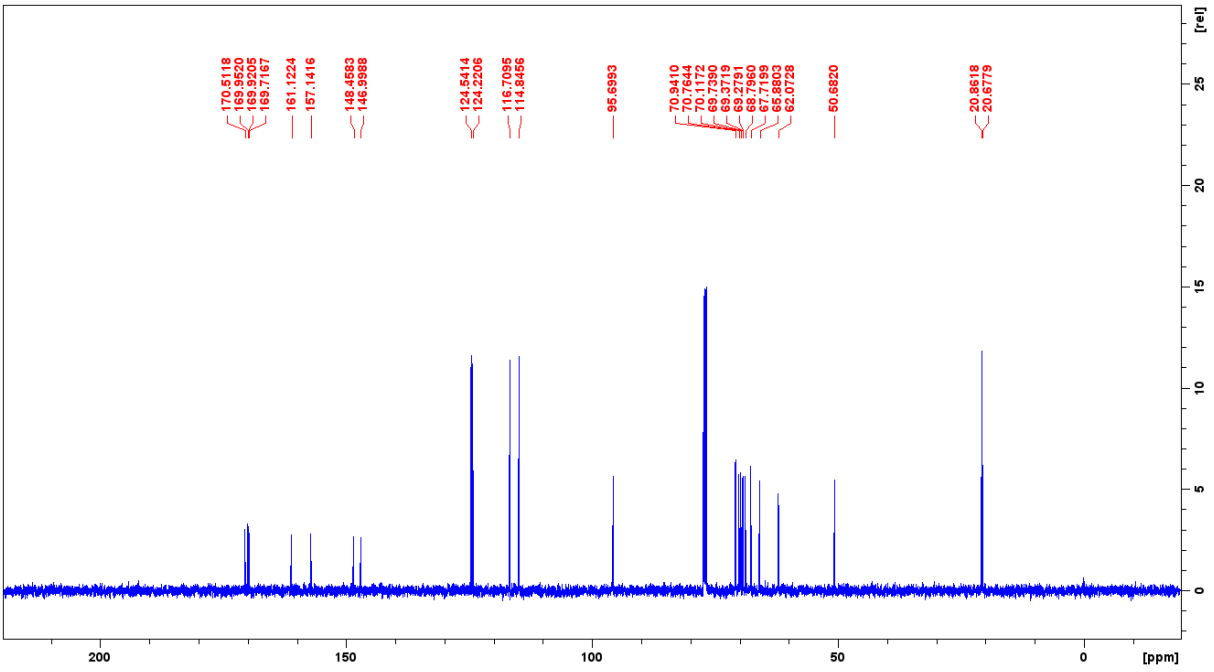


Figure 5.4-70: ^{13}C NMR spectrum of compound **41** (126 MHz, $CDCl_3$, 298 K).

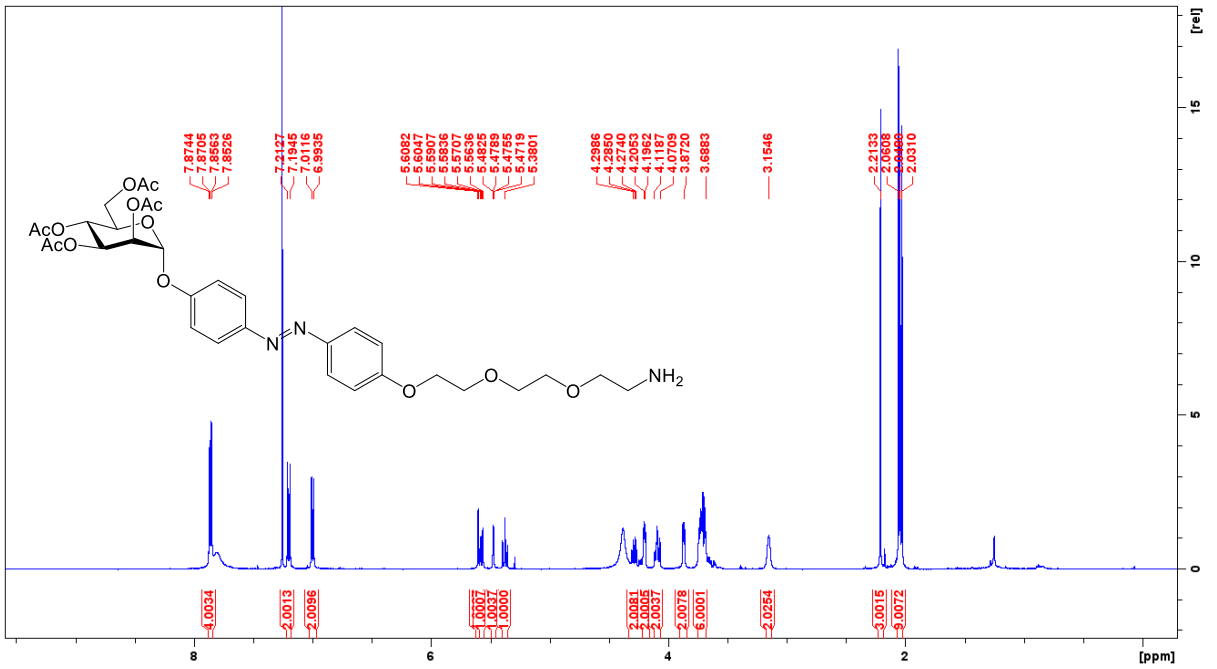


Figure 5.4-71: ^{1}H NMR spectrum of crude compound **43** (500 MHz, $CDCl_3$, 298 K).

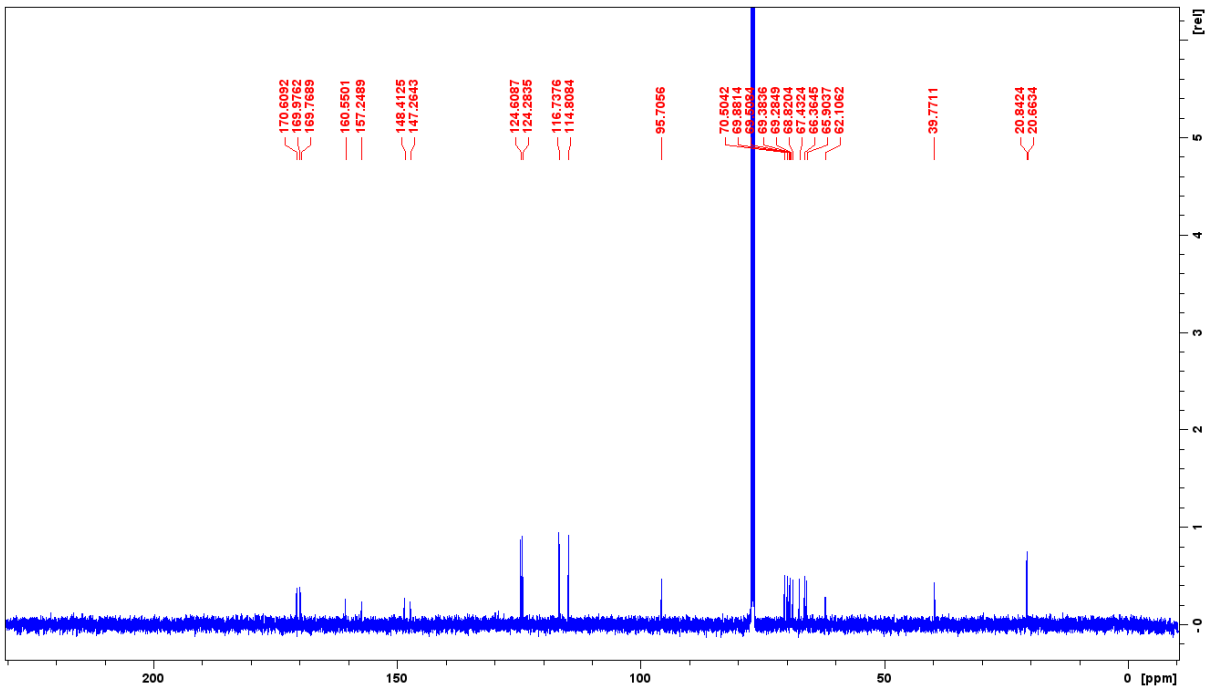


Figure 5.4-72: ^{13}C NMR spectrum of crude compound **43** (126 MHz, $CDCl_3$, 298 K).

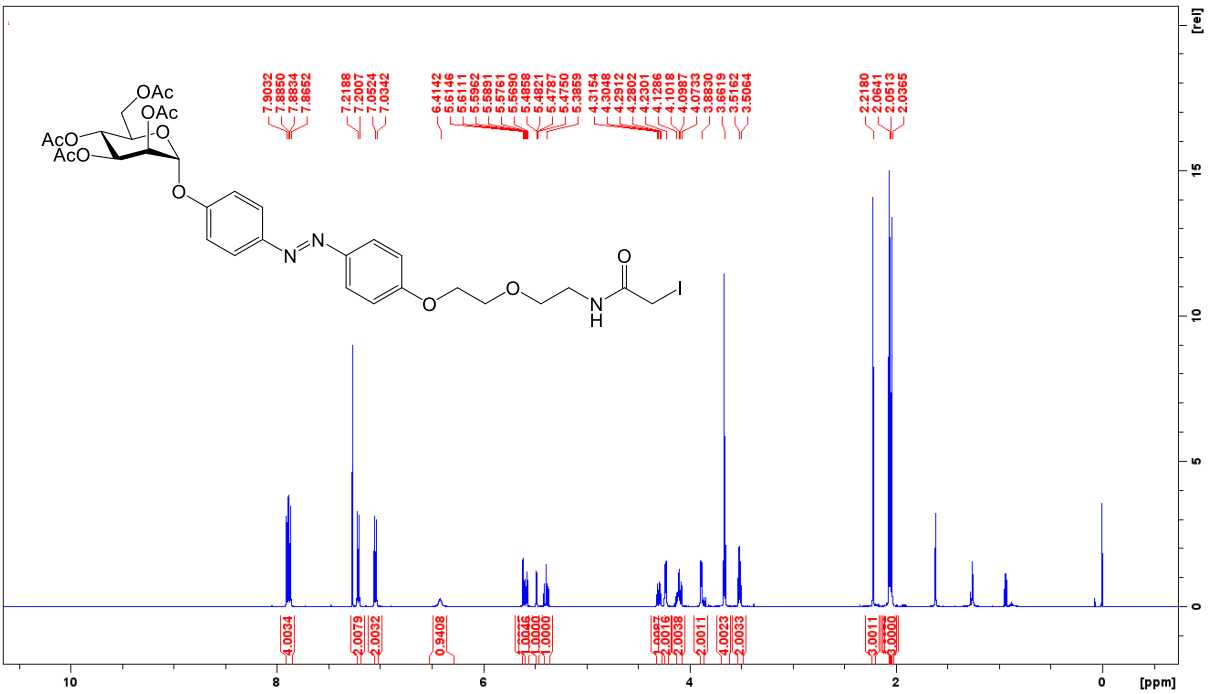


Figure 5.4-73: ^{1}H NMR spectrum of compound **44** (500 MHz, $CDCl_3$, 298 K).

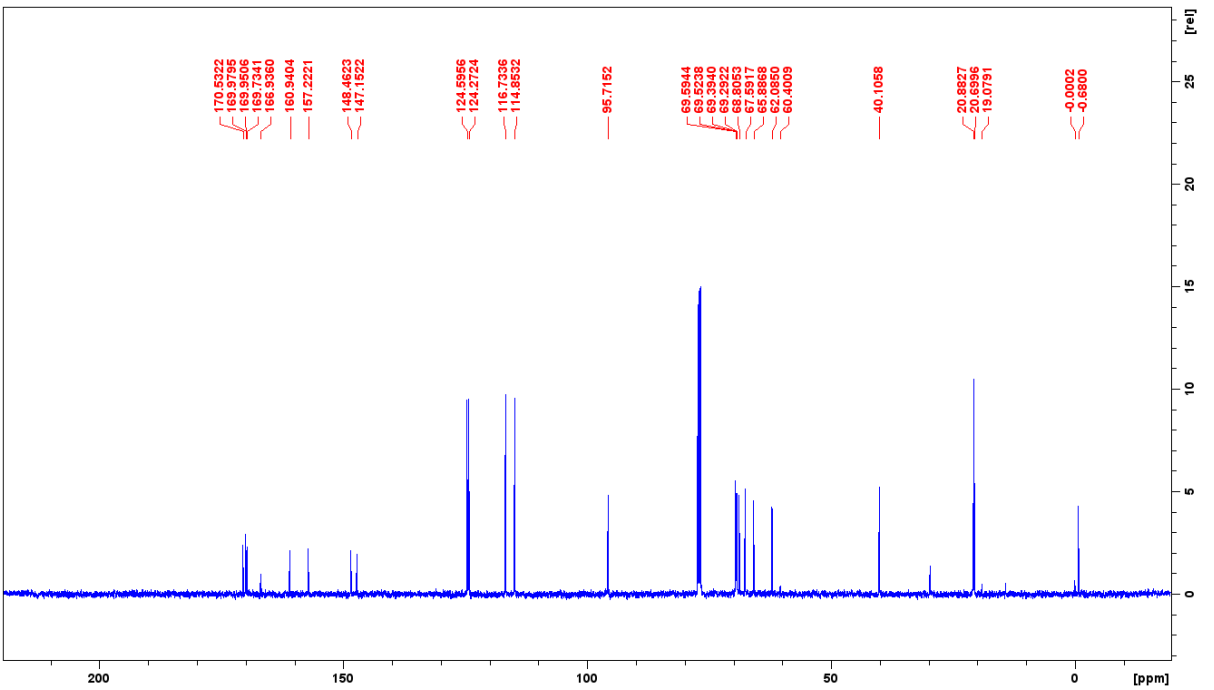


Figure 5.4-74: ^{13}C NMR spectrum of crude compound **44** (126 MHz, $CDCl_3$, 298 K).

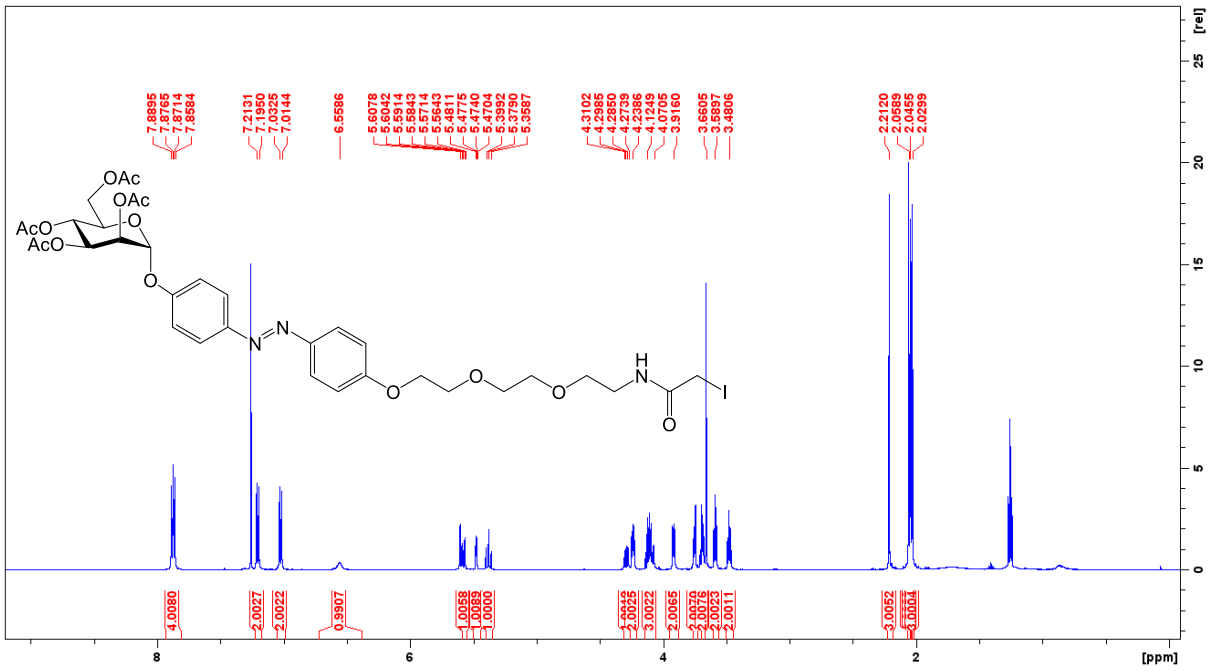


Figure 5.4-75: ^{1}H NMR spectrum of compound **45** (500 MHz, $CDCl_3$, 298 K).

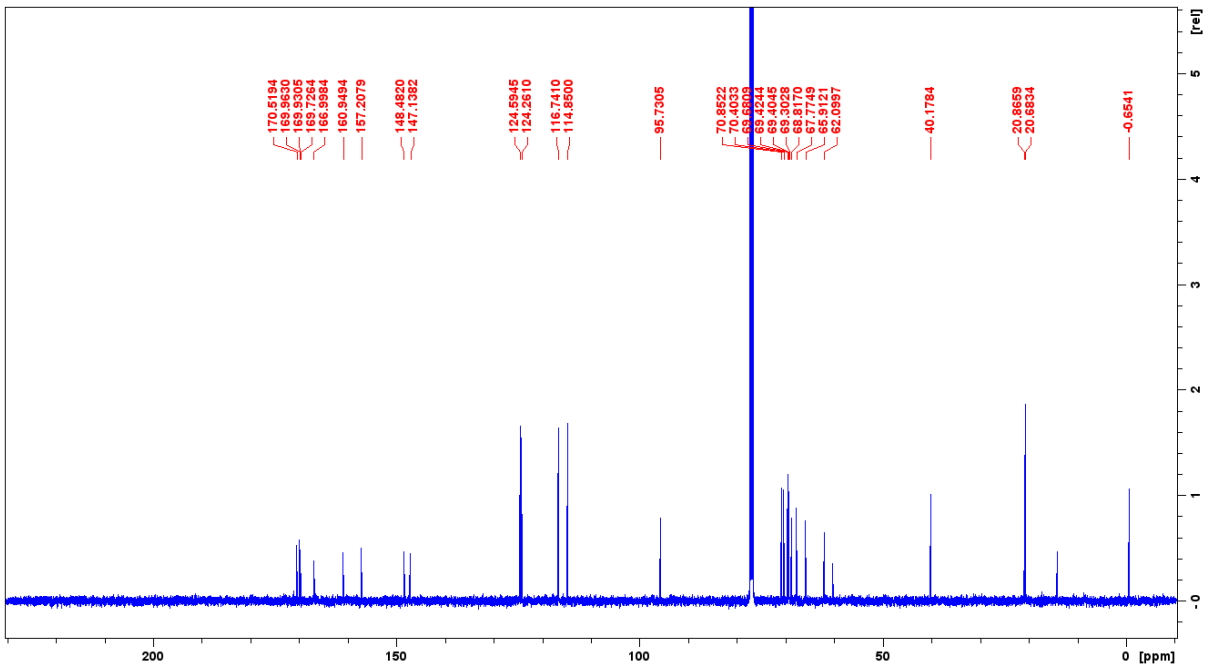


Figure 5.4-76: ^{13}C NMR spectrum of compound **45** (126 MHz, $CDCl_3$, 298 K).

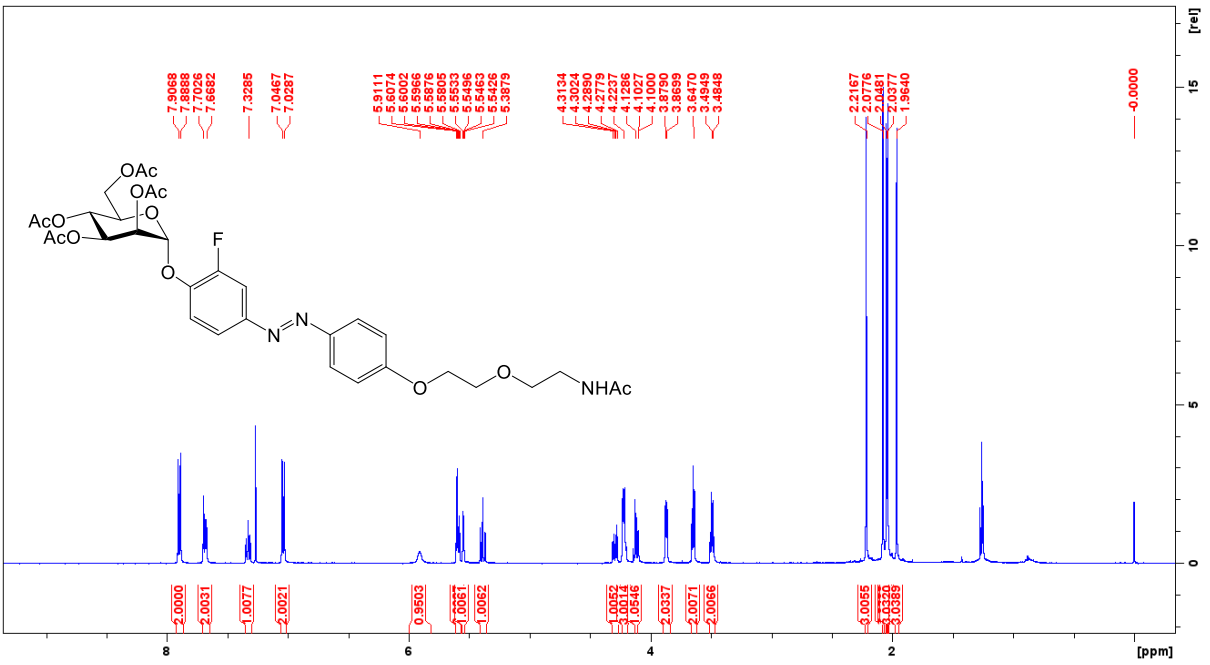


Figure 5.4-77: ^{1}H NMR spectrum of compound **46** (500 MHz, $CDCl_3$, 298 K).

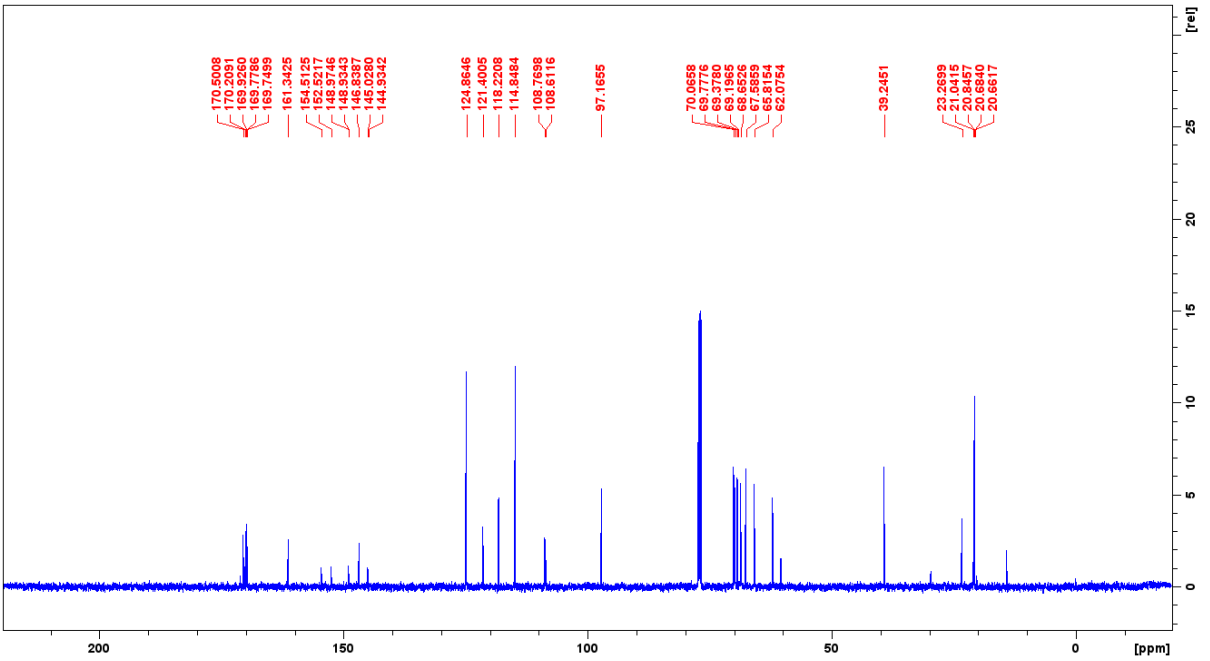


Figure 5.4-78: ^{13}C NMR spectrum of compound **46** (126 MHz, $CDCl_3$, 298 K).

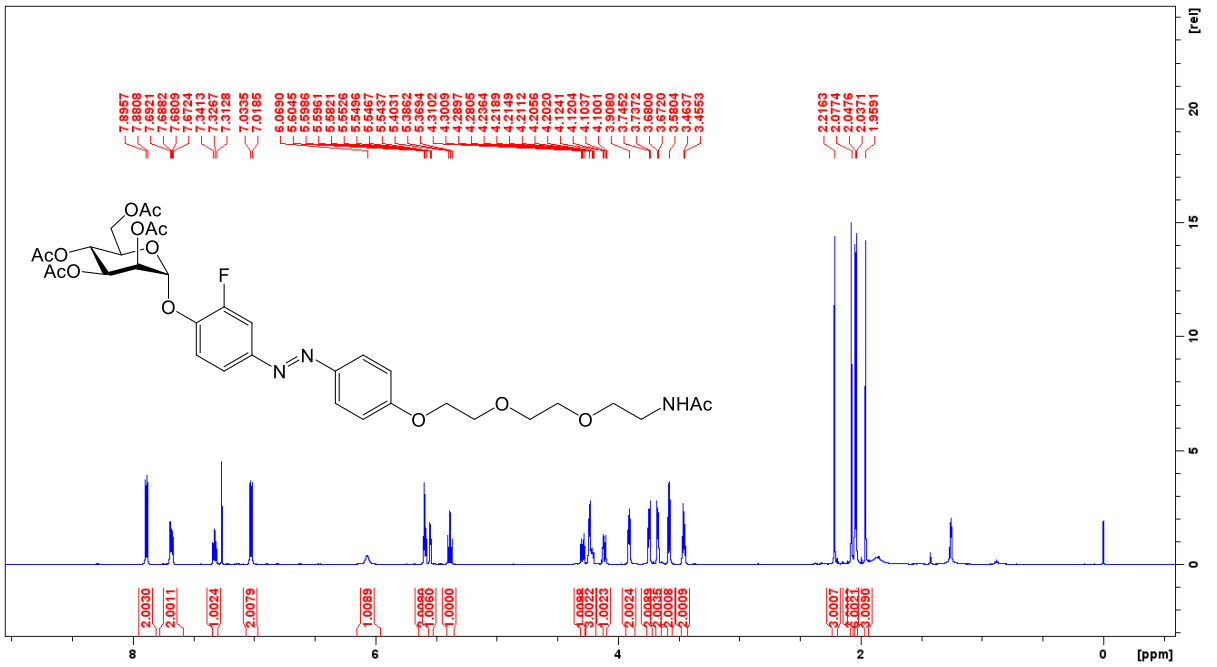


Figure 5.4-79: ^{1}H NMR spectrum of compound **47** (600 MHz, $CDCl_3$, 298 K).

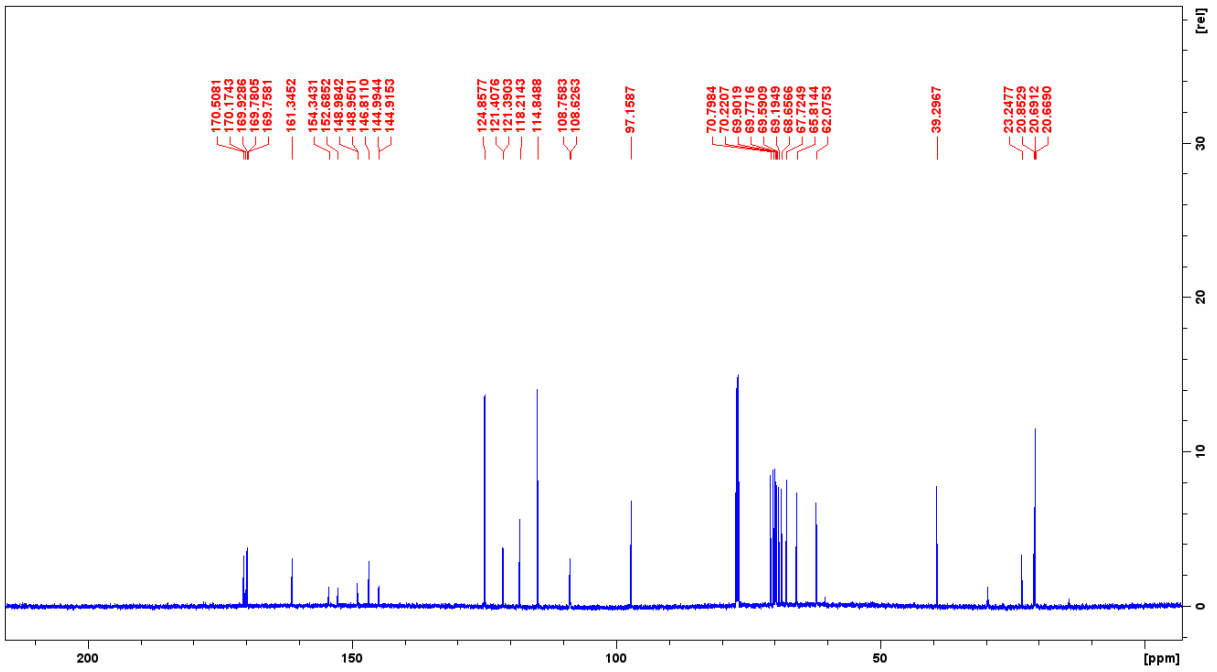


Figure 5.4-80: ^{13}C NMR spectrum of compound **47** (151 MHz, CD_3OD, 298 K).

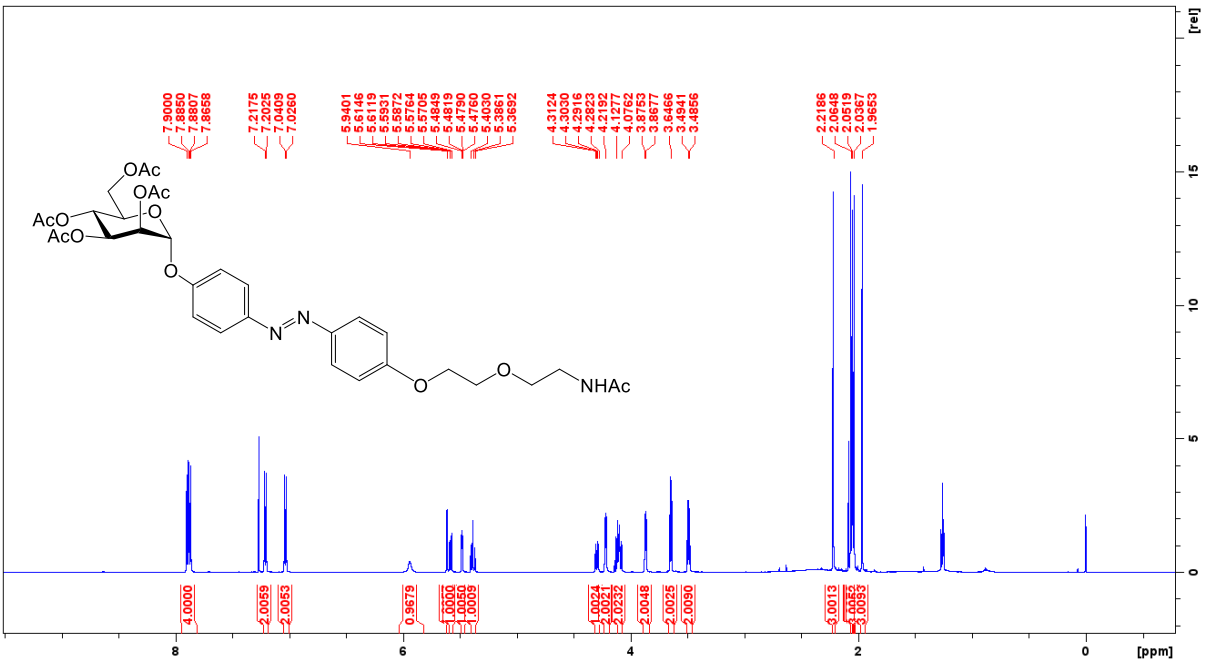


Figure 5.4-81: ^{1}H NMR spectrum of compound **48** (500 MHz, $CDCl_3$, 298 K).

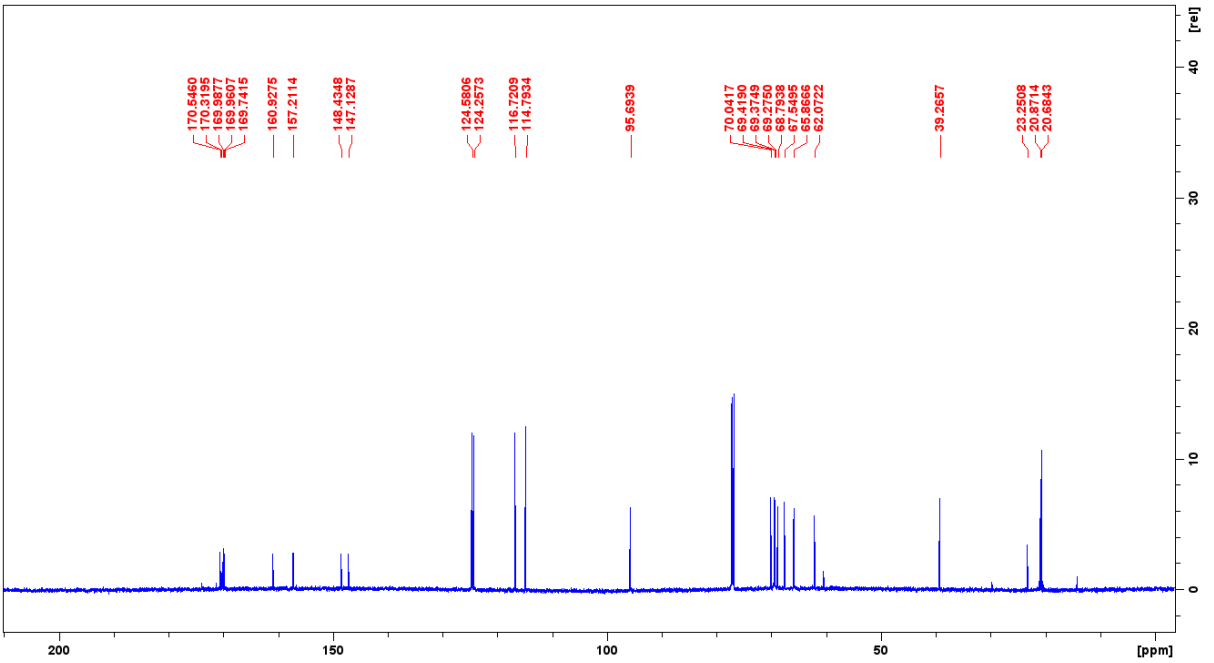


Figure 5.4-82: ^{13}C NMR spectrum of compound **48** (126 MHz, $CDCl_3$, 298 K).

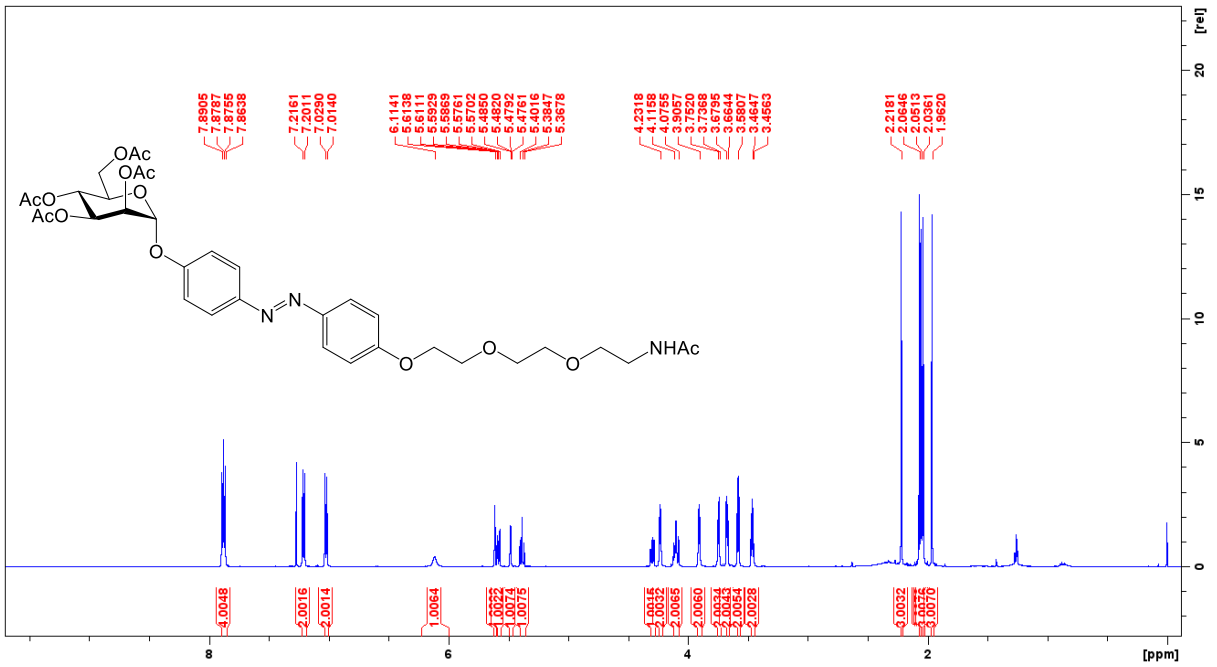


Figure 5.4-83: ^{1}H NMR spectrum of compound **49** (600 MHz, $CDCl_3$, 298 K).

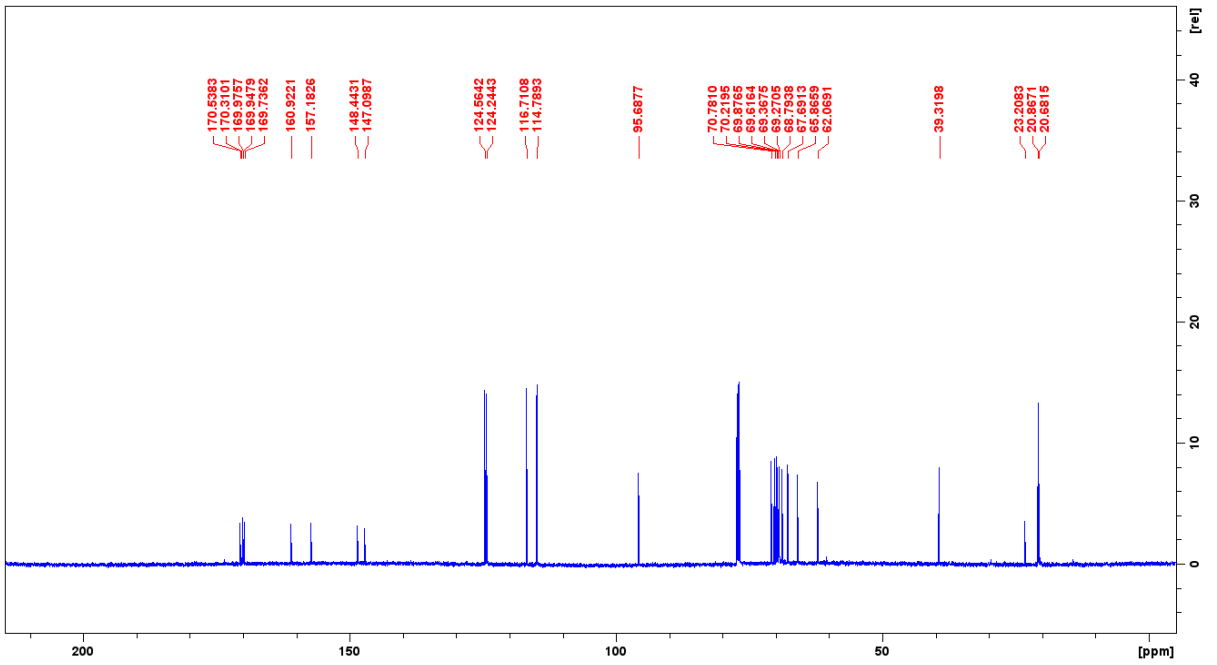


Figure 5.4-84: ^{13}C NMR spectrum of compound **49** (151 MHz, $CDCl_3$, 298 K).

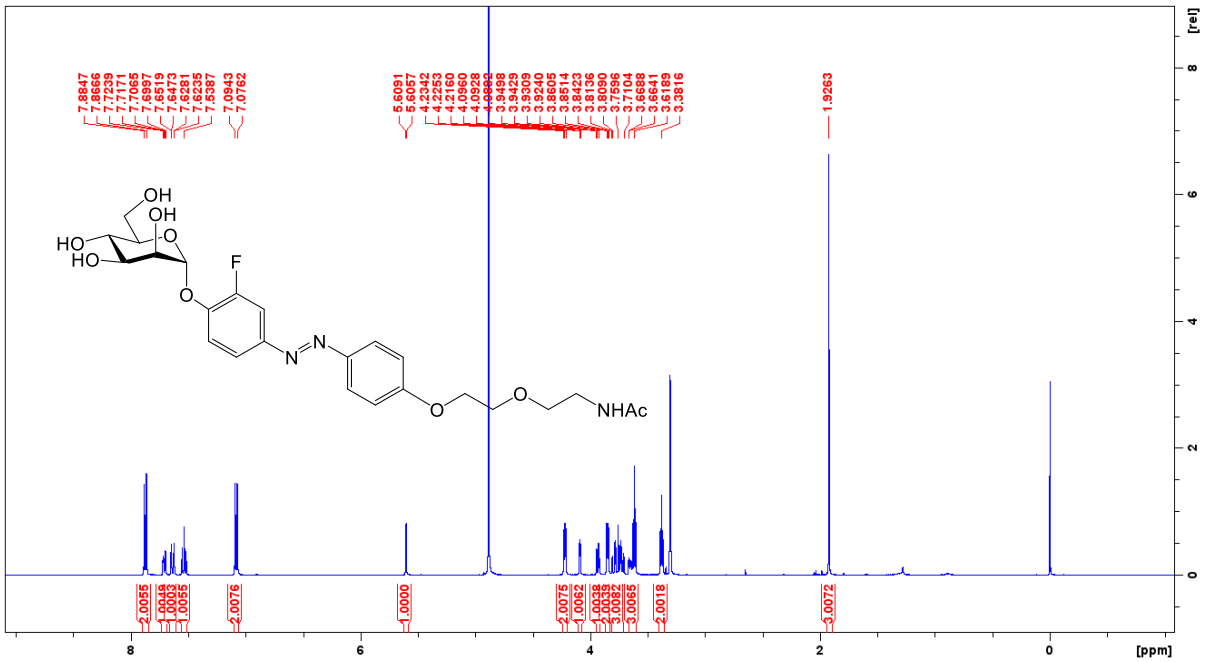


Figure 5.4-85: ^{1}H NMR spectrum of compound **51** (500 MHz, CD_3OD, 298 K).

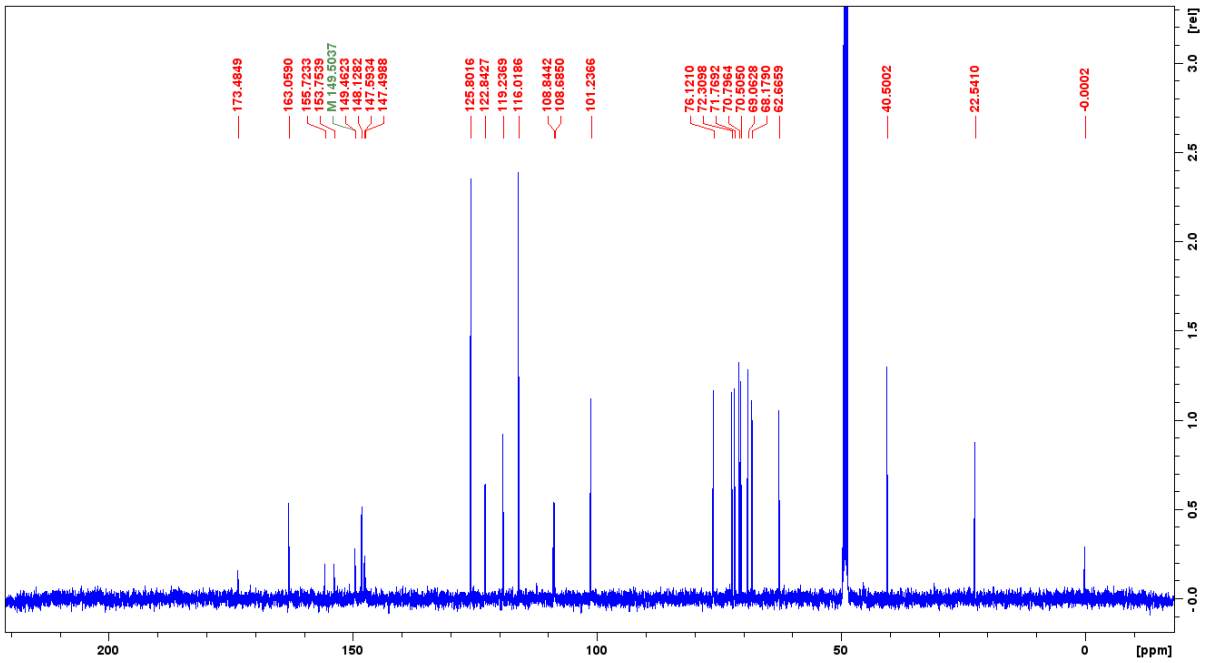


Figure 5.4-86: ^{13}C NMR spectrum of compound **51** (126 MHz, CD_3OD, 298 K).
The second signal of the doublet at 149.5 ppm (green) has been manually amended.

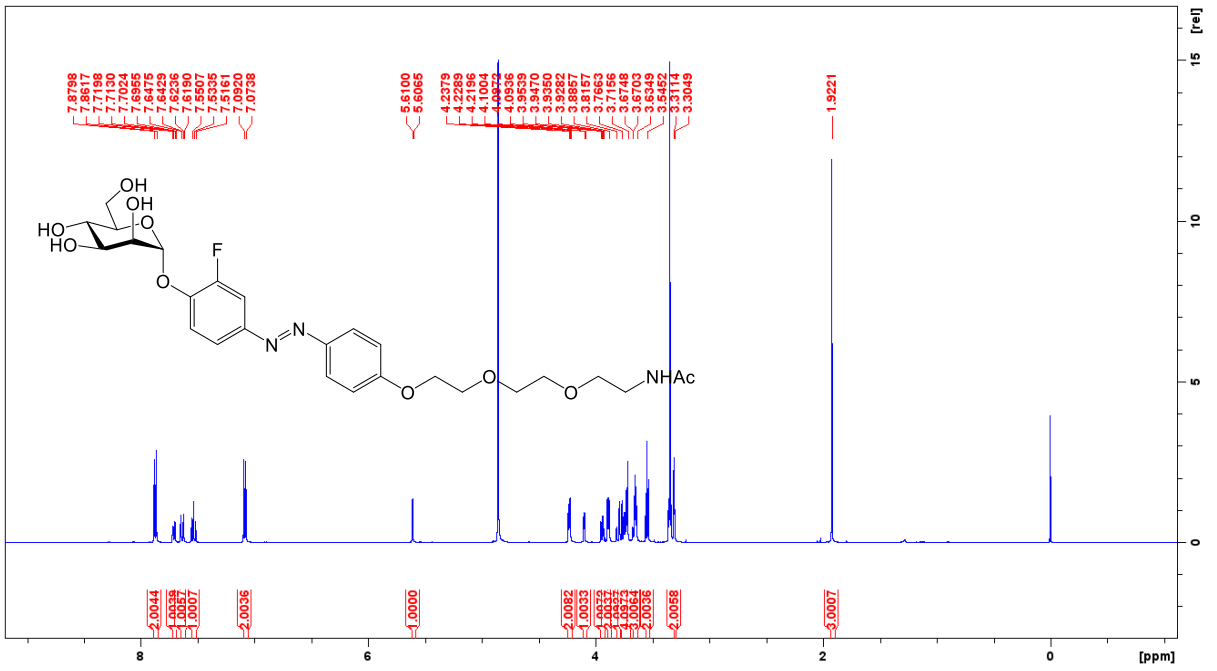


Figure 5.4-87: ^{1}H NMR spectrum of compound **52** (500 MHz, $CD3_3OD$, 298 K).

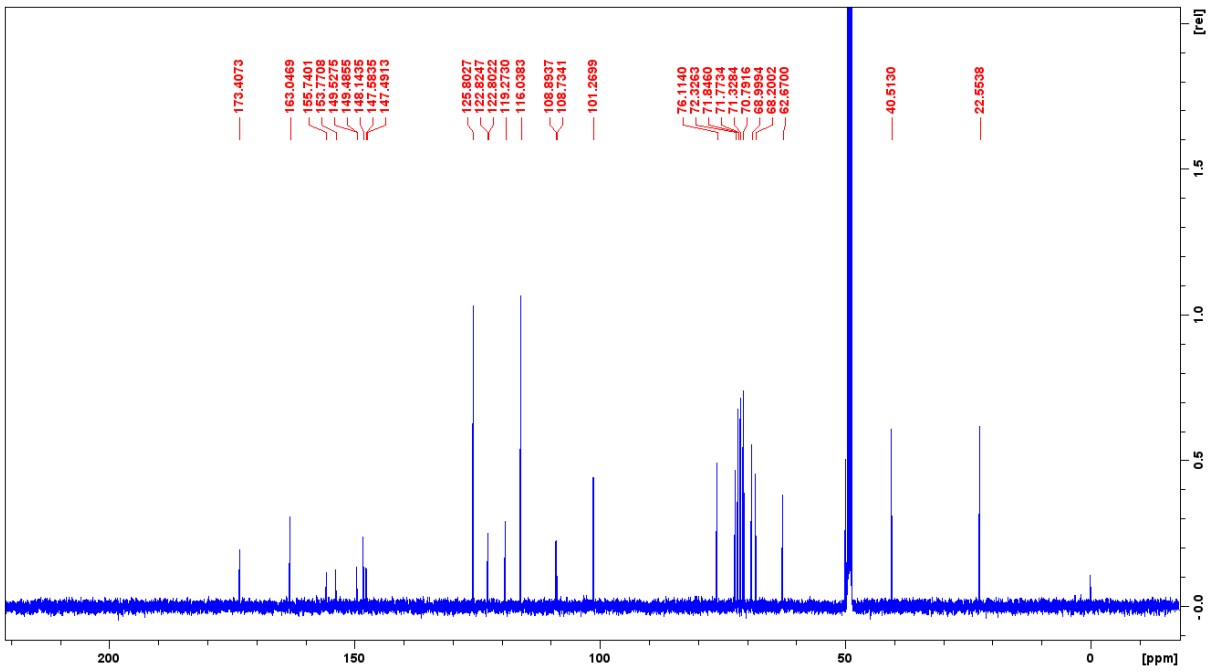


Figure 5.4-88: ^{13}C NMR spectrum of compound **52** (126 MHz, CD_3OD, 298 K).

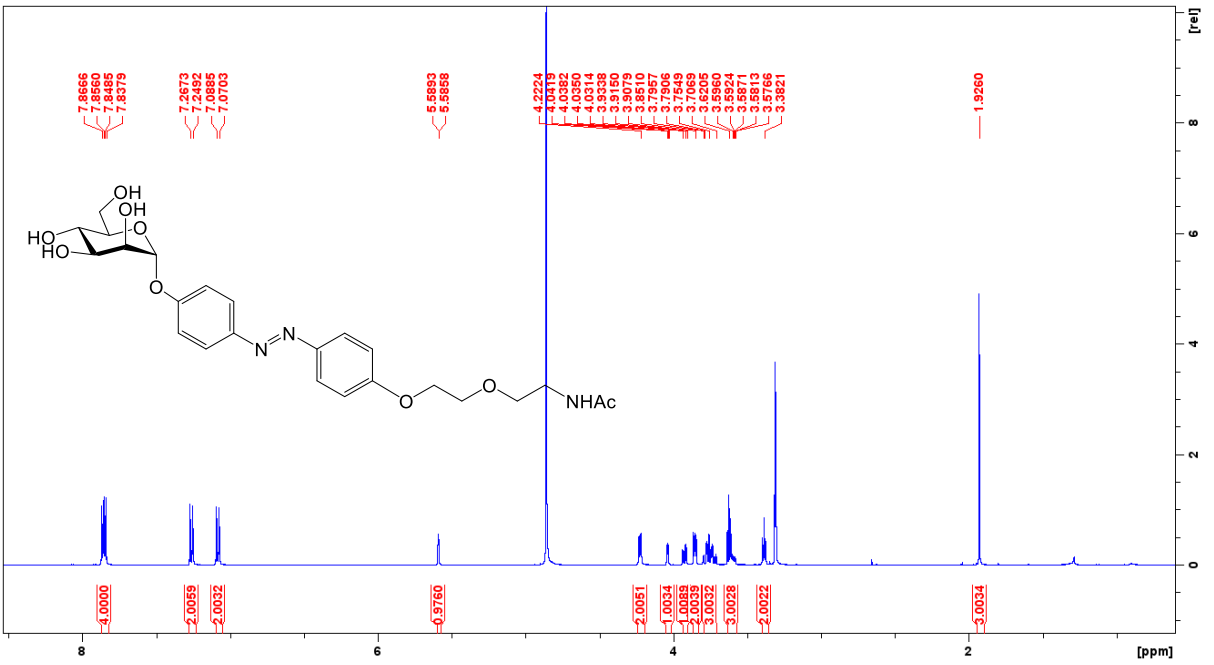


Figure 5.4-89: ^{1}H NMR spectrum of compound **53** (500 MHz, CD_3OD, 298 K).

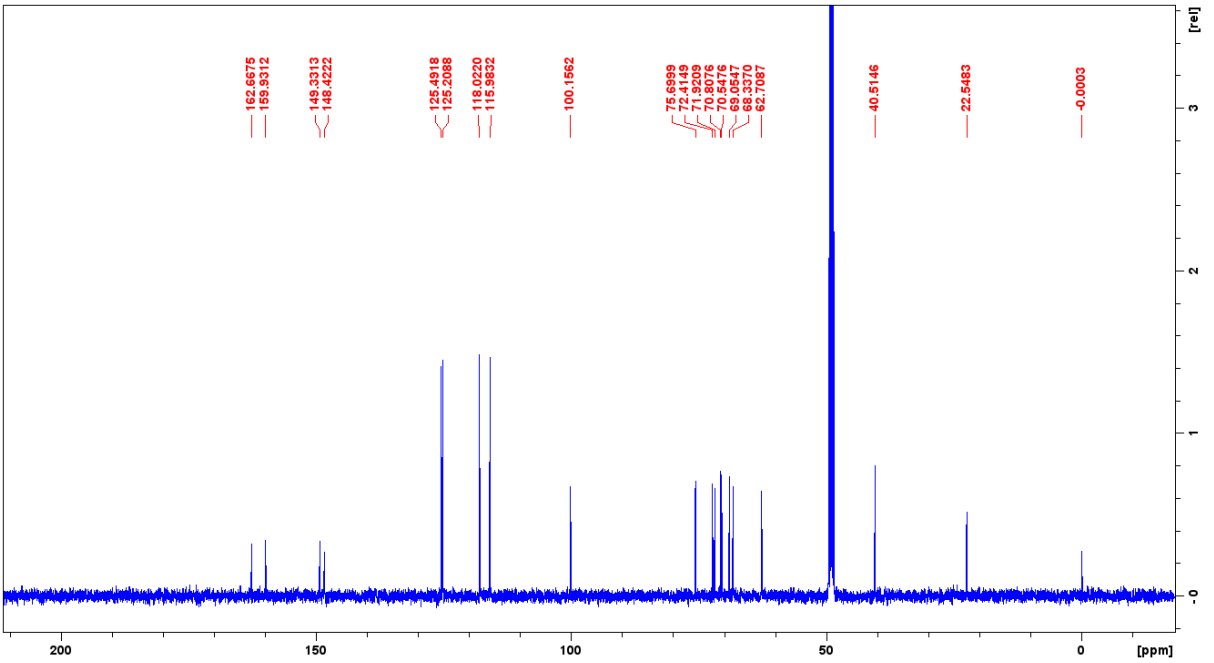


Figure 5.4-90: ^{13}C NMR spectrum of compound **53** (126 MHz, CD_3OD, 298 K).

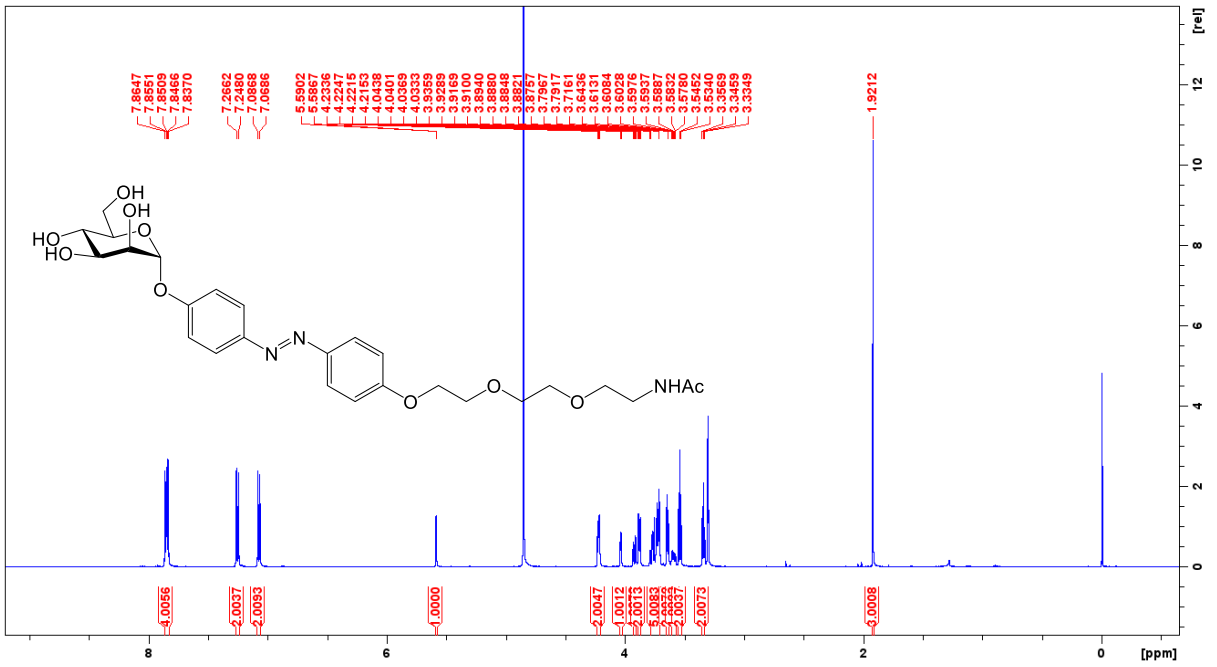


Figure 5.4-91: ^{1}H NMR spectrum of compound **54** (500 MHz, CD_3OD, 298 K).

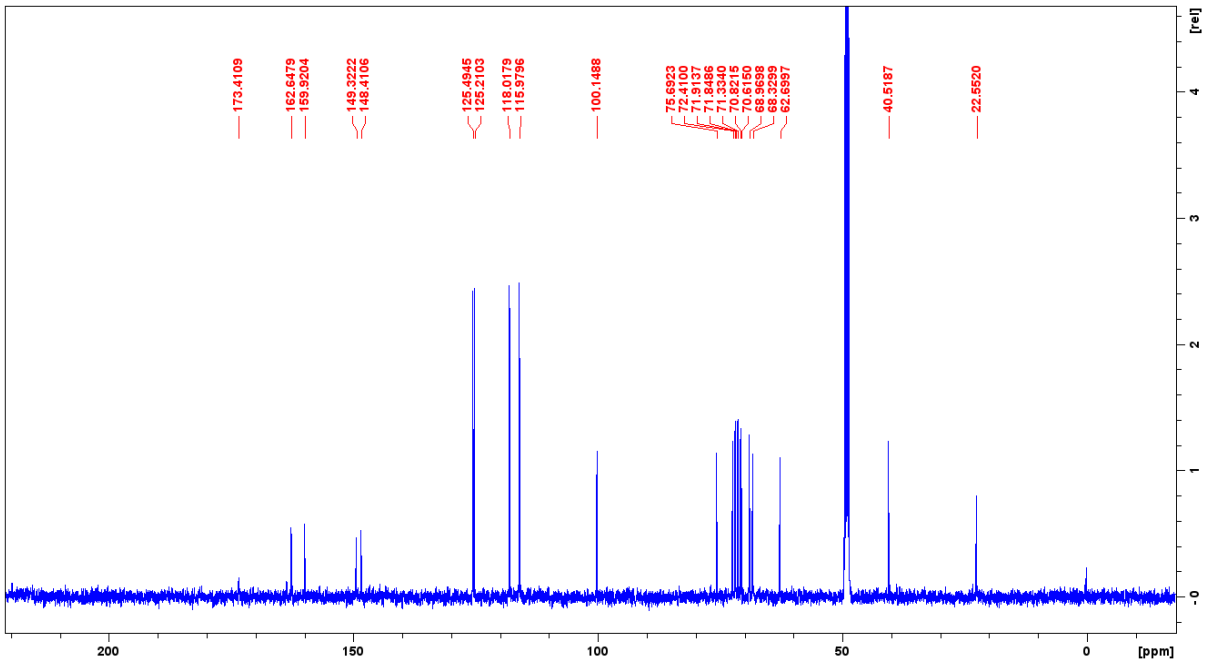


Figure 5.4-92: ^{13}C NMR spectrum of compound **54** (126 MHz, CD_3OD, 298 K).

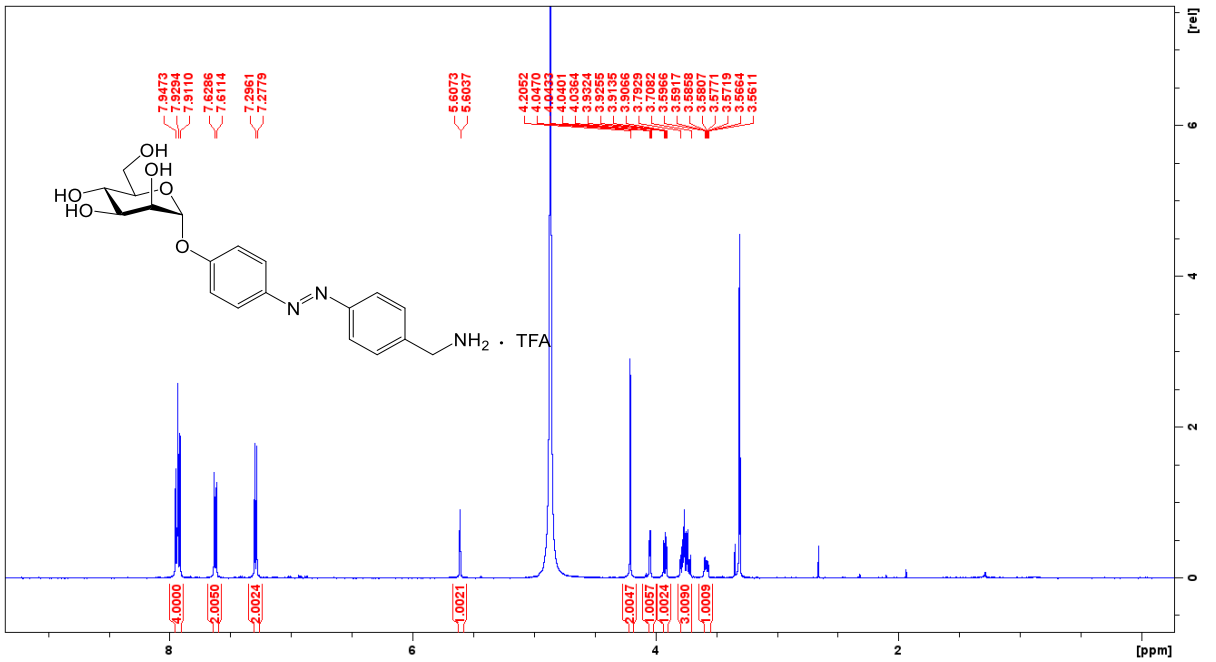


Figure 5.4-93: ^{1}H NMR spectrum of compound **56** (500 MHz, CD_3OD, 298 K).

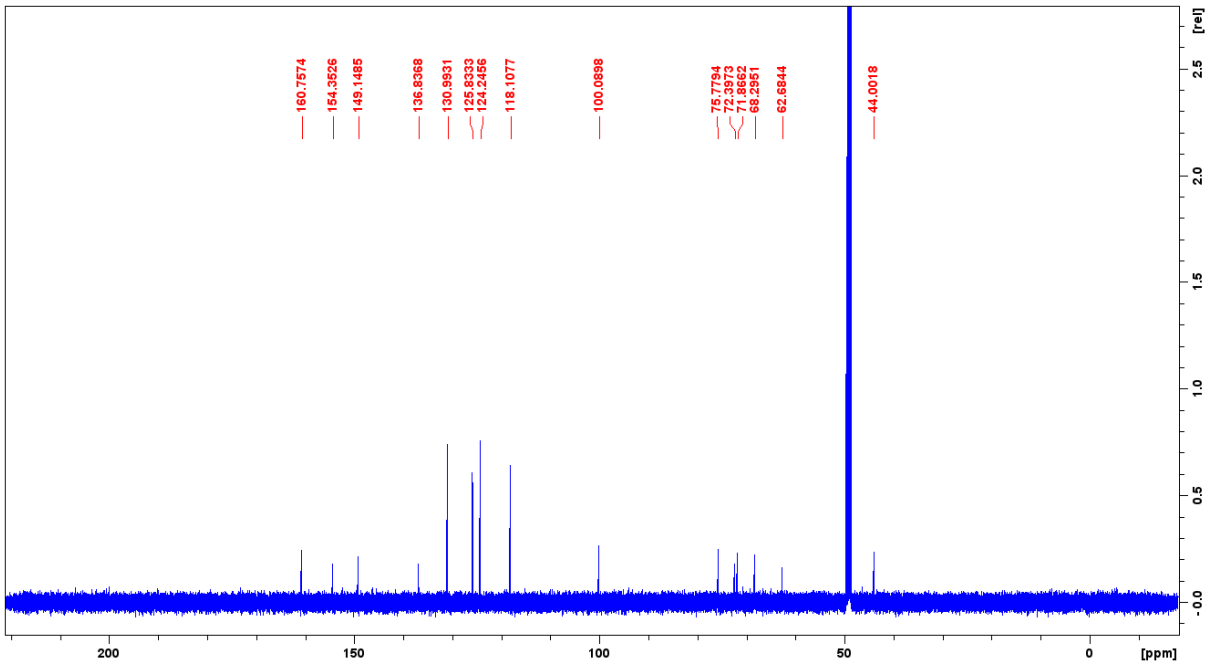


Figure 5.4-94: ^{13}C NMR spectrum of compound **56** (126 MHz, CD_3OD, 298 K).

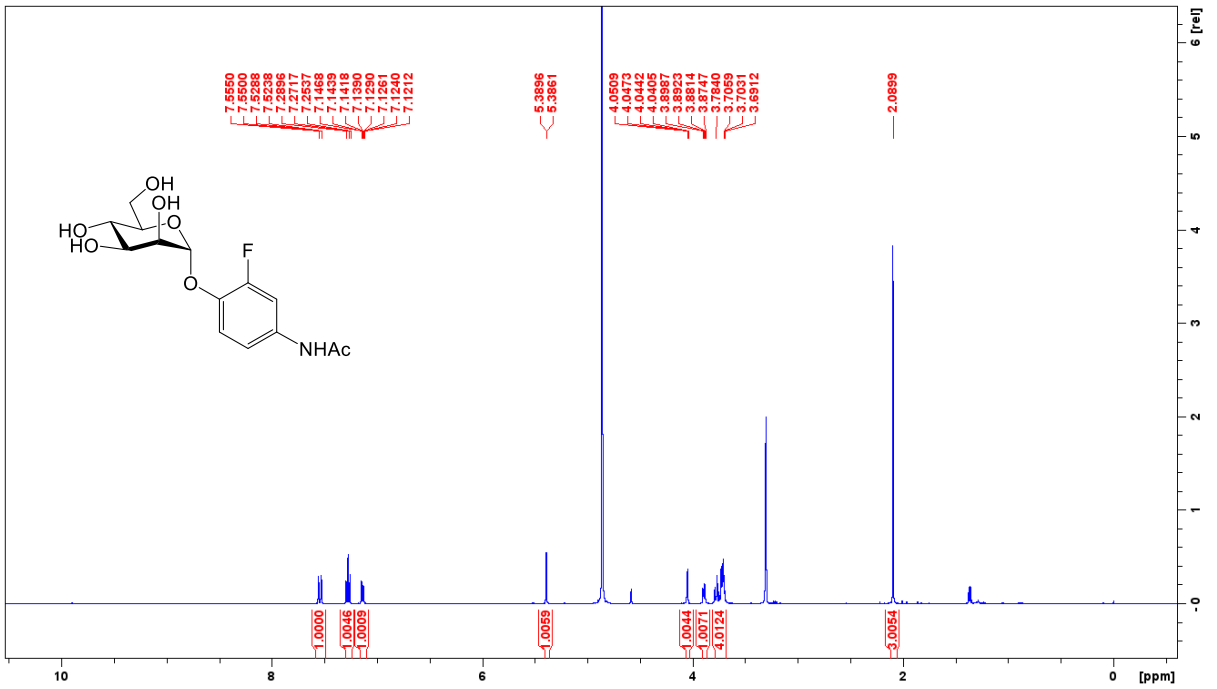


Figure 5.4-95: ^{1}H NMR spectrum of compound **57** (500 MHz, CD_3OD, 298 K).

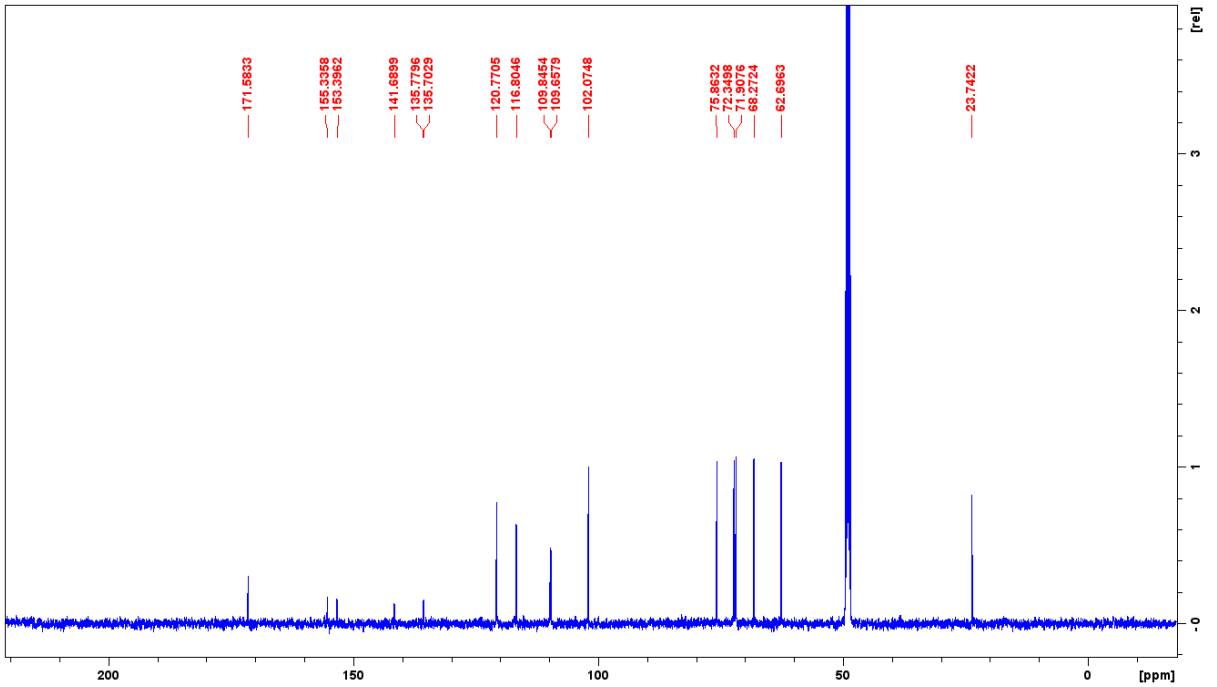


Figure 5.4-96: ^{13}C NMR spectrum of compound **57** (126 MHz, CD_3OD, 298 K).

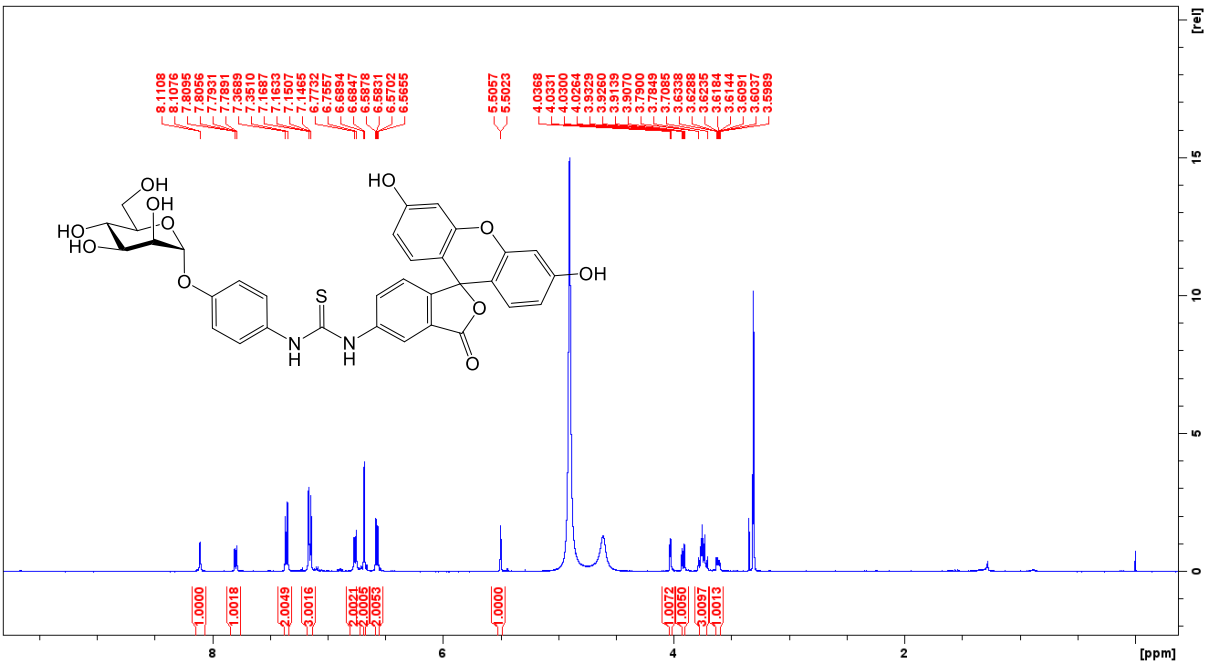


Figure 5.4-97: ^{1}H NMR spectrum of compound **59** (500 MHz, CD_3OD, 298 K).

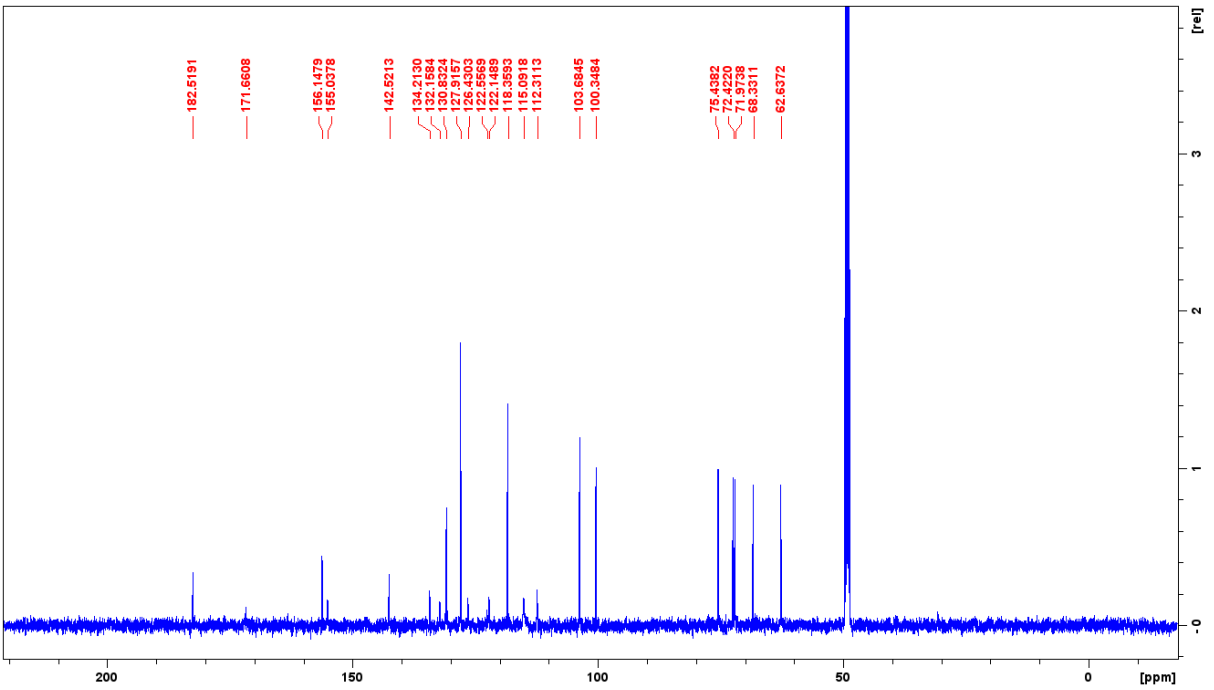


Figure 5.4-98: ^{13}C NMR spectrum of compound **59** (126 MHz, CD_3OD, 298 K).

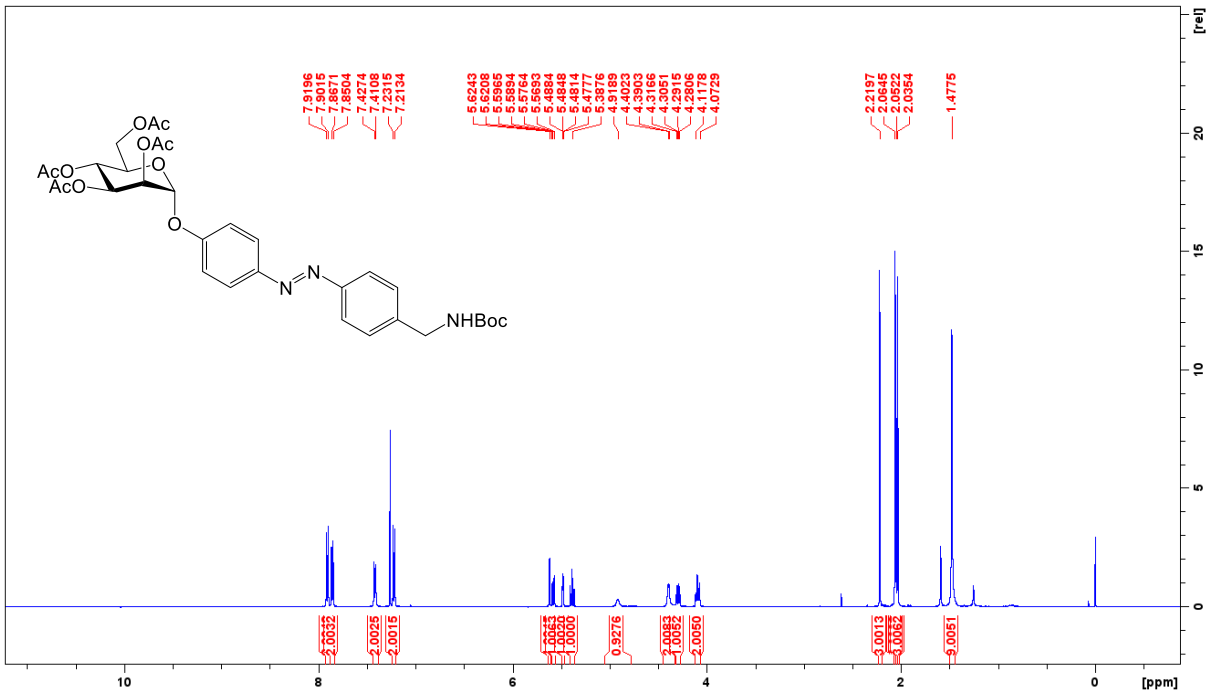


Figure 5.4-99: ^{1}H NMR spectrum of compound **66** (500 MHz, $CDCl_3$, 298 K).

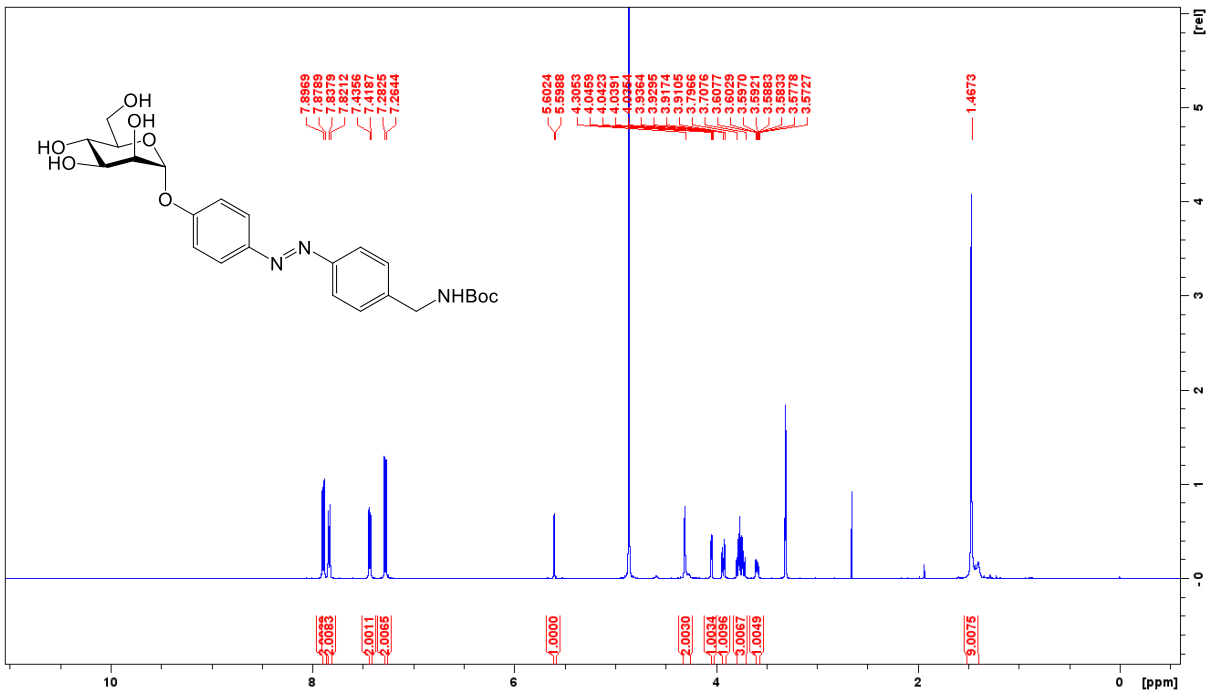


Figure 5.4-100: ^{1}H NMR spectrum compound **67** (500 MHz, CD_3OD, 298 K).

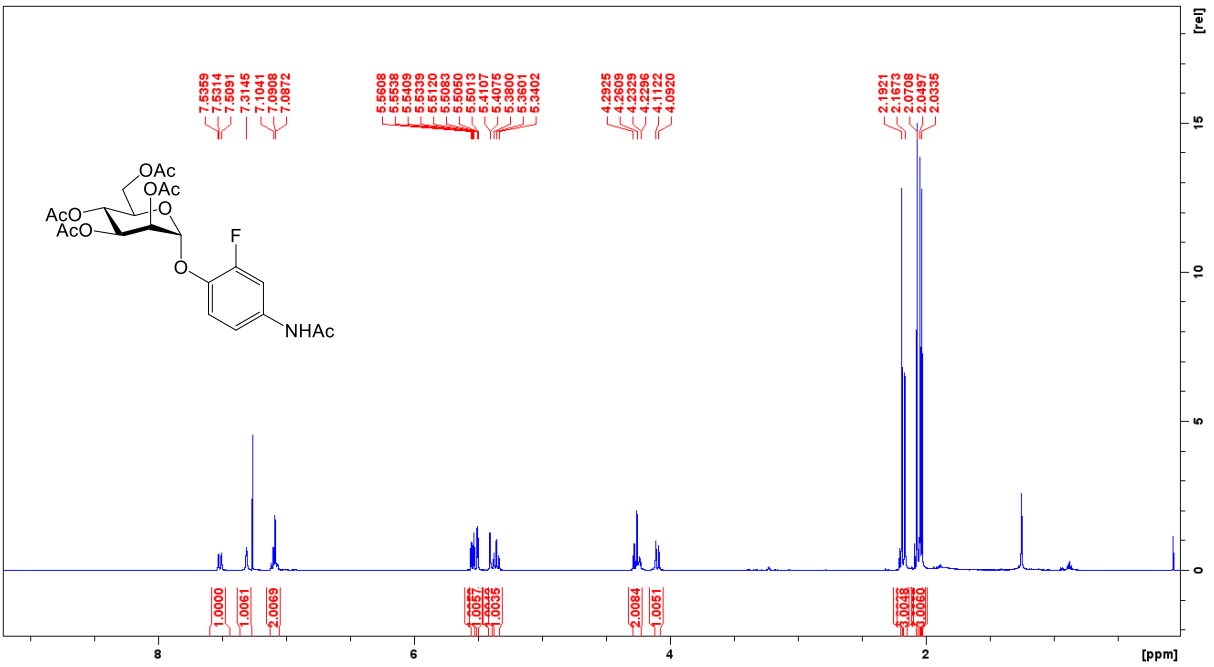


Figure 5.4-101: ^{1}H NMR spectrum of compound **68** (500 MHz, $CDCl_3OD$, 298 K)

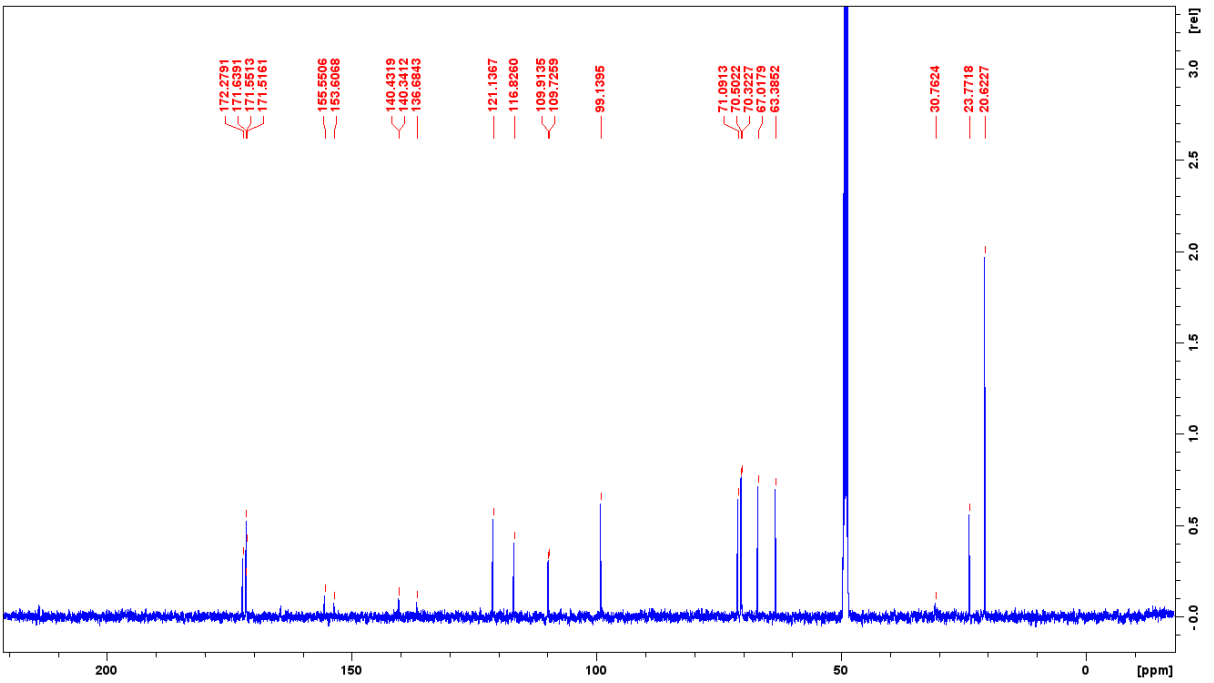


Figure 5.4-102: ^{13}C NMR spectrum of compound **68** (126 MHz, CD_3OD, 298 K).

5.5 Investigation of photochromic properties

5.5.1 Photoirradiation experiments

For photoirradiation experiments of azobenzene derivatives, the samples were irradiated at the appropriate wavelength for 3 – 10 min (unless otherwise stated) at a distance of about 1 – 3 cm from the light source. For *trans* → *cis* photoisomerisation, a UV light emitting LED NC4U133AT (λ = 365 nm, Nichia Corporation, 10 LEDs/device) with a full width at half maximum (FWHW) of 10 nm and an intensity of 235 mW/cm² per LED was used. The *cis* → *trans* photoisomerisation was performed, using a blue light emitting LED (λ = 440 nm, Nichia Corporation, 10 LEDs/device) with a FWHW of 45 nm and an intensity of 1 mW/cm^2 per LED. Furthermore, a green light-emitting LED NCSE119AT (λ = 500 nm, Nichia Corporation, 5 LEDs/device) was used for *cis* → *trans* photoisomerisation. Thermal relaxation into the *trans*-isomers was achieved by storing the sample solutions in darkness, at 45 °C for at least 18 h.

Photostationary states at 365 nm and half-life of the thermal *cis* → *trans* relaxation were determined using ^{1}H NMR spectroscopy. Fully relaxed samples were irradiated for 5 – 10 min at 365 nm and immediately submitted to the ^{1}H NMR measurement. The resulting isomer ratios were usually ascertained by integrating the H-1 proton signals of the mannoside moieties. Attainment of the photostationary equilibrium was verified by a second NMR measurement after further irradiation at 365 nm for 1 min. Analogously, ^{19}F NMR measurements were performed at 471 MHz prior and after irradiation in order to demonstrate the discriminability of the *cis*- and *trans*-fluoroazobenzene signals. For the determination of the thermal relaxation kinetics, the alteration of the ^{1}H NMR signal integrals (ensuing from the PSS) was monitored in regular intervals (e.g. 30 min) over 2 to 3 days under exclusion of light. The half-life ($\tau_{1/2}$) is given as $\ln 2/k$ whereby the decay constant (k) can be determined as the slope resulting from a semi-logarithmic (log-linear) plot of $\ln(I/I_0)$ as a function of time. I describes the signal intensity (integral) of the *cis*-isomer.

Absorption maxima (λ_{max}), photostationary states and extinction coefficients (ε) were determined by UV/vis spectroscopy in quartz cuvettes on a Shimadzu UV-2600i UV/vis spectrophotometer. UV/vis spectra were recorded with 1 nm resolution from 250 to 700 nm. Solutions of thermally relaxed azobenzene derivatives were diluted in the respective solvents in order to approach an absorbance of ≤ 1 at ~ 353 nm wavelength (π → π* transition) in the UV/vis spectra. Subsequently, the probes were irradiated for 3 min at 365 nm to record UV/vis spectra of the respective *cis*-isomers. Attainment of the photostationary equilibrium was confirmed by recording of a second UV/vis spectrum after 1 min of further irradiation at 365 nm. Analogously, the UV/vis spectra of partially restored *trans*-isomers were recorded after irradiation at 440 nm and 500 nm for 3 min, respectively. The corresponding *trans*:*cis* ratios at the PSS were calculated by linear regression of the absorbance against varying sample concentrations (at least 7 different concentrations in a range of 0.05 mM – 5.9 µM). At this, the maximal absorbance wavelength λ_{max} of the π → π* transition was used as reference value. The calculation of molar extinction coefficients ε was performed under consideration of the Beer-Lambert law. Hence, the extinction coefficient can be directly deduced from linear fitting of the slope in a plot showing the absorbance as a function of varying sample concentrations.

The photochromic properties of acetamido-functionalised azobenzene mannosides **51**, **52**, **53**, and **54**, as well as azobenzene mannoside derivative **56** were also investigated on an Infinite® 200 Pro M Nano+ dual mode microplate reader (Tecan). In this case, UV/vis spectra were recorded with 2 nm resolution between 250 – 600 nm at 37 °C. Solutions of the respective samples (100 µL in PBS:MeOH, 9:1) were each applied onto UV-Star® F96-well microplates (Greiner Bio-One GmbH) and incubated under exclusion of light for 63 h at room temperature. After measuring the UV/vis spectra of the resulting *trans*-isomers, the plates bearing the sample solutions were each irradiated at 365 nm, 440 nm or 500 nm, respectively, and the corresponding absorbance spectra were also recorded. The PSSs were confirmed by further irradiation for 3 min at the respective wavelengths and the isomeric ratios at PSS_{440} and PSS_{500} were calculated by linear regression of the absorbance at 365 nm as a function of sample concentration.

In the case of thermal relaxation experiments, the samples were irradiated at 365 nm for 5 min and the PSS was controlled as described before. Subsequently, UV/vis absorbance spectra were recorded in the kinetic cycle mode at 37 °C in a timescale of 65 min. After each cycle, the plates were moderately agitated for 50 s (orbital mode, amplitude: 6, frequency: 142 rpm). The *trans* → *cis* ratios after approx. 20 min (± 5 min) of relaxation were calculated by linear regression of the absorbance at 365 nm as a function of sample concentration.

5.5.2 NMR spectra of synthesised azobenzene compounds prior and after irradiation

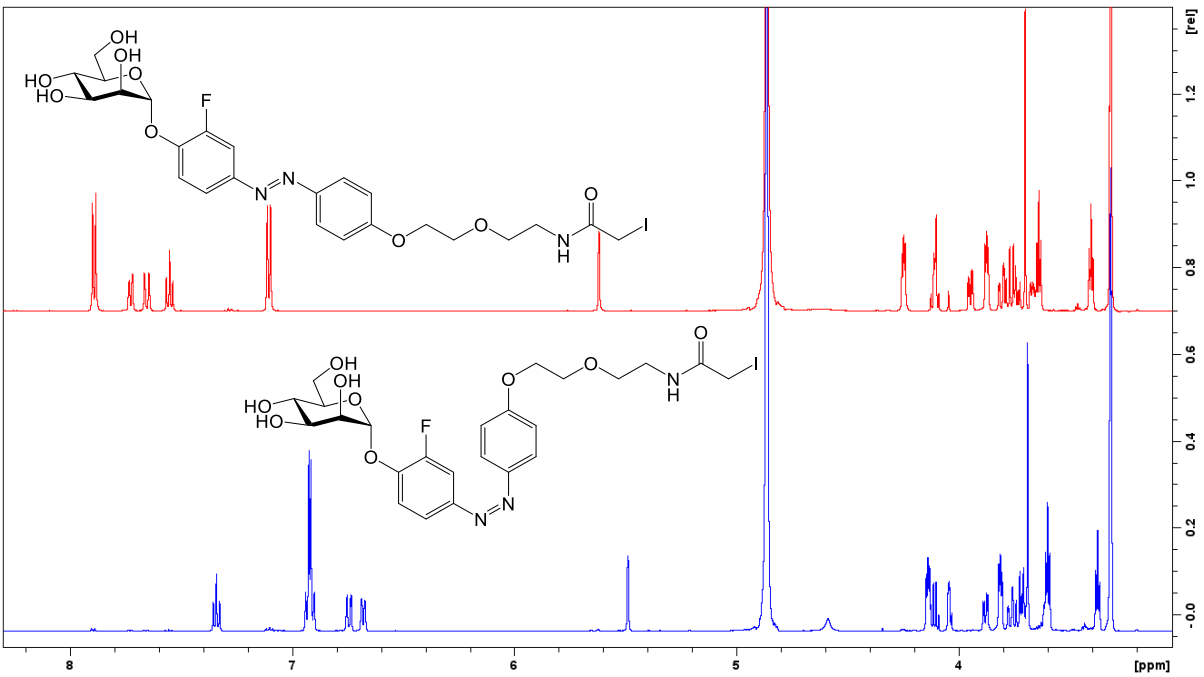


Figure 5.5-1: ^{1}H NMR spectra (600 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **F1** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 4:96.

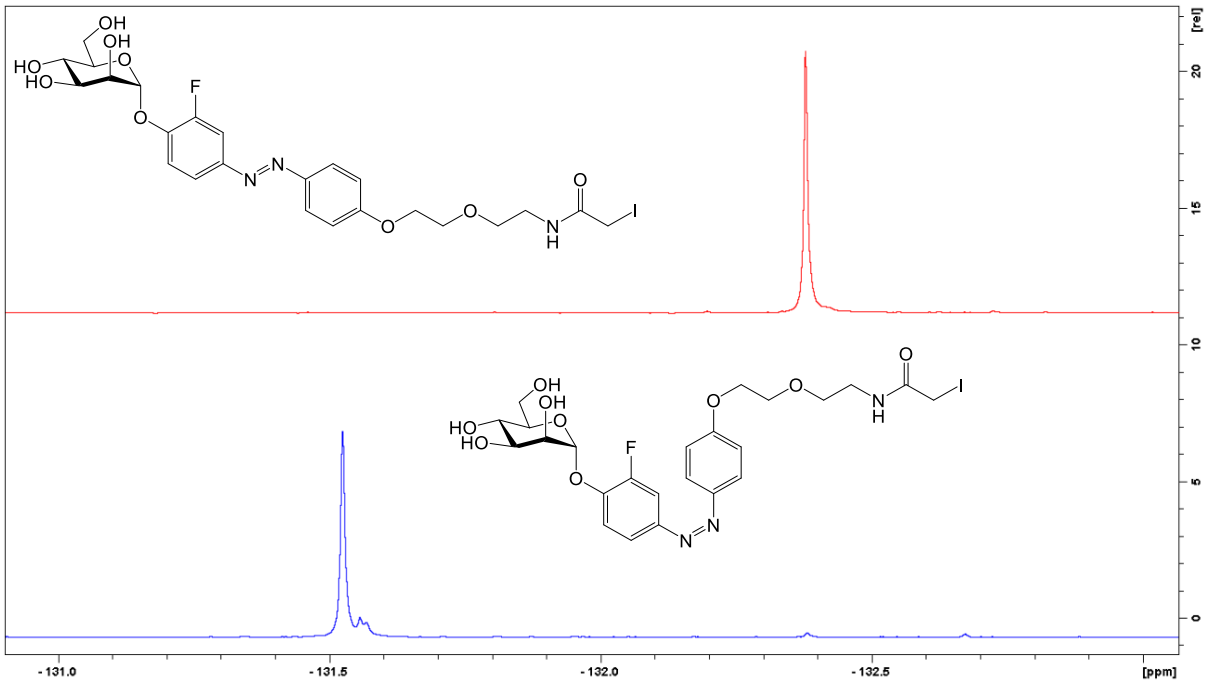


Figure 5.5-2: ^{19}F NMR spectra (471 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **F1** (red) and after irradiation at 365 nm for 3 min (blue). PSS_{365} (*trans:cis*): 2:98.

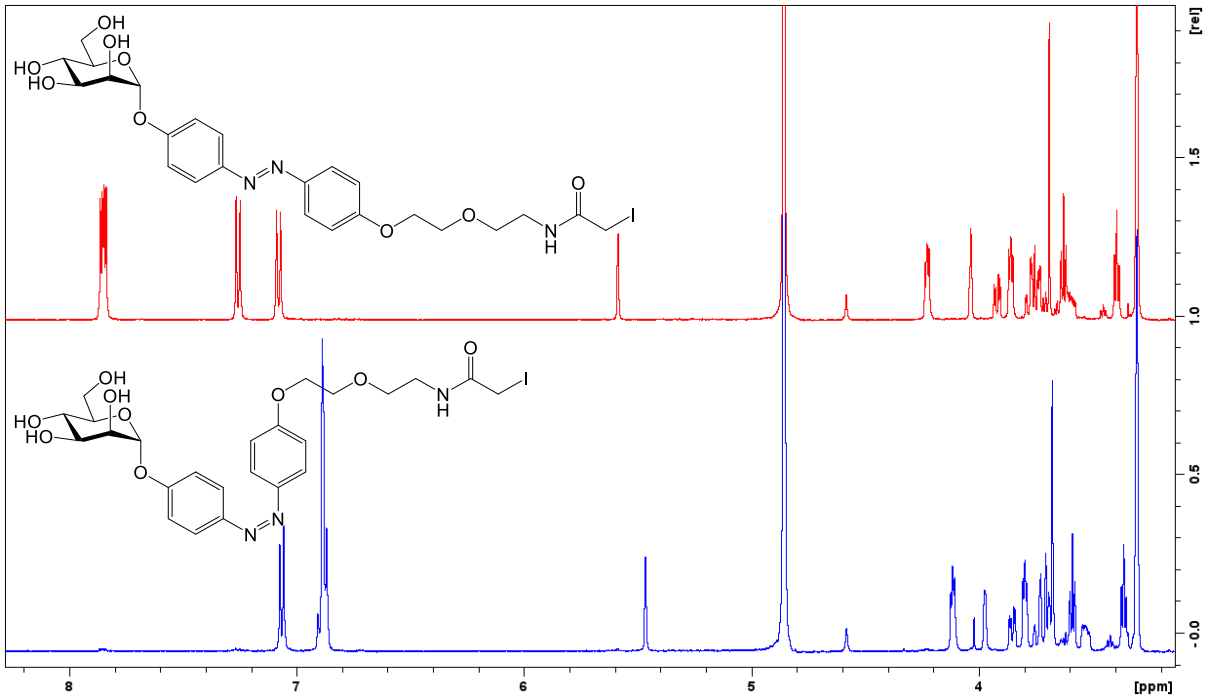


Figure 5.5-3: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed azobenzene derivative **H1** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 5:95.

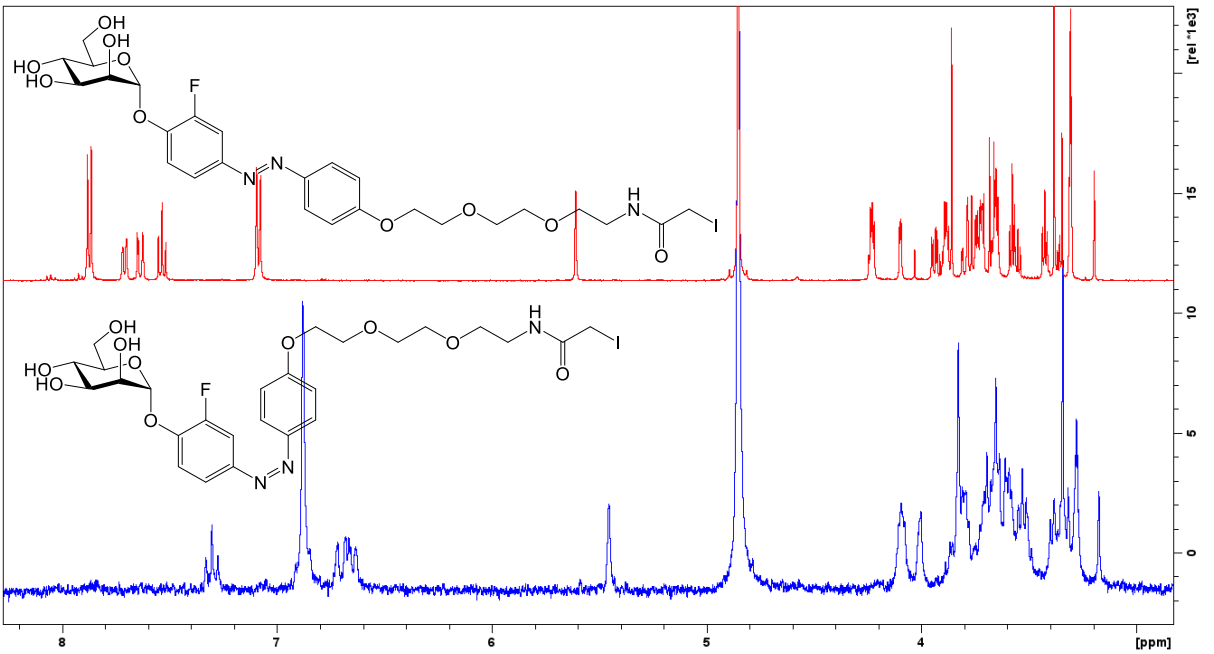


Figure 5.5-4: ^{1}H NMR spectra of thermally relaxed fluoroazobenzene derivative **F2** (red, 500 MHz, CD_3OD, 298 K) and after irradiation at 365 nm for 5 min (blue, 200 MHz, CD_3OD, 298 K). PSS_{365} (*trans:cis*): 5:95.

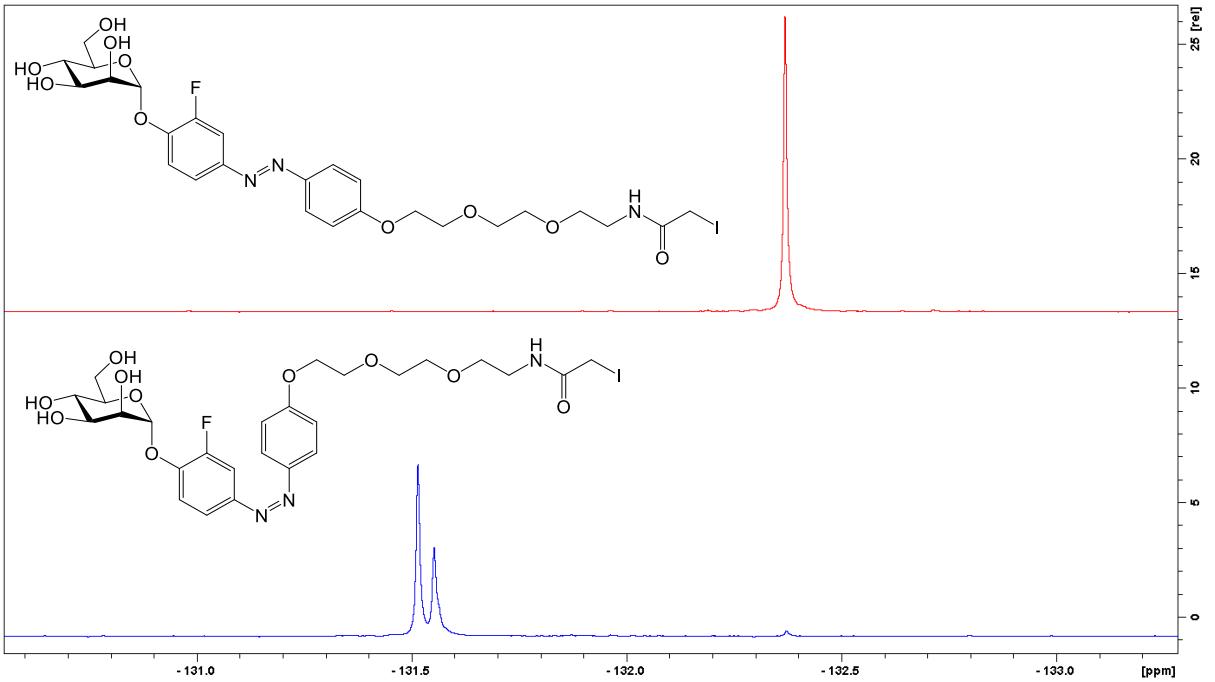


Figure 5.5-5: ^{19}F NMR spectra (471 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **F2** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 2:98.

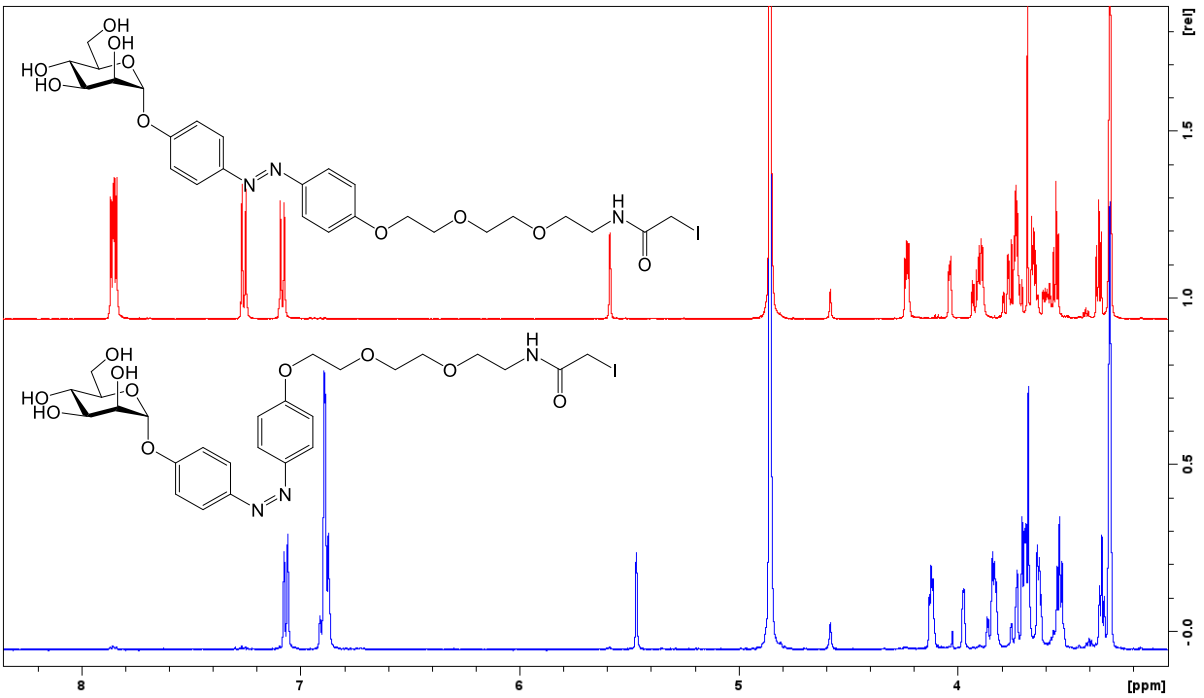


Figure 5.5-6: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed azobenzene derivative **H2** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 7:93.

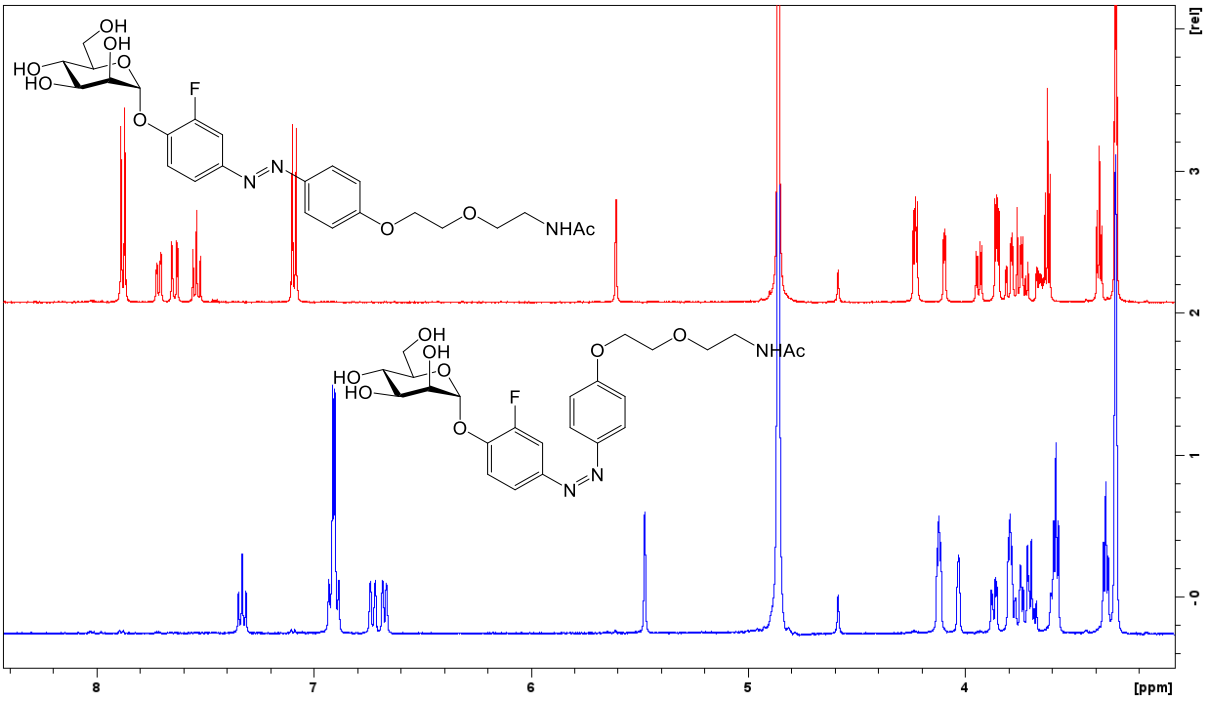


Figure 5.5-7: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **51** (red) and after irradiation at 365 nm for 10 min (blue). PSS_{365} (*trans:cis*): 5:95.

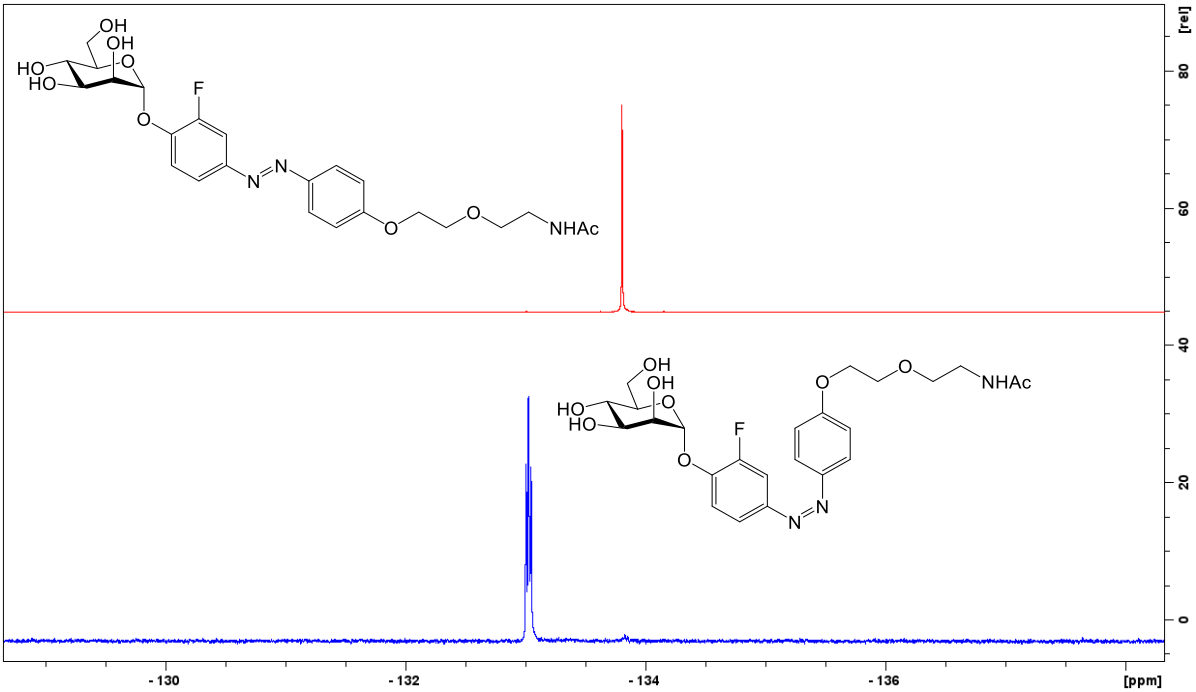


Figure 5.5-8: ^{19}F NMR spectra (471 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **51** (red) and after irradiation at 365 nm for 10 min (blue). The spectrum of the *cis*-isomer was recorded without proton decoupling. PSS_{365} (*trans:cis*): 4:96.

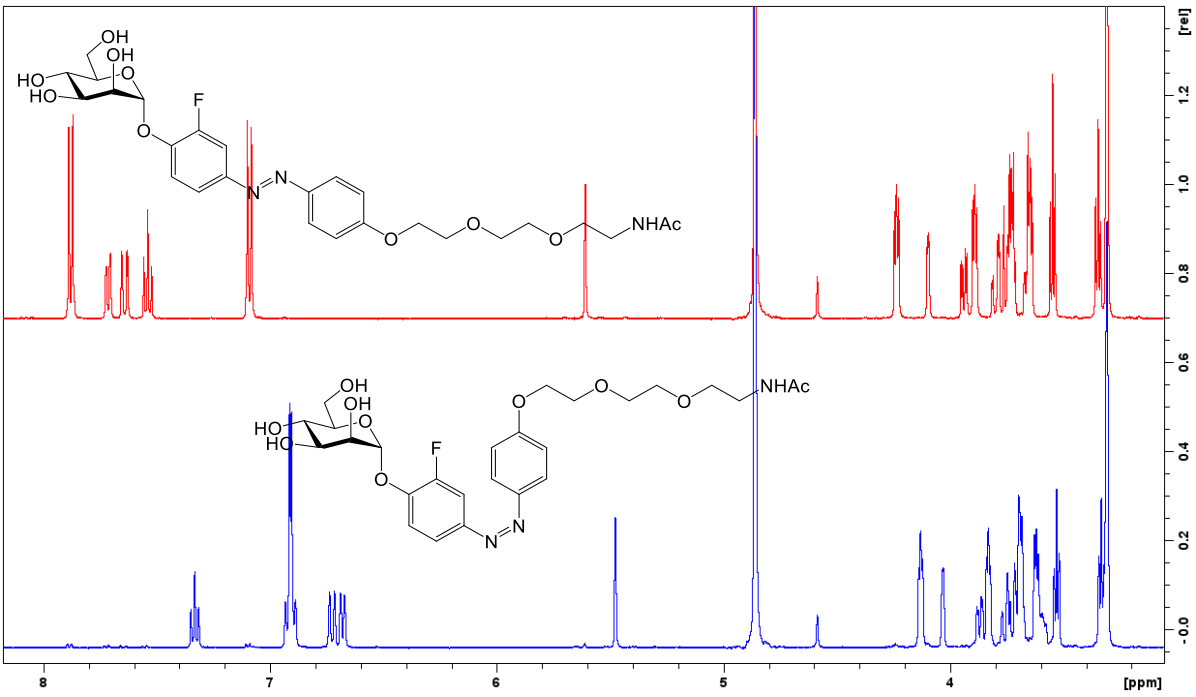


Figure 5.5-9: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **52** (red) and after irradiation at 365 nm for 10 min (blue). PSS_{365} (*trans:cis*): 5:95.

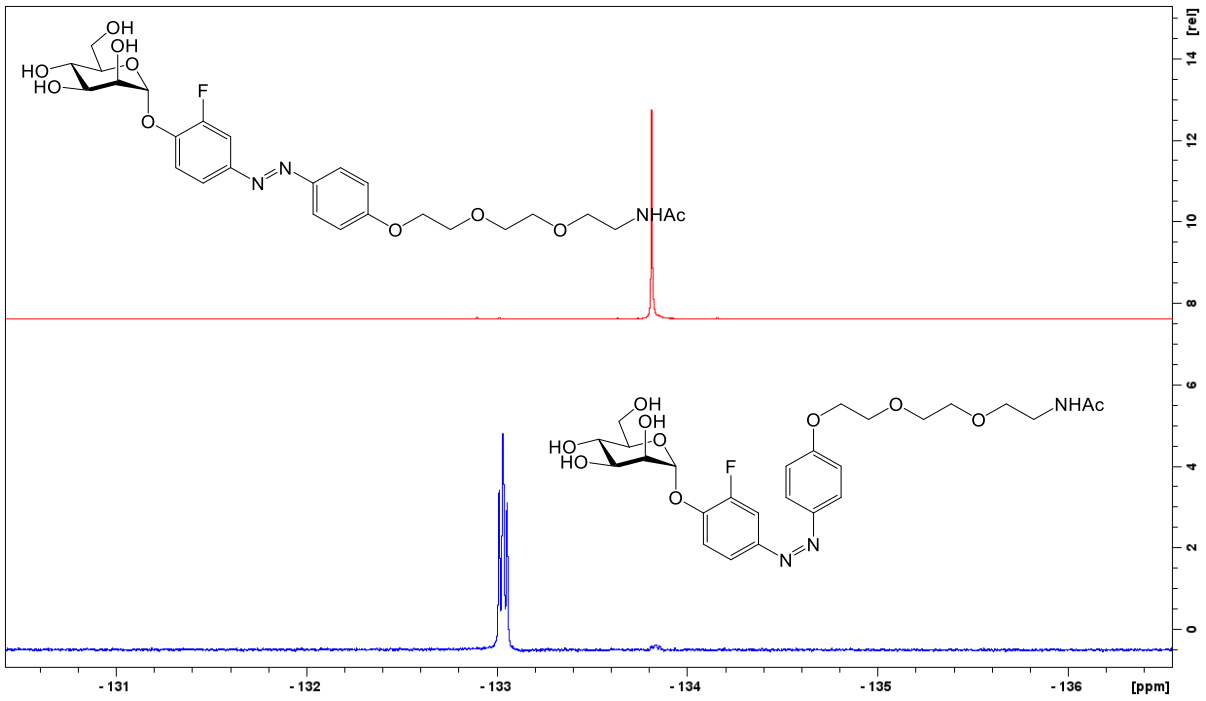


Figure 5.5-10: ^{19}F NMR spectra (471 MHz, CD_3OD, 298 K) of thermally relaxed fluoroazobenzene derivative **52** (red) and after irradiation at 365 nm for 10 min (blue). The spectrum of the *cis*-isomer was recorded without proton decoupling. PSS_{365} (*trans:cis*): 3:97.

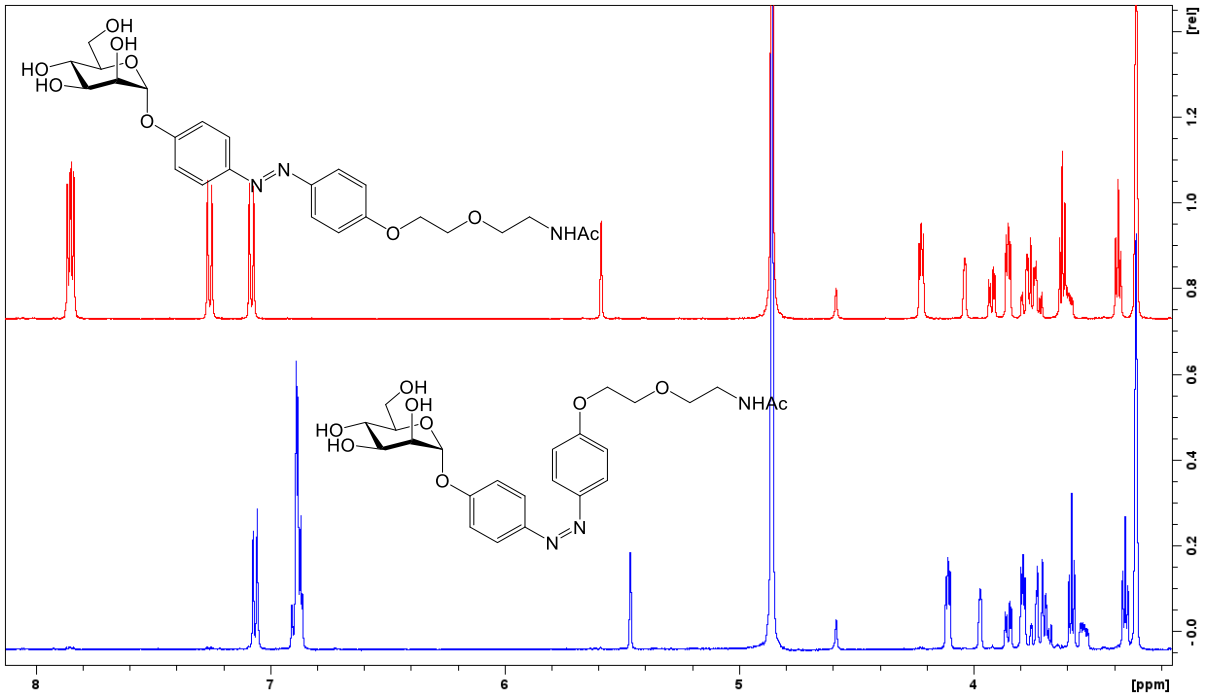


Figure 5.5-11: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed azobenzene derivative **53** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 5:95.

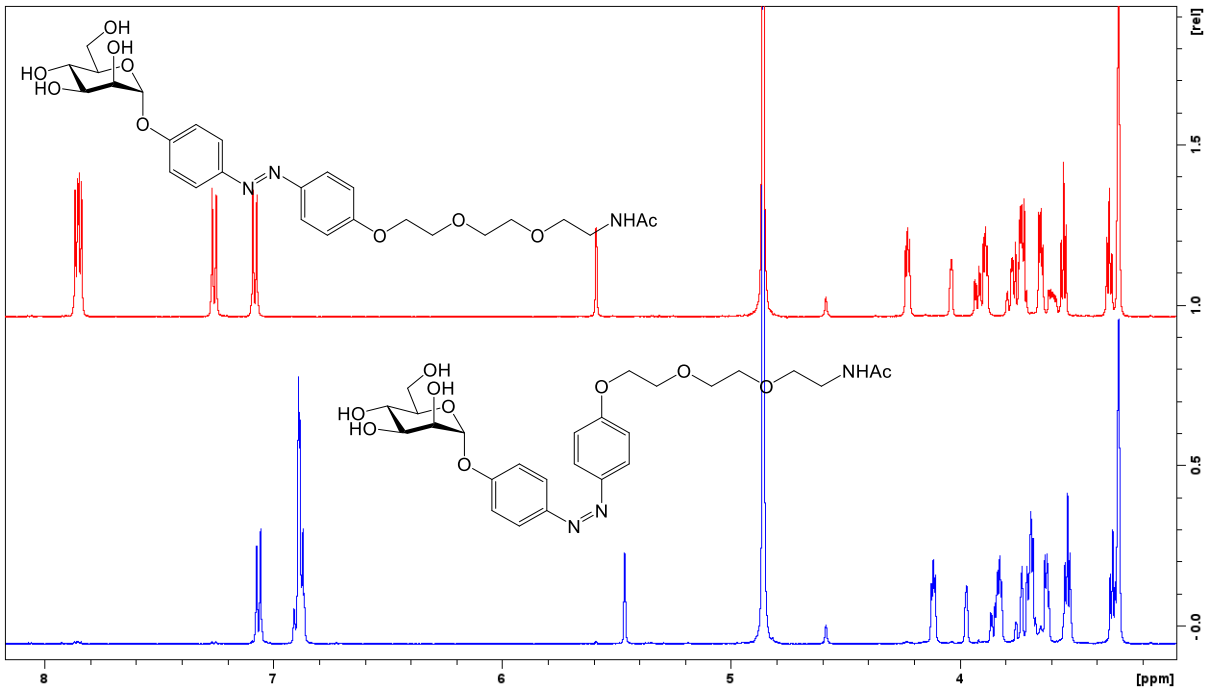


Figure 5.5-12: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed azobenzene derivative **54** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 5:95.

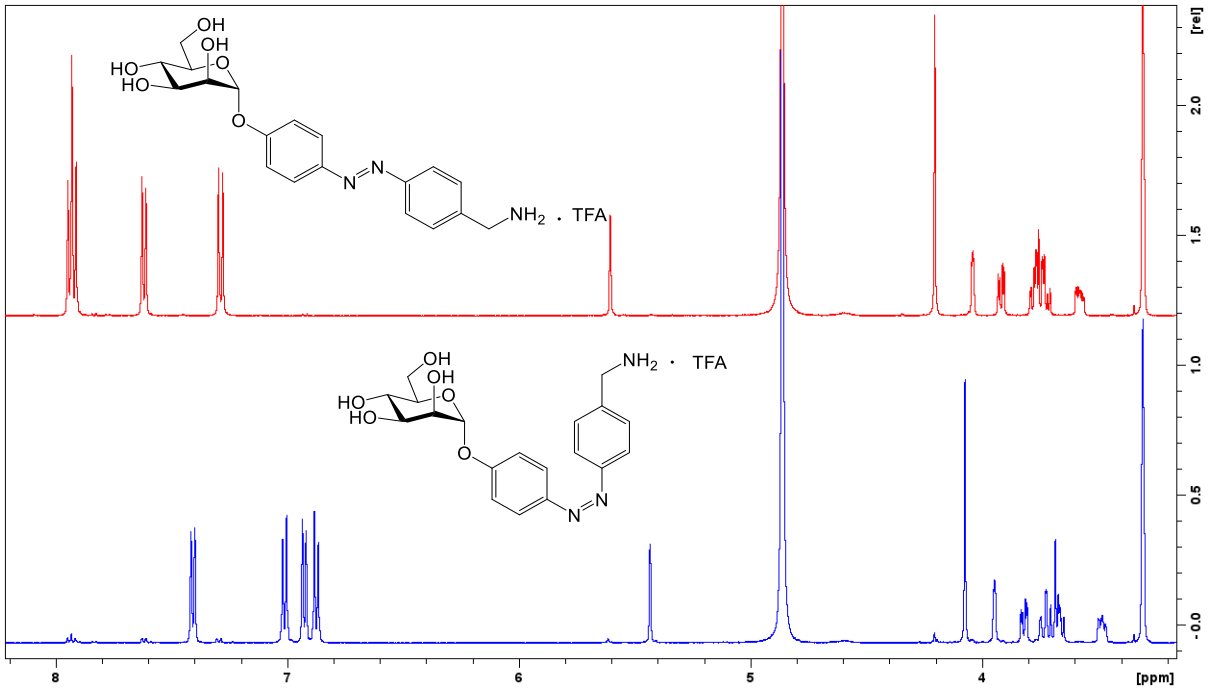


Figure 5.5-13: ^{1}H NMR spectra (500 MHz, CD_3OD, 298 K) of thermally relaxed azobenzene derivative **56** (red) and after irradiation at 365 nm for 5 min (blue). PSS_{365} (*trans:cis*): 7:93.

5.5.3 UV/vis absorption spectra of synthesised azobenzene compounds prior and after irradiation

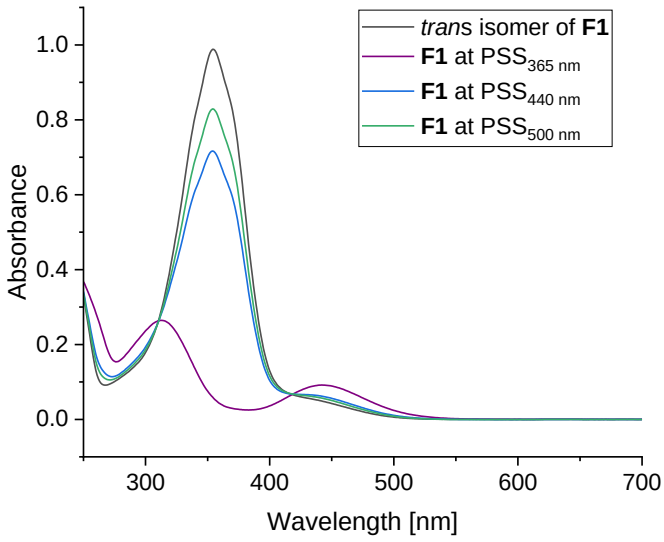


Figure 5.5-14: UV/vis spectrum of **F1** (0.025 mM) after thermal relaxation (black) and after *trans* → *cis* photoisomerisation by irradiation at 365 nm for 4 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-15b: 4:96 (PSS_{365}), 71:29 (PSS_{440}) and 82:18 (PSS_{500}). Measured in methanol at 298 K.

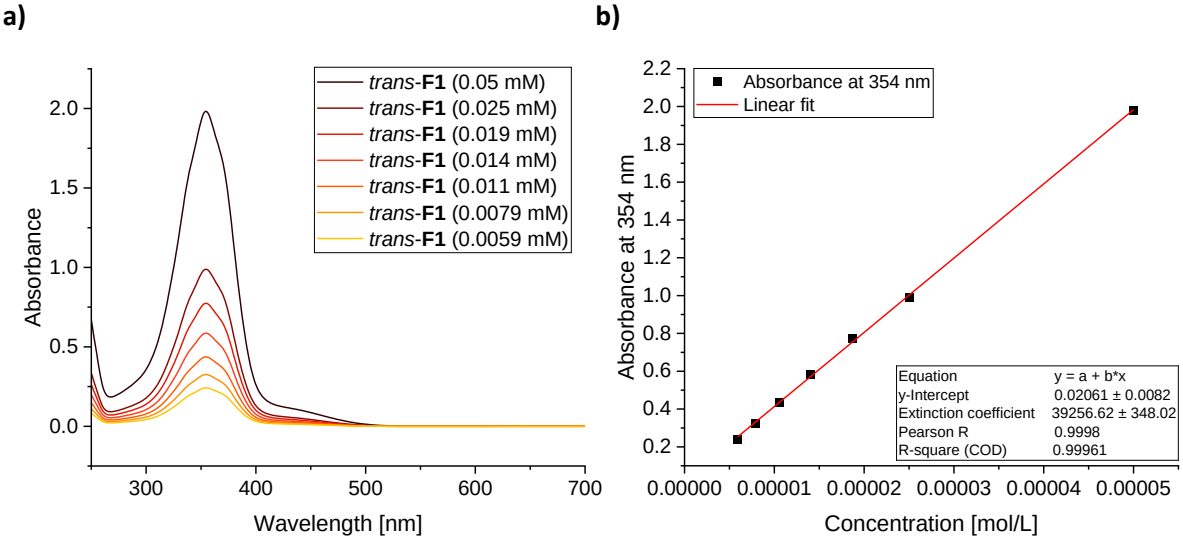


Figure 5.5-15: a) UV/vis spectra of thermally relaxed **F1** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **F1** in varying concentrations at the λ_{max} (354 nm).

The molar extinction coefficient of **F1**(ε = 39256.6 L · mol^{-1} · cm^{-1}) is given as the slope of the linear regression line. Measured in methanol at 298 K.

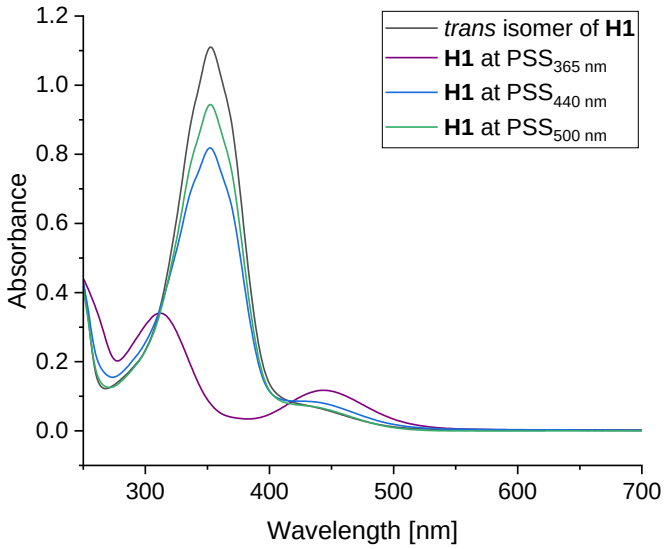


Figure 5.5-16: UV/vis spectrum of **H1** (0.033 mM) after thermal relaxation (black) and after *trans* ⟶ *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-17b: 5:95 (PSS_{365}), 78:22 (PSS_{440}) and 90:10 (PSS_{500}). Measured in methanol at 298 K.

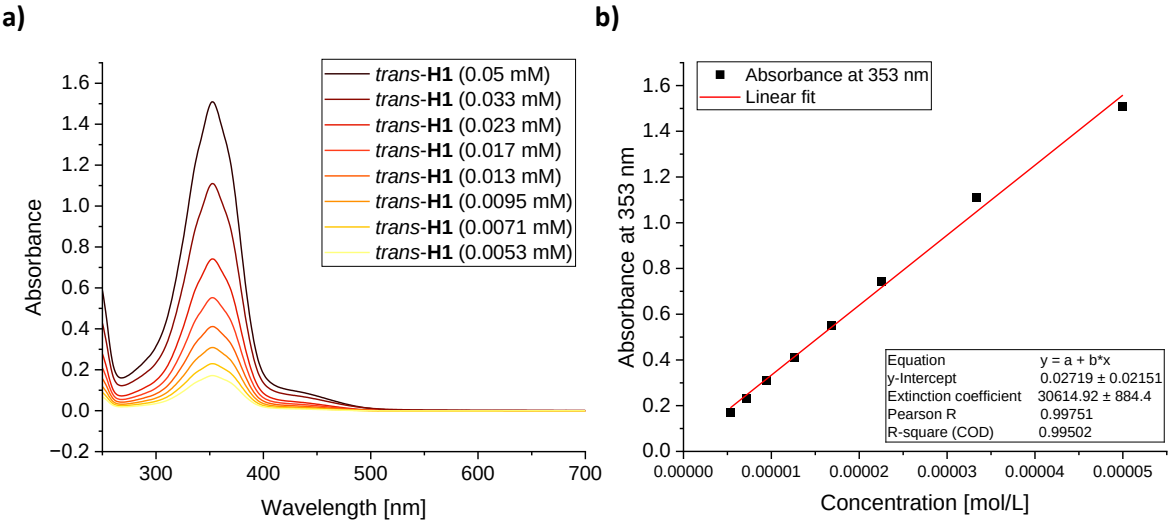


Figure 5.5-17: a) UV/vis spectra of thermally relaxed **H1** in different concentrations for the calculation of extinction coefficient**. b)** Linear fitting of the absorbance of thermally relaxed **H1** in varying concentrations at the λ_{max} (353 nm).

The molar extinction coefficient of **H1** (ε = 30614.9 $L \cdot mol^{-1} \cdot cm^{-1}$) is given as the slope of the linear regression line. Measured in methanol at 298 K.

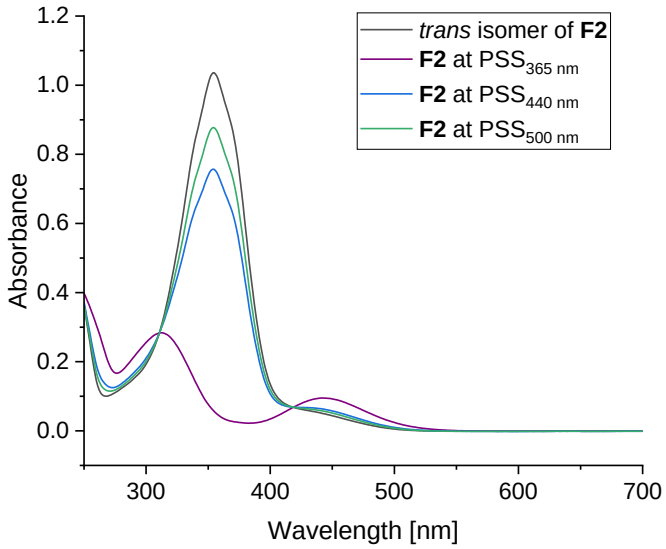


Figure 5.5-18: UV/vis spectrum of **F2** (0.025 mM) after thermal relaxation (black) and after *trans* → *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-19b: 4:96 (PSS_{365}), 74:26. (PSS_{440}) and 86:14 (PSS_{500}). Measured in methanol at 298 K.

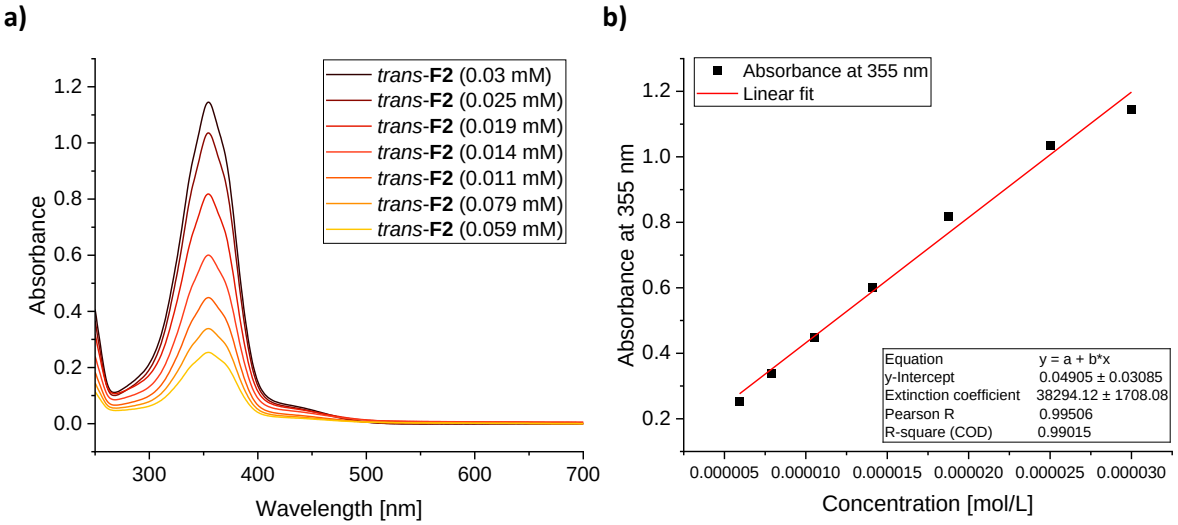


Figure 5.5-19: a) UV/vis spectra of thermally relaxed **F2** in different concentrations for the calculation of extinction coefficient. b) Linear fitting of the absorbance of thermally relaxed **F2** in varying concentrations at the λ_{max} (355 nm).

The molar extinction coefficient of **F2** (ε = 38294.12 L · mol^{-1} · cm^{-1}) is given as the slope of the linear regression line. Measured in methanol at 298 K.

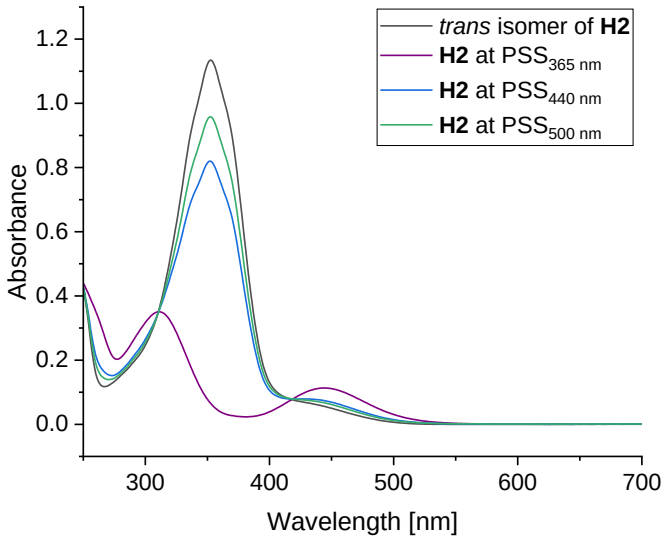


Figure 5.5-20: UV/vis spectrum of **H2** (0.025 mM) after thermal relaxation (black) and after *trans* ⟶ *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-21b: 5:95 (PSS_{365}), 74:26 (PSS_{440}) and 87:13 (PSS_{500}). Measured in methanol at 298 K.

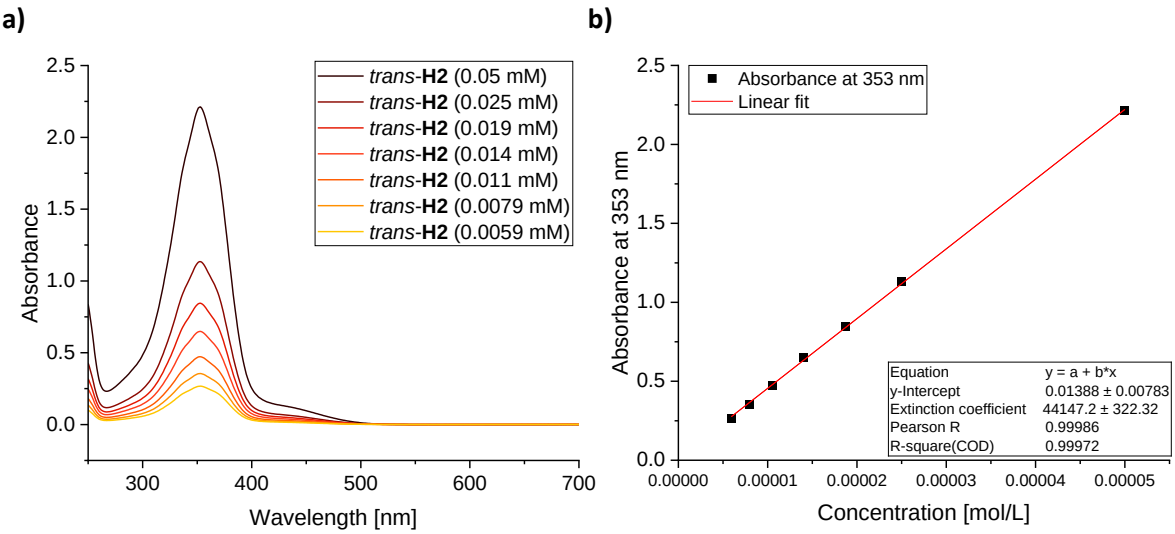


Figure 5.5-21: a) UV/vis spectra of thermally relaxed **H2** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **H2** in varying concentrations at the λ_{max} (353 nm).

The molar extinction coefficient of **H2** (ε = 44147.2 L · mol−1 · cm−1) is given as the slope of the linear regression line. Measured in methanol at 298 K.

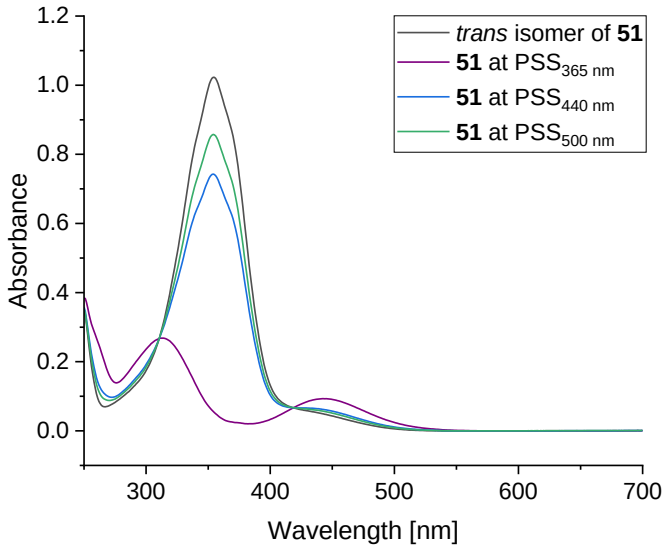


Figure 5.5-22: UV/vis spectrum of **51** (0.032 mM) after thermal relaxation (black) and after *trans* → *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-23b: 5:95 (PSS_{365}), 72:28 (PSS_{440}) and 84:16 (PSS_{500}). Measured in methanol at 298 K.

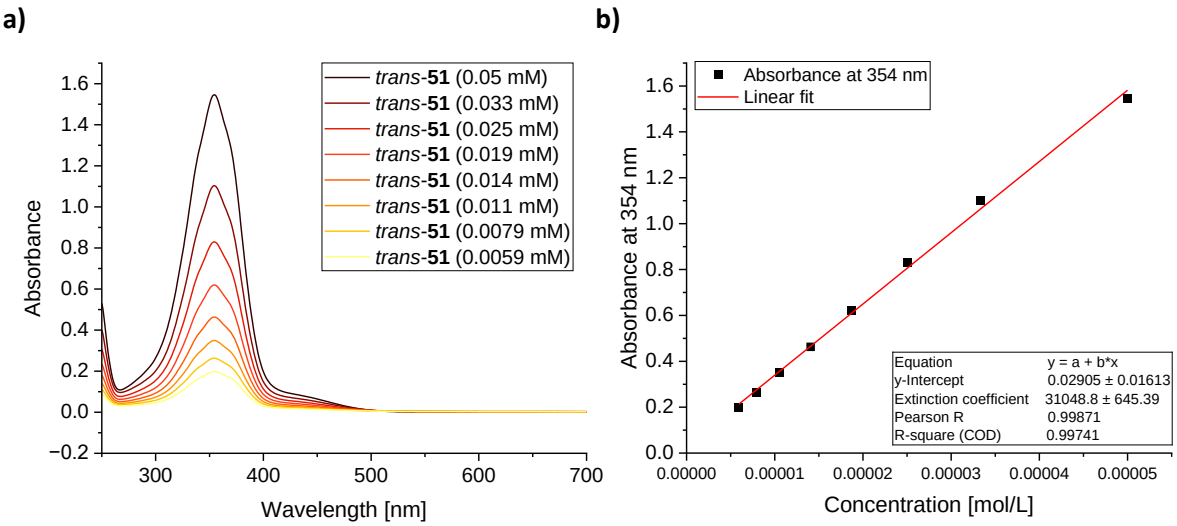


Figure 5.5-23: a) UV/vis spectra of thermally relaxed **51** in different concentrations for the calculation of extinction coefficient**. b)** Linear fitting of the absorbance of thermally relaxed **51** in varying concentrations at the λ_{max} (354 nm).

The molar extinction coefficient of **51** (ε = 31048.8 L · mol^{-1} · cm^{-1}) is given as the slope of the linear regression line. Measured in methanol at 298 K.

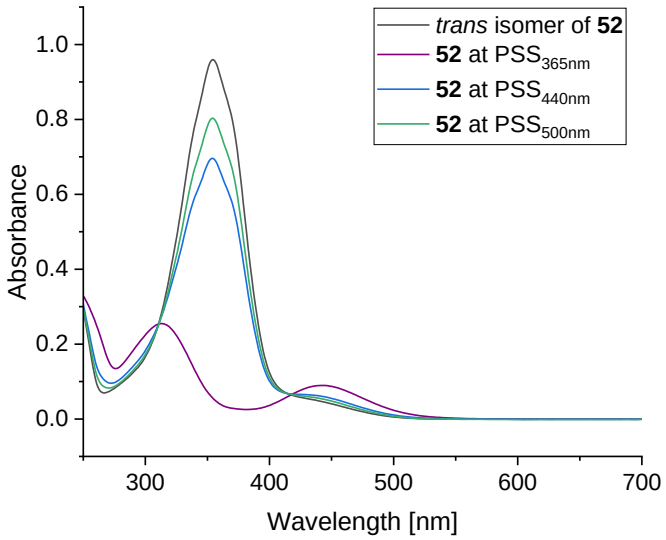


Figure 5.5-24: UV/vis spectrum of **52** (0.025 mM) after thermal relaxation (black) and after *trans* ⟶ *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-25b: 6:94 (PSS_{365}), 70:30 (PSS_{440}) and 81:19 (PSS_{500}). Measured in methanol at 298 K.

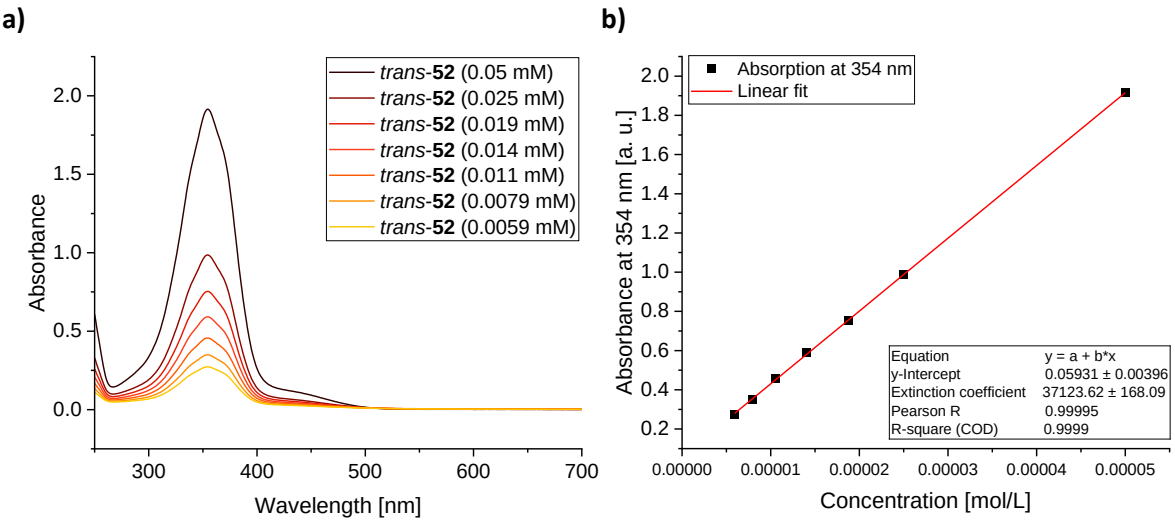


Figure 5.5-25: a) UV/vis spectra of thermally relaxed **52** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **52** in varying concentrations at the λ_{max} (354 nm).

The molar extinction coefficient of **52** (ε = 37123.6 L · mol^{-1} · cm^{-1}) is given as the slope of the linear regression line. Measured in methanol at 298 K.

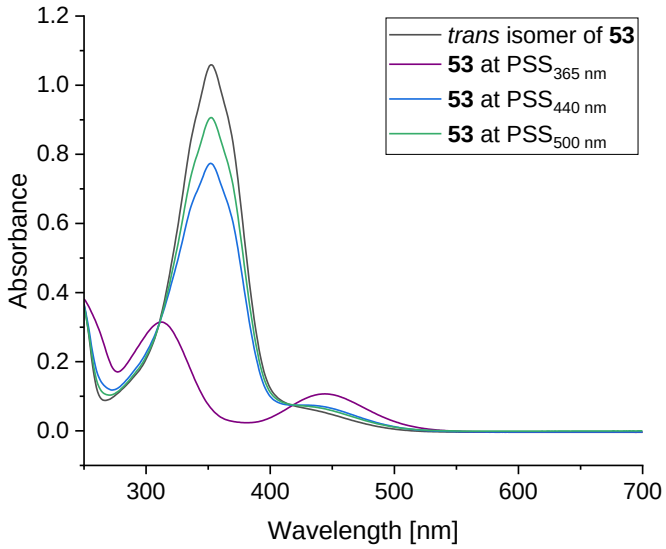


Figure 5.5-26: UV/vis spectrum of **53** (0.03 mM) after thermal relaxation (black) and after *trans* → *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 4 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-27b: 6:94 (PSS_{365}), 72:28 (PSS_{440}) and 84:16 (PSS_{500}). Measured in methanol at 298 K.

a) **b)**

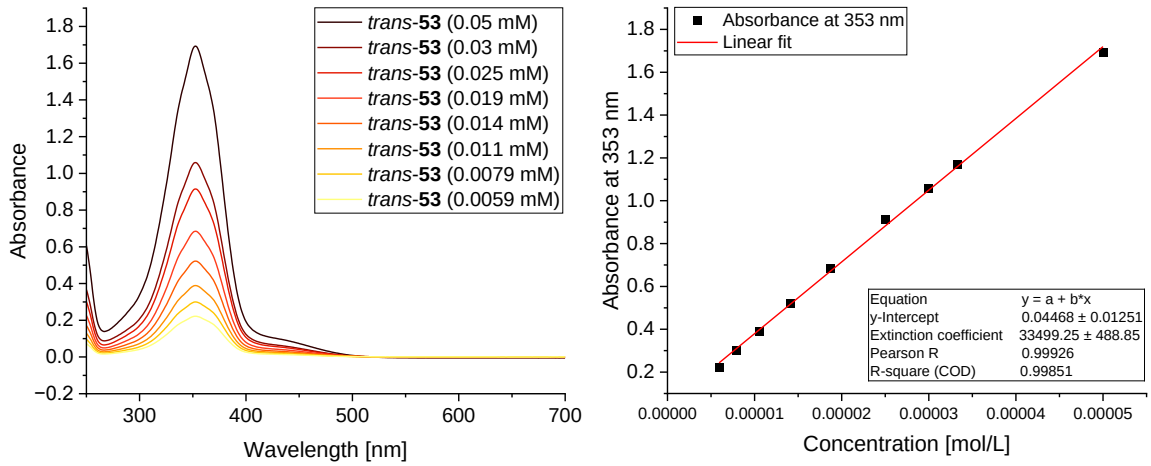


Figure 5.5-27: a) UV/vis spectra of thermally relaxed **53** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **53** in varying concentrations at the λ_{max} (353 nm).

The molar extinction coefficient of **53** (ε = 33499.3 $L \cdot mol^{-1} \cdot cm^{-1}$) is given as the slope of the linear regression line. Measured in methanol at 298 K.

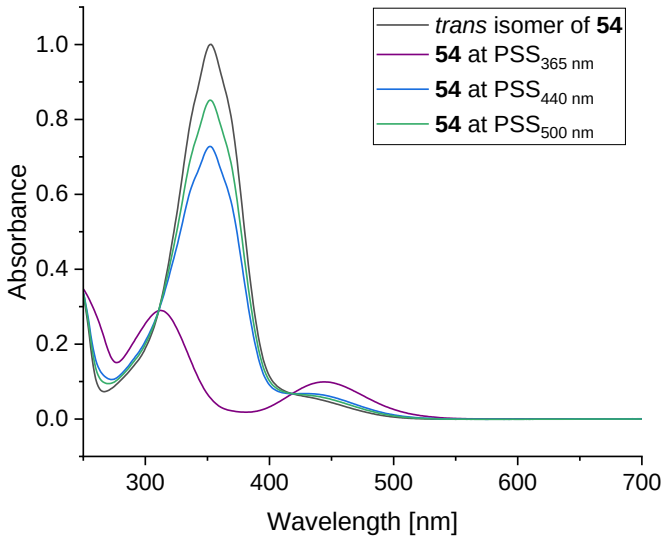


Figure 5.5-28: UV/vis spectrum of **54** (0.03 mM) after thermal relaxation (black) and after *trans* ⟶ *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-29b: 4:96 (PSS_{365}), 67:33 (PSS_{440}) and 79:21 (PSS_{500}). Measured in methanol at 298 K.

a) **b)**

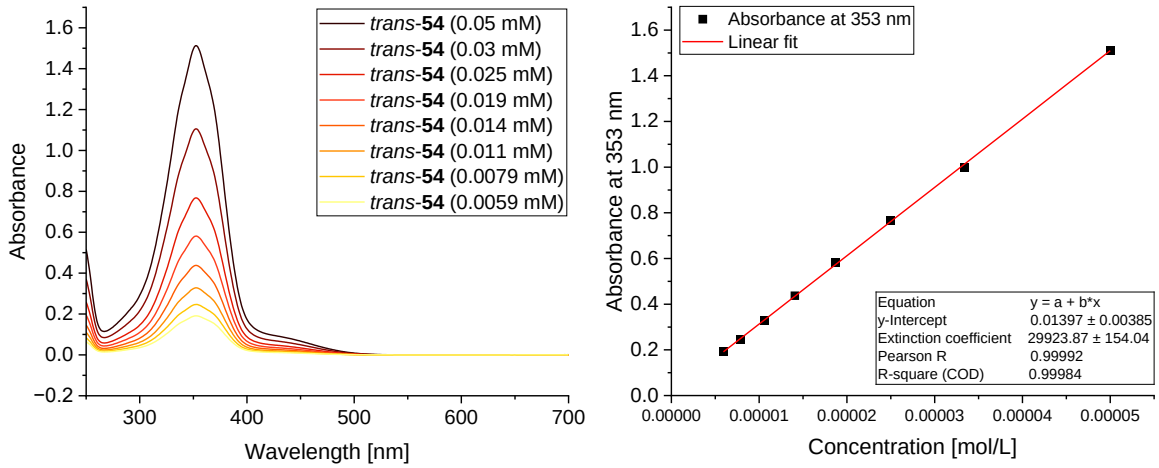


Figure 5.5-29: a) UV/vis spectra of thermally relaxed **54** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **54** in varying concentrations at the λ_{max} (353 nm).

The molar extinction coefficient of **54** (ε = 29923.9 L · mol^{-1} · cm^{-1}) is given as the slope of the linear regression line. Measured in methanol at 298 K.

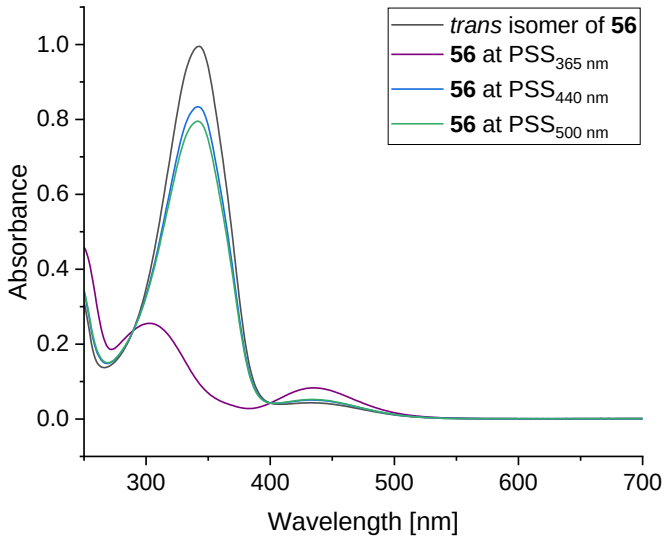


Figure 5.5-30: UV/vis spectrum of **56** (0.035 mM) after thermal relaxation (black) and after *trans* → *cis* photoisomerisation by irradiation at 365 nm for 3 min (violet).

The *trans*-isomer was partially restored by irradiation with 440 nm for 3 min (blue) and 500 nm for 3 min (green). The PSSs were verified by further irradiation for 1 min at the respective wavelengths and the corresponding *trans*:*cis* ratios were calculated using the linear regression from Figure 5.4-31b: 8:92 (PSS_{365}), 78:22 (PSS_{440}) and 74:26 (PSS_{500}). Measured in methanol at 298 K.

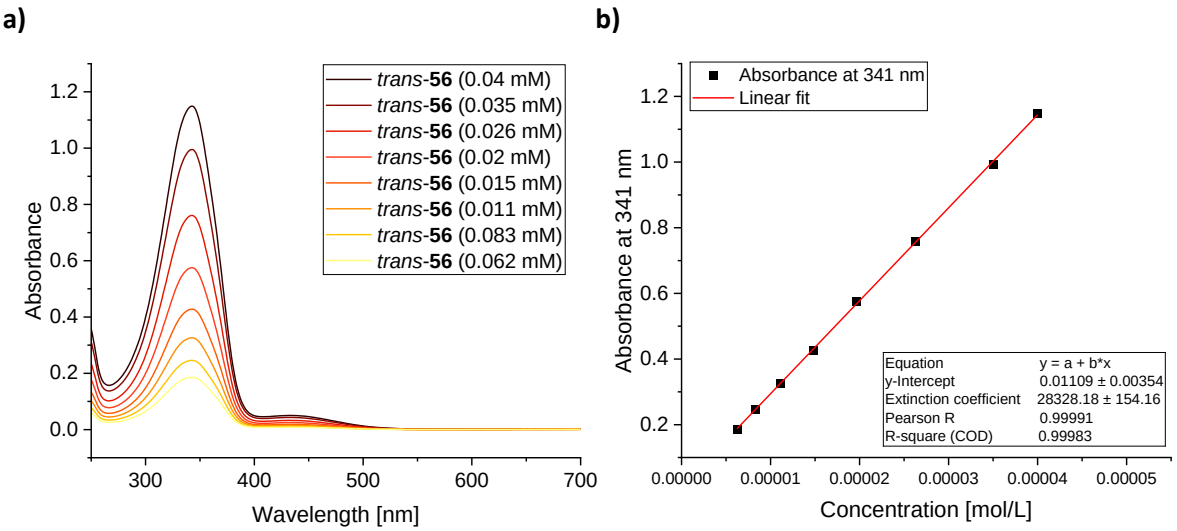


Figure 5.5-31: a) UV/vis spectra of thermally relaxed **56** in different concentrations for the calculation of extinction coefficient. **b)** Linear fitting of the absorbance of thermally relaxed **56** in varying concentrations at the λ_{max} (341 nm).

The molar extinction coefficient of **56** (ε = 28328.2 $L \cdot mol^{-1} \cdot cm^{-1}$) is given as the slope of the linear regression line. Measured in methanol at 298 K.

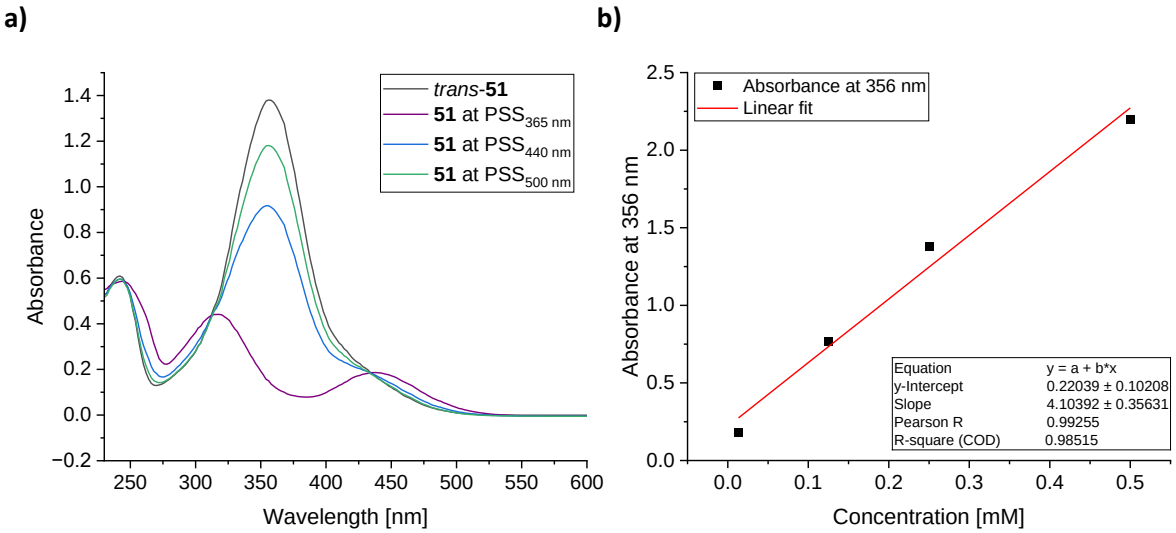


Figure 5.5-32: a) UV/vis spectra of **51** (0.25 mM) at 37 °C after thermal relaxation (black), irradiation at 365 nm for 5 min (violet), at 440 nm for 5 min (blue) and 500 nm for 5 min (green). **b)** Linear regression of absorbance at λ = 365 nm and different concentrations of **51** (0.5 mM – 18 µM) for the calculation of *trans:cis* ratios: 4:96 (PSS_{365}), 66:34 (PSS_{440}), 87:13 (PSS_{500}).
The PSSs were verified by further irradiation for 3 min at the respective wavelengths. Measured at 310 K in PBS/methanol, 9:1, on a Tecan microplate reader.

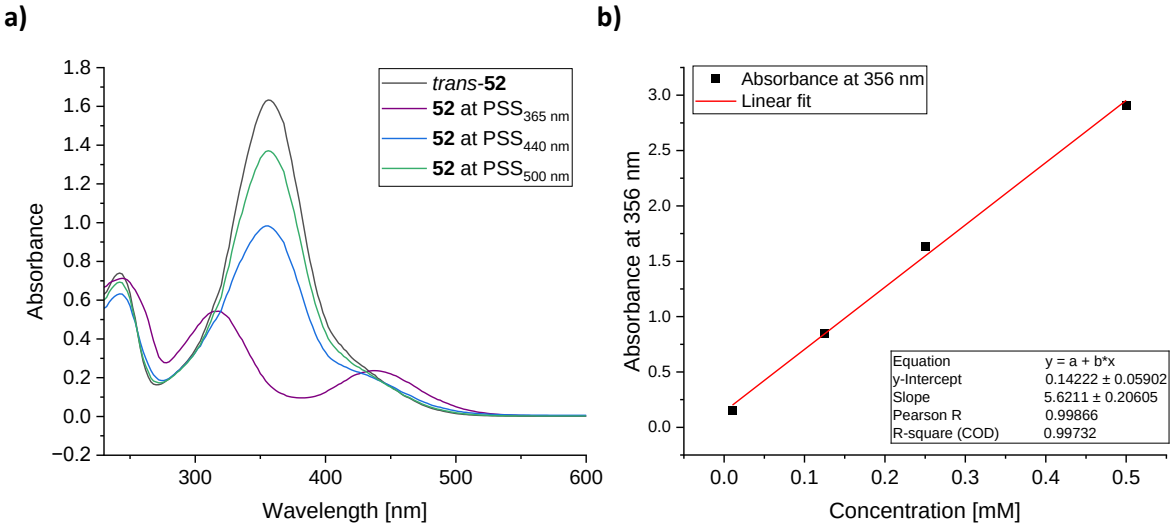


Figure 5.5-33: a) UV/vis spectra of **52** (0.25 mM) at 37 °C after thermal relaxation (black), irradiation at 365 nm for 5 min (violet), at 440 nm for 5 min (blue) and 500 nm for 5 min (green) **b)** Linear regression of absorbance at λ = 365 nm and different concentrations of **52** (0.5 mM – 10 µM) for the calculation of *trans:cis* ratios: 4:96 (PSS_{365}), 64:36 (PSS_{440}), 88:12 (PSS_{500}).
The PSSs were verified by further irradiation for 3 min at the respective wavelengths. Measured at 310 K in PBS/methanol, 9:1, on a Tecan microplate reader.

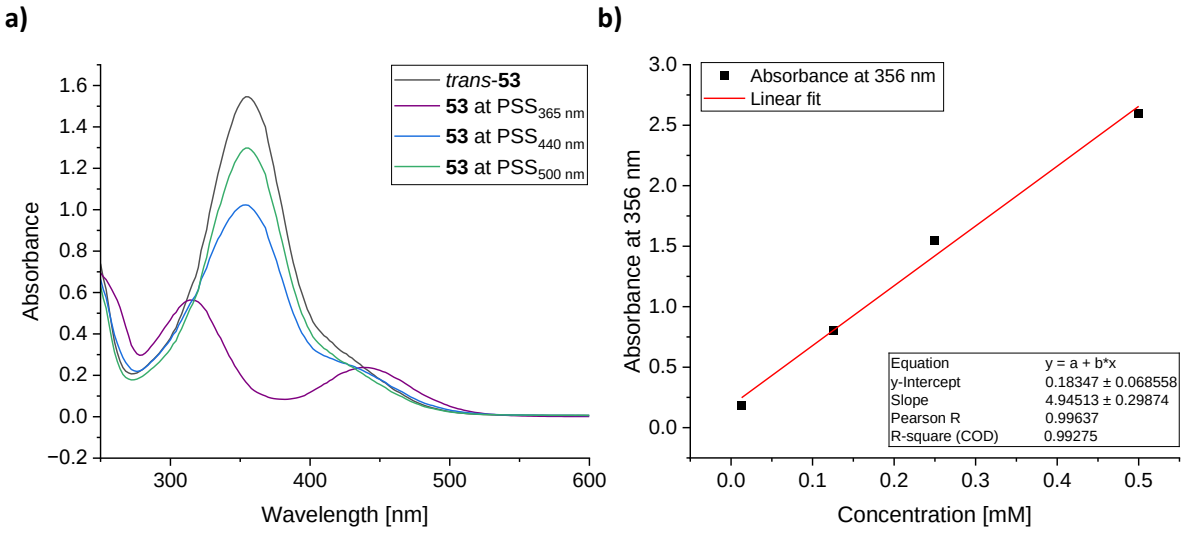


Figure 5.5-34: a) UV/vis spectra of **53** (0.25 mM) at 37 °C after thermal relaxation (black), irradiation at 365 nm for 5 min (violet), at 440 nm for 5 min (blue) and 500 nm for 5 min (green). **b)** Linear regression of absorbance at λ = 365 nm and different concentrations of **53** (0.5 mM – 11.8 µM) for the calculation of *trans*:*cis* ratios: 4:96 (PSS_{365}), 66:34 (PSS_{440}), 84:16 (PSS_{500}).

The PSSs were verified by further irradiation for 3 min at the respective wavelengths. Measured at 310 K in PBS/methanol, 9:1, on a Tecan microplate reader.

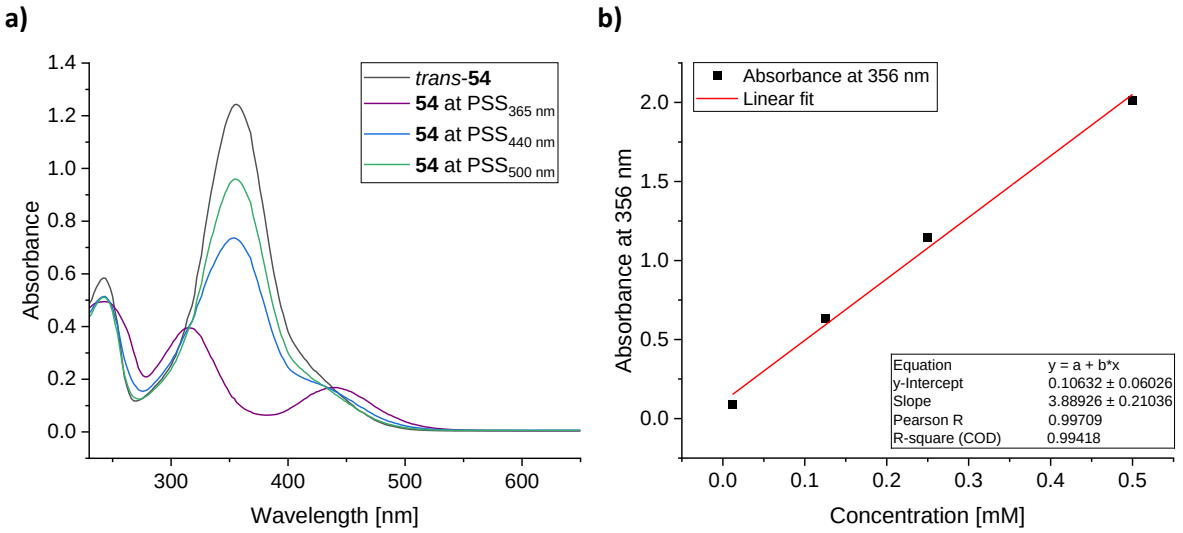


Figure 5.5-35: a) UV/vis spectra of **54** (0.25 mM) at 37 °C after thermal relaxation (black), irradiation at 365 nm for 5 min (violet), at 440 nm for 5 min (blue) and 500 nm for 5 min (green). **b)** Linear regression of absorbance at λ = 365 nm and different concentrations of **54** (0.5 mM – 11.9 µM) for the calculation of *trans*:*cis* ratios: 5:95 (PSS_{365}), 63:34 (PSS_{440}), 83:17 (PSS_{500}).

The PSSs were verified by further irradiation for 3 min at the respective wavelengths. Measured at 310 K in PBS/methanol, 9:1, on a Tecan microplate reader.

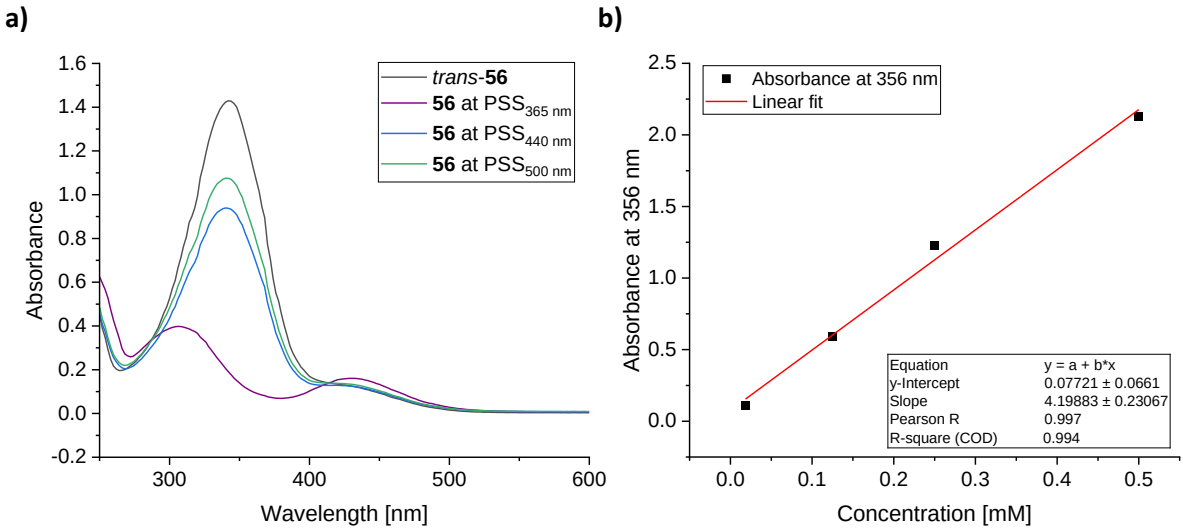


Figure 5.5-36: a) UV/vis spectra of **56** (0.25 mM) at 37 °C after thermal relaxation (black), irradiation at 365 nm for 5 min (violet), at 440 nm for 5 min (blue) and 500 nm for 5 min (green). **b)** Linear regression of absorbance at λ = 365 nm and different concentrations of **56** (0.5 mM – 11.2 µM) for the calculation of *trans*:*cis* ratios: 8:92 (PSS_{365}), 65:35 (PSS_{440}), 75:25 (PSS_{500}).

The PSSs were verified by further irradiation for 3 min at the respective wavelengths. Measured at 310 K in PBS/methanol, 9:1, on a Tecan microplate reader.

5.5.4 Thermal relaxation experiments

The kinetics of thermal *cis* → *trans* relaxation was monitored by ^{1}H NMR spectroscopy. The signal intensity (I) of the *cis*-isomer undergoes an exponential decay that can be described by the following equation:

$$I(t) = I_0 \cdot e^{-k \cdot t},$$

given $I(t)$ as the signal intensity at the time (t), I_0 as the initial signal intensity at $PSS_{365\,nm}$ and k as the exponential decay constant. Following first order kinetics, the half-life ($\tau_{1/2}$) can be derived as:

$$\tau_{1/2} = \frac{\ln(2)}{k}$$

Thus, the half-life exclusively depends on k, which can be deduced as the slope of a semi-logarithmic plot of $\ln(I/I_0)$ as a function of time.

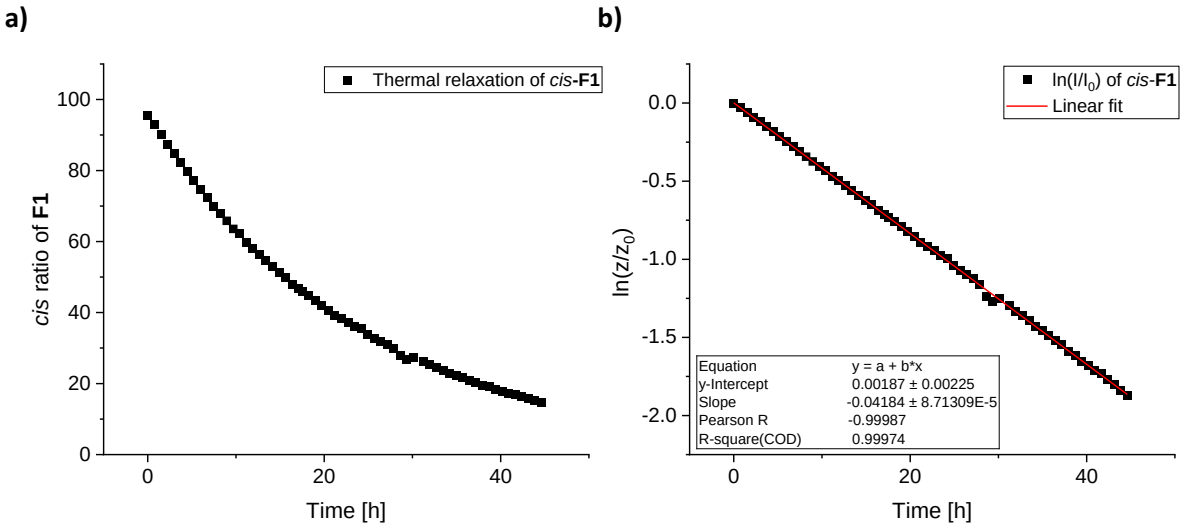


Figure 5.5-37: a) Thermal *cis* → *trans* relaxation of **F1** (17.6 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\tau_{1/2}$ = 16 h 34 min.

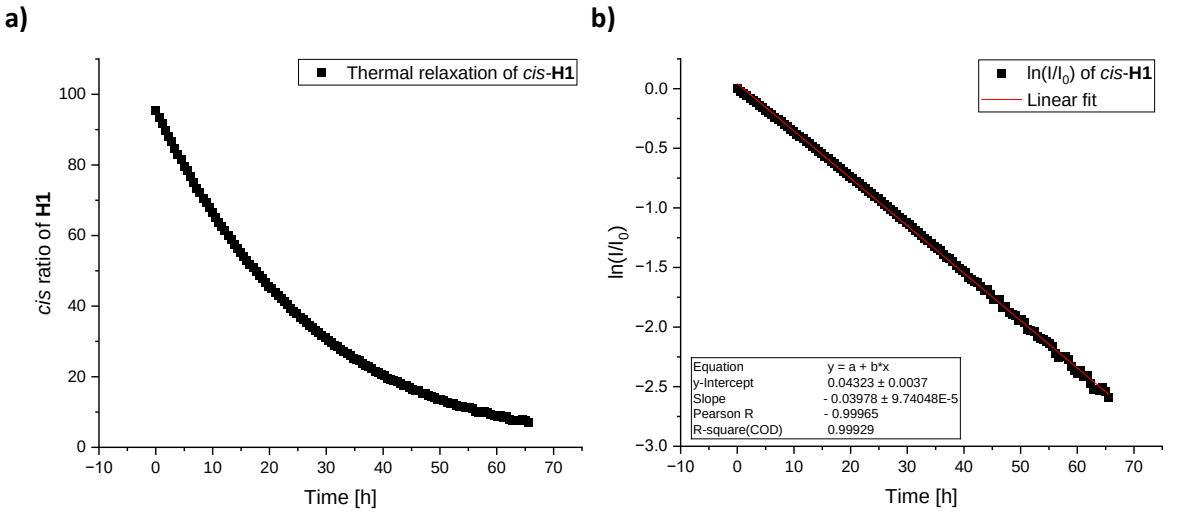


Figure 5.5-38: a) Thermal *cis* → *trans* relaxation of **H1** (20.5 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\tau_{1/2}$ = 17 h 25 min.

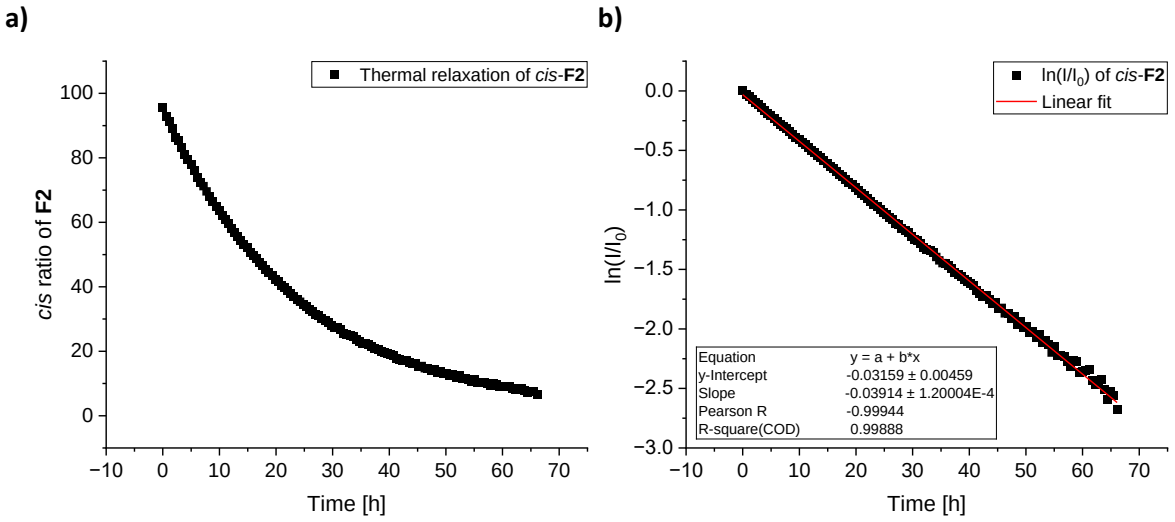


Figure 5.5-39: a) Thermal *cis* ⟶ *trans* relaxation of **F2** (24.4 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 17 h 43 min.

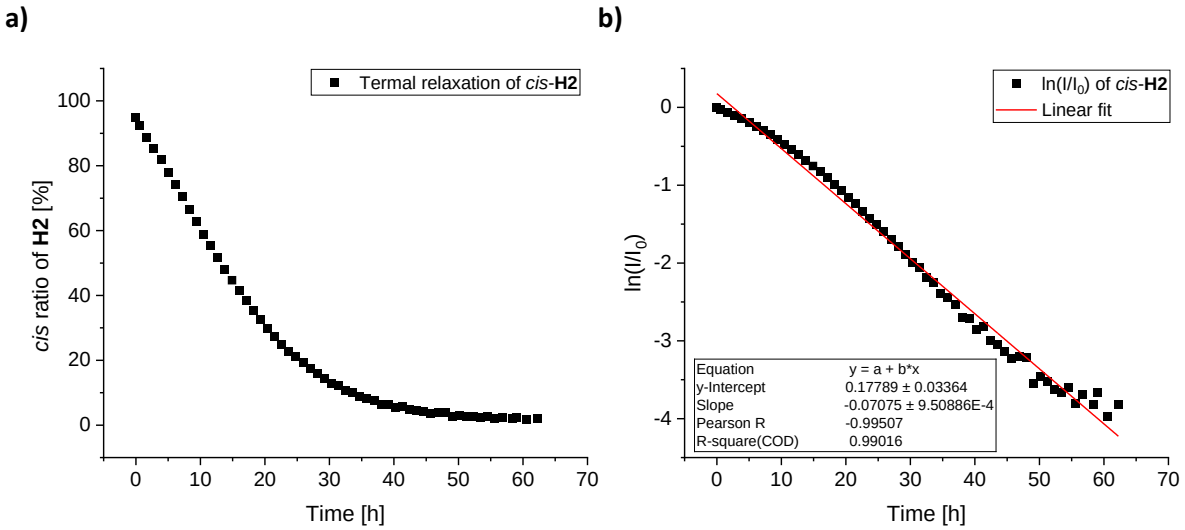


Figure 5.5-40: a) Thermal *cis* ⟶ *trans* relaxation of **H2** (17.8 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 9 h 48 min.

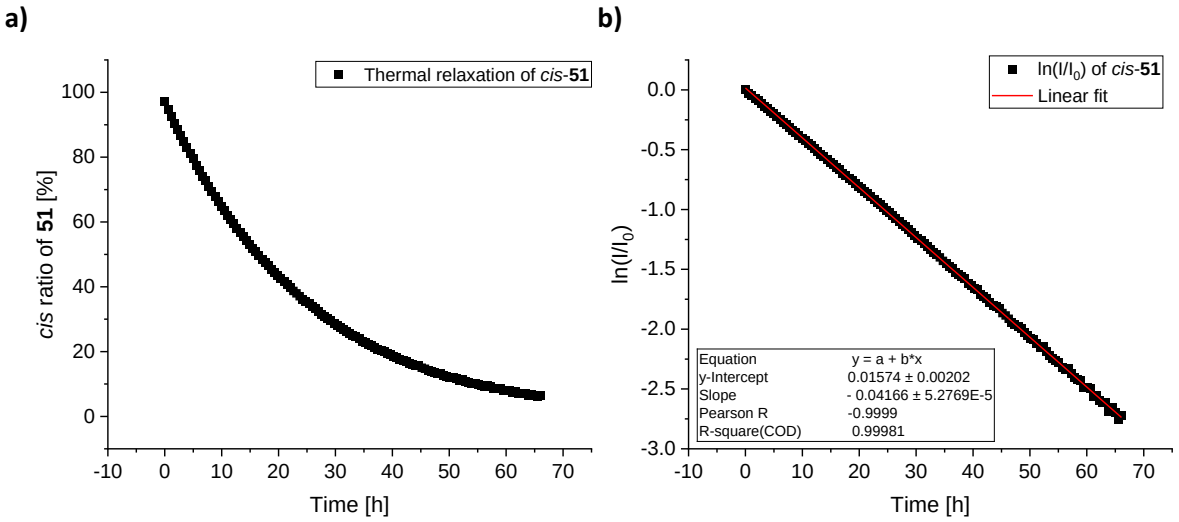


Figure 5.5-41: a) Thermal *cis* → *trans* relaxation of **51** (22.9 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 16 h 38 min.

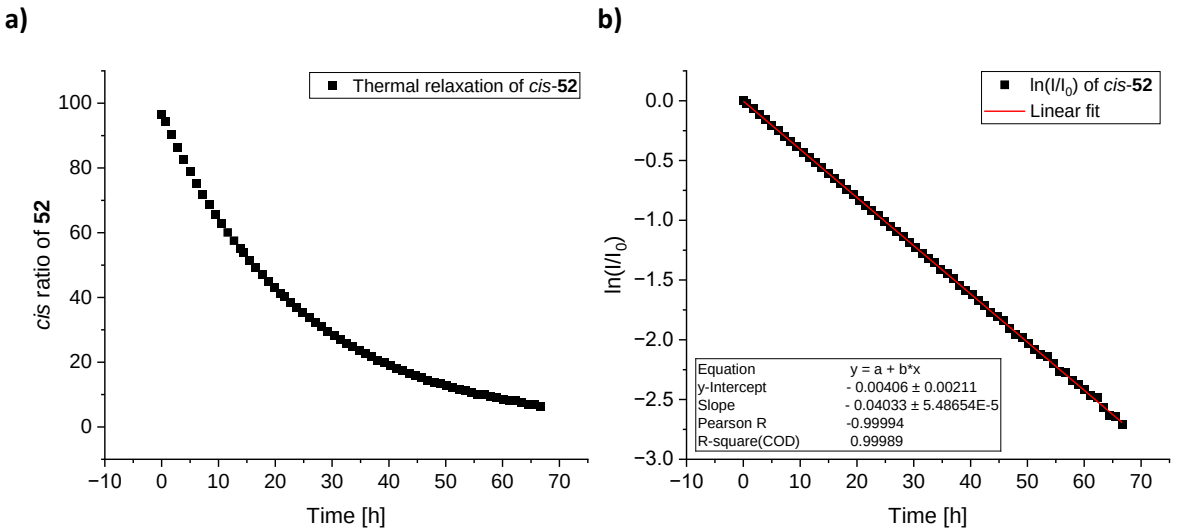


Figure 5.5-42: a) Thermal *cis* → *trans* relaxation of **52** (23.7 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 17 h 11 min.

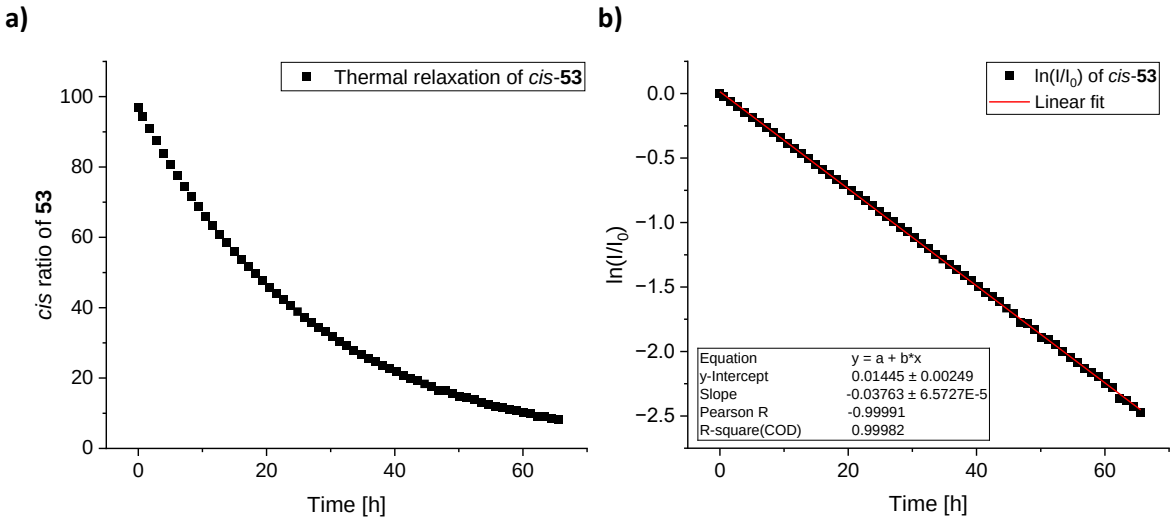


Figure 5.5-43: a) Thermal *cis* ⟶ *trans* relaxation of **53** (22.3 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 18 h 25 min.

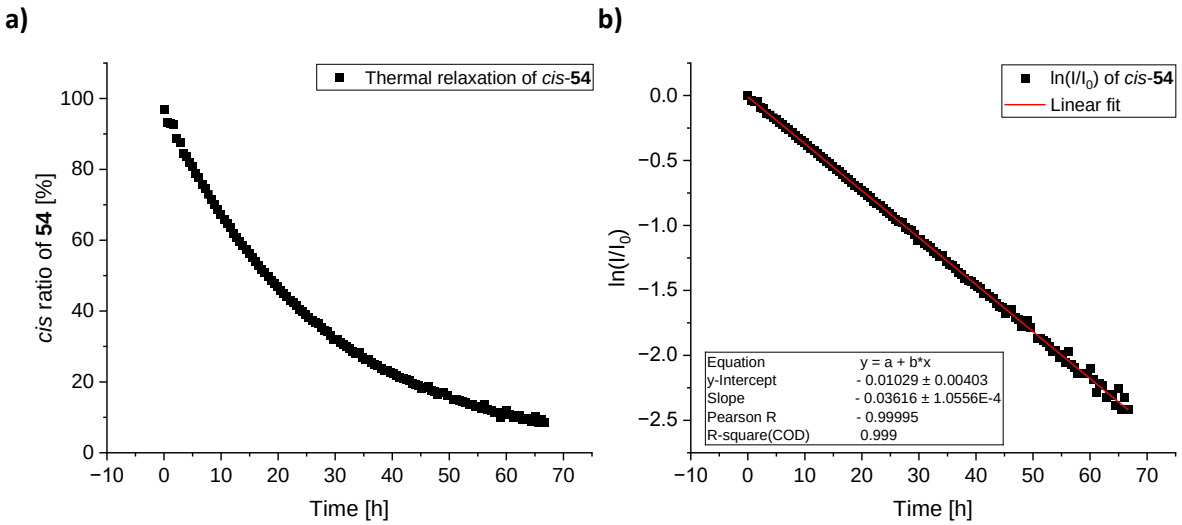


Figure 5.5-44: Thermal *cis* ⟶ *trans* relaxation of **54** (22.8 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 19 h 10 min.

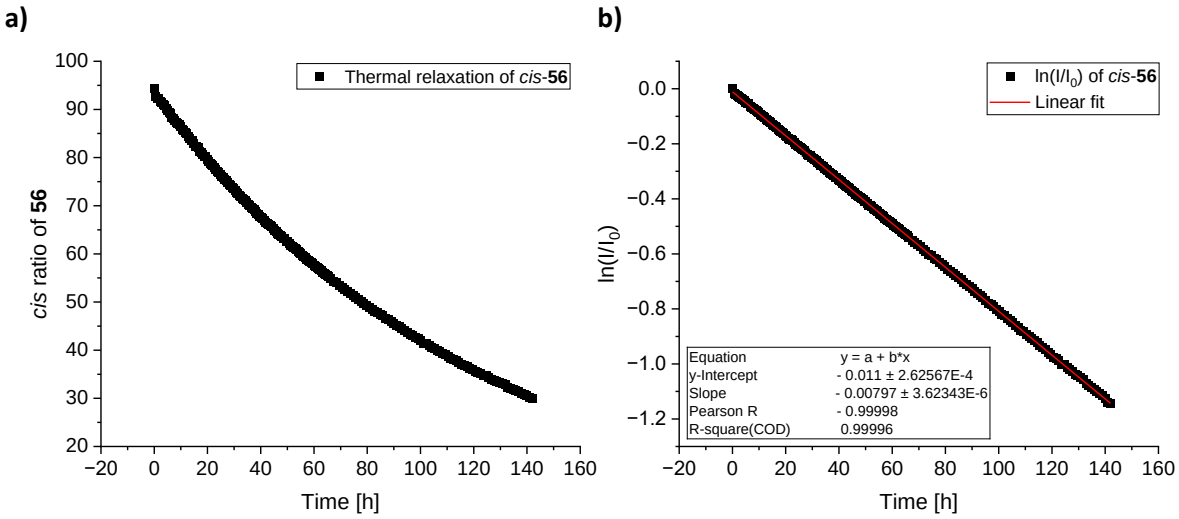


Figure 5.5-45: a) Thermal *cis* ⟶ *trans* relaxation of **56** (26.7 mM) at 298 K, monitored by ^{1}H NMR spectroscopy (600 MHz, CD_3OD). **b)** Log-linear plot for the calculation of $\boldsymbol{\tau_{1/2}}$ = 86 h 58 min.

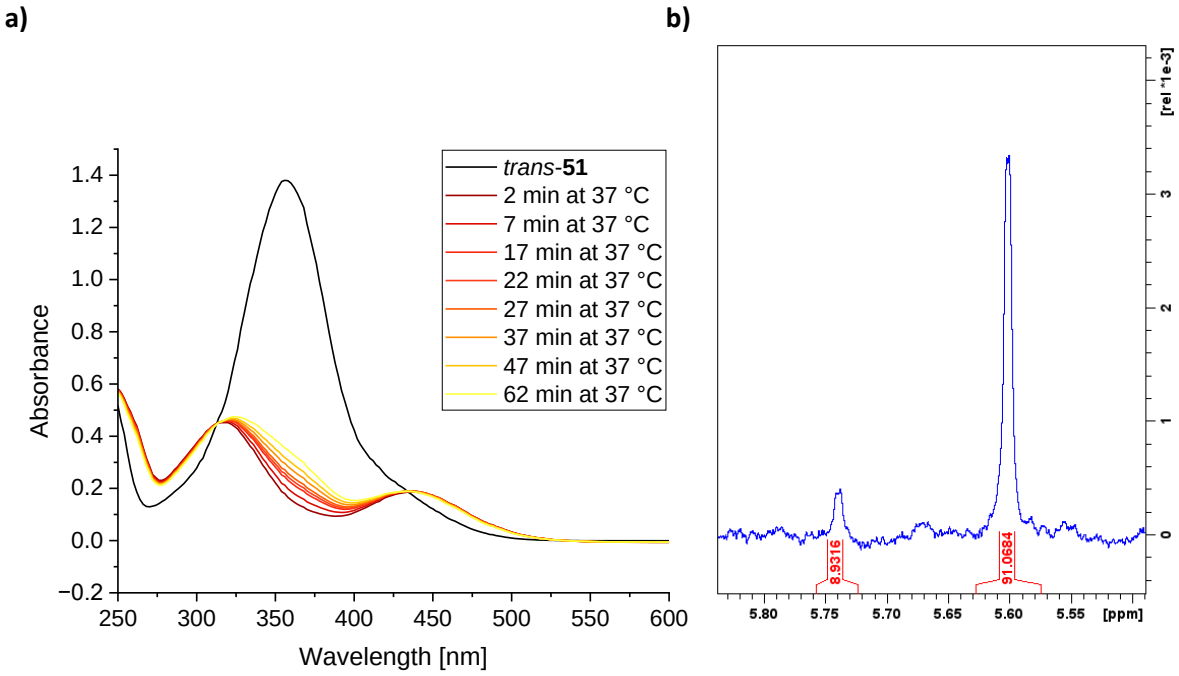


Figure 5.5-46: a) Thermal *cis* ⟶ *trans* relaxation of **51** (0.025 mM) at 37 °C monitored by UV/vis spectroscopy. The thermally relaxed sample was irradiated at 365 nm for 5 min and the deviation of UV/vis absorbance was monitored for 1 h at 37 °C. *Trans*:*cis* ratio after 17 min of relaxation according to the linear regression from Figure 5.4-32b: 8:92. Measured in PBS/methanol, 9:1, on a Tecan microplate reader. **b)** Section of an ^{1}H NMR spectrum of **51** (3.8 mM) after 14 min of relaxation at 37 °C, showing the H-1 integrals of the *trans*- and *cis*-isomer (600 MHz, D_2O/CD_3OD, 9:1, 310 K).

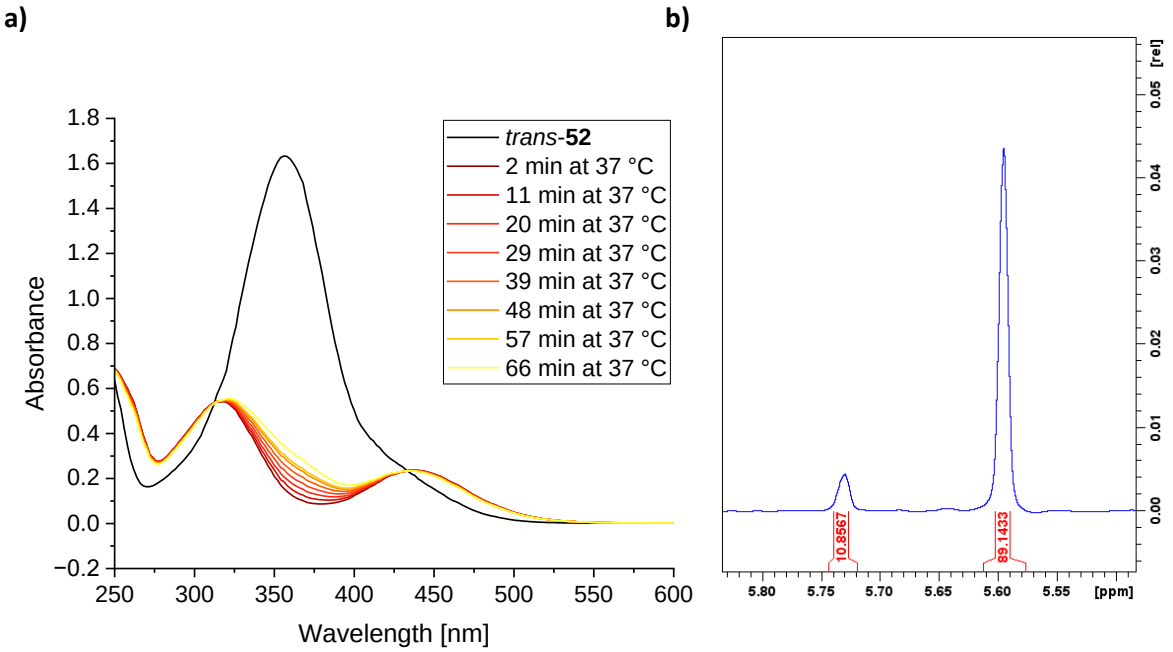


Figure 5.5-47: a) Thermal *cis* → *trans* relaxation of **52** (0.025 mM) at 37 °C monitored by UV/vis spectroscopy. The thermally relaxed sample was irradiated at 365 nm for 5 min and the deviation of UV/vis absorbance was monitored for 1 h at 37 °C. *Trans*:*cis* ratio after 25 min of relaxation according to the linear regression from Figure 5.4-33b: 9:91. Measured in PBS/methanol, 9:1, on a Tecan microplate reader. **b)** Section of an ^{1}H NMR spectrum of **52** (4.2 mM) after 40 min of relaxation at 37 °C, showing the H-1 integrals of the *trans*- and *cis*-isomer (600 MHz, D_2O/CD_3OD, 9:1, 310 K).

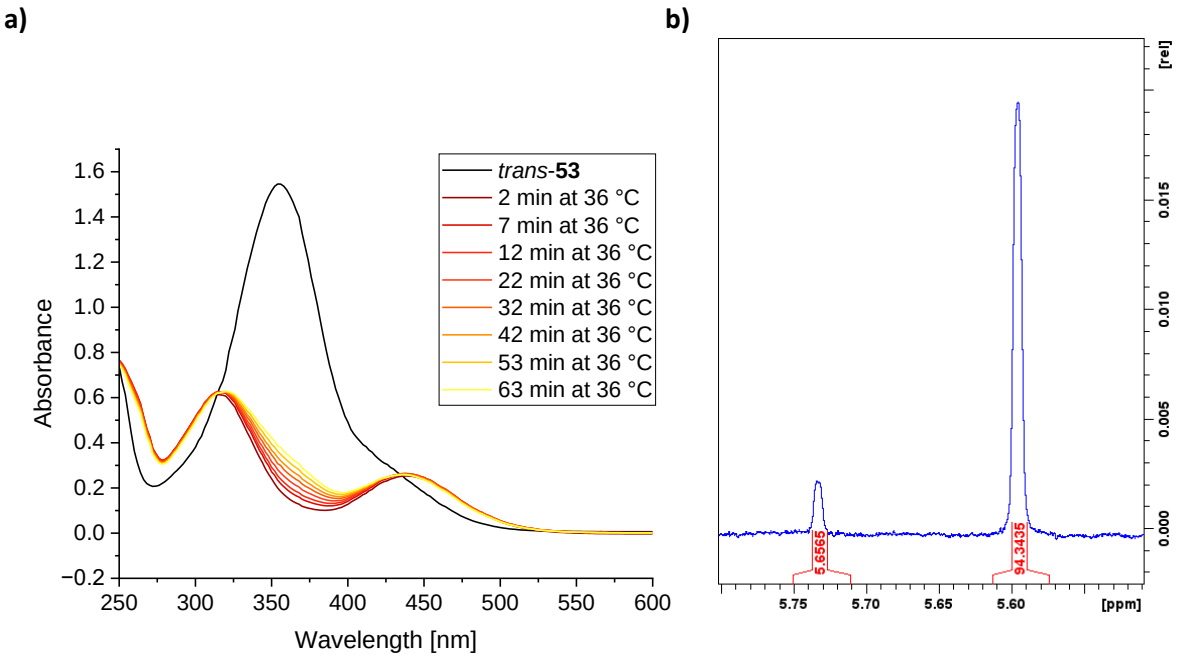


Figure 5.5-48: a) Thermal *cis* → *trans* relaxation of **53** (0.25 mM) at 37 °C monitored by UV/vis spectroscopy. The thermally relaxed sample was irradiated at 365 nm for 5 min and the deviation of UV/vis absorbance was monitored for 1 h at 37 °C. *Trans*:*cis* ratio after 22 min of relaxation according to the linear regression from Figure 5.4-34b: 7:93. Measured in PBS/methanol, 9:1, on a Tecan microplate reader. **b)** Section of an ^{1}H NMR spectrum of **53** (4 mM) after 28 min of relaxation at 37 °C, showing the H-1 integrals of the *trans*- and *cis*-isomer (600 MHz, D_2O/CD_3OD, 9:1, 310 K).

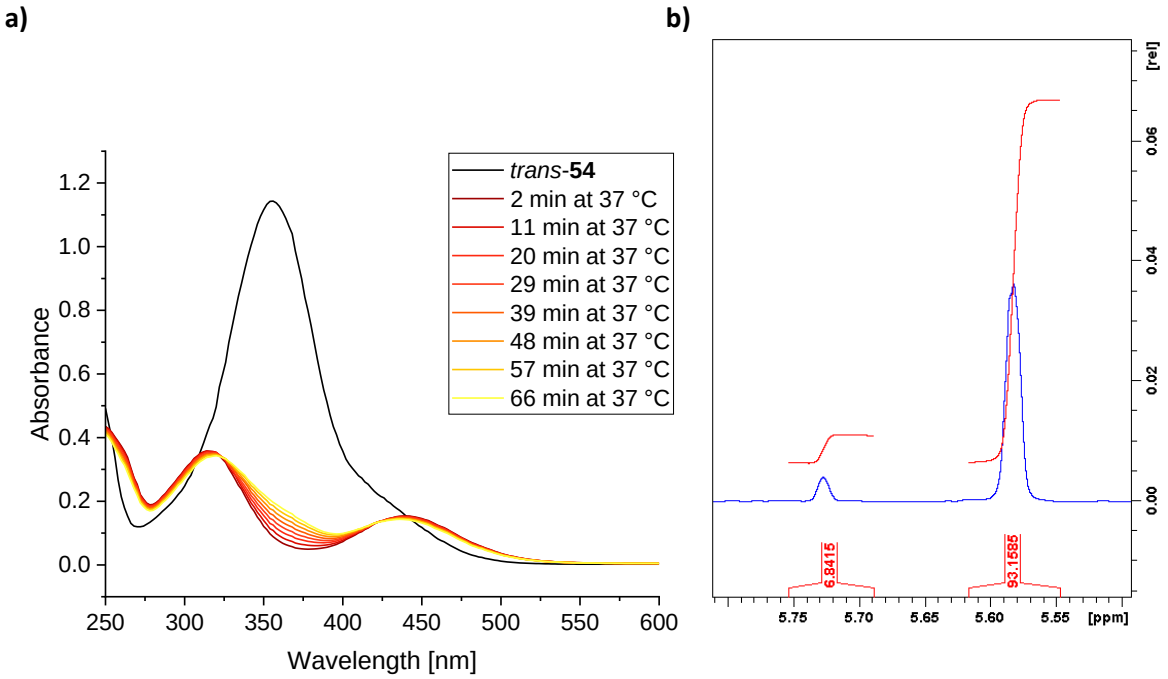


Figure 5.5-49: a) Thermal *cis* → *trans* relaxation of **54** (0.25 mM) at 37 °C monitored by UV/vis spectroscopy. The thermally relaxed sample was irradiated at 365 nm for 5 min and the deviation of UV/vis absorbance was monitored for 1 h at 37 °C. *Trans*:*cis* ratio after 20 min of relaxation according to the linear regression from Figure 5.5-35b: 7:93. Measured in PBS/methanol, 9:1, on a Tecan microplate reader. **b)** Section of an ^{1}H NMR spectrum of **54** (4.3 mM) after 25 min of relaxation at 37 °C, showing the H-1 integrals of the *trans*- and *cis*-isomer (600 MHz, D_2O/CD_3OD, 9:1, 310 K).

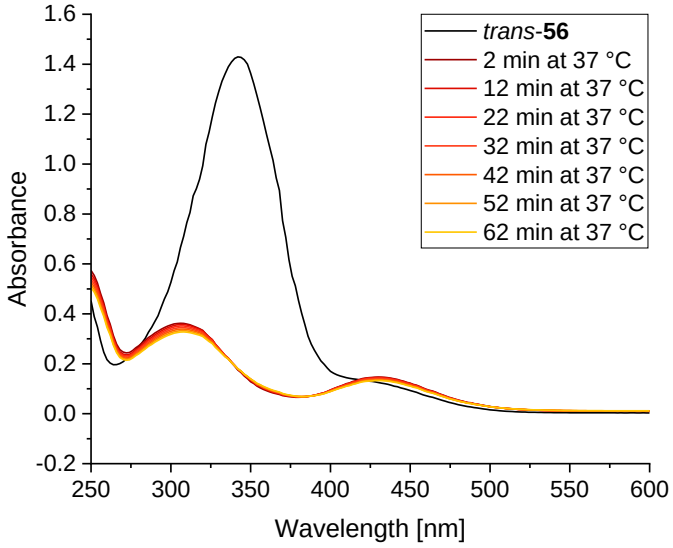


Figure 5.5-50: Thermal *cis* → *trans* relaxation of **56** (0.25 mM) at 37 °C monitored by UV/vis spectroscopy. The thermally relaxed sample was irradiated at 365 nm for 5 min and the deviation of UV/vis absorbance was monitored for 1 h at 37 °C. *Trans*:*cis* ratio after 22 min of relaxation according to the linear regression from Figure 5.4-36b: 8:92.
Measured in PBS/methanol, 9:1, on a Tecan microplate reader.

5.6 Computational methods

5.6.1 Preparation of X-ray crystal structures and ligands

Illustrations of X-ray crystal structures and *in silico*-simulations were performed using the Maestro suite for molecular modelling (Schrödinger Inc.). The structures of azobenzene mannoside derivatives **3**, **4**, **H1**, **H2**, **F1**, and **F2** were drawn with the implemented 2D Sketcher function. Minimisation and conformational search was performed using the MacroModel application[431] with the OPLS-2005[427,428] force field in implicit water and in consideration of the calculated distances, dihedral angles and torsions of *trans*-azobenzene (N=N 1.259 Å, C-N 1.438 Å, C-C 1.387 Å, NNC 114.2°, CNNC 180°, NNCC 0°) and *cis*-azobenzene (N=N 1.24 Å, C-N 1.472 Å, C-C 1.387 Å, NNC 121.9°, CNNC 0°, NNCC 60.8°).[684] The FimH PDB structures 1UWF[373] (closed-gate conformation) and 1KLF[318] (open-gate conformation, chain B) were imported and prepared with the Protein Preparation Wizard tool. Excess co-crystallised Fim proteins and water molecules were removed except for the entries chain A, W2255 (1UWF) and chain B, W1506 (1KLF). The two lectin domain structures share the identical protein sequence and both adopt the elongated high-affinity (HA) structure, with the exception of the flexible tyrosine gate formed by Tyr48 and Tyr137.[371] The cysteine residues on positions Asn136 and Ser139 were introduced manually using the mutate residue tool, followed by a minimisation via MacroModel.

5.6.2 Glide docking, *in silico*-alkylation and energy minimisation

Receptor grids around the binding pocket were defined using the Glide application[426] in the OPLS-2005 force field. For this purpose, the following residues and molecules were specified as binding site: Ile13 – Gly14, Asp47 – Tyr48, Glu50 – Ile52, Gln133, Asn135 – 138, Asp 140 – 141, as well as the co-crystallised water molecules chain A: W2255 (1UWF) and chain B: W1506 (1KLF). The tyrosine gate residues Tyr48 and Tyr137 were defined as rotatable. Standard precision ligand docking was performed with the respective function in Glide, setting the ligand sampling as flexible with the exception of the above-mentioned geometries of *trans*- and *cis*-azobenzene. Sampling of nitrogen conversions and ring conformations was set as allowed. As output, a maximum of 20 poses per ligand were calculated and a post-docking minimisation was performed with a rejection threshold of 0.50 kcal · mol^{-1}. A selection of docking scores, ligand interaction diagrams and poses from standard docking simulations are given in Tables 5.5-1 – 5.5-10, and Figures 5.5-1 – 5.5-30. The resulting poses from docking simulations with the cysteine variants of FimH in 1KLF and 1UWF were evaluated and distances between the thiol-sulfur atom and the iodoacetamidomethyl-carbon atom were measured. Selected poses with preferably small S – C distances (< 4 Å) were chosen and alkylation was performed manually using the 3D Builder tool. The *S*-carbamidoalkylated structures were once again energy-minimised using the MacroModel application with the OPLS-2005 force field in implicit water and with constrained geometry of *trans*- and *cis*-azobenzene. The resulting structures were assessed and categorised according to whether the mannose moiety was correctly bound in the binding pocket (bound), whether the mannose moiety was not bound (unbound) or rejected, if the structure of ligands or lectins was illogical or incorrect (wrong, Tables 5.5-11 – 5.5-12). Ligand interaction diagrams and illustrations of energy-minimised poses with correct docking are shown in Figures 5.5-31 – 5.5-41. The following assignment is used in ligand interaction diagrams: amino acids and ribbons in red, blue and green symbolise negatively charged, polar or hydrophobic properties; arrows in magenta, ochre yellow, green and black indicate hydrogen bonds, halogen bonds, π-π stacking or covalent bonds, respectively. Solvent exposure is visualised by grey outlines around the respective atoms.

Table 5.6-1: Selected scoring values from docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the crystal structure of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal
trans-**F1**	-6.144	-0.943	-26.351	-32.645	-78.286	-58.996	14.670
trans-**F1**	-6.115	-1.013	-21.884	-40.523	-82.855	-62.407	12.109
trans-**F1**	-6.030	-0.931	-23.946	-31.755	-75.029	-55.701	13.609
trans-**F1**	-5.882	-0.730	-23.972	-29.197	-66.995	-53.168	23.333
trans-**F1**	-5.639	-0.878	-23.445	-24.002	-65.867	-47.447	12.035
trans-**F1**	-5.198	-0.696	-22.547	-22.868	-60.878	-45.415	14.133
trans-**F1**	-4.672	-0.920	-26.191	-16.650	-58.642	-42.841	8.888
cis-**F1**	-6.471	-0.924	-28.236	-40.232	-91.537	-68.468	15.979
cis-**F1**	-6.359	-0.466	-23.617	-31.718	-80.966	-55.335	9.774
cis-**F1**	-6.195	-0.832	-21.213	-40.247	-84.731	-61.460	12.920
cis-**F1**	-6.190	-0.936	-26.133	-33.498	-83.041	-59.631	12.594
cis-**F1**	-6.059	-0.750	-20.455	-31.195	-76.710	-51.650	8.032
cis-**F1**	-5.970	-0.905	-24.909	-33.065	-82.110	-57.974	9.028
cis-**F1**	-5.910	-0.547	-25.715	-33.040	-82.360	-58.756	9.522
cis-**F1**	-5.727	-0.938	-19.955	-24.203	-65.751	-44.158	11.783
cis-**F1**	-5.704	-0.528	-21.250	-31.225	-75.875	-52.474	7.951
cis-**F1**	-5.579	-0.351	-22.986	-31.465	-77.389	-54.452	7.673
trans-**F2**	-6.741	-1.149	-27.460	-36.109	-87.341	-63.569	12.322
trans-**F2**	-6.379	-0.917	-25.935	-31.801	-79.951	-57.736	12.217
trans-**F2**	-6.211	-1.034	-25.082	-35.164	-81.313	-60.246	13.026
trans-**F2**	-6.115	-0.982	-22.276	-39.402	-80.623	-61.679	16.418
trans-**F2**	-6.049	-1.007	-23.218	-38.193	-77.630	-61.411	21.283
trans-**F2**	-5.865	-0.858	-27.223	-27.068	-68.964	-54.292	22.740
trans-**F2**	-5.417	-0.831	-15.153	-33.568	-65.079	-48.721	15.393
trans-**F2**	-4.866	-0.115	-14.633	-27.430	-55.969	-42.063	15.402
trans-**F2**	-4.781	0.000	-15.753	-33.478	-66.006	-49.231	8.910
cis-**F2**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.

a) No correct docking results and/or azobenzene structures observed.

Table 5.6-2: Selected scoring values from docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the crystal structure of 1KLF (open-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal
trans-**F1**	-5.212	-0.551	-18.888	-27.188	-56.825	-46.076	11.209
trans-**F1**	-5.173	-0.607	-15.753	-27.000	-53.503	-42.753	10.963
trans-**F1**	-5.136	-0.630	-15.044	-26.688	-53.970	-41.732	7.759
trans-**F1**	-4.709	-0.669	-14.483	-23.716	-49.319	-38.199	7.312
trans-**F1**	-4.524	-0.181	-1.453	-31.756	-43.162	-33.209	8.486
trans-**F1**	-4.260	-0.085	-9.855	-33.245	-50.699	-43.101	11.541
trans-**F1**	-4.239	-0.079	-7.392	-33.587	-48.654	-40.979	11.254
trans-**F1**	-4.211	-0.083	-7.777	-33.550	-50.834	-41.327	7.417
trans-**F1**	-4.066	-0.205	-2.117	-33.458	-44.757	-35.574	7.153
trans-**F1**	-3.920	-0.382	-15.142	-19.001	-41.559	-34.143	9.769
cis-**F1**	-6.121	-0.805	-21.780	-38.094	-81.256	-59.873	10.498
cis-**F1**	-5.824	-0.548	-19.930	-39.522	-78.938	-59.452	11.711
cis-**F1**	-5.549	-0.557	-17.729	-29.711	-68.987	-47.440	5.193
cis-**F1**	-5.522	-0.531	-17.801	-35.942	-73.534	-53.744	8.469
cis-**F1**	-5.514	-0.717	-20.157	-33.179	-72.865	-53.336	8.923
cis-**F1**	-5.511	-0.190	-21.706	-30.897	-70.974	-52.603	11.213
cis-**F1**	-5.500	-0.238	-25.815	-33.285	-76.793	-59.100	12.475
cis-**F1**	-5.495	-0.312	-20.512	-29.313	-70.702	-49.825	6.060
cis-**F1**	-5.471	-0.107	-22.550	-30.332	-73.223	-52.883	6.927
cis-**F1**	-5.412	-0.645	-19.712	-32.429	-71.323	-52.141	8.730
trans-**F2**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**F2**	-4.576	-0.221	-17.246	-24.259	-51.709	-41.506	11.637
cis-**F2**	-4.560	-0.185	-13.857	-34.280	-57.735	-48.136	12.730

a) No correct docking results and/or azobenzene structures observed.

Table 5.6-3: Selected scoring values and S – C distances from docking simulations of azobenzene derivatives **H1** and **H2** in the N136C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**H1**	-6.027	-0.995	-21.142	-37.020	-78.066	-58.162	12.539	11.01[a]
trans-**H1**	-5.976	-1.157	-16.294	-33.050	-71.003	-49.344	8.586	13.44
trans-**H1**	-5.958	-1.288	-23.759	-23.173	-67.404	-46.932	10.801	13.74
trans-**H1**	-5.899	-0.975	-18.240	-27.999	-66.006	-46.239	11.694	14.18
trans-**H1**	-5.813	-0.821	-13.260	-37.011	-70.862	-50.272	9.304	12.58
trans-**H1**	-5.801	-0.945	-16.292	-33.532	-71.024	-49.824	7.982	15.77
trans-**H1**	-5.728	-1.008	-14.443	-27.632	-63.407	-42.075	7.082	11.96
trans-**H1**	-5.721	-0.939	-19.348	-40.836	-79.096	-60.184	11.858	11.36
trans-**H1**	-5.637	-0.872	-17.357	-41.414	-76.554	-58.772	13.390	17.10
trans-**H1**	-5.583	-0.830	-15.705	-32.118	-66.587	-47.822	10.961	12.36
cis-**H1**	-6.354	-1.014	-21.546	-34.166	-82.363	-55.712	7.725	20.26
cis-**H1**	-6.308	-0.845	-20.002	-42.345	-88.338	-62.348	8.601	18.26
cis-**H1**	-6.238	-1.045	-19.817	-33.964	-80.319	-53.782	6.841	19.84
cis-**H1**	-6.142	-0.811	-17.392	-42.874	-85.556	-60.266	8.408	18.73
cis-**H1**	-5.898	-0.262	-22.219	-37.023	-82.726	-59.242	9.676	15.66
cis-**H1**	-5.693	-0.958	-21.584	-28.862	-73.056	-50.446	9.454	19.15
cis-**H1**	-5.645	-1.007	-19.417	-31.648	-73.616	-51.065	9.117	16.42
cis-**H1**	-5.620	-0.801	-17.569	-44.491	-84.072	-62.060	9.949	18.07
cis-**H1**	-5.613	-0.900	-20.541	-28.948	-72.537	-49.489	7.817	19.76
cis-**H1**	-5.571	-0.737	-23.300	-22.565	-66.683	-45.865	11.872	13.49[b]
trans-**H2**	-5.726	-0.681	-22.653	-34.189	-74.988	-56.841	13.596	14.31
trans-**H2**	-5.657	-0.581	-23.066	-33.604	-73.815	-56.670	14.994	15.15
trans-**H2**	-5.015	-0.231	-17.206	-38.875	-71.176	-56.081	13.506	14.09
trans-**H2**	-4.745	-0.048	-16.723	-27.883	-56.411	-44.606	17.733	12.94[c]
trans-**H2**	-4.525	-0.143	-14.821	-28.995	-56.332	-43.816	14.389	16.84
trans-**H2**	-4.366	-0.011	-11.132	-30.255	-54.995	-41.387	10.817	12.74
trans-**H2**	-4.311	-0.000	-13.070	-30.349	-58.580	-43.419	7.234	14.19
trans-**H2**	-4.181	-0.011	-10.915	-29.223	-50.430	-40.138	15.837	12.96
trans-**H2**	-4.092	-0.041	-14.048	-24.714	-49.423	-38.763	14.327	12.60
cis-**H2**	-6.346	-0.918	-20.290	-39.531	-80.906	-59.821	17.533	14.31
cis-**H2**	-6.265	-0.691	-26.776	-29.105	-78.597	-55.881	13.507	14.96[d]
cis-**H2**	-6.090	-0.926	-24.885	-39.578	-85.474	-64.463	15.273	13.81
cis-**H2**	-6.038	-0.833	-19.237	-35.272	-75.279	-54.509	15.270	14.65
cis-**H2**	-5.954	-0.810	-23.892	-31.094	-77.092	-54.986	11.805	13.71
cis-**H2**	-5.873	-0.863	-21.968	-39.930	-84.021	-61.898	11.011	14.41
cis-**H2**	-5.871	-0.950	-27.238	-24.566	-71.986	-51.804	14.868	15.73
cis-**H2**	-5.831	-0.829	-20.815	-31.199	-76.393	-52.015	6.103	13.58
cis-**H2**	-5.821	-0.848	-19.885	-27.191	-71.616	-47.076	5.686	14.88
cis-**H2**	-5.810	-0.816	-21.343	-40.856	-83.653	-62.199	11.755	11.28

a) Ligand interaction diagram and illustration shown in Figure 5.6-1. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-2. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-3. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-4.

Table 5.6-4: Selected scoring values and S – C distances from docking simulations of azobenzene derivatives **H1** and **H2** in the S139C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**H1**	-6.184	-1.310	-22.728	-24.152	-68.400	-46.880	10.146	15.17
trans-**H1**	-6.003	-0.932	-19.709	-29.924	-71.223	-49.633	8.451	21.72
trans-**H1**	-5.857	-0.834	-13.959	-36.478	-69.994	-50.437	11.258	14.59[a,e]
trans-**H1**	-5.839	-1.306	-22.600	-22.797	-66.639	-45.397	7.737	21.55
trans-**H1**	-5.836	-1.082	-14.978	-33.649	-69.479	-48.627	8.489	16.91
trans-**H1**	-5.835	-0.959	-17.535	-28.240	-67.033	-45.776	7.667	17.84
trans-**H1**	-5.605	-0.841	-15.241	-34.150	-69.868	-49.390	7.252	16.35
trans-**H1**	-5.589	-1.096	-21.776	-25.744	-67.449	-47.520	8.213	17.12
trans-**H1**	-5.559	-0.851	-13.828	-33.357	-66.406	-47.185	9.367	17.17
trans-**H1**	-5.548	-0.756	-15.897	-28.755	-64.097	-44.653	8.832	23.49
cis-**H1**	-6.354	-1.016	-21.555	-34.170	-81.237	-55.725	8.365	18.30
cis-**H1**	-6.317	-0.577	-25.830	-29.951	-79.339	-55.780	11.927	3.24[b,e]
cis-**H1**	-6.158	-1.041	-18.298	-37.880	-81.491	-56.178	6.927	18.09
cis-**H1**	-6.145	-0.846	-16.458	-43.106	-83.313	-59.564	9.940	20.64
cis-**H1**	-5.857	-0.912	-20.098	-41.422	-84.756	-61.520	8.271	18.75
cis-**H1**	-5.636	-0.893	-21.005	-28.762	-72.575	-49.767	7.060	14.75
cis-**H1**	-5.626	-1.006	-19.130	-32.439	-73.421	-51.569	8.879	9.19
cis-**H1**	-5.436	-0.853	-19.078	-28.141	-68.666	-47.219	7.920	16.20
cis-**H1**	-4.472	-0.780	-24.212	-17.572	-57.442	-41.783	10.484	3.70
cis-**H1**	-4.328	-0.699	-28.415	-16.144	-59.340	-44.558	10.892	3.37
trans-**H2**	-5.604	-0.568	-22.263	-35.347	-72.025	-57.611	20.074	3.50
trans-**H2**	-5.475	-0.621	-22.880	-32.447	-69.371	-55.327	19.695	4.01
trans-**H2**	-5.319	-0.446	-18.899	-35.127	-70.411	-54.026	13.644	4.66
trans-**H2**	-5.013	-0.141	-19.305	-28.222	-60.869	-47.528	17.062	3.66[c,e]
trans-**H2**	-4.765	-0.299	-15.876	-36.972	-66.052	-52.848	15.175	15.44
trans-**H2**	-4.497	-0.017	-11.392	-32.574	-58.506	-43.966	10.164	4.38
trans-**H2**	-4.424	-0.071	-15.112	-26.335	-52.893	-41.446	15.714	4.94
trans-**H2**	-4.358	-0.001	-13.938	-29.438	-58.686	-43.375	7.408	3.75
trans-**H2**	-4.345	-0.076	-12.268	-30.481	-54.609	-42.749	14.201	14.70
trans-**H2**	-4.312	-0.004	-13.296	-27.342	-50.489	-40.638	17.927	5.30
cis-**H2**	-6.355	-0.766	-26.535	-28.192	-76.562	-54.727	16.121	21.55
cis-**H2**	-6.033	-0.859	-24.866	-26.574	-73.584	-51.440	12.472	7.78
cis-**H2**	-5.994	-0.888	-24.035	-33.727	-79.035	-57.762	13.853	3.24[d,e]
cis-**H2**	-5.881	-0.887	-21.153	-40.927	-84.088	-62.079	11.314	3.60
cis-**H2**	-5.875	-0.867	-21.364	-39.398	-81.446	-60.762	13.908	8.02
cis-**H2**	-5.821	-0.843	-20.304	-26.774	-71.759	-47.079	5.403	5.00
cis-**H2**	-5.804	-0.850	-19.677	-40.935	-80.673	-60.612	14.482	6.49
cis-**H2**	-5.799	-1.053	-22.042	-32.935	-74.936	-54.978	14.646	6.79
cis-**H2**	-5.728	-0.951	-18.886	-30.162	-70.606	-49.048	10.772	8.35
cis-**H2**	-5.719	-0.716	-13.948	-34.391	-67.834	-48.339	14.813	3.86

a) Ligand interaction diagram and illustration shown in Figure 5.6-5. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-6. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-7. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-8. **e)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-11).

a) **b)**

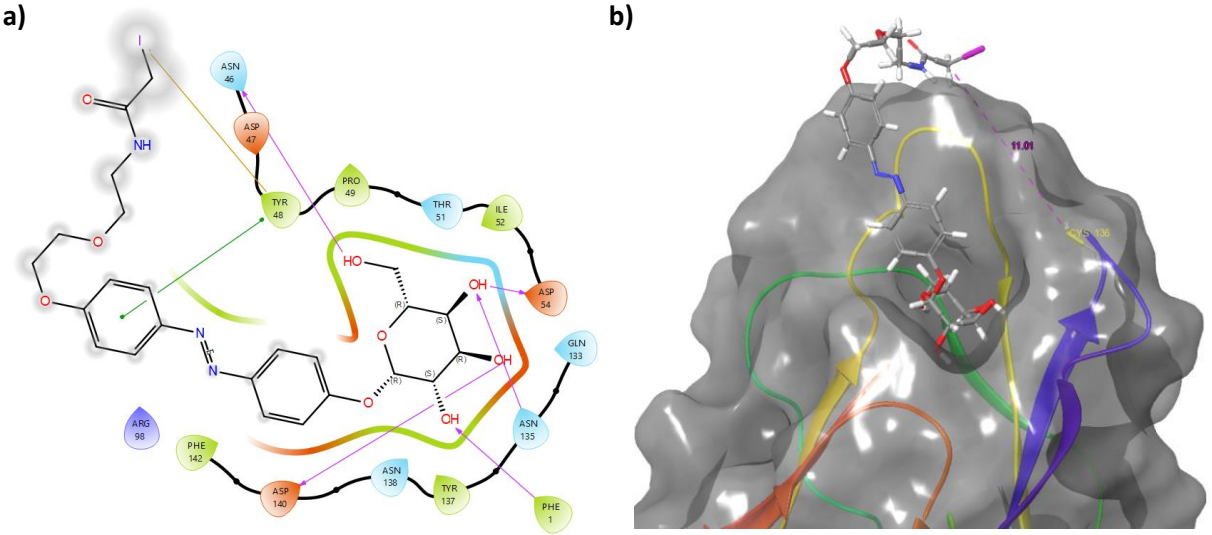


Figure 5.6-1: **a)** Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H1** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

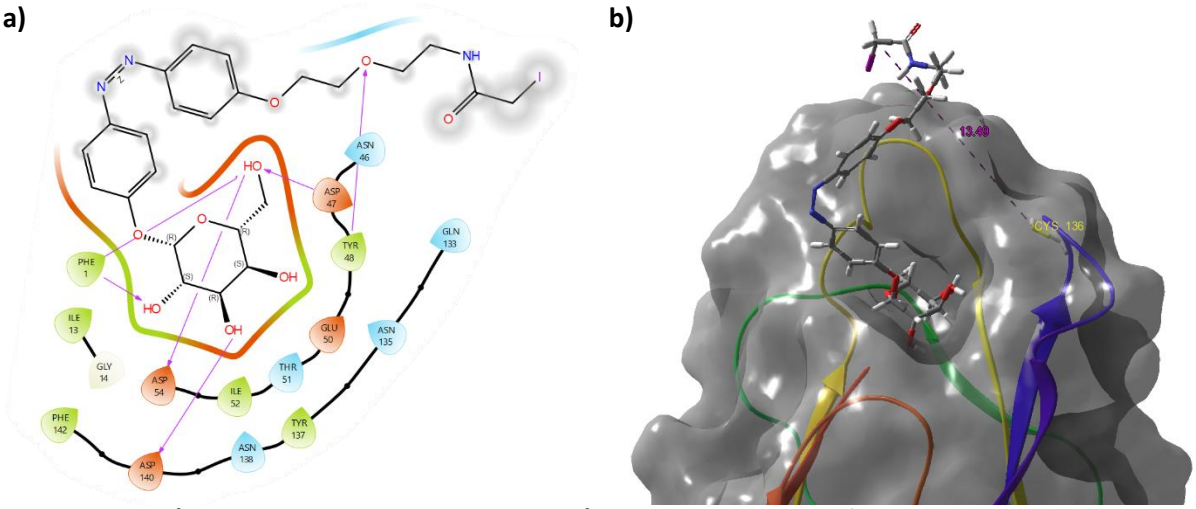


Figure 5.6-2: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H1** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

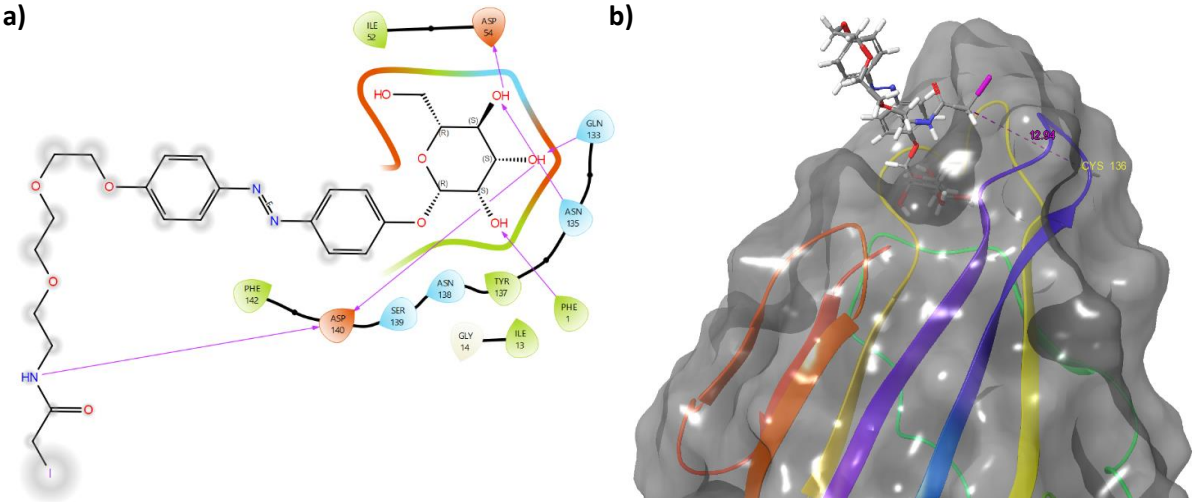


Figure 5.6-3: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

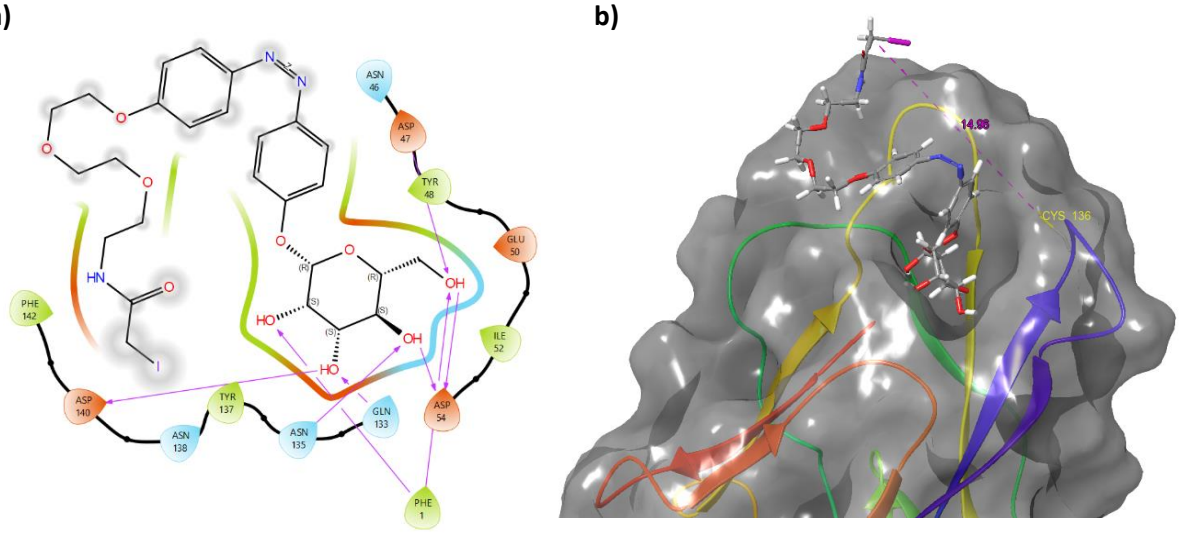


Figure 5.6-4: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

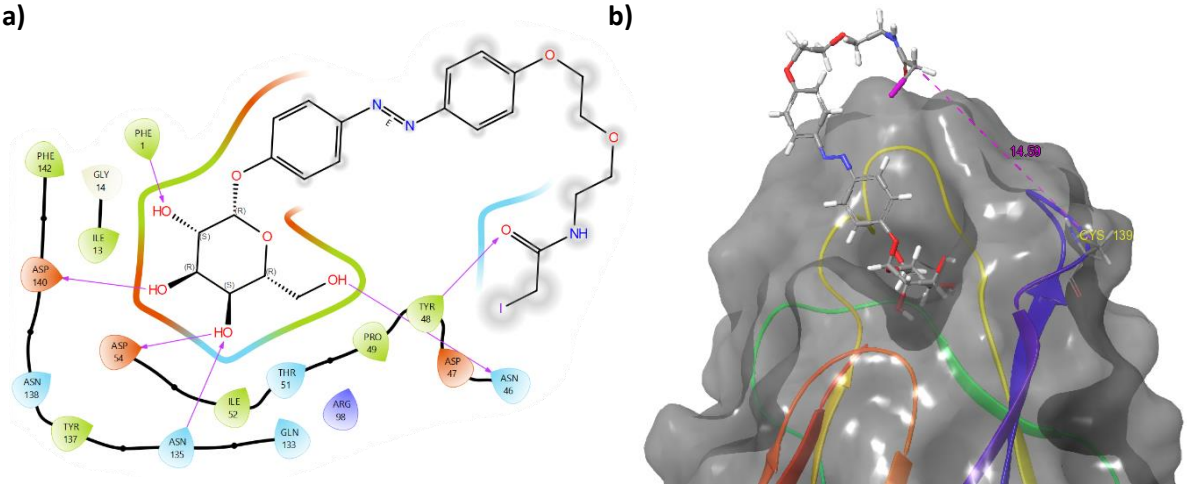


Figure 5.6-5: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H1** in the S139C variant of 1UWF (closed-gate conformation).

a) **b)**

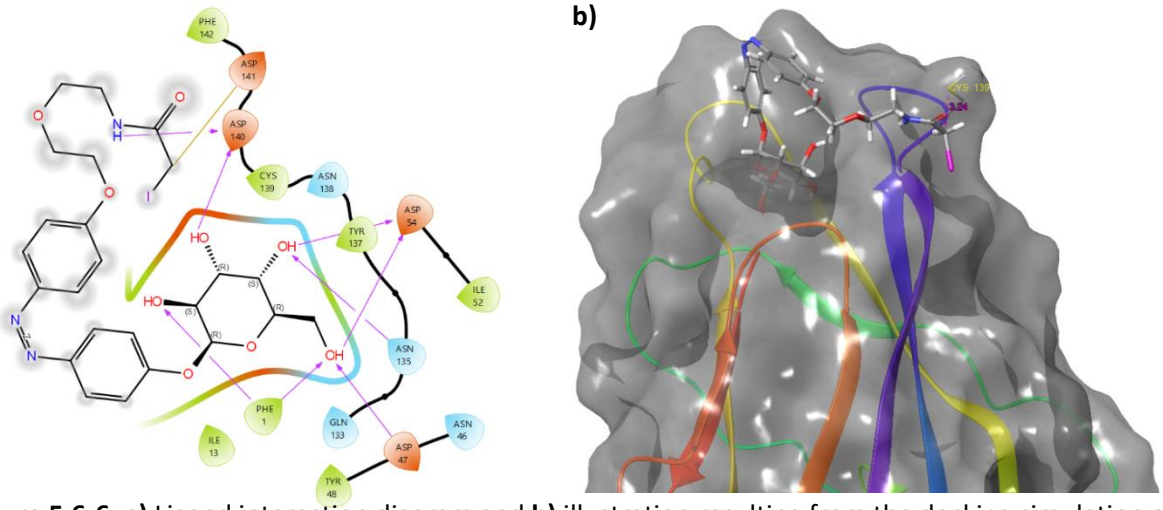


Figure 5.6-6: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H1** in the S139C variant of 1UWF (closed-gate conformation).

a) **b)**

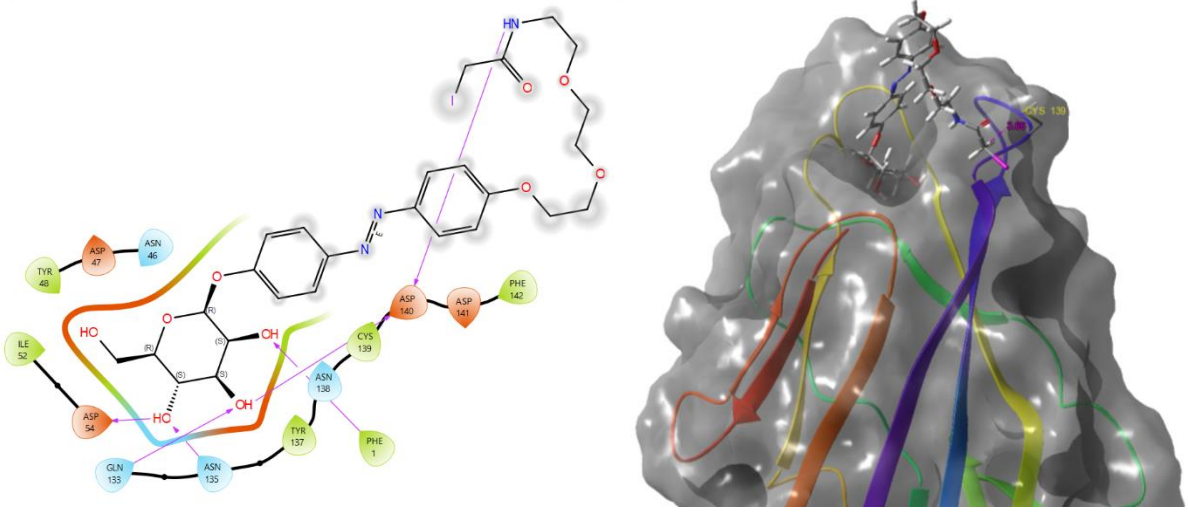


Figure 5.6-7: a) Ligand interaction diagram and **b)** illustrations resulting from the docking simulation of ligand *trans*-**H2** in the S139C variant of 1UWF (closed-gate conformation).

a) **b)**

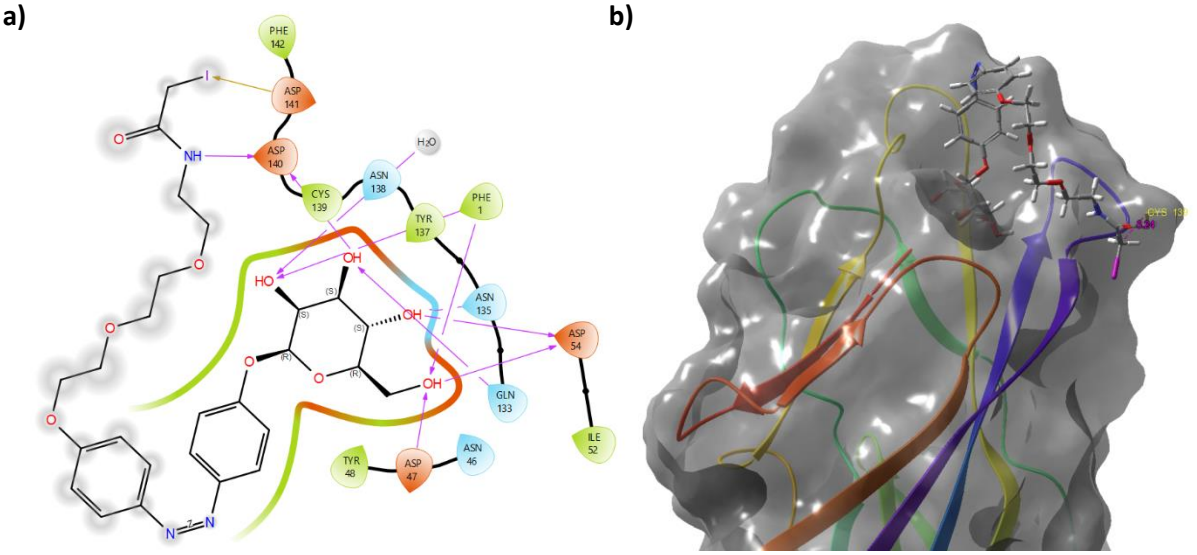


Figure 5.6-8: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H2** in the S139C variant of 1UWF (closed-gate conformation).

Table 5.6-5: Selected scoring values and S – C distances from docking simulations of azobenzene derivatives **H1** and **H2** in the N136C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**H1**	-4.681	-0.287	-16.773	-21.715	-48.714	-38.487	7.335	22.16
trans-**H1**	-4.321	-0.383	-10.156	-25.545	-43.594	-35.701	9.865	13.36[a,e]
trans-**H1**	-4.095	-0.113	-6.916	-30.540	-43.014	-37.456	13.195	18.07
trans-**H1**	-3.989	-0.249	-3.968	-30.328	-40.593	-34.296	11.089	20.02
trans-**H1**	-3.863	-0.172	-3.577	-30.738	-40.952	-34.315	9.658	20.10
trans-**H1**	-3.826	-0.149	-3.819	-26.445	-36.895	-30.264	9.452	19.66
trans-**H1**	-3.807	-0.183	-3.622	-26.154	-36.155	-29.775	9.843	19.95
trans-**H1**	-3.674	-0.307	-8.134	-22.452	-36.971	-30.586	9.038	18.97
cis-**H1**	-7.009	-0.476	-31.790	-26.158	-80.505	-57.948	10.318	3.28[b,e]
cis-**H1**	-6.372	-0.548	-28.276	-22.400	-71.750	-50.676	8.244	6.47
cis-**H1**	-6.335	-0.513	-28.818	-22.190	-71.790	-51.008	8.539	6.19
cis-**H1**	-6.219	-0.418	-28.034	-23.144	-71.345	-51.178	8.854	6.11
cis-**H1**	-6.120	-0.628	-30.099	-21.849	-69.393	-51.948	13.510	6.01
cis-**H1**	-5.688	-0.617	-18.557	-34.910	-69.988	-53.468	11.943	18.99
cis-**H1**	-5.618	-0.606	-13.401	-34.059	-65.698	-47.460	7.954	20.02
cis-**H1**	-5.570	-0.489	-33.952	-18.776	-68.597	-52.728	12.318	5.92
cis-**H1**	-5.561	-0.425	-25.938	-20.920	-65.204	-46.857	7.285	8.56
cis-**H1**	-5.458	-0.493	-13.149	-33.449	-62.902	-46.598	10.563	16.87
trans-**H2**	-5.120	-0.396	-17.174	-33.324	-60.428	-50.497	17.072	11.46
trans-**H2**	-4.584	-0.216	-12.460	-29.069	-52.515	-41.529	11.093	12.36
trans-**H2**	-4.472	-0.277	-11.462	-33.159	-56.439	-44.622	8.623	11.72
trans-**H2**	-4.433	-0.276	-9.935	-31.562	-54.198	-41.498	6.573	11.52
trans-**H2**	-4.431	-0.304	-7.470	-33.169	-50.037	-40.638	13.159	12.32[c]
trans-**H2**	-4.334	-0.144	-3.909	-32.401	-44.536	-36.310	14.811	13.19
trans-**H2**	-4.309	-0.215	-11.368	-32.037	-52.238	-43.404	13.411	10.83
trans-**H2**	-4.292	-0.266	-9.557	-29.635	-47.318	-39.192	14.704	11.99
trans-**H2**	-4.290	-0.568	-20.602	-21.289	-48.789	-41.891	17.147	11.20
trans-**H2**	-4.278	-0.303	-8.622	-30.246	-47.513	-38.868	13.566	12.00
cis-**H2**	-5.983	-0.500	-24.842	-28.762	-74.978	-53.604	14.123	15.11
cis-**H2**	-5.868	-0.852	-22.563	-35.382	-80.639	-57.945	10.388	14.59
cis-**H2**	-5.611	-0.117	-24.501	-41.652	-86.478	-66.153	12.687	17.26
cis-**H2**	-5.461	-0.181	-21.165	-42.078	-82.979	-63.243	12.437	17.62
cis-**H2**	-5.419	-0.590	-21.926	-32.336	-72.430	-54.262	15.174	13.18
cis-**H2**	-5.396	-0.387	-19.525	-36.818	-75.040	-56.343	13.894	10.24
cis-**H2**	-5.357	-0.551	-20.694	-27.446	-67.097	-48.140	13.007	15.99
cis-**H2**	-5.350	-0.525	-18.850	-34.370	-74.273	-53.220	8.749	14.73
cis-**H2**	-5.256	-0.433	-18.439	-34.745	-73.280	-53.184	9.763	14.56
cis-**H2**	-5.156	-0.138	-18.238	-36.919	-72.862	-55.157	13.597	16.32[d]

a) Ligand interaction diagram and illustration shown in Figure 5.5-9. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-10. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-11. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-12. **e)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-12).

Table 5.6-6: Selected scoring values and S – C distances from docking simulations of azobenzene derivatives **H1** and **H2** in the S139C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**H1**	-4.920	-0.411	-11.973	-25.534	-47.393	-37.507	10.964	8.92[a]
trans-**H1**	-4.794	-0.326	-17.833	-21.472	-50.333	-39.305	7.895	18.71
trans-**H1**	-4.392	-0.241	-10.523	-28.915	-48.186	-39.438	9.943	10.66
trans-**H1**	-4.355	-0.236	-9.892	-29.009	-45.412	-38.901	14.188	19.89
trans-**H1**	-4.322	-0.205	-9.786	-28.897	-46.193	-38.683	11.981	14.14
trans-**H1**	-4.040	-0.233	-3.724	-33.058	-43.694	-36.782	11.413	23.42
trans-**H1**	-3.989	-0.162	-5.959	-29.840	-43.779	-35.799	8.958	11.74
cis-**H1**	-6.305	-0.463	-29.061	-22.652	-73.028	-51.712	8.750	16.30
cis-**H1**	-6.033	-0.419	-34.576	-20.501	-74.473	-55.076	10.373	13.76
cis-**H1**	-5.838	-0.503	-29.232	-21.687	-67.703	-50.919	14.011	16.21
cis-**H1**	-5.749	-0.507	-28.416	-20.770	-68.902	-49.186	7.422	13.76
cis-**H1**	-5.678	-0.644	-27.288	-22.194	-67.052	-49.482	11.129	12.81
cis-**H1**	-5.674	-0.706	-27.834	-21.357	-65.358	-49.191	13.909	10.77
cis-**H1**	-5.526	-0.539	-20.033	-34.054	-70.307	-54.087	12.596	10.93
cis-**H1**	-5.460	-0.490	-18.640	-36.356	-71.094	-54.996	12.299	10.87
cis-**H1**	-5.434	-0.654	-9.502	-29.656	-57.427	-39.159	7.748	16.34[b]
cis-**H1**	-5.402	-0.742	-34.061	-17.199	-66.760	-51.261	13.026	17.40
trans-**H2**	-4.895	-0.332	-11.990	-33.799	-51.340	-45.789	19.000	8.00
trans-**H2**	-4.825	-0.586	-20.570	-21.425	-48.654	-41.994	16.358	18.10
trans-**H2**	-4.437	-0.042	-11.599	-36.588	-54.750	-48.186	14.161	3.18[c,e]
trans-**H2**	-4.309	-0.317	-8.924	-31.567	-47.281	-40.491	12.920	9.33
trans-**H2**	-4,294	-0.125	-7.459	-35.372	-52.435	-42.831	7.202	13.12
trans-**H2**	-4.260	-0.150	-5.380	-33.749	-48.989	-39.128	6.477	13.40
trans-**H2**	-4.202	-0.350	-7.040	-30.818	-47.618	-37.859	6.327	14.70
trans-**H2**	-4.171	-0.269	-6.418	-34.368	-47.629	-40.785	11.963	9.14
cis-**H2**	-5.556	-0.207	-21.851	-42.479	-84.729	-64.330	11.228	18.86
cis-**H2**	-5.447	-0.275	-25.026	-36.313	-78.310	-61.339	17.062	19.23
cis-**H2**	-5.370	-0.205	-23.061	-39.372	-80.518	-62.433	14.120	19.11
cis-**H2**	-5.325	-0.165	-17.285	-44.833	-82.180	-62.117	9.736	21.57
cis-**H2**	-5.210	-0.195	-20.888	-39.475	-79.043	-60.362	11.427	18.55
cis-**H2**	-4.966	-0.409	-18.652	-29.793	-67.237	-48.445	8.920	13.86[d,e]
cis-**H2**	-4.675	-0.155	-17.267	-34.002	-65.641	-51.270	15.032	20.51
cis-**H2**	-4.518	-0.112	-14.990	-31.908	-63.005	-46.898	10.093	21.36
cis-**H2**	-4.385	-0.451	-25.691	-19.705	-59.338	-45.396	13.178	20.96

a) Ligand interaction diagram and illustration shown in Figure 5.6-13. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-14. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-15. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-16. **e)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-12).

a) **b)**

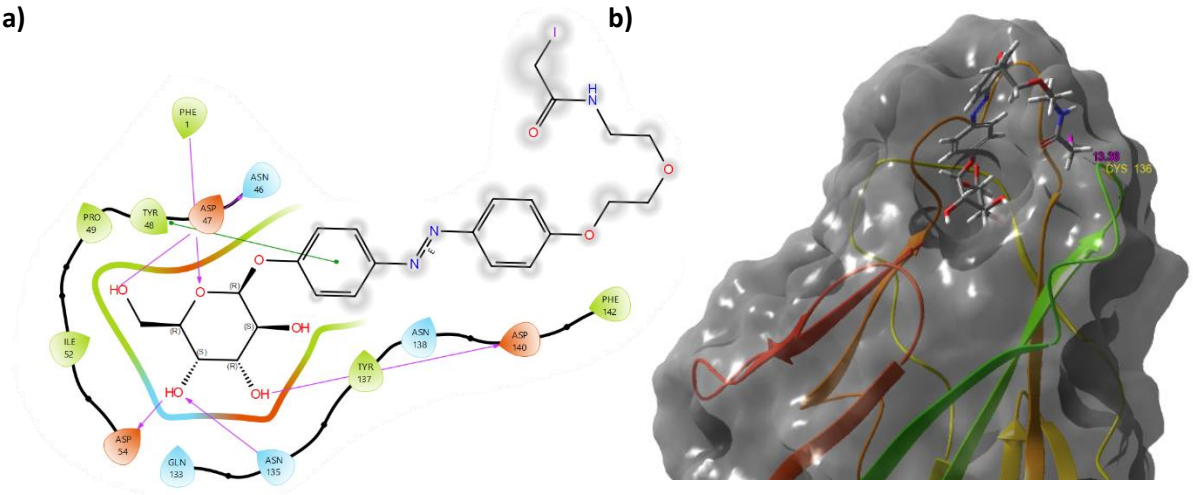


Figure 5.6-9: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

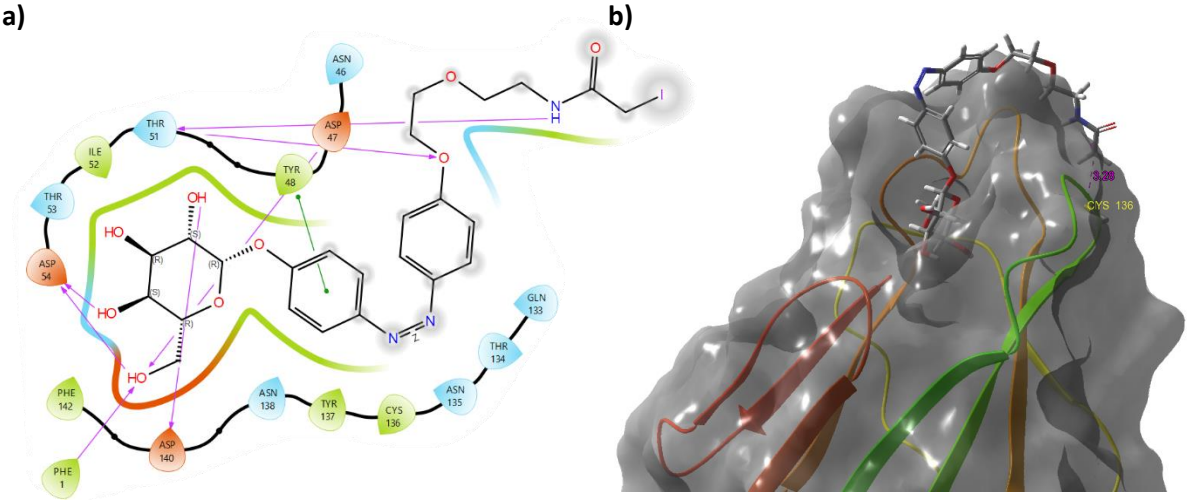


Figure 5.6-10: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

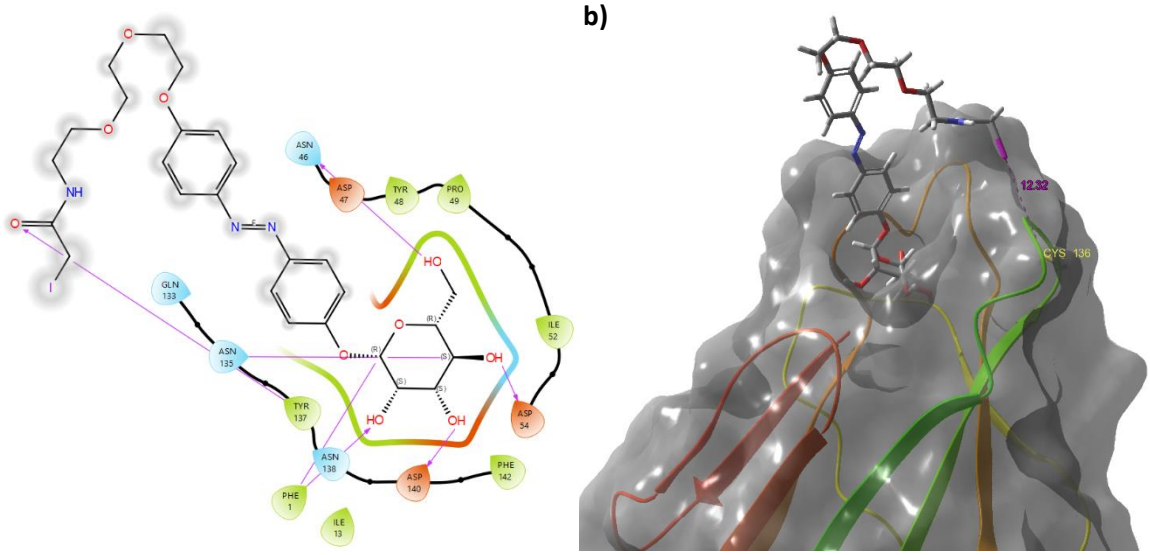


Figure 5.6-11: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H2** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

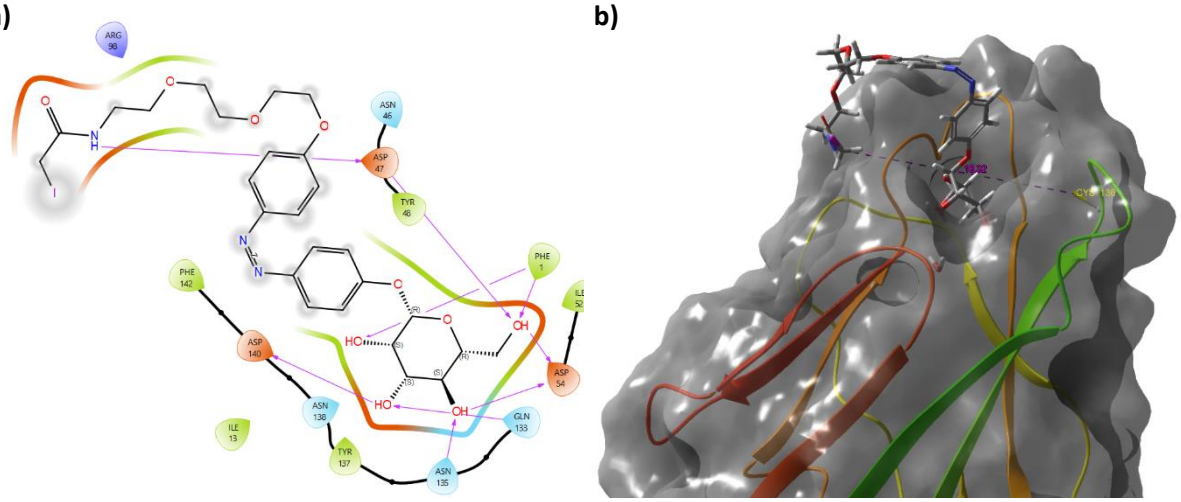


Figure 5.6-12: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H2** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

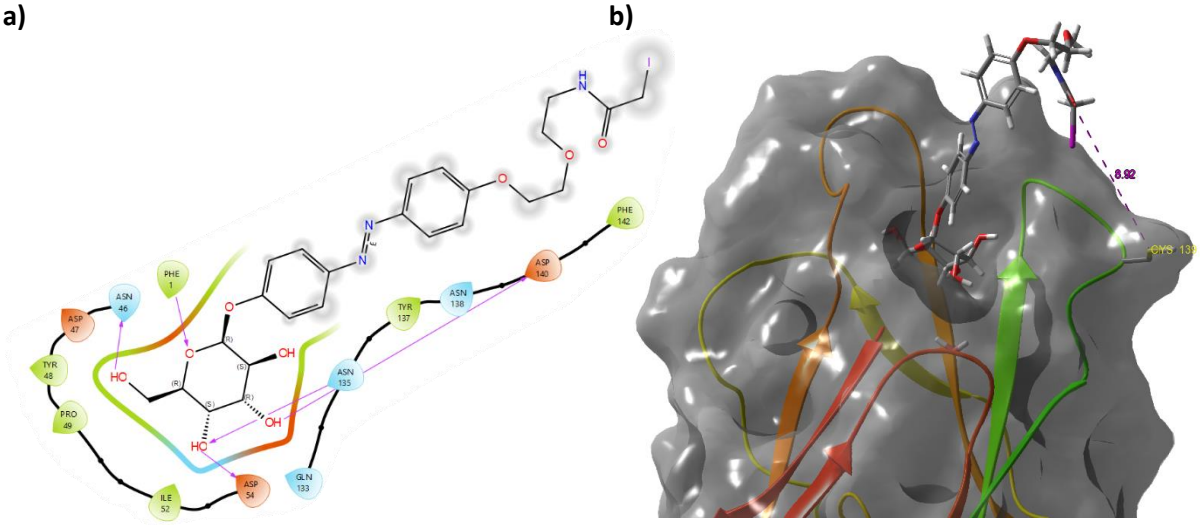


Figure 5.6-13: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**H1** in the S139C variant of 1KLF (open-gate conformation).

a) **b)**

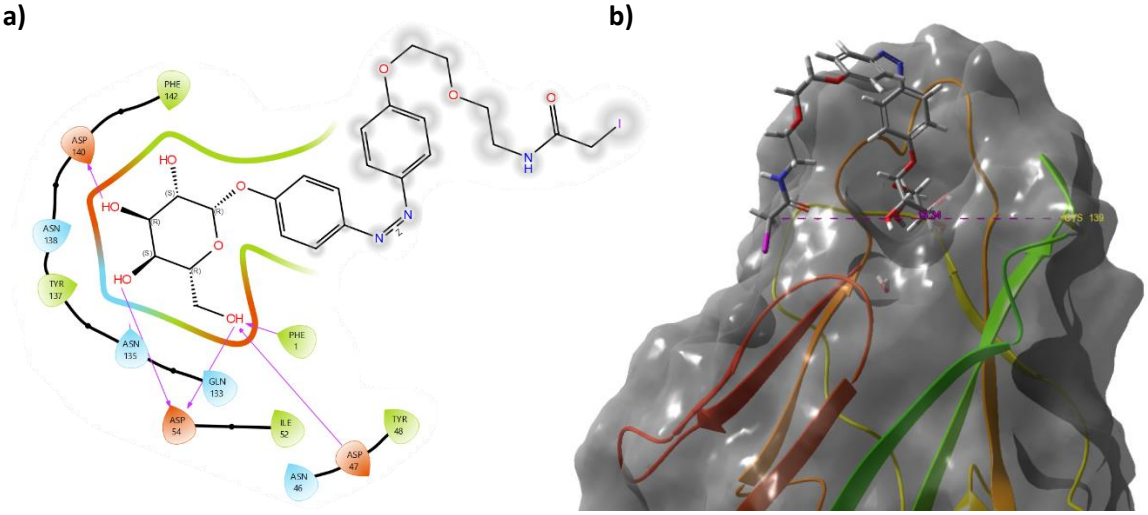


Figure 5.6-14: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H1** in the S139C variant of 1KLF (open-gate conformation).

a) **b)**

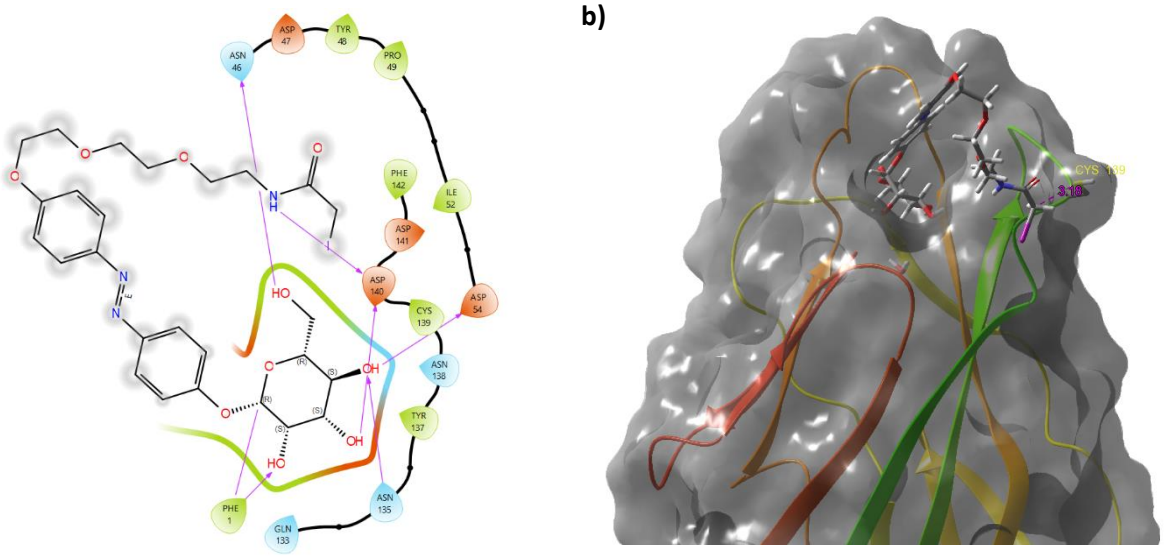


Figure 5.6-15: a) Ligand interaction diagram and **b)** and illustration resulting from the docking simulation of ligand *trans*-**H2** in the S139C variant of 1KLF (open-gate conformation).

a) **b)**

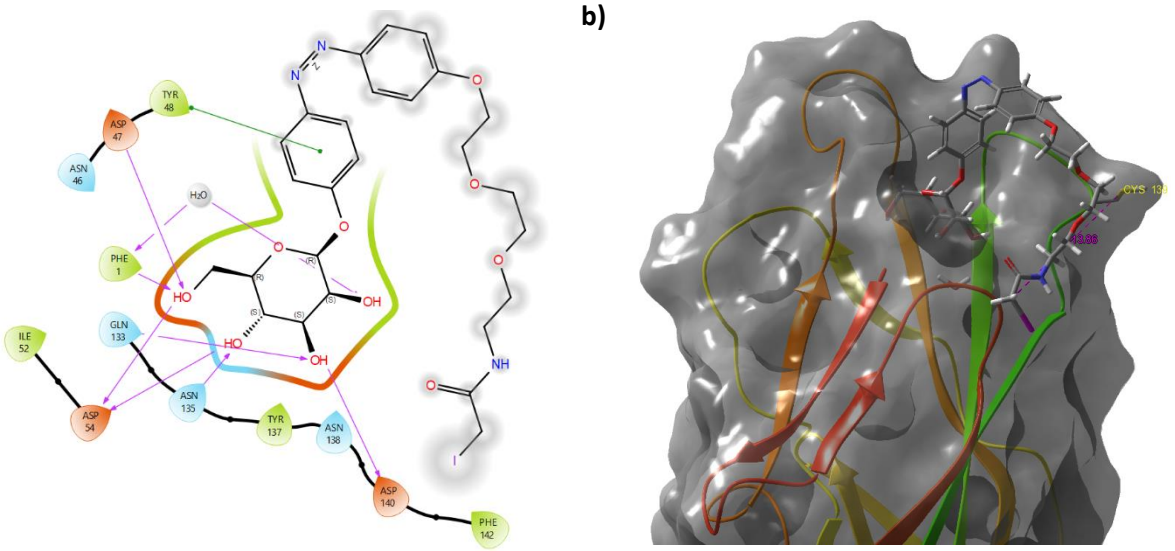


Figure 5.6-16: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**H2** in the S139C variant of 1KLF (open-gate conformation).

Table 5.6-7: Selected scoring values and S – C distances from docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the N136C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**F1**	-6.728	-1.383	-25.087	-38.141	-86.499	-63.228	11.670	14.55[a]
trans-**F1**	-5.954	-0.902	-20.431	-39.864	-81.900	-60.294	8.300	17.30
trans-**F1**	-5.341	-0.691	-20.454	-24.575	-60.986	-45.030	14.294	14.94
trans-**F1**	-5.251	-0.731	-22.603	-22.851	-61.166	-45.454	14.011	14.61
cis-**F1**	-6.125	-0.890	-26.496	-33.491	-80.728	-59.987	12.017	13.78
cis-**F1**	-6.104	-0.787	-20.365	-31.056	-72.937	-51.420	10.280	13.03
cis-**F1**	-6.018	-0.818	-24.530	-32.356	-78.708	-56.886	8.919	13.10
cis-**F1**	-5.955	-0.907	-24.475	-33.751	-79.642	-58.226	9.179	10.65
cis-**F1**	-5.952	-0.717	-21.929	-29.823	-71.361	-51.752	12.764	13.61
cis-**F1**	-5.830	-0.521	-24.716	-33.304	-78.914	-58.020	9.132	17.09
cis-**F1**	-5.793	-0.475	-24.248	-24.383	-70.074	-48.631	7.708	14.34
cis-**F1**	-5.741	-0.495	-21.622	-28.727	-71.357	-50.349	8.128	14.57
cis-**F1**	-5.685	-0.433	-24.810	-32.434	-77.447	-57.244	9.251	15.52
cis-**F1**	-5.554	-0.554	-19.463	-26.688	-66.924	-46.151	6.964	9.49[b]
trans-**F2**	-6.719	-1.147	-27.337	-36.027	-86.973	-63.364	11.718	15.47
trans-**F2**	-6.225	-0.961	-21.292	-38.974	-81.110	-60.266	12.912	10.64[c]
trans-**F2**	-5.990	-0.984	-19.579	-41.482	-78.728	-61.061	17.208	15.31
trans-**F2**	-5.383	-0.821	-15.136	-32.729	-64.344	-47.865	14.252	17.74
trans-**F2**	-5.259	-0.764	-17.141	-27.885	-56.449	-45.026	23.277	15.57
trans-**F2**	-4.789	-0.058	-15.119	-34.456	-65.551	-49.576	10.057	14.26
trans-**F2**	-4.030	-0.008	-10.352	-29.384	-49.519	-39.736	15.777	15.79
cis-**F2**	-6.133	-0.332	-30.430	-39.905	-91.800	-70.335	17.673	18.85
cis-**F2**	-6.127	-0.428	-24.627	-40.092	-87.279	-64.719	15.422	19.18
cis-**F2**	-5.676	-0.590	-22.165	-34.274	-77.704	-56.440	13.552	16.77
cis-**F2**	-5.487	-0.192	-21.537	-41.699	-80.065	-63.236	20.561	18.31
cis-**F2**	-5.348	-0.227	-20.504	-41.756	-80.698	-62.260	15.961	17.90
cis-**F2**	-5.244	-0.431	-18.367	-26.868	-65.876	-45.235	10.528	13.16[d]
cis-**F2**	-5.232	-0.404	-16.467	-38.516	-74.048	-54.983	13.568	16.96
cis-**F2**	-5.222	-0.411	-23.022	-25.784	-67.309	-48.806	14.590	14.79
cis-**F2**	-5.171	-0.243	-16.819	-41.063	-76.420	-57.882	14.021	17.64
cis-**F2**	-5.090	-0.304	-19.472	-26.224	-66.900	-45.696	7.886	15.50

a) Ligand interaction diagram and illustration shown in Figure 5.6-17. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-18. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-19. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-20.

Table 5.6-8: Selected scoring values and S – C distances from docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the S139C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**F1**	-6.211	-1.102	-24.553	-36.697	-83.321	-61.251	10.447	19.38
trans-**F1**	-6.176	-1.044	-22.054	-39.534	-83.809	-61.587	9.835	25.42
trans-**F1**	-5.984	-0.992	-23.207	-36.684	-80.802	-59.891	10.771	19.51
trans-**F1**	-5.924	-0.765	-22.474	-33.915	-74.666	-56.389	15.507	14.54
trans-**F1**	-5.811	-0.825	-23.483	-30.983	-73.688	-54.465	12.625	22.53
trans-**F1**	-5.494	-0.656	-19.916	-25.929	-61.666	-45.846	16.643	21.54
trans-**F1**	-5.345	-0.661	-22.449	-23.936	-61.666	-46.386	16.411	21.56
trans-**F1**	-4.508	-0.108	-15.831	-24.943	-55.879	-40.774	9.410	5.97[a,e]
cis-**F1**	-6.310	-0.474	-22.923	-31.651	-78.816	-54.575	10.034	4.01[b,e]
cis-**F1**	-6.057	-0.800	-19.441	-31.031	-73.898	-50.472	9.324	6.02
cis-**F1**	-5.982	-0.916	-24.941	-33.370	-81.237	-58.311	9.626	6.55
cis-**F1**	-5.944	-0.571	-25.928	-32.925	-81.981	-58.854	8.872	8.66
cis-**F1**	-5.895	-0.896	-14.973	-29.560	-67.412	-44.533	8.915	7.44
cis-**F1**	-5.563	-0.343	-22.801	-30.247	-74.528	-53.048	8.633	10.41
cis-**F1**	-5.562	-0.442	-20.720	-30.868	-74.002	-51.588	6.755	3.24[e]
cis-**F1**	-5.485	-0.561	-18.662	-30.962	-70.578	-49.623	8.961	10.99
cis-**F1**	-5.461	-0.868	-22.431	-22.683	-64.655	-45.114	11.568	8.03
cis-**F1**	-5.460	-0.893	-16.505	-25.796	-64.286	-42.301	6.666	18.07
trans-**F2**	-6.353	-0.918	-23.636	-39.016	-85.065	-62.652	10.847	19.39
trans-**F2**	-6.254	-0.951	-21.884	-39.914	-82.397	-61.797	13.606	20.96
trans-**F2**	-6.063	-1.001	-21.280	-39.893	-79.458	-61.173	16.568	22.60
trans-**F2**	-5.893	-0.815	-14.920	-33.708	-66.507	-48.628	15.889	20.12
trans-**F2**	-4.709	-0.002	-14.987	-33.992	-64.713	-48.979	9.806	3.46[c,e]
trans-**F2**	-4.251	-0.000	-10.794	-29.674	-51.432	-40.469	15.328	3.64[e]
cis-**F2**	-5.696	-0.611	-22.775	-33.187	-76.607	-55.963	12.018	20.35
cis-**F2**	-5.324	-0.275	-19.397	-41.890	-77.701	-61.287	16.995	20.89
cis-**F2**	-5.300	-0.198	-19.396	-42.583	-76.434	-61.979	20.692	21.25
cis-**F2**	-5.280	-0.207	-20.045	-41.817	-80.850	-61.862	11.440	21.64
cis-**F2**	-5.269	-0.429	-19.868	-23.960	-61.339	-43.828	14.283	4.19[d,e,f]
cis-**F2**	-5.173	-0.458	-22.323	-25.307	-64.803	-47.630	14.061	21.62
cis-**F2**	-5.109	-0.600	-21.685	-20.908	-56.910	-42.593	19.178	4.69[e]
cis-**F2**	-4.875	-0.362	-14.533	-30.018	-62.258	-44.551	10.204	21.27
cis-**F2**	-4.844	-0.250	-20.160	-24.515	-62.824	-44.675	9.038	20.75
cis-**F2**	-4.830	-0.187	-15.339	-26.778	-61.681	-42.118	6.073	10.83

a) Ligand interaction diagram and illustration shown in Figure 5.6-21. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-22. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-23. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-24. **e)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-11). **f)** The mannoside moiety appeared slightly misangled in the binding site.

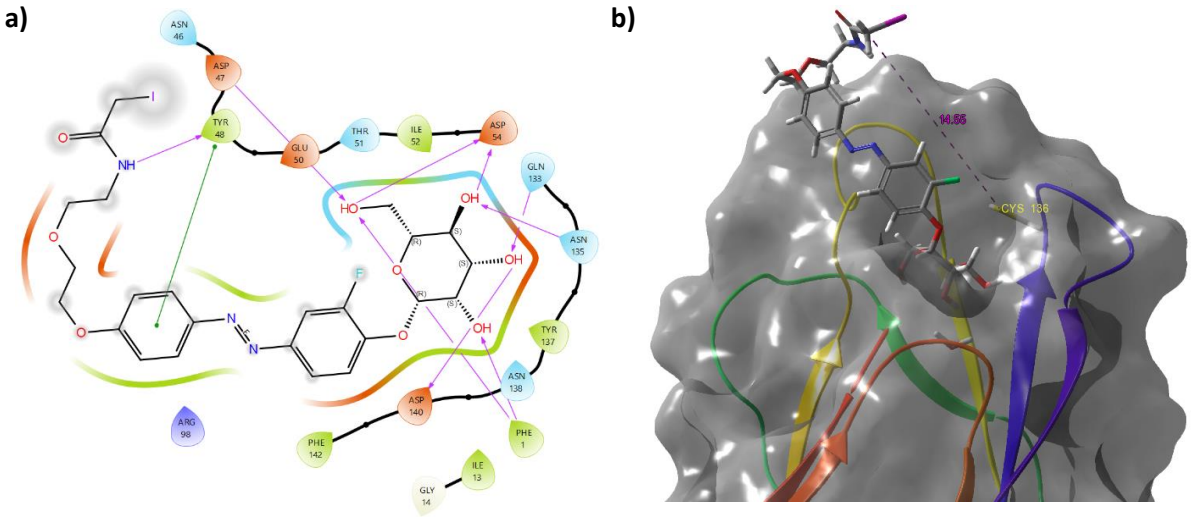


Figure 5.6-17: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F1** in the N136C variant of 1UWF (closed-gate conformation).

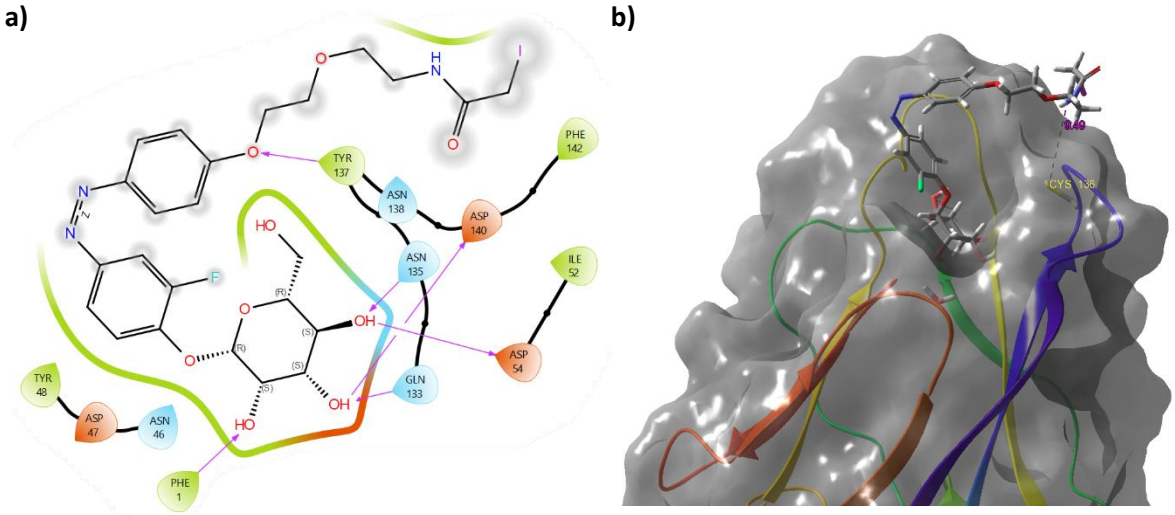


Figure 5.6-18: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F1** in the N136C variant of 1UWF (closed-gate conformation).

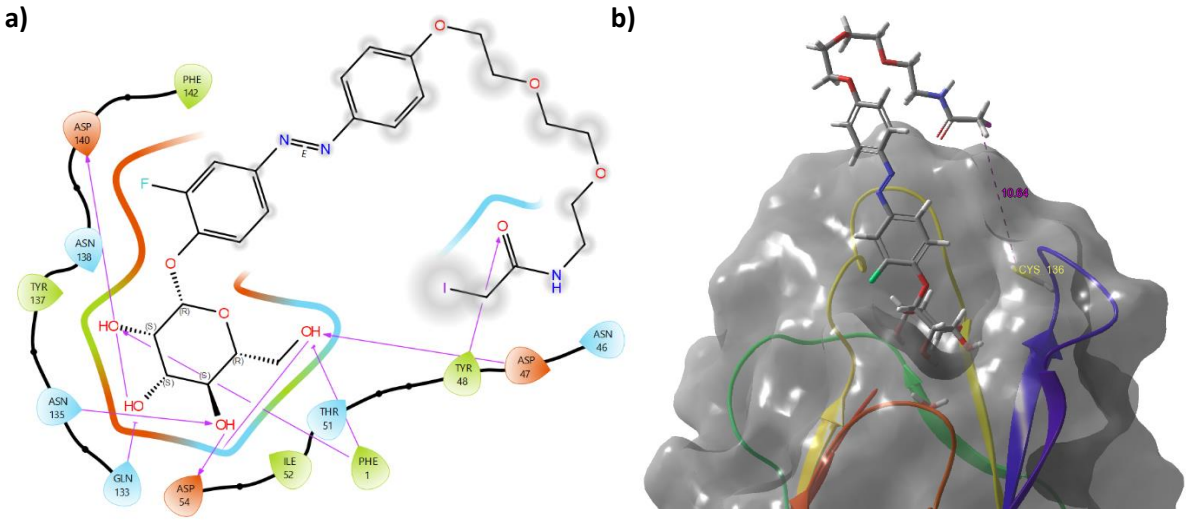


Figure 5.6-19: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

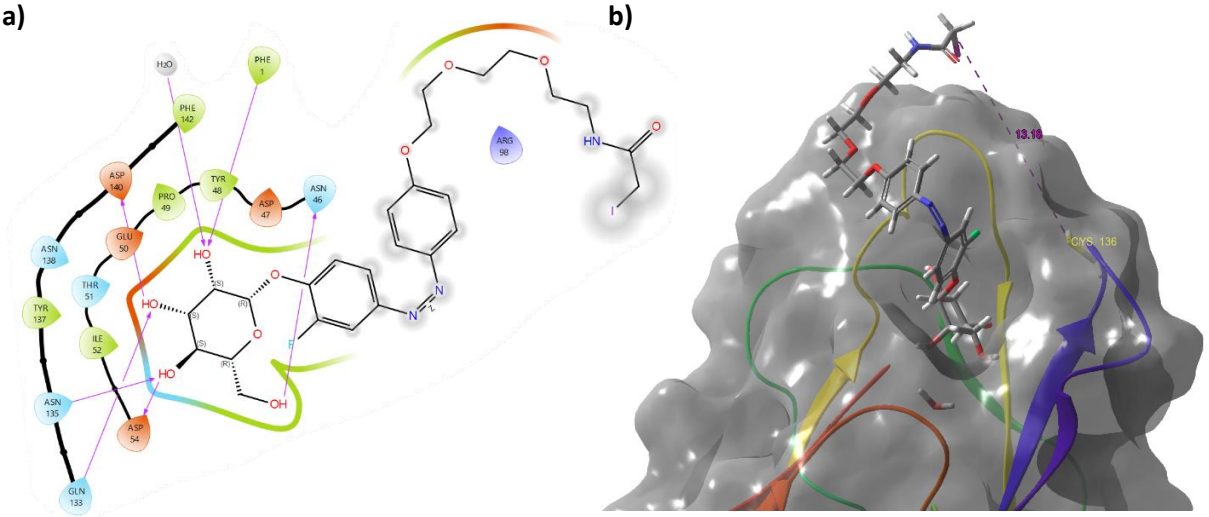


Figure 5.6-20: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

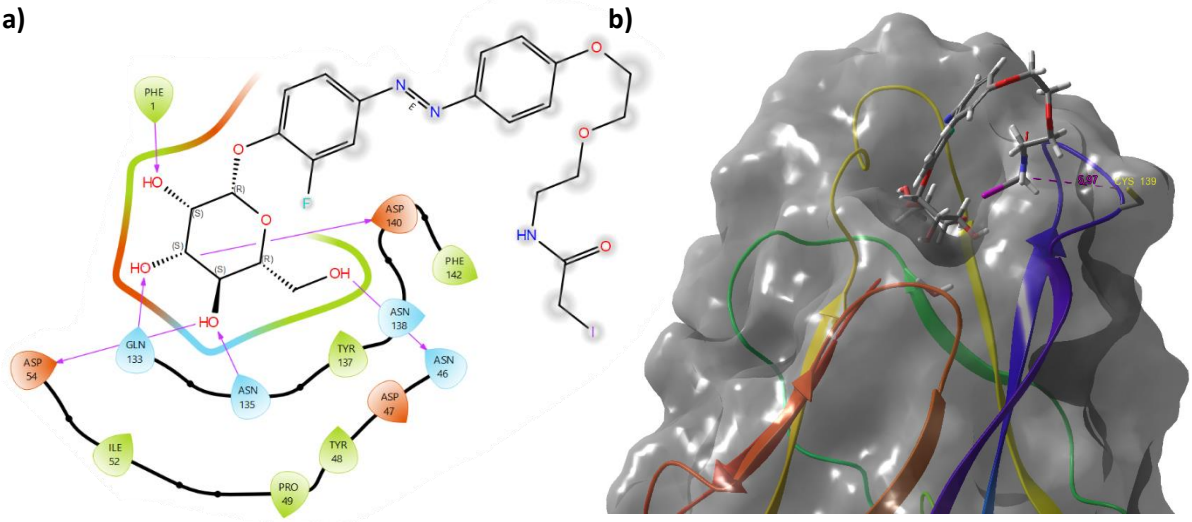


Figure 5.6-21: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F1** in the S139C variant of 1UWF (closed-gate conformation).

a) **b)**

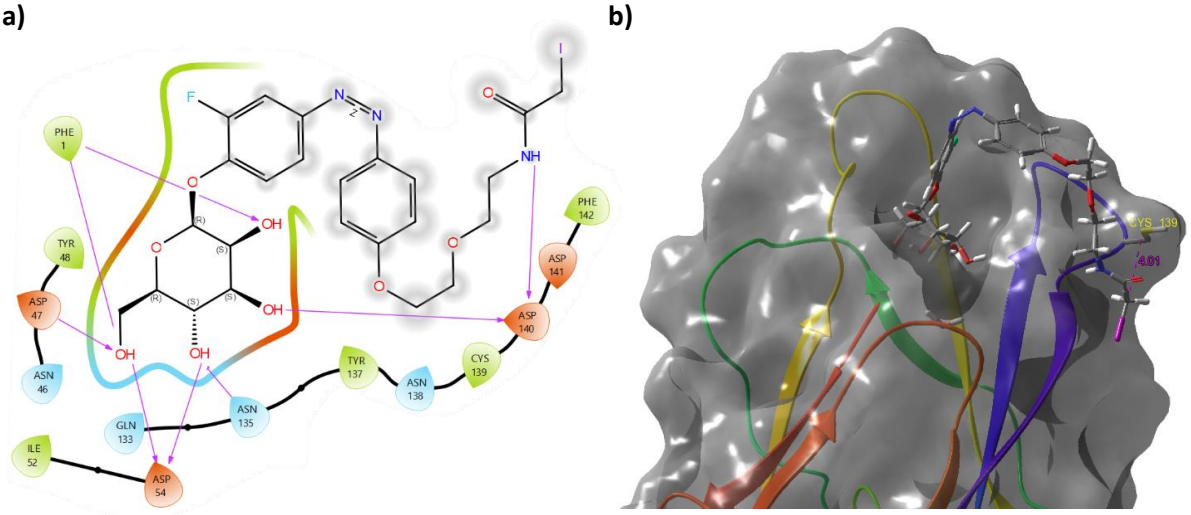


Figure 5.6-22: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F1** in the S139C variant of 1UWF (closed-gate conformation).

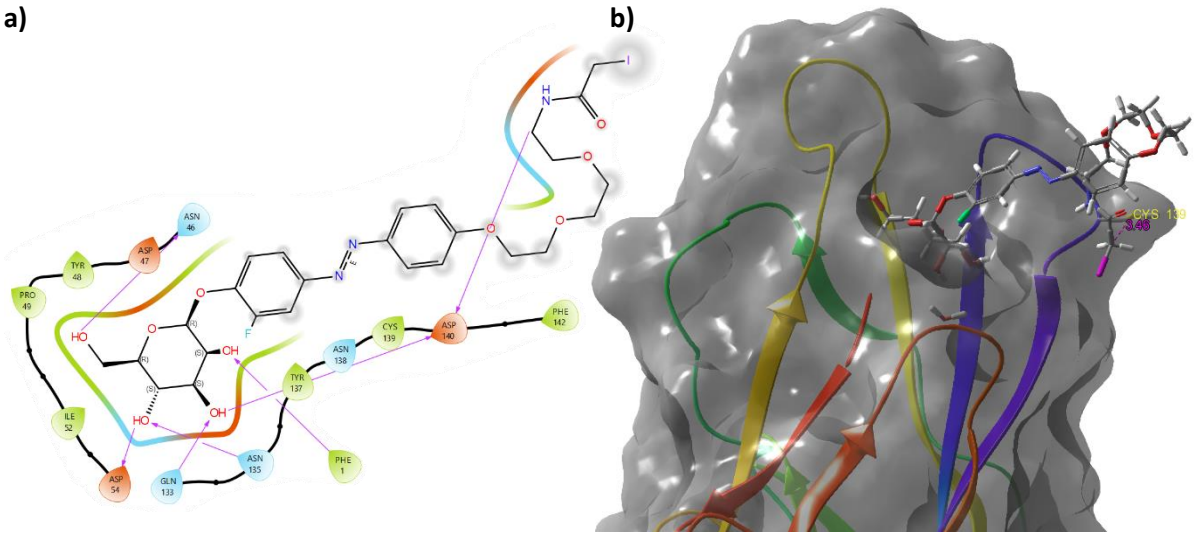


Figure 5.6-23: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F2** in the S139C variant of 1UWF (closed-gate conformation).

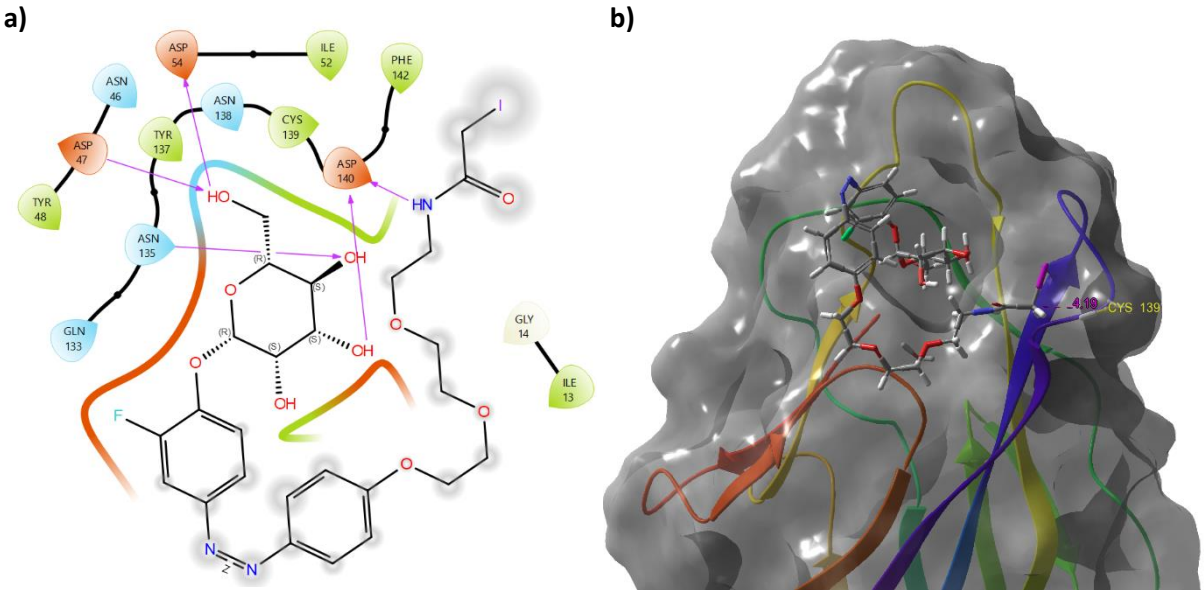


Figure 5.6-24: a) Ligand interaction diagram and **b)** illustrations resulting from the docking simulation of ligand *cis*-**F2** in the S139C variant of 1UWF (closed-gate conformation).

Table 5.6-9: Selected scoring values and S – C distances from docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the N136C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**F1**	-5.104	-0.655	-13.979	-26.711	-53.538	-40.690	7.478	16.60
trans-**F1**	-5.008	-0.721	-17.147	-25.682	-54.221	-42.829	9.768	13.30[a,d]
trans-**F1**	-4.226	-0.079	-7.544	-33.502	-48.997	-41.046	11.568	19.26
trans-**F1**	-4.210	-0.109	-7.917	-34.531	-49.619	-42.448	13.021	18.09
trans-**F1**	-4.204	-0.080	-7.607	-33.426	-49.400	-41.033	10.590	24.24
trans-**F1**	-4.193	-0.083	-7.790	-33.535	-50.888	-41.326	8.133	23.37a
trans-**F1**	-4.042	-0.175	-2.853	-30.945	-42.986	-33.799	7.899	19.46
cis-**F1**	-5.867	-0.525	-20.778	-39.460	-80.308	-60.239	9.723	13.88
cis-**F1**	-5.716	-0.586	-19.939	-37.299	-76.600	-57.238	9.848	10.53
cis-**F1**	-5.606	-0.187	-24.923	-27.615	-70.400	-52.538	11.914	15.70
cis-**F1**	-5.604	-0.546	-17.238	-35.840	-72.392	-53.077	8.991	9.56
cis-**F1**	-5.561	-0.479	-15.264	-28.612	-62.631	-43.876	9.746	12.27
cis-**F1**	-5.486	-0.599	-17.007	-29.902	-67.176	-46.909	6.084	9.62[b,d]
cis-**F1**	-5.471	-0.674	-20.069	-32.834	-71.785	-52.903	8.732	11.43
cis-**F1**	-5.464	-0.186	-21.458	-29.766	-71.064	-51.224	6.757	9.95
cis-**F1**	-5.386	-0.351	-19.228	-36.989	-70.231	-56.217	17.741	10.40
cis-**F1**	-5.358	-0.620	-19.400	-33.565	-71.841	-52.966	7.779	13.38
trans-**F2**[c]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**F2**[c]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.

a) Ligand interaction diagram and illustration shown in Figure 5.6-25. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-26. **c)** No correct docking results and/or azobenzene structures observed. **d)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-12).

Table 5.6-10: Selected scoring values and S – C distances from docking simulations of fluoro-azobenzene derivatives **F1** and **F2** in the S139C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	S – C distance [Å]
trans-**F1**	-5.535	-0.591	-17.252	-33.588	-64.252	-50.841	11.939	19.91[a,e]
trans-**F1**	-5.531	-0.555	-17.929	-35.839	-66.891	-53.768	12.488	20.02
trans-**F1**	-5.287	-0.616	-13.922	-25.931	-54.372	-39.852	7.990	20.09
trans-**F1**	-5.071	-0.720	-17.770	-25.875	-56.380	-43.645	10.045	19.86
trans-**F1**	-5.061	-0.601	-13.897	-26.786	-54.349	-40.683	8.111	22.62
trans-**F1**	-5.023	-0.585	-13.600	-28.478	-55.980	-42.078	7.376	21.65
trans-**F1**	-4.958	-0.601	-18.062	-25.405	-56.241	-43.467	9.176	20.47
trans-**F1**	-4.950	-0.778	-21.885	-23.321	-57.557	-45.206	9.965	19.95
trans-**F1**	-4.744	-0.626	-21.214	-23.095	-56.892	-44.310	8.062	21.69
trans-**F1**	-4.697	-0.600	-20.290	-23.258	-55.575	-43.549	8.841	21.38
cis-**F1**	-6.169	-0.834	-21.945	-38.135	-81.686	-60.080	10.366	6.40
cis-**F1**	-5.841	-0.561	-20.451	-39.197	-79.805	-59.649	10.414	7.99
cis-**F1**	-5.839	-0.575	-22.679	-27.645	-73.045	-50.324	5.272	8.93
cis-**F1**	-5.709	-0.591	-19.409	-38.813	-77.092	-58.221	11.841	3.79[b,e]
cis-**F1**	-5.651	-0.603	-19.084	-34.408	-73.486	-53.492	9.091	6.40
cis-**F1**	-5.631	-0.215	-24.228	-28.165	-73.197	-52.393	7.299	5.98
cis-**F1**	-5.597	-0.582	-17.300	-23.587	-60.015	-40.887	10.353	14.59
cis-**F1**	-5.453	-0.672	-19.015	-33.033	-70.829	-52.047	9.794	3.35[c]
cis-**F1**	-5.388	-0.095	-22.414	-29.822	-71.577	-52.236	8.114	6.68
cis-**F1**	-5.351	-0.472	-17.720	-26.580	-63.888	-44.299	7.291	11.25
trans-**F2**	-5.294	-0.251	-21.632	-33.883	-68.012	-55.515	16.650	4.74[c,e]
trans-**F2**	-5.109	-0.397	-16.802	-34.385	-64.401	-51.188	13.759	14.10
trans-**F2**	-5.005	-0.830	-22.431	-23.107	-53.197	-45.538	16.278	17.05
trans-**F2**	-4.456	-0.161	-10.827	-30.102	-49.006	-40.929	11.978	6.88
cis-**F2**	-4.659	-0.506	-8.261	-24.346	-38.717	-32.607	0.000	12.14
cis-**F2**	-4.455	-0.772	4.786	-25.521	-26.576	-20.735	0.000	12.18
cis-**F2**	-4.190	-0.339	-3.814	-23.768	-33.076	-27.582	0.000	13.10
cis-**F2**	-4.175	-0.680	0.353	-22.883	-28.004	-22.530	0.000	11.86[d,e]
cis-**F2**	-4.171	-0.265	-4.545	-23.924	-34.851	-28.469	0.000	16.03
cis-**F2**	-3.969	-0.426	-17.671	-17.029	-40.459	-34.699	0.000	8.06
cis-**F2**	-3.555	-0.427	0.965	-20.447	-24.641	-19.482	0.000	7.46

a) Ligand interaction diagram and illustration shown in Figure 5.6-27. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-28. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-29. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-33. **e)** The respective lectin structures were manually alkylated with the ligand and energy minimised (see Table 5.6-12).

a) **b)**

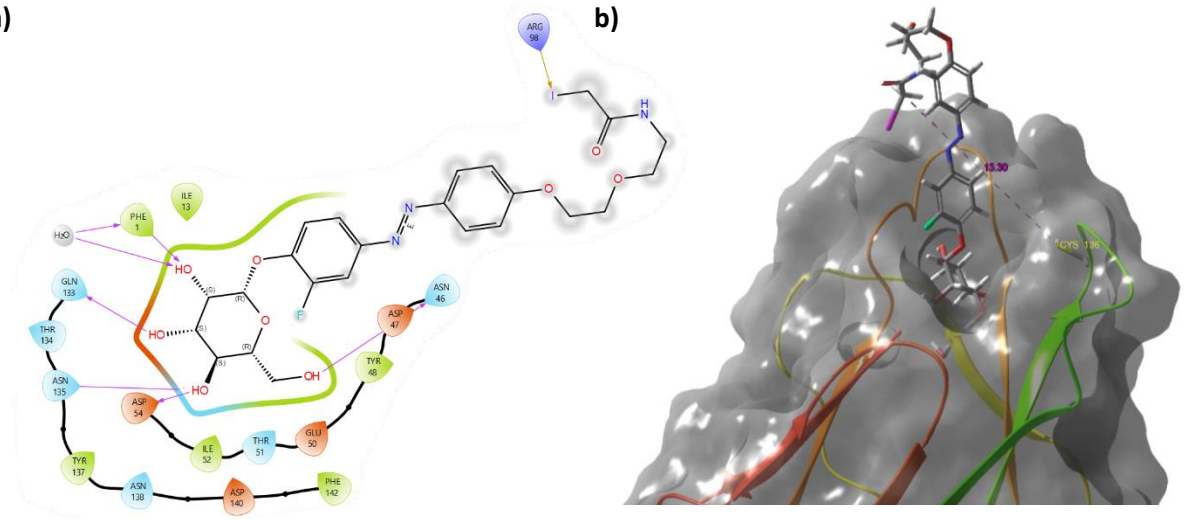


Figure 5.6-25: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

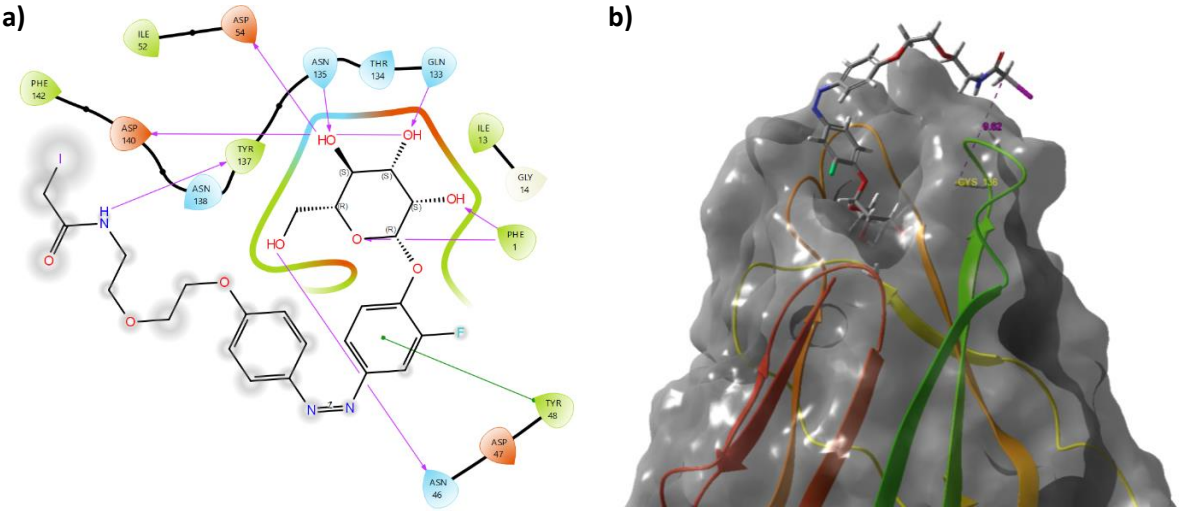


Figure 5.6-26: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

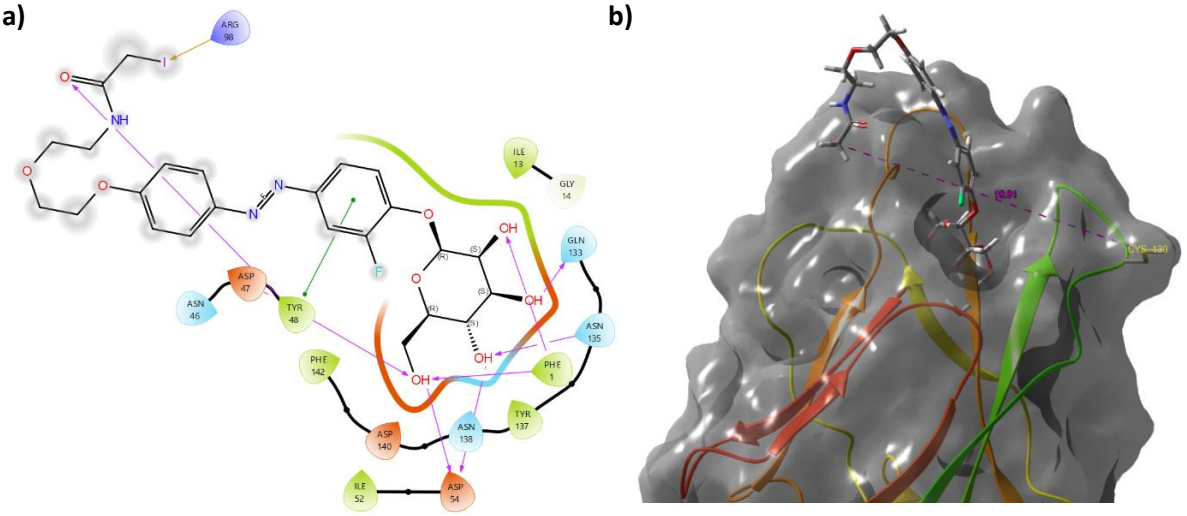


Figure 5.6-27: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F1** in the S139C variant of 1KLF (open-gate conformation).

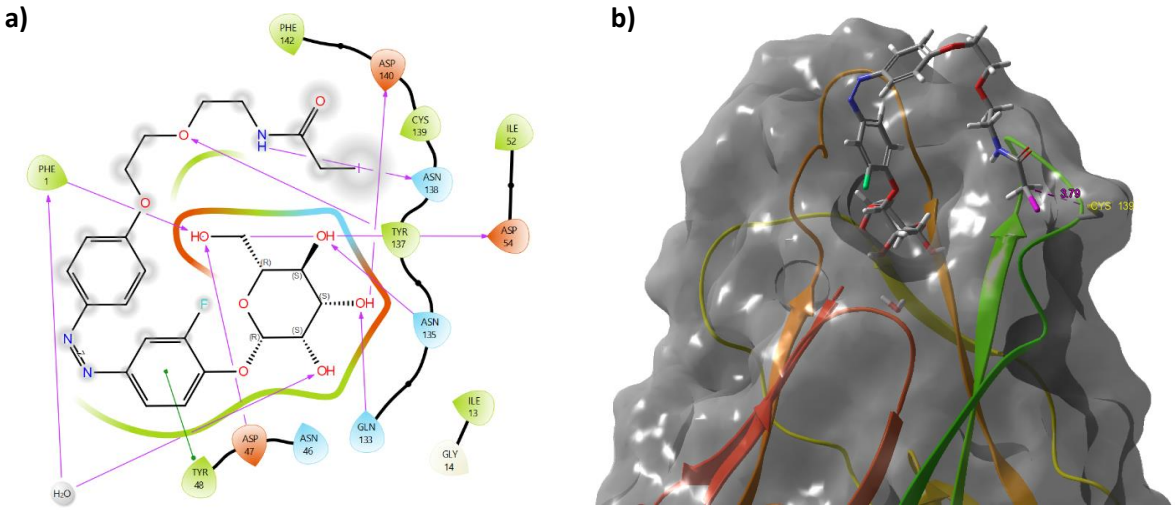


Figure 5.6-28: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F1** in the S139C variant of 1KLF (open-gate conformation).

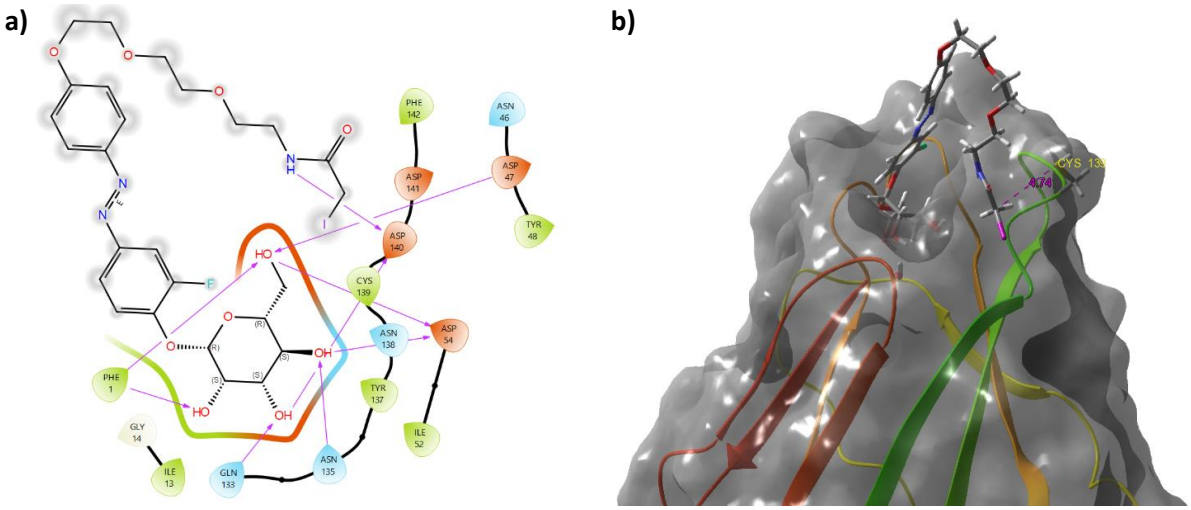


Figure 5.6-29: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *trans*-**F2** in the S139C variant of 1KLF (open-gate conformation).

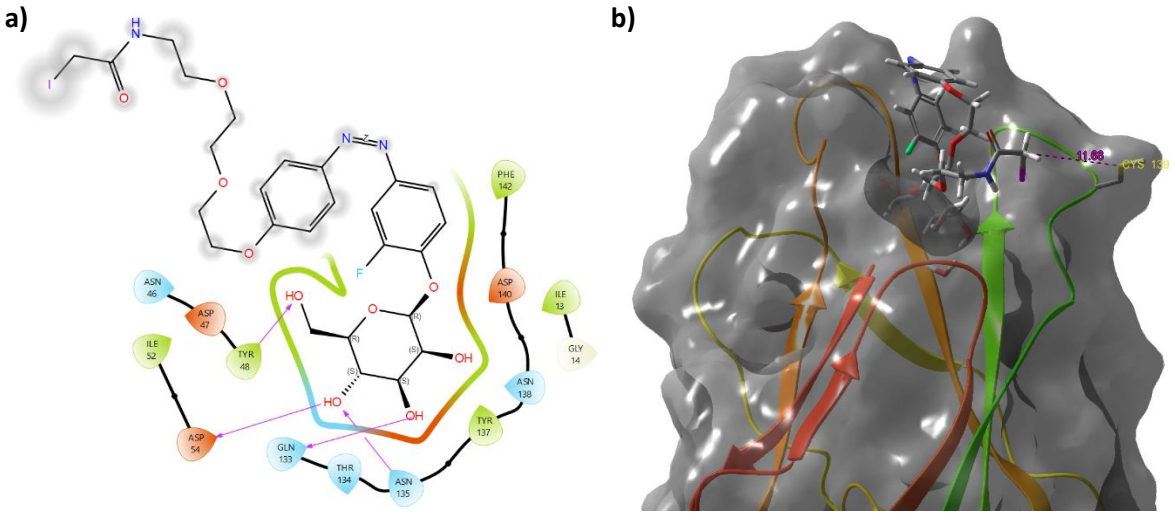


Figure 5.6-30: a) Ligand interaction diagram and **b)** illustration resulting from the docking simulation of ligand *cis*-**F2** in the S139C variant of 1KLF (open-gate conformation).

Table 5.6-11: Scoring values from docking simulations of covalently linked azobenzene derivatives with the S139C variant of the 1UWF crystal structure (closed-gate conformation) after energy-minimisation.

Lectin structure	Alkylated ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	Binding mode
S139C	*trans*-**H1**	-5.857	-0.834	-13.959	-36.478	-69.994	-50.437	11.258	unbound
S139C	*cis*-**H1**	-6.317	-0.577	-25.830	-29.951	-79.339	-55.780	11.927	bound
S139C	*trans*-**H2**	-5.013	-0.141	-19.305	-28.222	-60.869	-47.528	17.062	bound
S139C	*cis*-**H2**	-5.994	-0.888	-24.035	-33.727	-79.035	-57.762	13.853	bound
S139C	*trans*-**F1**	-3.184	-0.464	-30.808	-9.580	-51.358	-40.388	6.045	unbound
S139C	*cis*-**F1**	-6.310	-0.474	-22.923	-31.651	-78.816	-54.575	10.034	bound
S139C	*cis*-**F1**	-5.562	-0.442	-20.720	-30.868	-74.002	-51.588	6.755	bound
S139C	*trans*-**F2**	-4.709	-0.002	-14.987	-33.992	-64.713	-48.979	9.806	bound
S139C	*trans*-**F2**	-4.251	-0.000	-10.794	-29.674	-51.432	-40.469	15.328	bound
S139C	*cis*-**F2**	-5.269	-0.429	-19.868	-23.960	-61.339	-43.828	14.283	bound
S139C	*cis*-**F2**	-5.109	-0.600	-21.685	-20.908	-56.910	-42.593	19.178	bound

The ligand interaction diagrams and illustrations of the energy-minimisation results highlighted in grey are depicted in the following Figures 5.5-31 – 5.5-36.

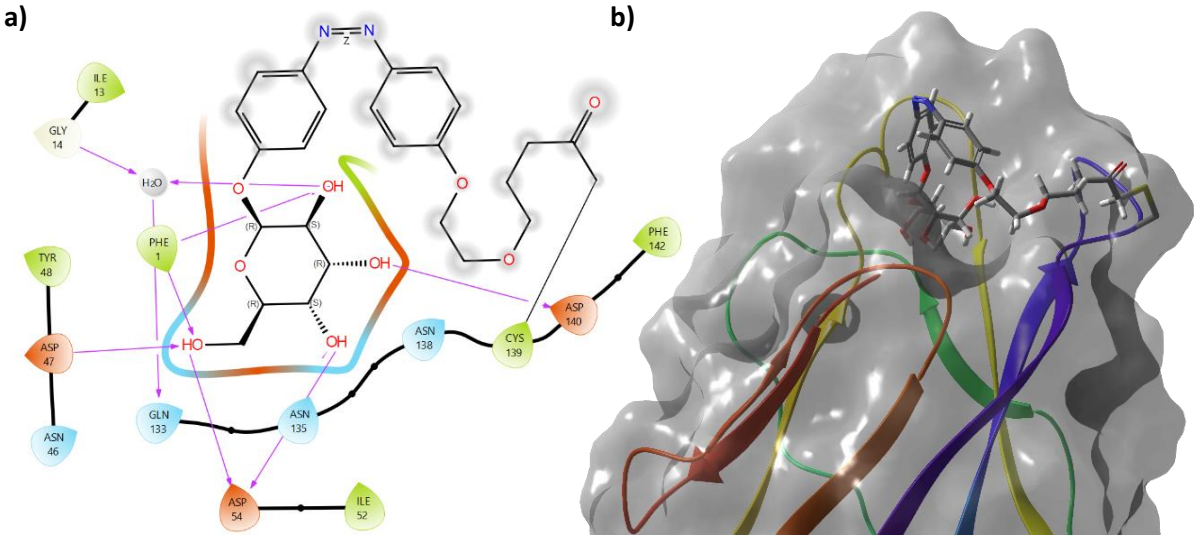


Figure 5.6-31: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**H1** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

a) **b)**

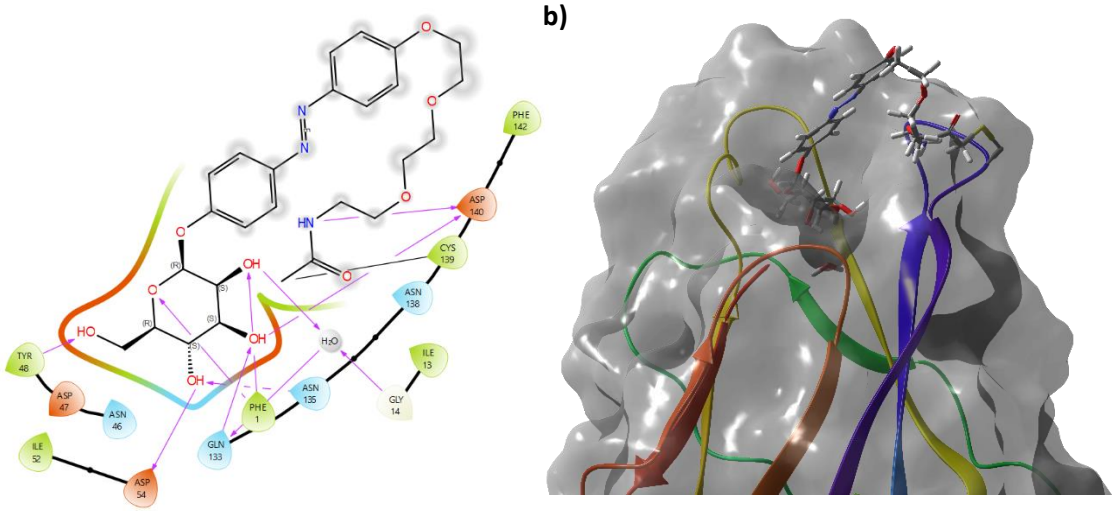


Figure 5.6-32: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *trans*-**H2** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

a) **b)**

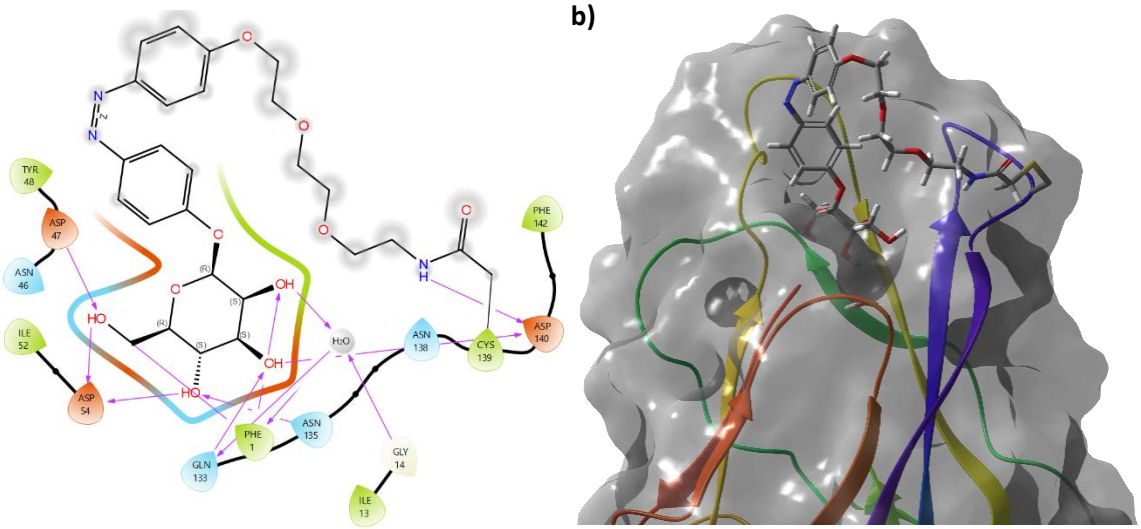


Figure 5.6-33: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**H2** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

a) **b)**

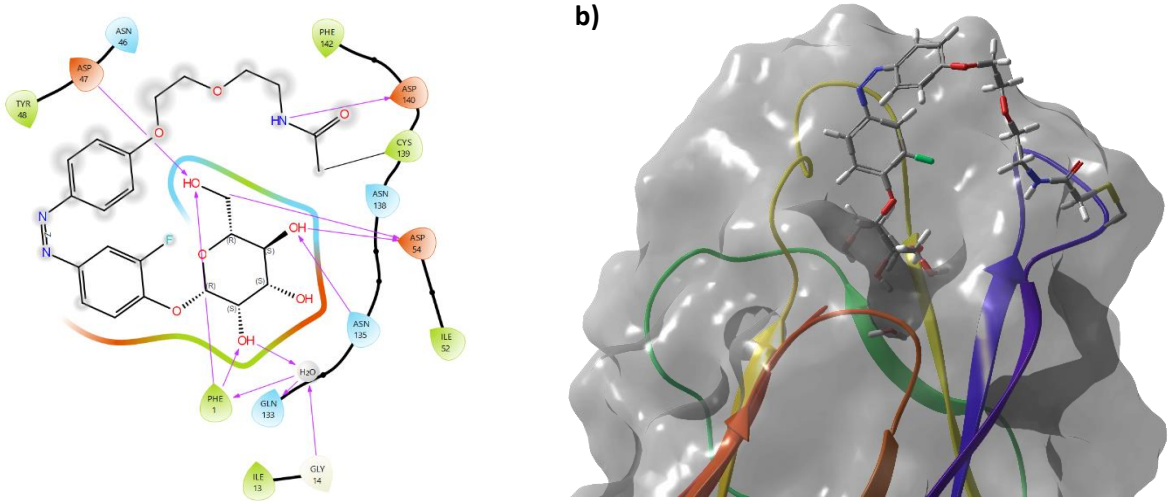


Figure 5.6-34: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**F1** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

a) **b)**

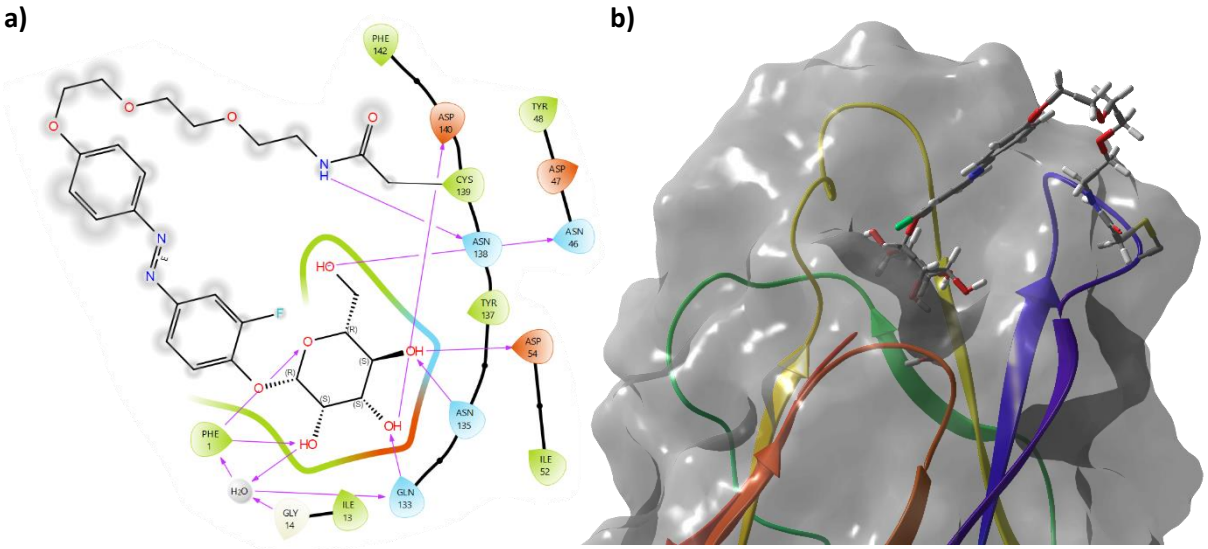


Figure 5.6-35:) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *trans*-**F2** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

a) **b)**

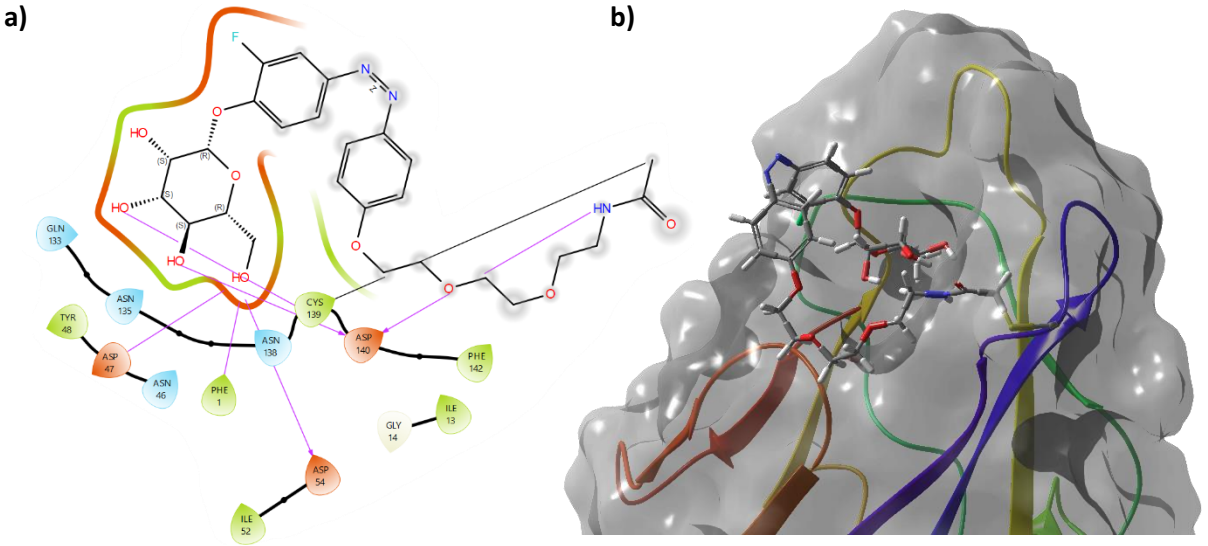


Figure 5.6-36:) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**F2** covalently alkylated to Cys139 in the 1UWF structure (closed-gate conformation).

Table 5.6-12: Scoring values from docking simulations of covalently linked azobenzene derivatives with cysteine variants of the 1KLF crystal structure (open-gate conformation) after energy-minimisation.

Lectin structure	Alkylated ligand	Docking score	Glide lipo	Glide evdw	Glide ecoul	Glide emodel	Glide energy	Glide einternal	Binding mode
N136C	*trans*-**H1**	-2.540	-0.448	-25.873	-8.397	-35.431	-34.270	12.755	wrong
N136C	*cis*-**H1**	-7.009	-0.476	-31.790	-26.158	-80.505	-57.948	10.318	bound
S139C	*trans*-**H2**	-4.437	-0.042	-11.599	-36.588	-54.750	-48.186	14.161	bound
S139C	*cis*-**H2**	-4.966	-0.409	-18.652	-29.793	-67.237	-48.445	8.920	unbound
N136C	*trans*-**F1**	-3.113	-0.369	-19.513	-12.384	-37.088	-31.898	9.853	wrong
1N136C	*cis*-**F1**	-5.486	-0.599	-17.007	-29.902	-67.176	-46.909	6.084	wrong
S139C	*trans*-**F1**	-2.965	-0.546	-19.734	-10.942	-35.986	-30.676	10.145	wrong
S139C	*cis*-**F1**	-5.453	-0.672	-19.015	-33.033	-70.829	-52.047	9.794	bound
S139C	*trans*-**F2**	-5.294	-0.251	-21.632	-33.883	-68.012	-55.515	16.650	bound
S139C	*cis*-**F2**	-4.175	-0.680	0.353	-22.883	-28.004	-22.530	0.000	bound

The ligand interaction diagrams and illustrations of the energy-minimisation results highlighted in grey are depicted in the following Figures 5.5-37 – 5.5-41.

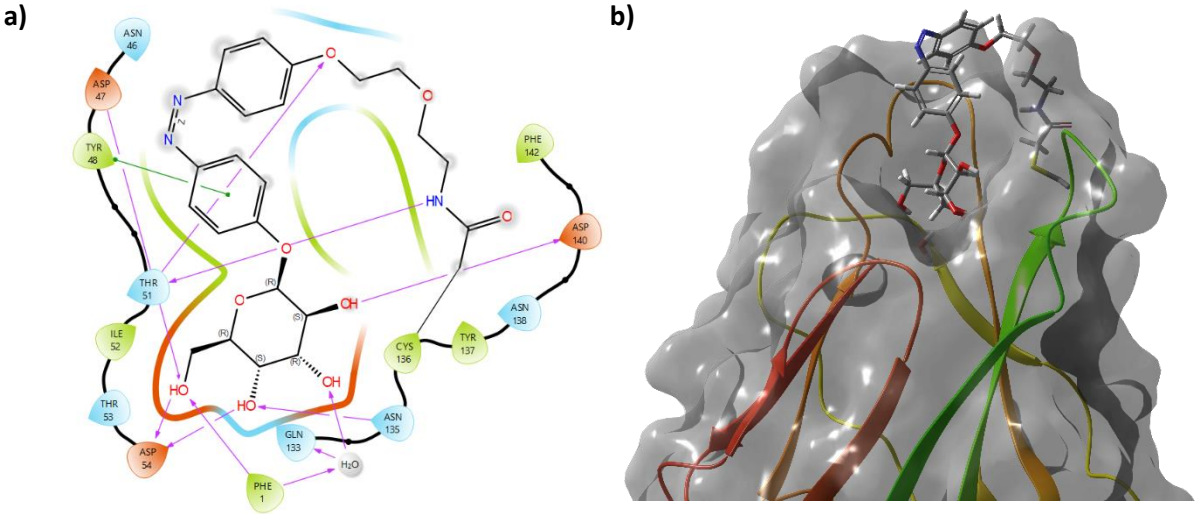


Figure 5.6-37: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**H1** covalently alkylated to Cys136 in the 1KLF structure (open-gate conformation).

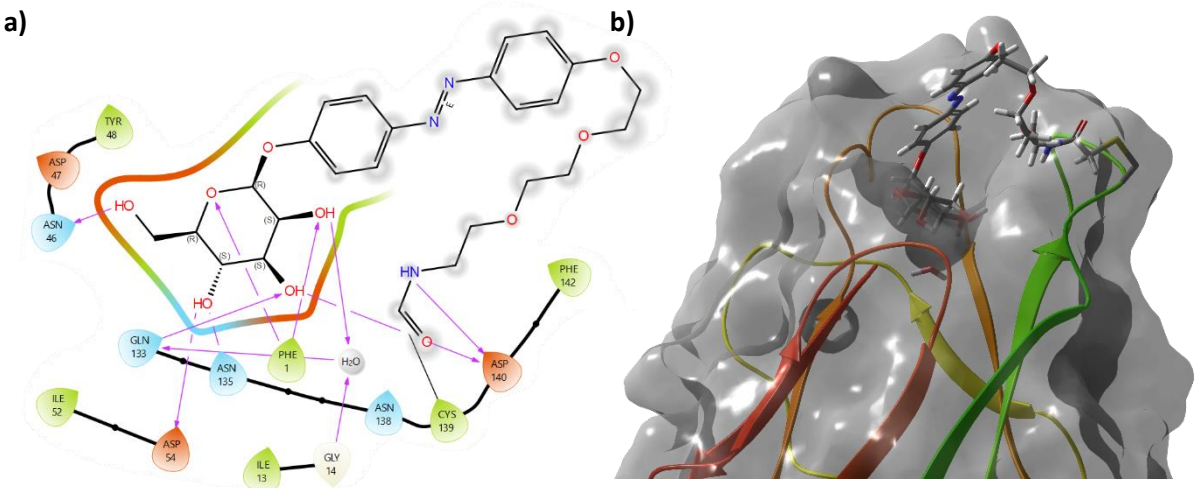


Figure 5.6-38: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *trans*-**H2** covalently alkylated to Cys139 in the 1KLF structure (open-gate conformation).

a) **b)**

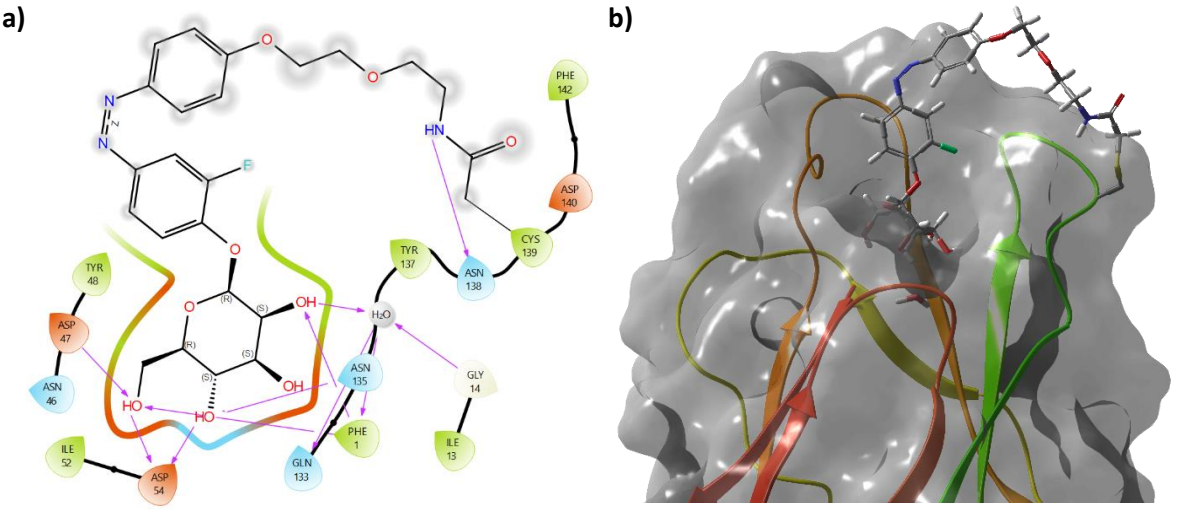


Figure 5.6-39: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**F1** covalently alkylated to Cys139 in the 1KLF structure (open-gate conformation).

a) **b)**

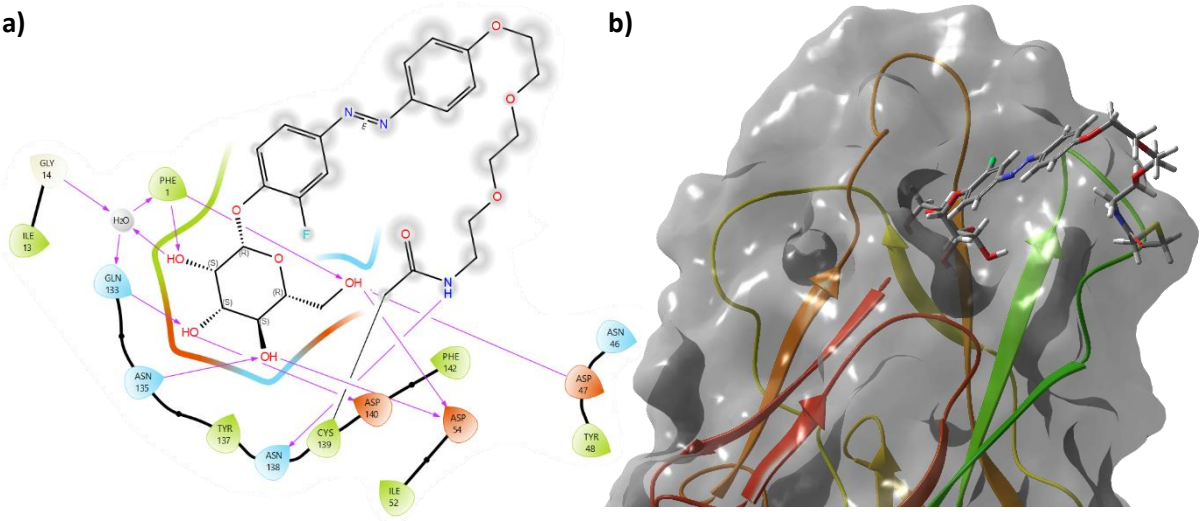


Figure 5.6-40: **a)** Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *trans*-**F2** covalently alkylated to Cys139 in the 1KLF structure (open-gate conformation).

a) **b)**

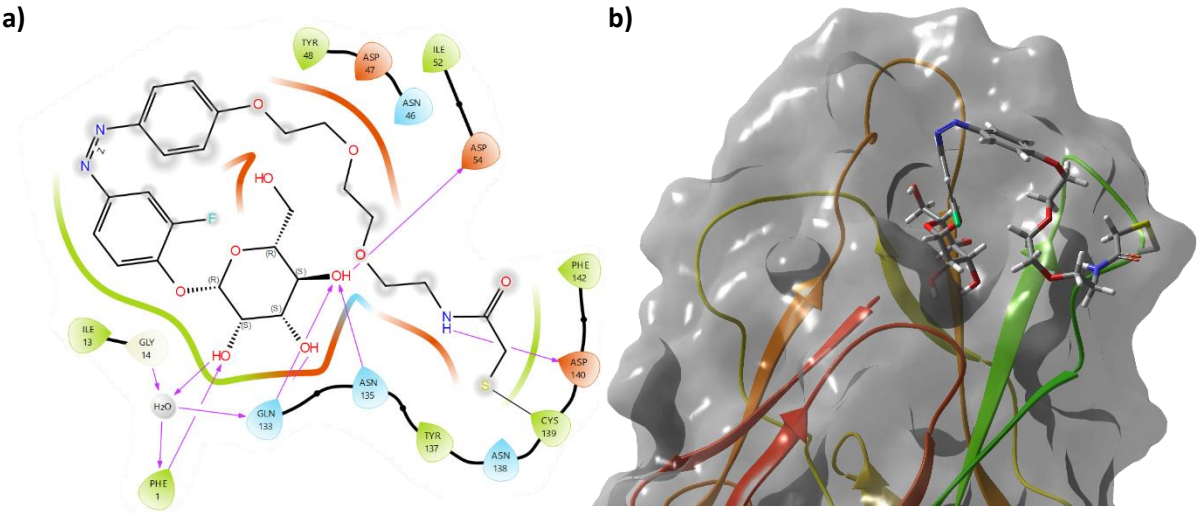


Figure 5.6-41: a) Ligand interaction diagram and **b)** illustration resulting from the energy-minimisation of *cis*-**F2** covalently alkylated to Cys139 in the 1KLF structure (open-gate conformation).

5.6.3 Covalent docking

The covalent docking protocol includes the mutation of predefined reactive amino acids to alanine with subsequent Glide docking in order to obtain suitable poses for the ensuing reaction. After restoring of the reactive side chain and formation of the covalent bond, a full energy-minimisation is calculated with Prime[437,685,686] and the outputs are ranked according to the respective Prime energies which are given in kcal · mol^{-1}. Since Prime energies are not well suited to compare different ligands, the docking score is provided as an expression for the approximate affinity of the respective ligand-receptor interaction.

Covalent docking simulations were performed using the CovDock[435] function in the Glide application[687] in a more recent OPLS4 force field[436] and in implicit water. The FimH crystal structures 1UWF (closed-gate conformation) and 1KLF (open-gate conformation) were prepared with the Protein Preparation Workflow as described in 5.6.1 and minimised with MacroModel.[688] An enclosing box with side lengths of 36 Å each was defined around the co-crystallised ligands within the respective crystal structure (chain A: W1159 in 1UWF; chain B: W1500 in 1KLF). The cysteine residues 136 or 139 were specified as reactive residues and nucleophilic substitution of an iodine atom, as part of the leaving group, was predetermined. The torsional constraints of *trans*- and *cis*-azobenzene were adopted as SMARTS patterns from energy-minimised structures with predefined geometry. Thorough pose prediction was performed with a cut-off of 2.5 kcal · mol^{-1} to retain poses for further refinement. A maximal output of 20 poses per ligand reaction site was calculated.

The resulting poses were evaluated with regard to correct docking of the *α*-D-mannoside moiety and geometry of *cis*- and *trans*-azobenzene isomers. If no correct covalent docking results were obtained, the negative results were confirmed by a further covalent docking simulation, in which the reference position of the ligand to be docked was restricted using the SMARTS pattern of the mannoside without aglycone. For this purpose, the respective co-crystallised ligand was selected as a reference with a deviation tolerance limit of 1 Å. Covalent docking outputs with the 10 highest docking scores are given in Tables 5.5-13 – 5.5-20. Corresponding ligand interaction diagrams and poses with the each highest Prime energies are depicted in Figures 5.5-42 – 5.5-69.

Table 5.6-13: Selected scoring values and Prime energies from covalent docking simulations of azobenzene derivatives **H1** and **H2** in the N136C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**H1**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**H1**	-7.170	-28.515	6.649	-73.787	-51.800	-23.284	-5435.94
cis-**H1**	-7.124	-29.636	13.752	-73.225	-54.661	-25.025	-5841.46[b]
cis-**H1**	-7.097	-28.862	11.857	-75.386	-52.806	-23.944	-5554.75
cis-**H1**	-6.996	-24.872	6.899	-74.524	-52.031	-27.159	-5445.87
cis-**H1**	-6.865	-28.381	6.936	-74.528	-53.568	-25.187	-5477.28
cis-**H1**	-6.663	-27.253	9.032	-72.775	-52.373	-25.120	-5424.59
cis-**H1**	-6.201	-27.329	7.355	-66.311	-51.593	-24.264	-5489.49
cis-**H1**	-6.007	-26.229	8.963	-75.057	-52.693	-26.464	-5832.28
trans-**H2**	-7.783	-31.192	7.649	-82.599	-55.839	-24.647	-5904.70[c]
trans-**H2**	-7.686	-27.759	6.881	-79.499	-52.875	-25.116	-5804.12
trans-**H2**	-7.624	-30.630	13.577	-76.628	-57.341	-26.711	-5845.46
trans-**H2**	-7.559	-30.836	10.504	-81.198	-55.586	-24.750	-5743.44
trans-**H2**	-7.480	-28.511	11.167	-81.023	-50.106	-21.595	-5577.14
trans-**H2**	-7.479	-27.901	6.875	-84.830	-56.976	-29.075	-5608.50
trans-**H2**	-7.308	-30.238	7.328	-75.886	-57.100	-26.862	-5734.10
trans-**H2**	-7.115	-28.508	9.428	-75.383	-53.121	-24.613	-5639.93
trans-**H2**	-7.024	-29.443	7.499	-77.738	-55.888	-26.445	-5654.92
trans-**H2**	-6.982	-29.897	11.304	-84.331	-53.367	-23.470	-5505.26
cis-**H2**	-8.347	-33.064	12.516	-89.729	-57.477	-24.414	-5889.46
cis-**H2**	-8.346	-33.683	12.698	-89.652	-60.340	-26.657	-5762.38
cis-**H2**	-8.313	-34.433	13.558	-90.782	-59.128	-24.696	-5779.26
cis-**H2**	-8.236	-28.687	13.668	-82.518	-56.197	-27.511	-5837.01
cis-**H2**	-8.233	-31.956	8.565	-86.807	-59.238	-27.281	-5436.92
cis-**H2**	-8.231	-33.413	10.881	-87.895	-59.852	-26.440	-5924.04[d]
cis-**H2**	-8.197	-29.629	7.164	-86.892	-58.830	-29.201	-5818.88
cis-**H2**	-8.125	-31.537	9.404	-89.179	-56.825	-25.288	-5868.17
cis-**H2**	-8.082	-32.849	8.198	-89.247	-57.166	-24.318	-5816.69
cis-**H2**	-7.790	-32.738	10.304	-81.838	-54.446	-21.709	-5502.18

a) No correct covalent docking results observed. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-42. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-43. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-44.

Table 5.6-14: Selected scoring values and Prime energies from covalent docking simulations of azobenzene derivatives **H1** and **H2** in the S139C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**H1**	-7.033	-27.670	6.411	-63.390	-48.825	-21.155	-5998.62[a]
trans-**H1**	-6.803	-27.667	7.333	-63.615	-48.284	-20.617	-5988.74
trans-**H1**	-6.728	-25.783	6.795	-74.582	-54.963	-29.180	-5981.30
trans-**H1**	-6.708	-28.688	7.601	-68.055	-48.976	-20.287	-5974.80
trans-**H1**	-6.627	-25.059	5.086	-75.598	-53.859	-28.800	-5997.61
trans-**H1**	-6.437	-26.784	6.146	-59.844	-47.398	-20.614	-5862.03
trans-**H1**	-6.225	-26.436	7.511	-62.842	-46.754	-20.318	-5962.81
trans-**H1**	-6.055	-25.472	6.390	-57.782	-45.031	-19.559	-5855.37
trans-**H1**	-5.875	-26.352	6.974	-58.517	-48.021	-21.670	-5923.04
trans-**H1**	-5.576	-26.961	9.581	-62.221	-52.466	-25.505	-5982.51
cis-**H1**	-7.450	-29.009	9.507	-76.430	-53.560	-24.551	-5978.99
cis-**H1**	-7.408	-27.763	10.176	-74.095	-51.843	-24.080	-5721.98
cis-**H1**	-7.398	-28.220	8.752	-74.881	-52.327	-24.106	-5980.99
cis-**H1**	-7.351	-28.969	9.948	-78.187	-52.655	-23.686	-5849.49
cis-**H1**	-7.333	-27.140	8.801	-71.333	-51.045	-23.905	-5830.63
cis-**H1**	-7.318	-25.964	0.000	-77.150	-52.039	-26.074	-5990.34
cis-**H1**	-7.297	-27.181	8.398	-79.462	-53.143	-25.962	-5832.10
cis-**H1**	-7.275	-29.578	7.682	-72.398	-52.195	-22.617	-5953.65
cis-**H1**	-7.171	-26.725	10.350	-84.269	-52.153	-25.428	-5881.15
cis-**H1**	-7.150	-28.904	5.213	-71.538	-50.444	-21.540	-5992.29[b]
trans-**H2**	-7.207	-27.608	0.000	-74.819	-50.806	-23.198	-5954.12[c]
trans-**H2**	-7.067	-29.975	10.136	-66.332	-51.924	-21.949	-5950.87
trans-**H2**	-6.993	-28.004	10.871	-67.469	-51.724	-23.720	-5773.01
trans-**H2**	-6.950	-29.486	15.198	-66.896	-48.527	-19.041	-5934.11
trans-**H2**	-6.937	-27.163	8.246	-65.958	-51.758	-24.595	-5933.52
trans-**H2**	-6.867	-27.406	8.122	-63.433	-48.420	-21.014	-5763.15
trans-**H2**	-6.472	-27.427	8.419	-63.429	-46.434	-19.007	-5805.15
trans-**H2**	-6.442	-25.739	7.122	-57.015	-45.131	-19.392	-5723.27
cis-**H2**	-8.329	-30.335	7.816	-79.426	-56.791	-26.457	-5988.02
cis-**H2**	-8.323	-31.438	8.511	-84.493	-57.795	-26.358	-5962.38
cis-**H2**	-8.277	-31.398	6.562	-88.593	-58.292	-26.895	-5763.33
cis-**H2**	-8.223	-31.269	0.000	-79.800	-57.831	-26.562	-5990.92
cis-**H2**	-8.210	-31.195	12.044	-77.041	-55.346	-24.152	-5985.84
cis-**H2**	-8.125	-30.252	9.592	-84.144	-56.099	-25.847	-5991.47
cis-**H2**	-8.090	-30.628	8.007	-80.966	-55.977	-25.349	-5923.46
cis-**H2**	-8.041	-30.919	8.672	-85.580	-55.592	-24.673	-5782.84
cis-**H2**	-8.026	-33.291	8.672	-80.178	-56.396	-23.105	-5992.42
cis-**H2**	-8.000	-32.342	7.384	-82.959	-59.710	-27.368	-5995.30[d]

a) Ligand interaction diagram and illustration shown in Figure 5.6-45. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-46. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-47. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-48.

a) **b)**

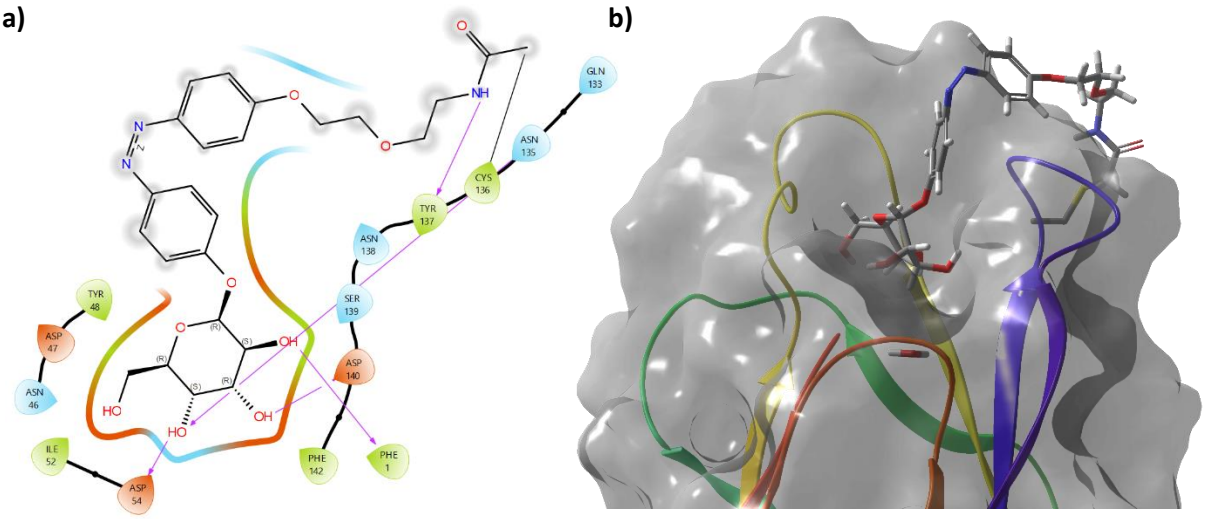


Figure 5.6-42: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H1** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

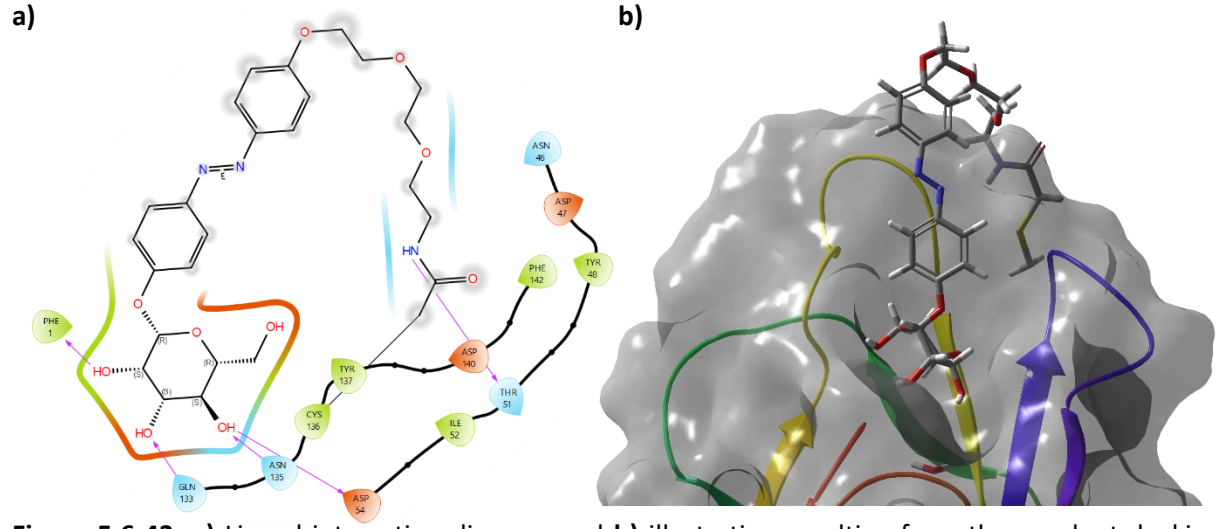


Figure 5.6-43: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

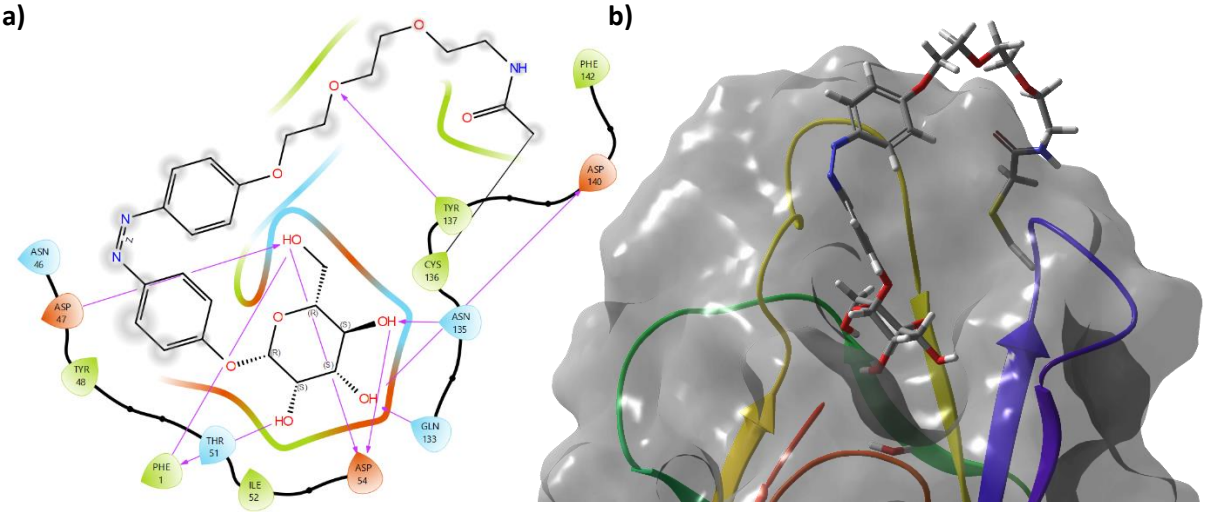


Figure 5.6-44: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H2** in the N136C variant of 1UWF (closed-gate conformation).

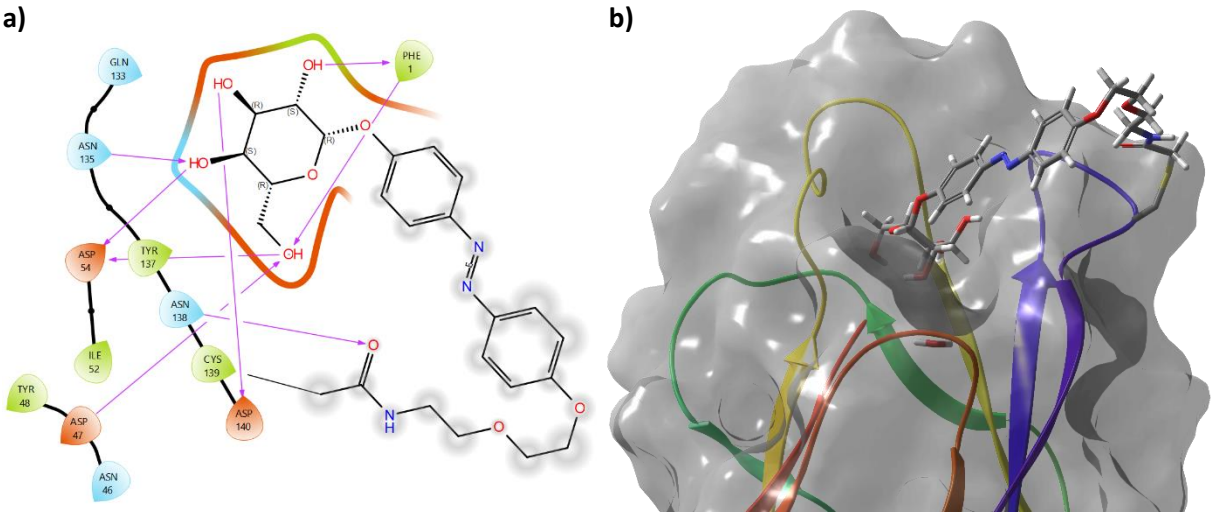


Figure 5.6-45: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H1** in the S139C variant of 1UWF (closed-gate conformation).

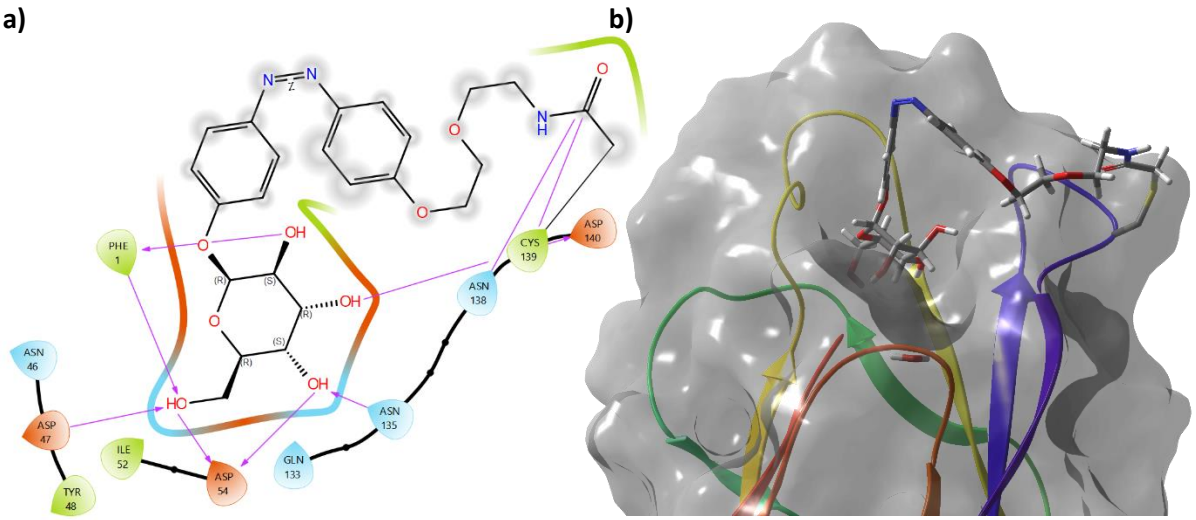


Figure 5.6-46: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H1** in the S139C variant of 1UWF (closed-gate conformation).

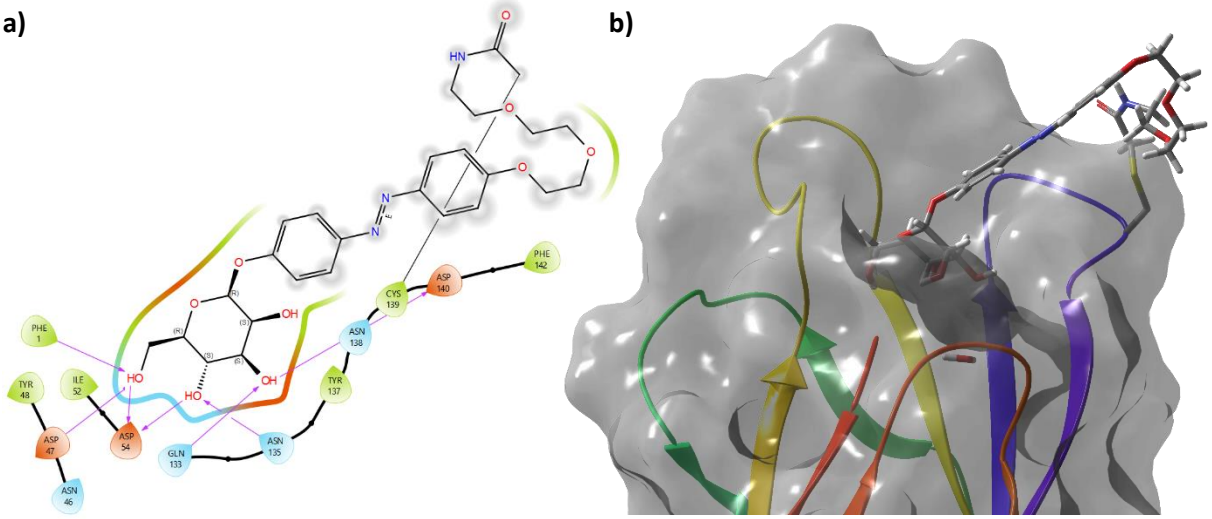


Figure 5.6-47: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H2** in the S139C variant of 1UWF (closed-gate conformation).

a) **b)**

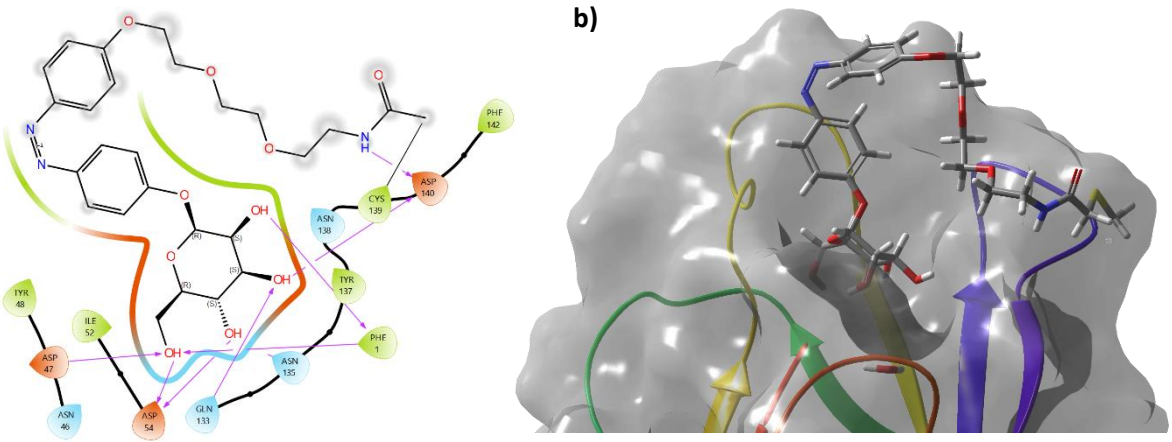


Figure 5.6-48: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H2** in the S139C variant of 1UWF (closed-gate conformation).

Table 5.6-15: Selected scoring values and Prime energies from covalent docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the N136C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**F1**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**F1**	-8.137	-26.640	8.335	-81.227	-53.490	-26.851	-5873.89
cis-**F1**	-8.037	-29.260	9.778	-85.048	-58.324	-29.064	-5862.67
cis-**F1**	-8.001	-25.757	7.535	-85.872	-57.874	-32.118	-5873.06
cis-**F1**	-7.995	-29.592	7.392	-85.351	-58.214	-28.621	-5875.58
cis-**F1**	-7.818	-31.714	9.749	-83.101	-55.297	-23.583	-5878.78[b]
cis-**F1**	-7.702	-27.342	8.438	-84.423	-56.103	-28.761	-5868.77
cis-**F1**	-7.659	-27.013	5.835	-74.137	-51.438	-24.425	-5857.05
cis-**F1**	-7.495	-28.912	6.684	-77.004	-51.132	-22.221	-5860.02
cis-**F1**	-7.350	-27.494	8.930	-77.273	-53.069	-25.575	-5870.47
cis-**F1**	-7.166	-25.780	14.281	-79.549	-54.296	-28.516	-5821.24
trans-**F2**	-8.804	-28.331	6.967	-90.075	-59.449	-31.118	-5912.68
trans-**F2**	-8.698	-27.746	7.400	-88.353	-57.720	-29.974	-5910.69
trans-**F2**	-8.481	-28.372	9.547	-90.116	-57.828	-29.455	-5912.12
trans-**F2**	-8.353	-28.423	9.788	-87.089	-57.769	-29.346	-5913.48[c]
trans-**F2**	-8.229	-25.909	11.664	-83.200	-56.201	-30.292	-5904.88
trans-**F2**	-5.283	-12.201	0.000	0.000	-18.920	-6.719	-5902.77
cis-**F2**	-8.820	-29.875	12.550	-87.745	-59.394	-29.519	-5944.48
cis-**F2**	-8.803	-31.809	13.309	-87.663	-58.899	-27.090	-5946.88
cis-**F2**	-8.701	-31.124	12.005	-91.550	-59.501	-28.377	-5948.67
cis-**F2**	-8.604	-27.509	10.304	-80.342	-58.985	-31.476	-5937.78
cis-**F2**	-8.567	-33.716	12.250	-86.573	-57.752	-24.036	-5936.36
cis-**F2**	-8.537	-30.778	14.838	-88.637	-60.043	-29.266	-5948.76
cis-**F2**	-8.401	-30.537	12.908	-78.143	-57.932	-27.395	-5918.53
cis-**F2**	-8.395	-29.174	11.626	-85.499	-60.607	-31.433	-5950.48[d]
cis-**F2**	-8.383	-28.658	7.146	-86.085	-57.449	-28.791	-5930.20
cis-**F2**	-8.373	-29.592	13.379	-87.517	-59.223	-29.631	-5942.56

a) No correct azobenzene structures observed. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-49. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-50. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-51.

Table 5.6-16: Selected scoring values and Prime energies from covalent docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the S139C variant of 1UWF (closed-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**F1**	-7.861	-32.206	8.395	-72.810	-58.155	-25.949	-5967.35
trans-**F1**	-7.552	-29.960	8.471	-72.481	-52.742	-22.782	-5968.41
trans-**F1**	-7.501	-29.061	9.206	-78.725	-58.623	-29.562	-5957.44
trans-**F1**	-7.374	-29.211	6.591	-74.160	-53.809	-24.598	-5969.85[a]
trans-**F1**	-7.339	-29.508	8.275	-76.078	-48.713	-19.206	-5967.60
trans-**F1**	-7.249	-27.535	9.735	-73.699	-52.139	-24.604	-5966.90
trans-**F1**	-7.112	-28.189	10.492	-72.686	-54.929	-26.739	-5966.03
trans-**F1**	-7.069	-30.383	9.845	-75.757	-52.501	-22.118	-5964.30
trans-**F1**	-5.259	-20.540	0.000	-58.954	-40.079	-19.538	-5943.71
cis-**F1**	-8.314	-30.911	8.853	-81.834	-53.648	-22.737	-6002.09[b]
cis-**F1**	-8.258	-30.009	8.167	-78.602	-53.268	-23.259	-5990.47
cis-**F1**	-8.255	-32.556	9.365	-77.410	-55.557	-23.001	-5998.54
cis-**F1**	-8.248	-29.778	9.239	-73.883	-52.876	-23.098	-5985.65
cis-**F1**	-8.225	-32.577	11.985	-78.055	-55.471	-22.894	-5988.62
cis-**F1**	-8.139	-32.943	9.781	-76.695	-55.300	-22.357	-5992.03
cis-**F1**	-8.091	-31.818	10.189	-83.704	-56.287	-24.469	-5986.52
cis-**F1**	-8.088	-29.813	8.596	-78.863	-53.587	-23.774	-5989.69
cis-**F1**	-8.080	-29.820	7.559	-82.373	-52.192	-22.373	-5986.15
cis-**F1**	-8.018	-31.495	7.375	-82.712	-54.752	-23.257	-5995.28
trans-**F2**	-7.773	-29.513	9.978	-87.650	-61.483	-31.970	-5953.65
trans-**F2**	-7.561	-29.696	7.909	-67.131	-55.048	-25.352	-5947.41
trans-**F2**	-7.506	-29.814	5.687	-72.879	-56.358	-26.544	-5946.38
trans-**F2**	-7.468	-31.706	9.375	-76.745	-53.782	-22.076	-5953.21
trans-**F2**	-7.369	-29.292	12.338	-78.227	-56.539	-27.247	-5949.24
trans-**F2**	-7.319	-28.090	12.899	-77.800	-55.814	-27.724	-5955.85[c]
trans-**F2**	-7.255	-31.033	11.664	-74.100	-54.881	-23.848	-5954.62
trans-**F2**	-7.243	-27.515	9.180	-75.891	-54.838	-27.323	-5949.75
trans-**F2**	-5.829	-12.580	0.000	0.000	-21.107	-8.528	-5933.94
cis-**F2**	-8.527	-33.192	14.726	-86.349	-59.843	-26.651	-5992.66
cis-**F2**	-8.513	-31.169	17.921	-78.091	-58.159	-26.990	-5987.54
cis-**F2**	-8.432	-31.138	14.256	-82.641	-57.471	-26.334	-5995.12[d]
cis-**F2**	-8.375	-31.603	12.589	-83.680	-59.135	-27.531	-5988.95
cis-**F2**	-8.367	-30.994	9.838	-79.340	-55.977	-24.983	-5984.29
cis-**F2**	-8.336	-31.645	13.566	-80.322	-57.594	-25.950	-5982.51
cis-**F2**	-8.316	-30.006	6.133	-80.333	-56.264	-26.258	-5991.65
cis-**F2**	-8.219	-30.260	19.839	-78.982	-56.724	-26.464	-5982.82
cis-**F2**	-8.214	-31.586	11.787	-74.704	-55.348	-23.761	-5991.72
cis-**F2**	-8.208	-33.099	12.910	-79.856	-57.138	-24.039	-5988.67

a) Ligand interaction diagram and illustration shown in Figure 5.6-52. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-53. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-54. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-55.

a) **b)**

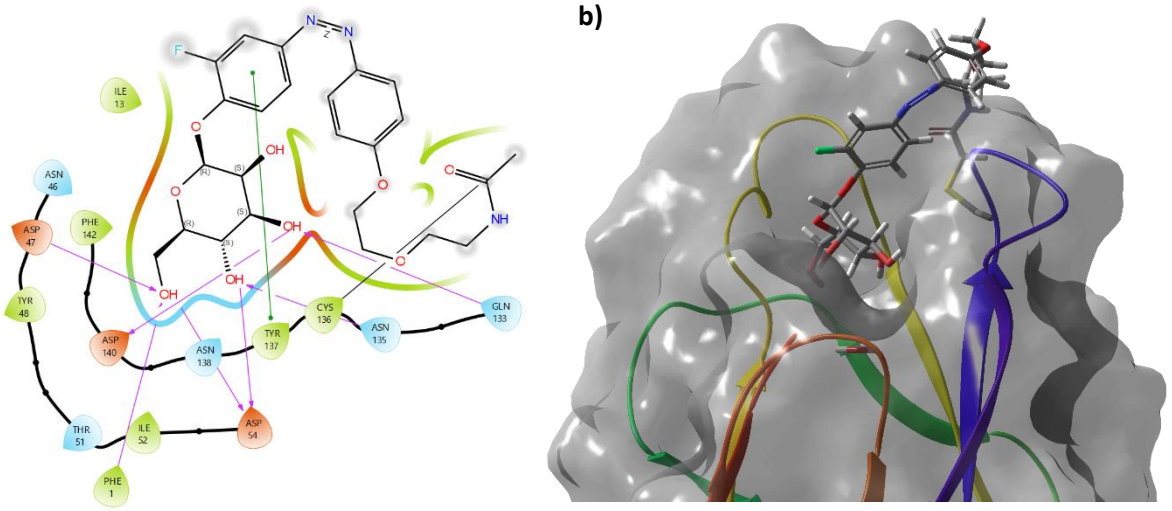


Figure 5.6-49: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F1** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

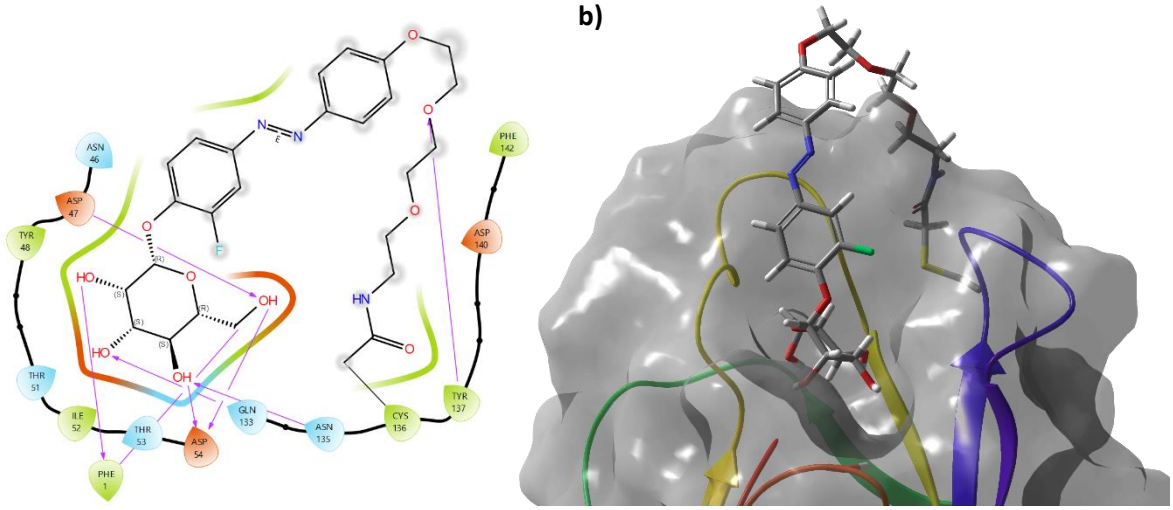


Figure 5.6-50: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F2** in the N136C variant of 1UWF (closed-gate conformation).

a) **b)**

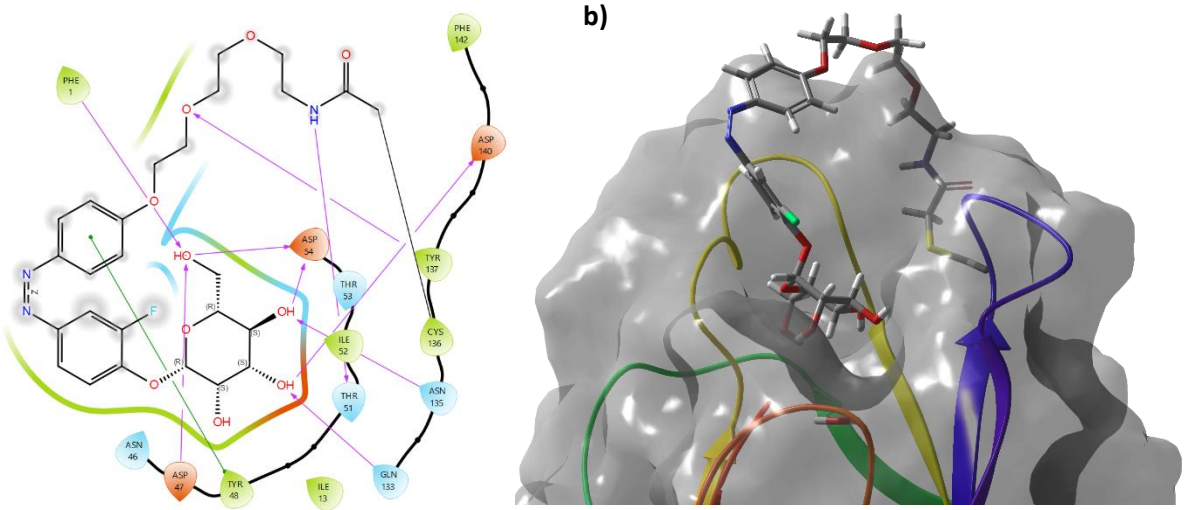


Figure 5.6-51: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F2** in the N136C variant of 1UWF (closed-gate conformation).

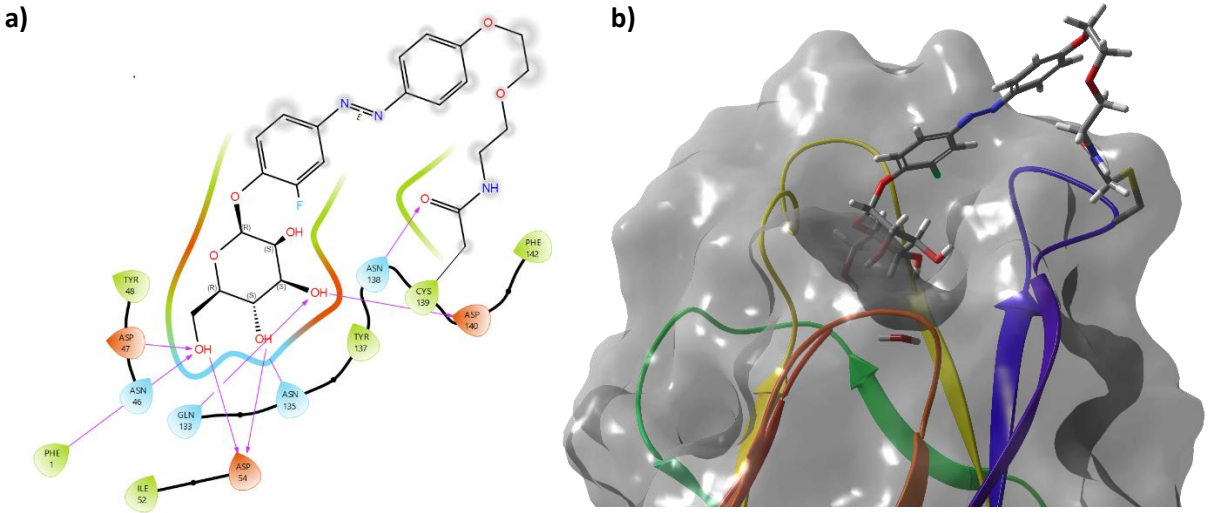


Figure 5.6-52: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F1** in the S139C variant of 1UWF (closed-gate conformation).

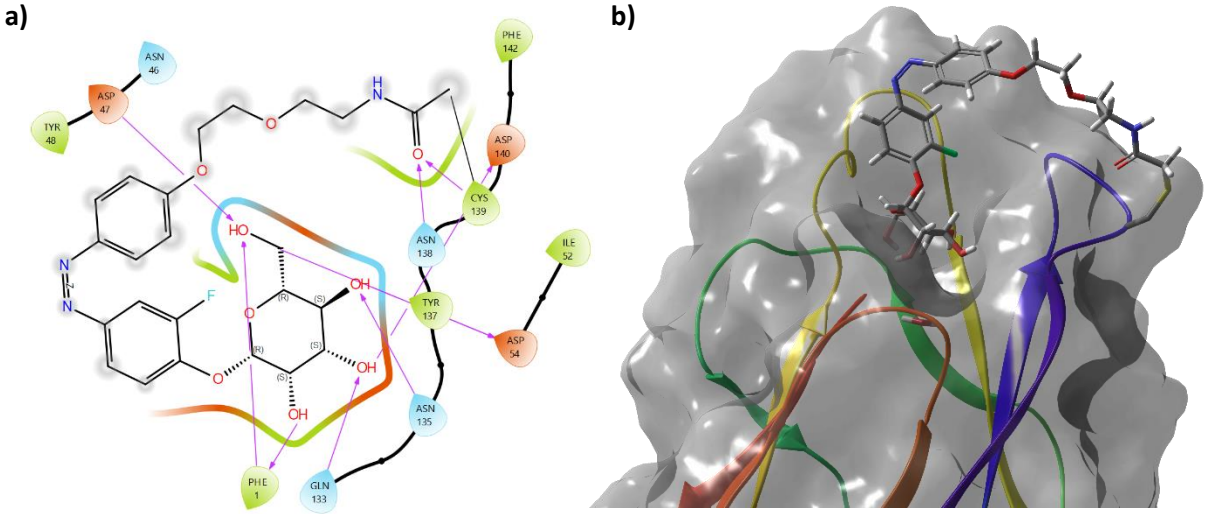


Figure 5.6-53: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F1** in the S139C variant of 1UWF (closed-gate conformation).

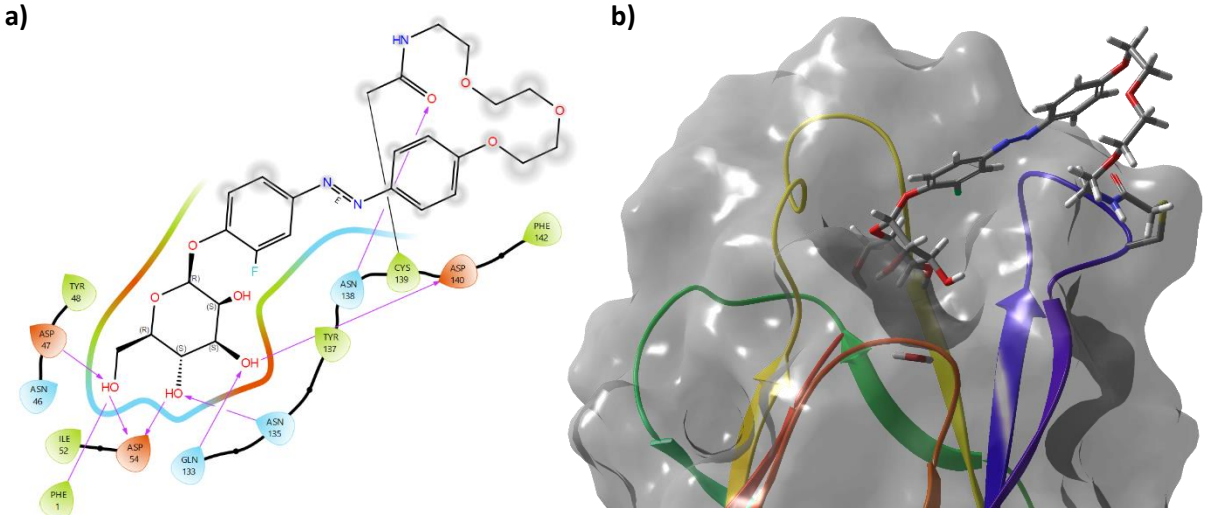


Figure 5.6-54: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F2** in the S139C variant of 1UWF (closed-gate conformation).

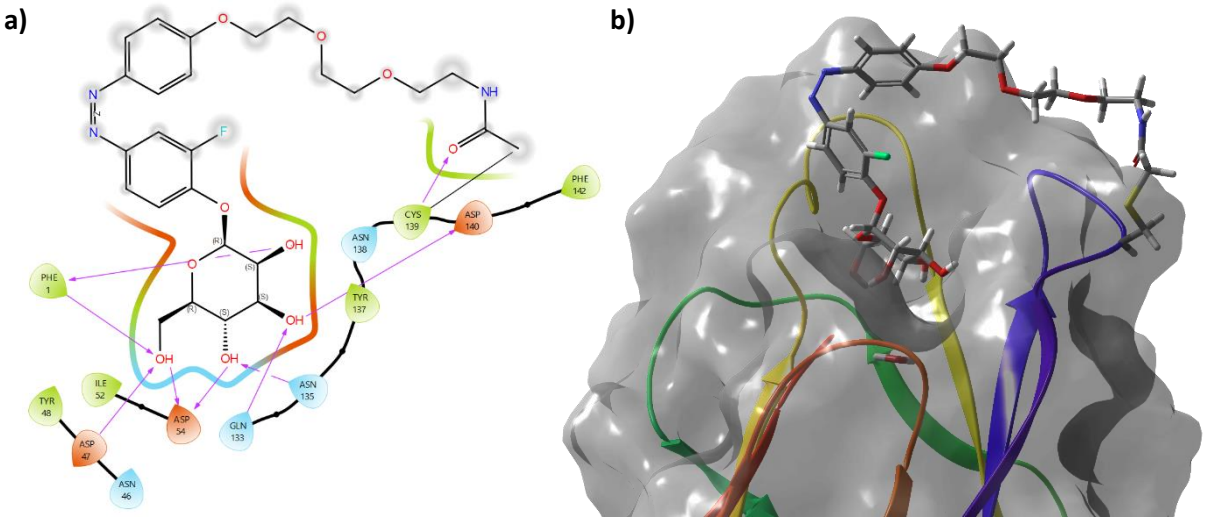


Figure 5.6-55: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F2** in the S139C variant of 1UWF (closed-gate conformation).

Table 5.6-17: Selected scoring values and Prime energies from covalent docking simulations of azobenzene derivatives **H1** and **H2** in the N136C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**H1**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**H1**	-7.443	-23.013	10.867	-77.595	-52.078	-29.065	-10200.10[b]
cis-**H1**	-7.430	-23.999	9.356	-77.887	-53.145	-29.146	-10088.09
cis-**H1**	-7.388	-27.179	7.792	-76.223	-52.049	-24.869	-10086.25
cis-**H1**	-7.326	-22.007	9.503	-72.137	-48.952	-26.945	-9978.36
cis-**H1**	-7.143	-24.679	5.999	-79.150	-52.394	-27.715	-9703.85
cis-**H1**	-7.110	-25.409	7.134	-76.286	-50.826	-25.417	-9718.07
cis-**H1**	-6.876	-23.431	6.475	-80.649	-53.477	-30.046	-10190.13
cis-**H1**	-6.719	-23.737	10.300	-76.121	-52.512	-28.775	-9770.65
cis-**H1**	-6.092	-26.013	10.908	-70.471	-50.732	-24.720	-9935.22
trans-**H2**	-7.410	-20.775	10.445	-82.881	-56.856	-36.081	-10097.83[c]
trans-**H2**	-6.874	-26.401	11.301	-77.258	-57.182	-30.781	-9688.87
trans-**H2**	-6.450	-26.534	11.898	-86.713	-58.099	-31.565	-9716.74
cis-**H2**	-8.681	-26.231	11.456	-84.997	-61.156	-34.925	-10202.73
cis-**H2**	-8.543	-26.714	11.480	-91.361	-60.573	-33.859	-10212.41
cis-**H2**	-8.392	-24.128	9.527	-87.181	-57.050	-32.921	-10145.35
cis-**H2**	-8.121	-28.857	10.506	-90.719	-57.320	-28.463	-9978.07
cis-**H2**	-7.988	-25.317	9.837	-93.119	-60.866	-35.549	-10216.28[d]
cis-**H2**	-7.938	-25.029	11.319	-86.600	-56.125	-31.097	-10210.83
cis-**H2**	-7.902	-25.267	9.195	-78.271	-54.869	-29.602	-9769.79
cis-**H2**	-7.859	-24.108	12.146	-86.235	-58.882	-34.774	-10118.28
cis-**H2**	-7.818	-28.864	8.139	-74.731	-53.788	-24.924	-9704.62
cis-**H2**	-7.745	-23.997	13.653	-83.015	-59.501	-35.504	-10208.40

a) No correct azobenzene structures and/or covalent docking results observed. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-56. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-57. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-58.

Table 5.6-18: Selected scoring values and Prime energies from covalent docking simulations of azobenzene derivatives **H1** and **H2** in the S139C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**H1**	-6.341	-25.459	8.079	-72.253	-47.610	-22.151	-10257.25[a]
trans-**H1**	-6.299	-23.031	7.718	-68.267	-47.139	-24.108	-10210.98
trans-**H1**	-6.217	-26.056	7.697	-62.423	-49.742	-23.686	-10256.37
trans-**H1**	-5.960	-25.422	8.599	-64.833	-48.056	-22.634	-10255.93
trans-**H1**	-5.857	-24.462	5.620	-57.746	-41.965	-17.503	-10089.69
trans-**H1**	-5.792	-23.677	9.136	-53.848	-44.556	-20.879	-10065.75
cis-**H1**	-7.258	-25.558	10.569	-69.844	-50.245	-24.687	-10253.83
cis-**H1**	-7.209	-26.681	6.969	-72.133	-50.445	-23.764	-10199.61
cis-**H1**	-7.207	-25.805	9.768	-70.813	-52.459	-26.654	-10163.88
cis-**H1**	-7.187	-28.033	12.263	-75.219	-52.779	-24.746	-10263.32[b]
cis-**H1**	-7.159	-25.489	7.817	-73.133	-50.841	-25.352	-10224.63
cis-**H1**	-7.113	-25.373	6.824	-75.521	-50.334	-24.961	-10249.89
cis-**H1**	-7.099	-24.010	13.895	-70.306	-51.669	-27.659	-10247.41
cis-**H1**	-7.086	-27.013	11.230	-74.480	-53.824	-26.811	-10218.94
cis-**H1**	-7.082	-25.426	6.852	-75.266	-51.114	-25.688	-10242.30
cis-**H1**	-6.979	-26.763	8.359	-71.082	-48.858	-22.096	-10243.55
trans-**H2**	-7.370	-28.116	10.722	-73.674	-52.211	-24.095	-10261.04[c]
trans-**H2**	-7.121	-26.247	7.450	-67.633	-49.302	-23.055	-10179.91
trans-**H2**	-6.939	-26.172	7.361	-66.313	-47.418	-21.246	-10118.80
trans-**H2**	-6.881	-28.355	7.017	-69.665	-51.625	-23.271	-10249.31
trans-**H2**	-6.822	-25.568	6.553	-60.871	-47.127	-21.559	-10074.64
trans-**H2**	-6.813	-25.393	8.560	-67.539	-48.635	-23.242	-10209.19
trans-**H2**	-6.764	-26.165	6.753	-68.376	-51.583	-25.418	-10204.50
trans-**H2**	-6.393	-24.423	8.661	-58.196	-45.637	-21.213	-10146.04
trans-**H2**	-6.080	-30.307	11.551	-75.915	-55.534	-25.228	-10251.86
cis-**H2**	-8.285	-28.155	8.579	-90.660	-60.726	-32.570	-10251.07
cis-**H2**	-8.267	-26.310	8.595	-88.362	-57.910	-31.599	-10251.95
cis-**H2**	-8.046	-27.690	9.813	-80.015	-52.742	-25.052	-10241.50
cis-**H2**	-7.968	-28.610	10.879	-82.930	-56.314	-27.704	-10041.31
cis-**H2**	-7.882	-29.862	10.987	-82.127	-56.142	-26.280	-10240.34
cis-**H2**	-7.862	-26.530	5.521	-89.237	-57.136	-30.607	-10223.62
cis-**H2**	-7.792	-25.889	8.443	-72.013	-49.836	-23.947	-10201.34
cis-**H2**	-7.787	-31.326	8.473	-78.925	-54.291	-22.965	-10235.12
cis-**H2**	-7.776	-26.308	9.114	-89.102	-58.808	-32.500	-10258.70[d]
cis-**H2**	-7.762	-30.003	7.197	-83.926	-56.496	-26.494	-10243.37

a) Ligand interaction diagram and illustration shown in Figure 5.6-59. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-60. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-61. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-62.

a) **b)**

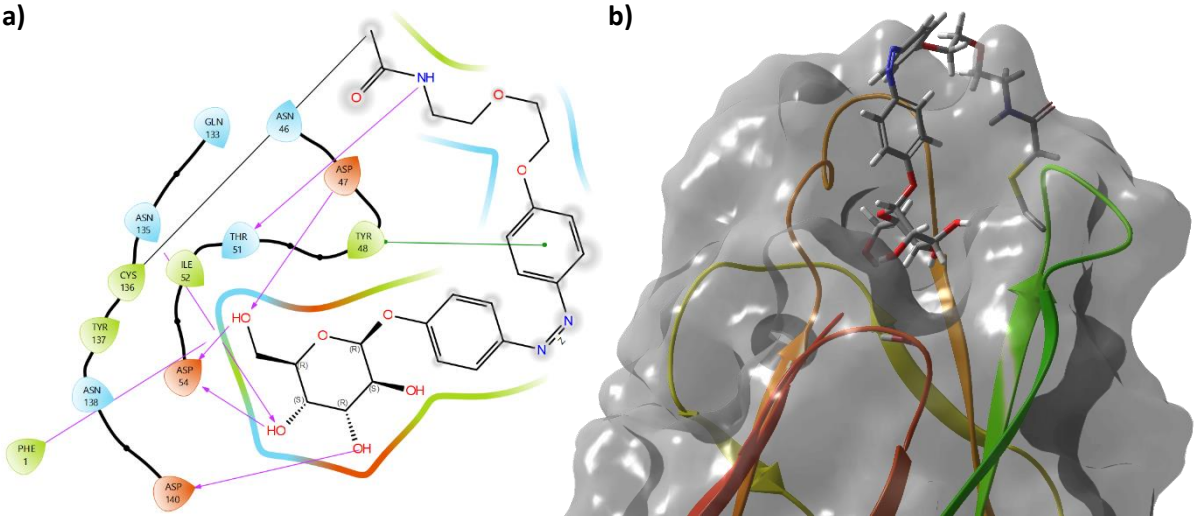


Figure 5.6-56: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

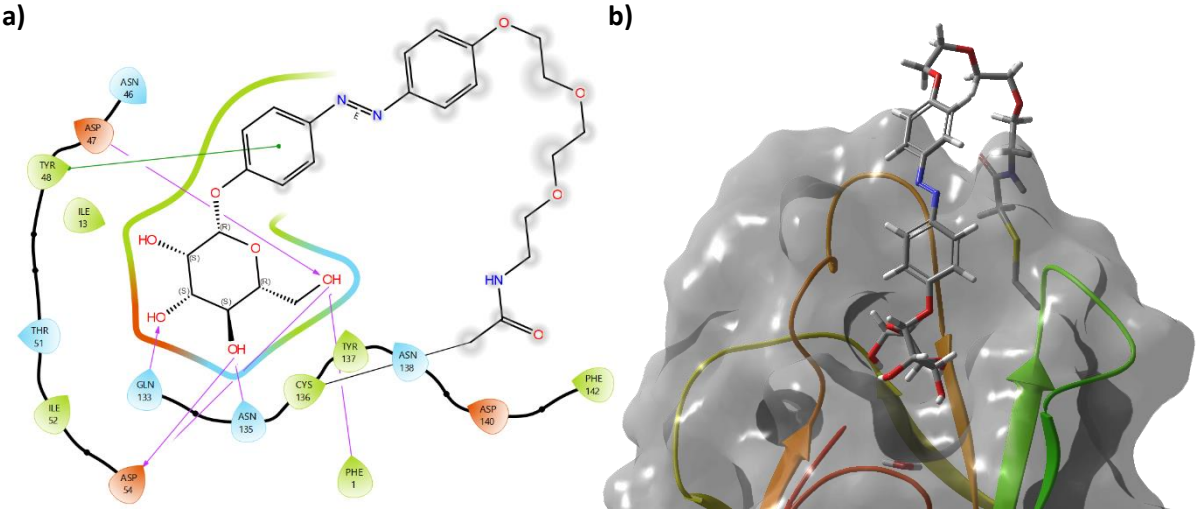


Figure 5.6-57: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H2** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

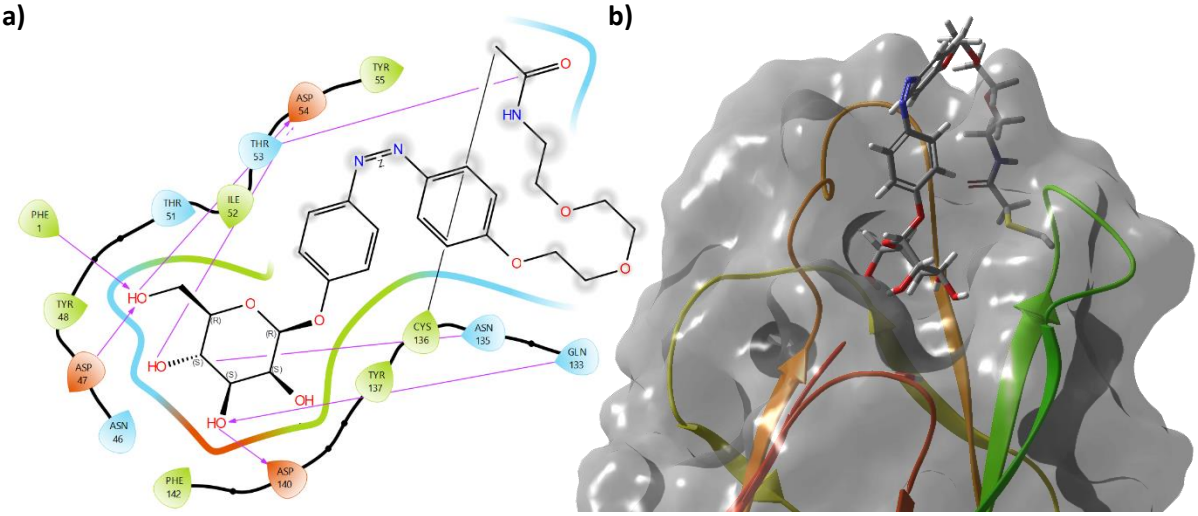


Figure 5.6-58: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H2** in the N136C variant of 1KLF (open-gate conformation).

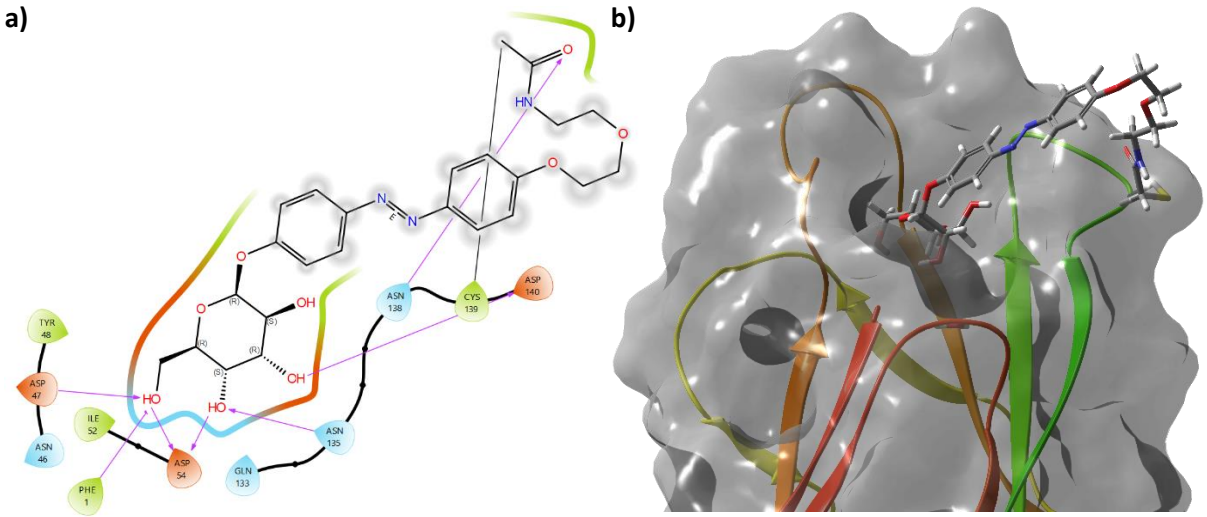


Figure 5.6-59: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H1** in the S139C variant of 1KLF (open-gate conformation).

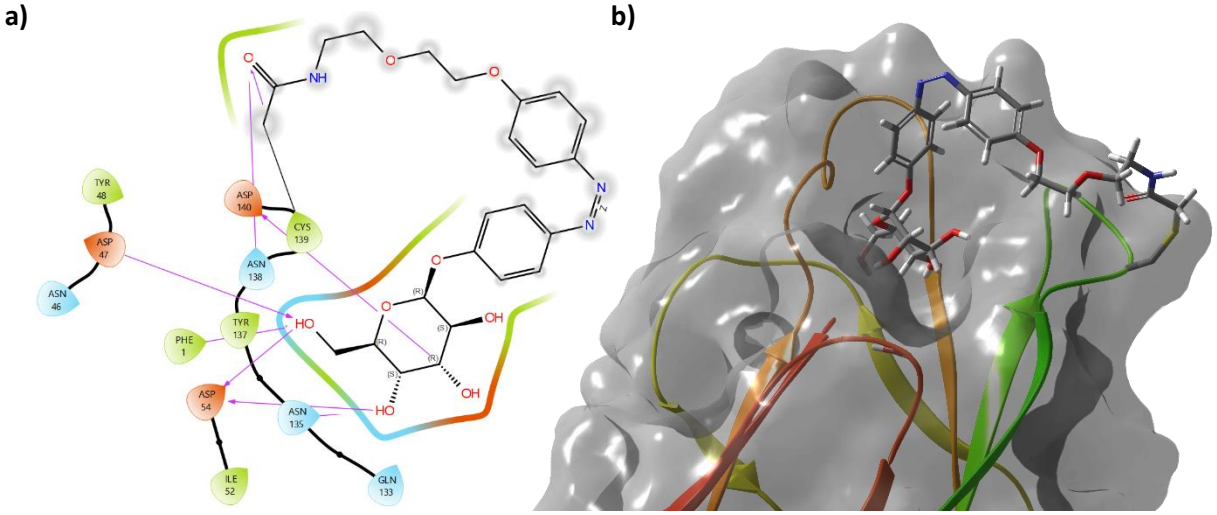


Figure 5.6-60: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H1** in the S139C variant of 1KLF (open-gate conformation).

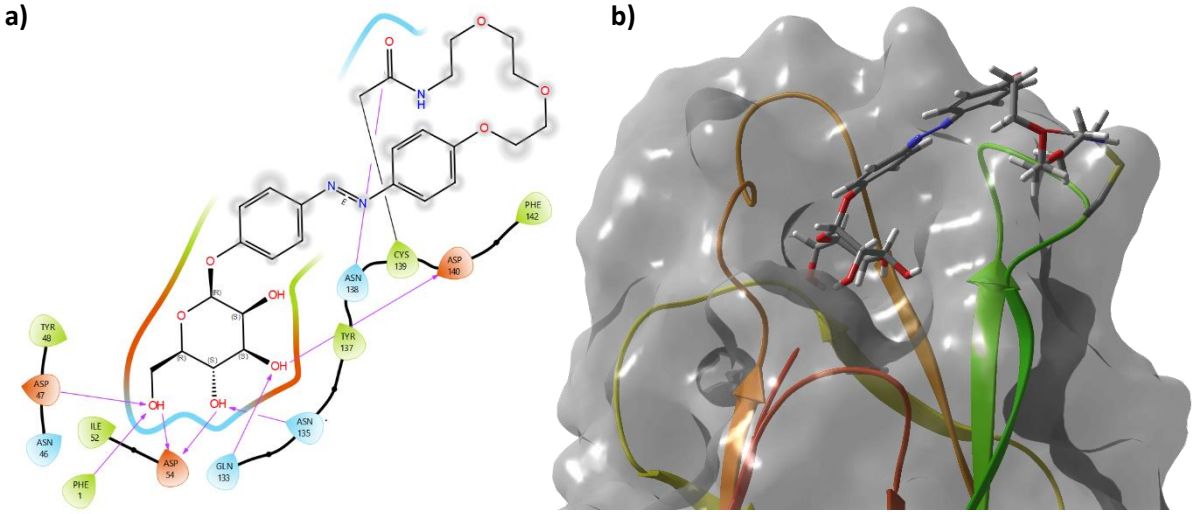


Figure 5.6-61: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**H2** in the S139C variant of 1KLF (open-gate conformation).

a) **b)**

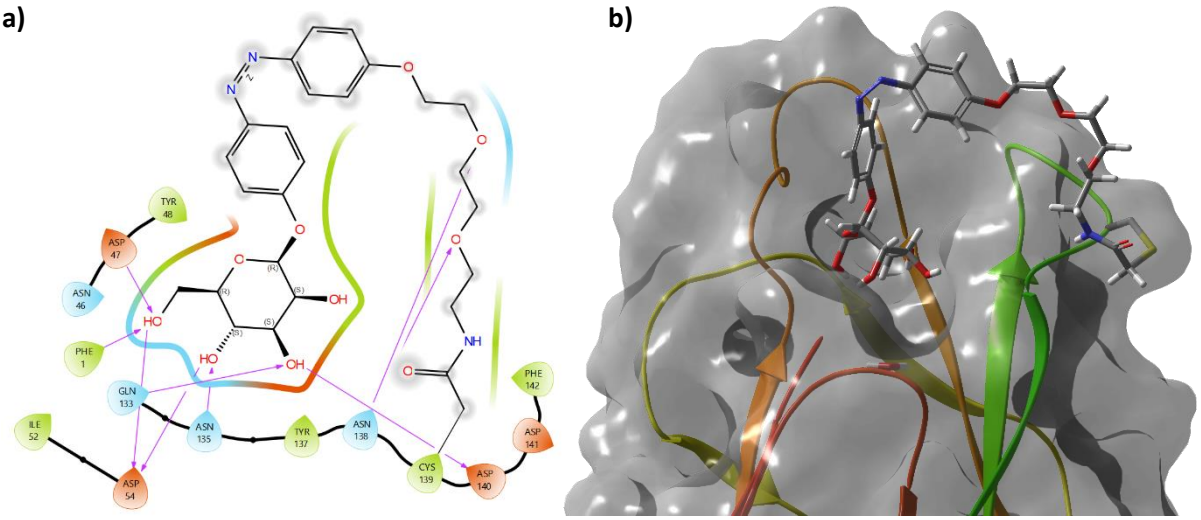


Figure 5.6-62: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**H2** in the S139C variant of 1KLF (open-gate conformation).

Table 5.6-19: Selected scoring values and Prime energies from covalent docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the N136C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**F1**[a]	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.	n. d.
cis-**F1**	-8.541	-24.796	9.452	-90.925	-57.439	-32.643	-10221.30[b]
cis-**F1**	-8.492	-23.638	8.964	-90.263	-55.734	-32.095	-10185.05
cis-**F1**	-8.392	-23.203	8.740	-85.087	-56.225	-33.022	-10206.87
cis-**F1**	-8.362	-23.294	7.907	-89.363	-55.733	-32.438	-10190.06
cis-**F1**	-8.337	-22.816	8.336	-84.787	-56.285	-33.469	-10219.53
cis-**F1**	-8.141	-22.523	9.860	-81.115	-54.261	-31.738	-10199.97
cis-**F1**	-8.119	-24.652	9.229	-90.952	-55.910	-31.258	-10181.20
cis-**F1**	-7.971	-22.461	9.388	-79.766	-54.750	-32.289	-10207.80
cis-**F1**	-7.930	-25.668	12.355	-87.607	-54.974	-29.305	-10190.65
cis-**F1**	-7.902	-22.451	8.067	-86.248	-55.788	-33.337	-10208.42
trans-**F2**	-5.979	-21.747	8.389	-79.011	-52.795	-31.047	-10185.13[c]
cis-**F2**	-9.218	-28.334	8.231	-95.148	-59.218	-30.884	-10217.50
cis-**F2**	-8.966	-25.690	7.623	-89.294	-61.077	-35.387	-10189.43
cis-**F2**	-8.752	-24.342	8.541	-85.358	-56.719	-32.377	-10211.30
cis-**F2**	-8.739	-23.679	8.339	-89.296	-58.194	-34.515	-10223.98[d]
cis-**F2**	-8.700	-23.205	9.709	-91.155	-57.994	-34.789	-10213.10
cis-**F2**	-8.561	-25.380	6.983	-91.156	-58.444	-33.064	-10211.44
cis-**F2**	-8.478	-27.445	9.658	-97.618	-62.637	-35.191	-10216.30
cis-**F2**	-8.422	-24.549	8.451	-94.329	-60.561	-36.012	-10203.24
cis-**F2**	-8.363	-24.546	11.469	-91.900	-58.535	-33.989	-10219.54
cis-**F2**	-8.348	-25.766	10.133	-84.867	-56.894	-31.128	-10213.32

a) No correct azobenzene structures and/or covalent docking results observed. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-63. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-64. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-65.

Table 5.6-20: Selected scoring values and Prime energies from covalent docking simulations of fluoroazobenzene derivatives **F1** and **F2** in the S139C variant of 1KLF (open-gate conformation).

Ligand	Docking score	Glide ecoul	Glide einternal	Glide emodel	Glide energy	Glide evdw	Prime Energy
trans-**F1**	-7.332	-27.591	7.453	-79.648	-55.604	-28.014	-10263.78
trans-**F1**	-7.237	-29.111	7.768	-72.008	-51.943	-22.832	-10268.52[a]
trans-**F1**	-6.895	-27.400	8.505	-80.316	-53.144	-25.744	-10267.07
trans-**F1**	-6.880	-27.220	9.128	-81.347	-53.791	-26.571	-10266.60
trans-**F1**	-6.546	-27.171	9.822	-72.952	-48.184	-21.014	-10265.23
cis-**F1**	-8.109	-30.713	6.989	-80.892	-53.251	-22.539	-10258.18[b]
cis-**F1**	-8.086	-30.367	7.316	-78.015	-51.821	-21.454	-10251.98
cis-**F1**	-8.031	-30.445	12.828	-74.207	-53.692	-23.247	-10257.38
cis-**F1**	-7.989	-25.976	9.371	-77.152	-52.663	-26.687	-10244.40
cis-**F1**	-7.944	-32.293	9.115	-82.214	-52.662	-20.369	-10257.92
cis-**F1**	-7.902	-28.945	9.571	-76.760	-52.715	-23.770	-10254.46
cis-**F1**	-7.844	-23.788	7.210	-75.557	-51.126	-27.337	-10255.11
cis-**F1**	-7.775	-26.935	13.210	-75.033	-53.245	-26.310	-10236.87
cis-**F1**	-7.767	-29.058	10.027	-79.464	-54.619	-25.561	-10243.31
cis-**F1**	-7.698	-26.822	8.992	-84.716	-54.727	-27.905	-10247.25
trans-**F2**	-7.640	-26.184	11.799	-79.358	-54.335	-28.151	-10250.29
trans-**F2**	-7.398	-25.385	5.522	-77.003	-53.197	-27.811	-10258.79
trans-**F2**	-7.277	-28.179	10.398	-81.804	-56.114	-27.936	-10258.63
trans-**F2**	-7.276	-27.105	9.335	-68.584	-51.313	-24.208	-10251.10
trans-**F2**	-7.181	-27.140	7.556	-73.623	-49.856	-22.716	-10254.03
trans-**F2**	-7.180	-29.936	9.605	-76.314	-52.347	-22.411	-10252.62
trans-**F2**	-7.145	-25.779	12.286	-81.772	-54.252	-28.472	-10259.19
trans-**F2**	-7.110	-24.215	10.294	-72.117	-50.358	-26.143	-10245.88
trans-**F2**	-7.054	-27.438	7.055	-72.702	-51.020	-23.583	-10260.80[c]
trans-**F2**	-6.756	-26.131	7.277	-67.886	-50.351	-24.220	-10260.30
cis-**F2**	-8.305	-33.646	12.055	-87.525	-54.889	-21.243	-10251.95
cis-**F2**	-8.125	-27.842	9.400	-91.720	-59.751	-31.909	-10256.83
cis-**F2**	-8.018	-29.958	10.652	-82.664	-58.192	-28.234	-10248.42
cis-**F2**	-7.968	-27.156	9.234	-73.291	-51.515	-24.360	-10248.88
cis-**F2**	-7.956	-29.908	10.631	-83.600	-56.321	-26.412	-10259.15[d]
cis-**F2**	-7.932	-25.283	8.628	-74.488	-55.241	-29.958	-10246.95
cis-**F2**	-7.822	-26.047	6.860	-73.369	-52.834	-26.786	-10249.96
cis-**F2**	-7.696	-26.214	8.361	-82.892	-56.255	-30.041	-10242.22
cis-**F2**	-7.688	-29.725	14.323	-74.658	-55.404	-25.679	-10240.66
cis-**F2**	-7.683	-25.122	7.395	-71.886	-54.671	-29.549	-10249.08

a) Ligand interaction diagram and illustration shown in Figure 5.6-66. **b)** Ligand interaction diagram and illustration shown in Figure 5.6-67. **c)** Ligand interaction diagram and illustration shown in Figure 5.6-68. **d)** Ligand interaction diagram and illustration shown in Figure 5.6-69.

a) **b)**

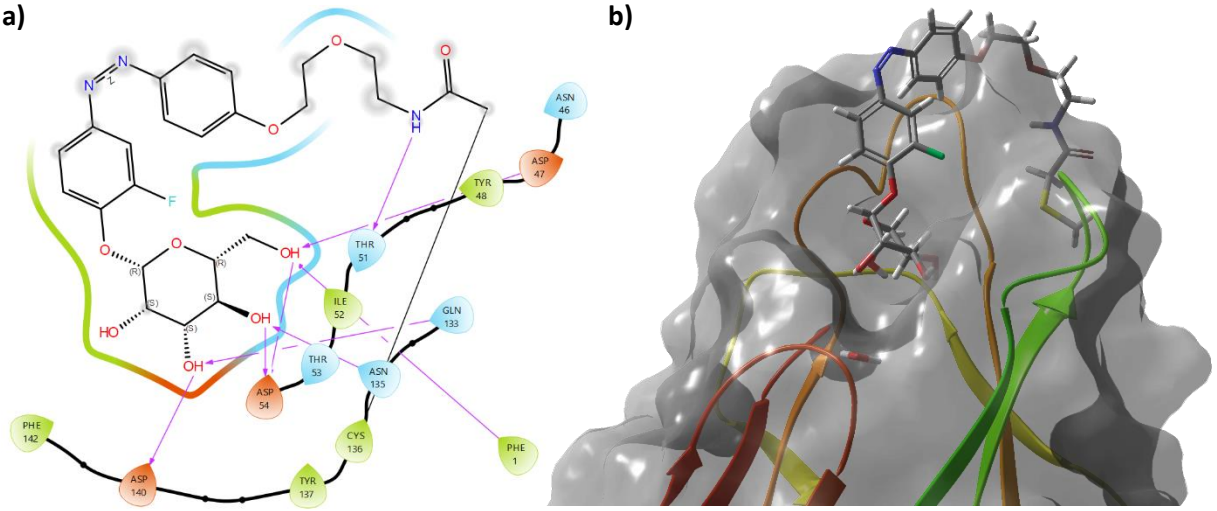


Figure 5.6-63: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F1** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

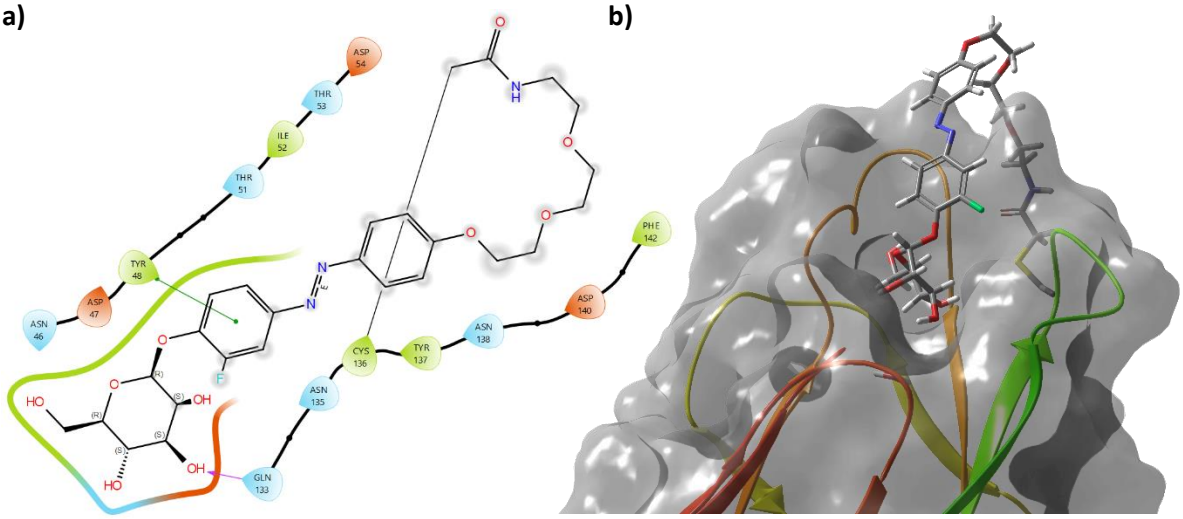


Figure 5.6-64: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F2** in the N136C variant of 1KLF (open-gate conformation).

a) **b)**

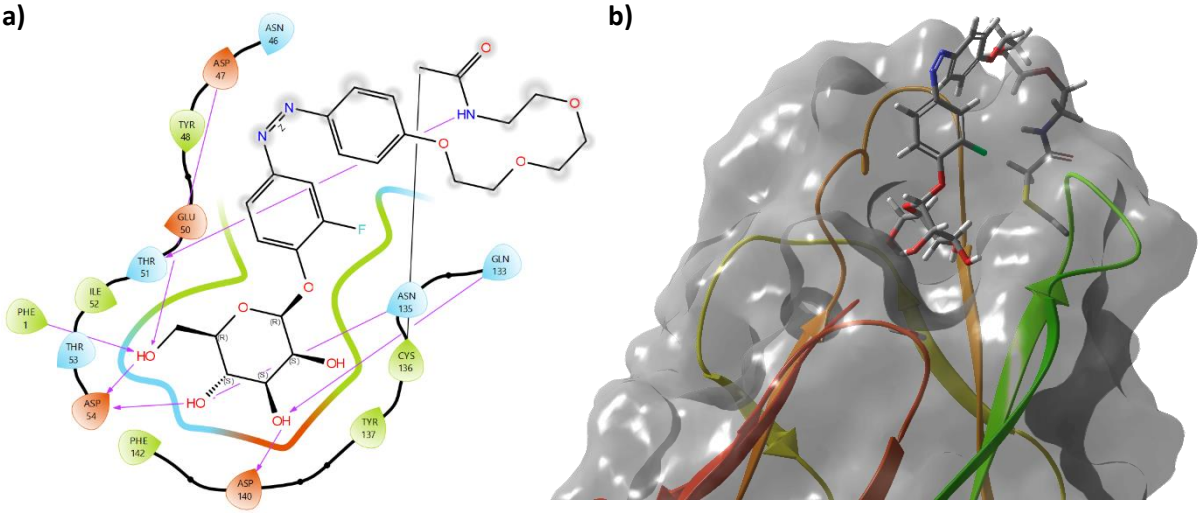


Figure 5.6-65: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F2** in the N136C variant of 1KLF (open-gate conformation).

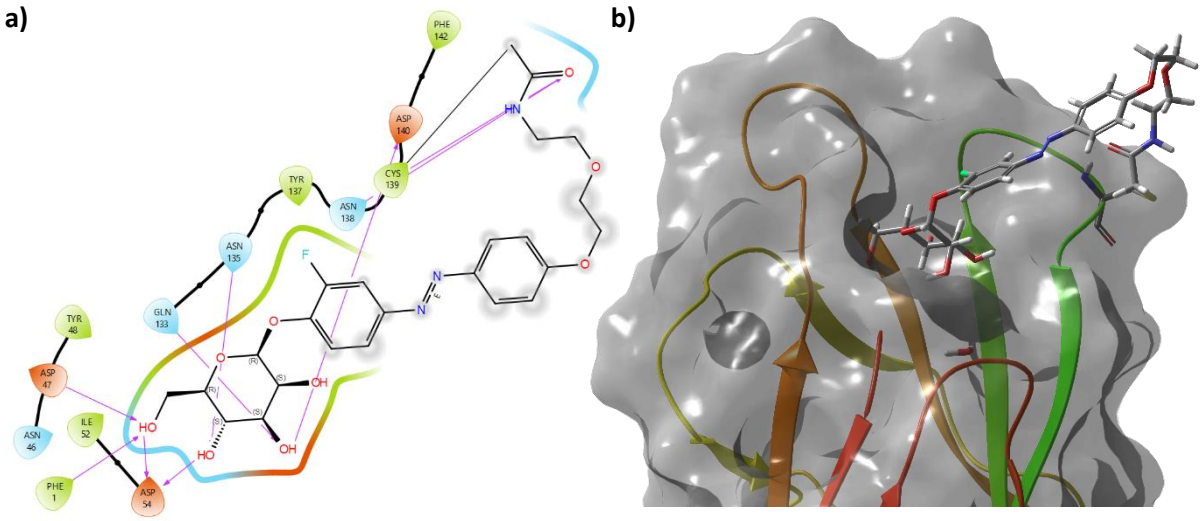


Figure 5.6-66: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F1** in the S139C variant of 1KLF (open-gate conformation).

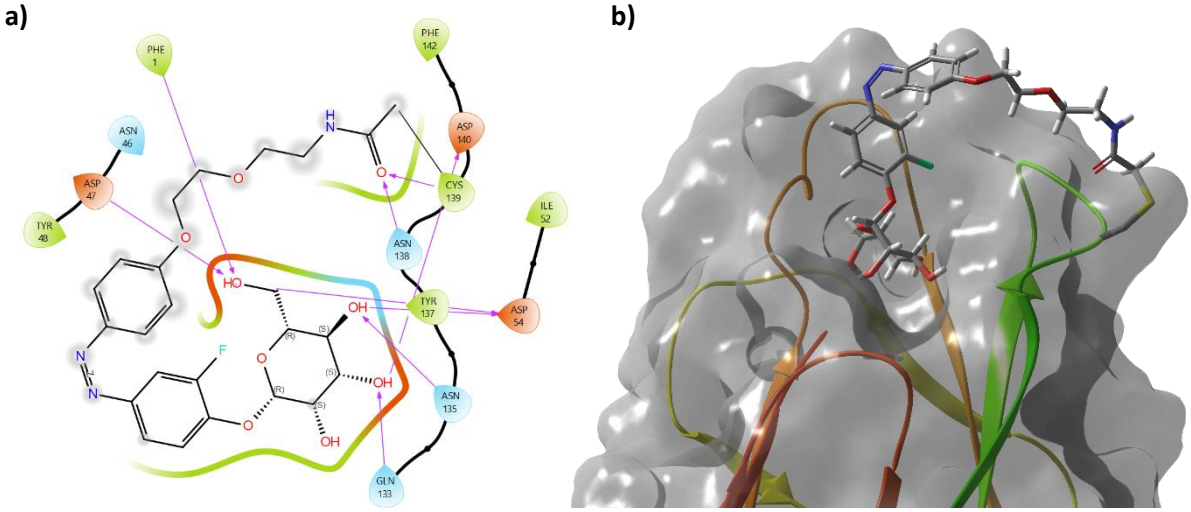


Figure 5.6-67: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F1** in the S139C variant of 1KLF (open-gate conformation).

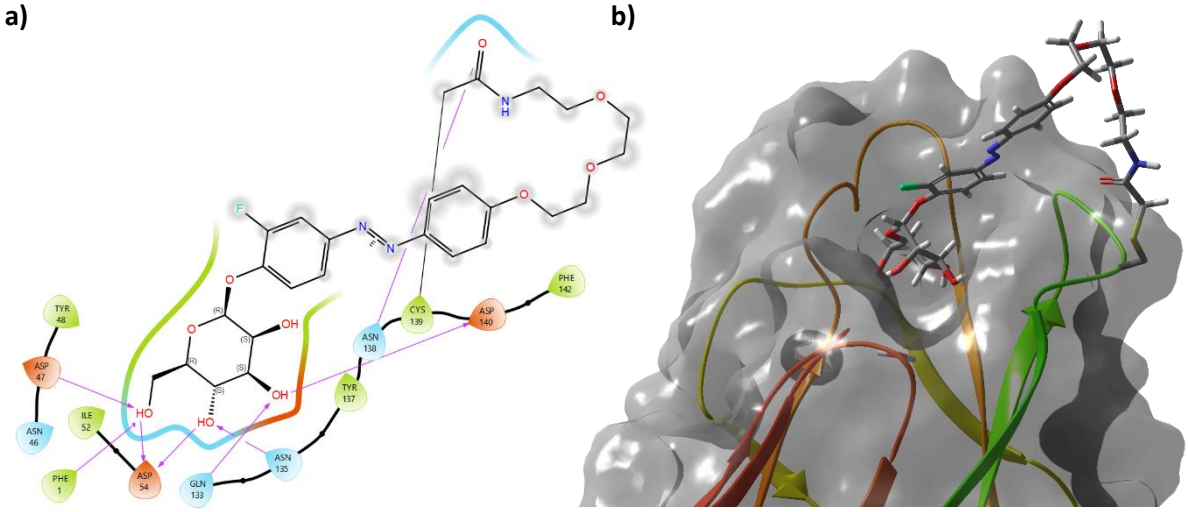


Figure 5.6-68: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *trans*-**F2** in the S139C variant of 1KLF (open-gate conformation).

a) **b)**

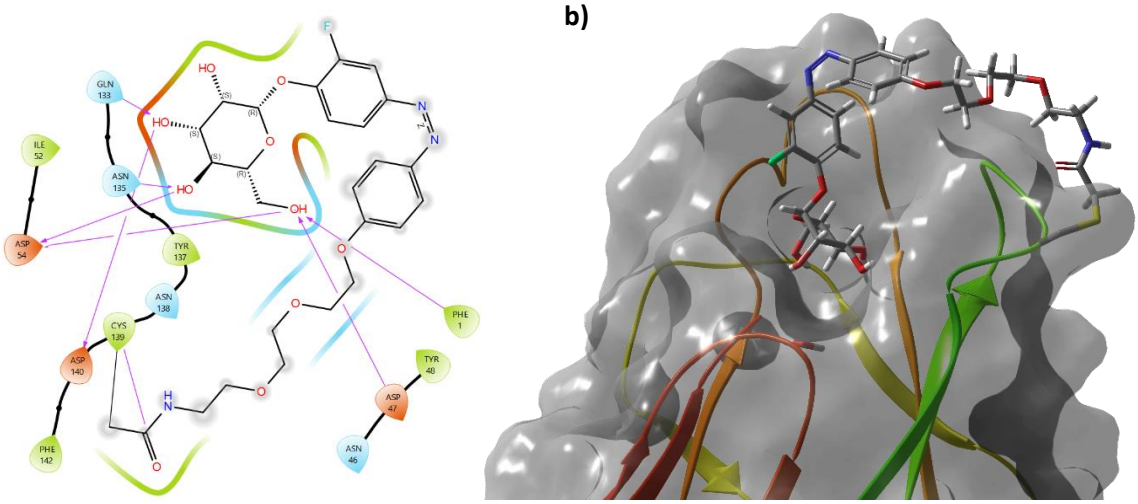


Figure 5.6-69: a) Ligand interaction diagram and **b)** illustration resulting from the covalent docking simulation of ligand *cis*-**F2** in the S139C variant of 1KLF (open-gate conformation).

5.7 General biological methods

5.7.1 Biochemicals

All biochemicals were purchased from commercial suppliers, such as Carl ROTH®, Sigma-Aldrich GmbH, Gibco™, Merck, New England Biolabs® and Thermo Scientific™, unless otherwise stated. DNA plasmids and oligonucleotides were ordered as high-purity salt-free lyophilisates from Eurofins Genomics (https://eurofinsgenomics.eu). New plasmids were either obtained from former or collaborating working groups or ordered from Addgene (https://www.addgene.org/) transformed into *E. coli* host strains.

5.7.2 Buffers and Media

All buffers and media were prepared in type 1 ultrapure water (cleansed via a Purelab® flex 3 system, ELGA Veolia). Media for cultivation of bacterial suspensions or on LB agar were sterilised (120 °C, 20 min) and antibiotics were added after cooling below ~ 45 °C. The pH values given in the following Table 5.7-1 were adjusted using aqueous solutions of 0.1 N HCl or 0.1 N NaOH respectively, if not otherwise stated. Commercially available buffers and media were used without pH adjustment and according to the supplier's recommendations.

Table 5.7-1: Composition of buffers, media and solutions.

Entry	Composition	Literature/ supplier
ROTIPHORESE® NF-Acrylamide/Bis solution	Acrylamide/*N*,*N*'-methylene bisacrylamide 29:1	ROTH®
Assay buffer	HEPES (20 mM), NaCl (150 mM), $CaCl \cdot 2\ H_2O$ (1 mM), pH 7.4 (at 22 °C)	Ernst et al.[486]
AEC high-salt buffer	Tris-HCl (10 mM, pH 8.0 at 22 °C), NaCl (1 M)	
AEC low-salt buffer	Tris-HCl (10 mM, pH 8.0 at 22 °C), NaCl (10 mM)	
Binding buffer	$NaH_2PO_4 \cdot 2\ H_2O$ (50 mM), NaCl (300 mM), imidazole (10 mM), pH 8.0 (at 22 °C) adjusted with phosphoric acid	Ernst et al.[486]
Carbonate buffer I	$NaCO_3$ (15 mM), $NaHCO_3$ (30 mM), pH 9.5 (at 22 °C)	Lindhorst et al.[467]
Carbonate buffer II	$NaCO_3$ (100 mM), $NaHCO_3$ (100 mM), pH 9.6 (at 22 °C)	Lindhorst et al.[467]
ConA buffer	$CaCl_2 \cdot 2\ H_2O$ (0.1 mM), $MnCl_2 \cdot 2\ H_2O$ (0.1 mM), dissolved in PBS, pH 7.4 (at 22 °C)	Despras et al.[689]
Coomassie destaining solution	Glacial acid (10 % v/v), ethanol (30 % v/v)	
Coomassie staining solution	Glacial acid (10 % v/v), ethanol (30 % v/v), Coomassie Brilliant Blue R-250	
CutSmart® Buffer	Potassium acetate (50 mM), Tris-acetate (20 mM), magnesium acetate (10 mM), BSA (100 µg/mL), pH 7.9 (at 25 °C)	New England Biolabs®
DNA Loading Dye (6 x)	Tris-HCl (10 mM, pH 7.6 at 22 °C) bromophenol blue (0.03 %), xylene cyanol FF (0.03 %), glycerol (60 %) EDTA (60 mM)	Thermo Scientific™ (cat. no. R0611)
Elution Buffer I	$NaH_2PO_4 \cdot 2\ H_2O$ (50 mM), NaCl (300 mM), imidazole (500 mM), pH 8.0 (at 22 °C)	
Elution Buffer II	$NaH_2PO_4 \cdot 2\ H_2O$ (50 mM), NaCl (300 mM), imidazole (250 mM), pH 8.0 (at 22 °C)	Ernst et al.[486]

FastAP™ Buffer (10 x)	Tris-HCl (100 mM, pH 8.0 at 372 °C), $MgCl_2$ (50 mM), KCl (1 M), Triton X-100 (0.2 %), BSA (1 mg/mL)	Thermo Scientific™
Laemmli buffer (4 x), non-reducing	Tris-HCl (250 mM, pH 6.8 at 22 °C), SDS (8 % w/v), glycerol (50 % v/v), bromophenol blue (0.05 % w/v)	
Laemmli buffer (4 x), reducing	Tris-HCl (250 mM, pH 6.8 at 22 °C), SDS (8 % w/v), glycerol (40 % v/v), β-mercaptoethanol (10 % v/v), bromophenol blue (0.05 % w/v)	
LB Agar [a]	Tryptone (10 g/L), yeast extract (5 g/L), NaCl (10 g/L), pH 7.0 (at 22 °C) + select agar (15 g/L)	ROTH®
LC-MS/MS loading buffer	Acetonitrile (3 % v/v), trifluoroacetic acid (0.1 % v/v)	Kaulich et al.[493]
Lysis buffer I	Tris-HCl (50 mM, pH 7.5 at 22 °C), NaCl (150 mM), EDTA (5 mM, pH 8.0 at 22 °C), polymyxin B sulfate is added (1 – 1.25 mg/mL)	Ernst et al.[486]
Lysis buffer II	Imidazole (5 mM), PMSF (1 mM), DNase I (1 U/mL) in PBS (pH 7.4 at 22 °C) or $NaH_2PO_4 \cdot 2\ H_2O$ (50 mM, pH 8.0 at 22 °C)	
Lysogeny Broth (LB)[a]	Tryptone (10 g/L), yeast extract (5 g/L), NaCl (10 g/L), pH 7.0 at 22 °C	ROTH®
Na-acetate buffer (DNA precipitation)	Na acetate (3 M), pH 5.5 adjusted with acetic acid	
Phosphate Buffered Saline (PBS)[a]	NaCl (8 g/L), KCl (0.2 g/L), Na_2HPO_4 (1.42 g/L), KH_2PO_4 (0.27 g/L), pH 7.4 at 22 °C	Gibco™
Phosphate Buffered Saline + Tween (PBST)	Tween®20 (0.05 % v/v), dissolved in PBS, pH 7.4 at 22 °C	
Phusion GC buffer [a] (10 x)	n. d.	Thermo Scientific™
T4 DNA Ligase buffer [a] (10 x)	Tris-HCl (300 mM, pH 7.8 at 25 °C), $MgCl_2$ (100 mM), DTT (100 mM), ATP (10 mM)	Promega (cat. no. C126B)
Terrific Broth (TB)[a]	Tryptically digested casein (12 g/L), yeast extract (24 g/L), K_2HPO_4 (12.54 g/L), KH_2PO_4 (2.31 g/L), glycerol (4 mL/L), pH 7.2 (at 22 °C)	ROTH®
TRIS-acetic acid-EDTA buffer (TAE buffer)	TRIS (40 mM), glacial acid (0.114 % v/v), EDTA (1 mM, pH 8.0 at 22 °C)	
TSS	PEG 6000 (10 % w/v), DMSO (10 % v/v), $MgSO_4$ (30 mM) dissolved in LB medium, pH 6.5 at 22 °C, sterilised by autoclaving or sterile filtration (0.2 µm)	Chung et al.[690]
SDS resolving gel buffer	Tris-HCl (1.5 M, pH 8.8 at 22 °C), SDS (0.4 % w/v)	
SDS running buffer (10 x)	Tris (0.25 M), glycine (1.92 M), SDS (1 % w/v), no pH adjustment required	
SDS stacking gel buffer	Tris-HCl (0.5 M, pH 6.8 at 22 °C), SDS (0.5 % w/v)	
Super optimal Broth with glucose (SOC) Outgrowth medium	Tryptone (2 % w/v), yeast extract (0.5 % w/v), NaCl (0.05 % w/v), KCl (2.5 mM), $MgCl_2$ (10 mM), glucose (20 mM), pH 7.0 at 22 °C	New England Biolabs

a) Buffers and/or media were prepared according to the manufacturers' specifications. **b)** Buffers were sterile-filtered (0.2 µm) and degassed for several minutes in an ultrasonic bath before use.

5.7.3 Bacterial strains

Table 5.7-2: *E. coli* strains used for plasmid storage, transformation and transduction.

***E. coli* strain**	**Genotype and features**	**Selection/cultivation**	**Literature/ supplier**
XL-1 Blue[a]	*rec*A1 *end*A1 *gyr*A96 *thi*$^{-}$1 *hsd*R17 ($r_K^- m_K^+$) *sup*E44 *rel*A1 *lac* [F' *pro*A$^+$B$^+$ *lacI*^{q}ZΔM15:: Tn*10* (Tet^R)]	Tetracycline	n. d.
HB101[a]	F$^-$ *thi*1 *hsd*S20 (r_B^-, m_B^-) *sup*E44 *rec*A13 *ara*14 *leu*B6 *pro*A2 *lac*Y1 *gal*K2 *rps*L20 (Str^R) *xyl*-5 *mtl*1 HB101 exhibits an extensive deletion in the chromosomal *fim* gene cluster and is therefore deficient in the assembly of type 1 fimbriae.[296]	Streptomycin (20 µg/mL),	Promega
SG13009[a]	*nal*^{S} Str^S Rif^S *thi*$^-$ *lac*$^-$ *ara*$^+$ *gal*$^+$ *mtl*$^-$ F$^-$ *rec*A$^+$ *uvr*$^+$ *lon*$^+$	Kanamycin (on pREP4), rifampicin (50 µg/mL) and streptomycin (100 µg/mL)	Qiagen
M15[a]	*nal*^{S} Str^S Rif^S *thi*$^-$ *lac*$^-$ *ara*$^+$ *gal*$^+$ *mtl*$^-$ F$^-$ *rec*A$^+$ *uvr*$^+$ *lon*$^+$ Δ*lac*ZM15	Kanamycin (on pREP4), rifampicin (50 µg/mL) and streptomycin (100 µg/mL)	Qiagen
C41 (DE3)[a]	F$^-$ *omp*T *hsd*S_B (r_B^- $_{mB}^-$) *gal dcm* (DE3) The strain contains the T7 RNA polymerase on plasmid DE3 and is thus suitable for expression under the control of the T7 promoter.	n. d.	n. d.
KS272 HM125[b]	KS272: F$^-$Δ*lac*X74 *gal*E *gal*K *thi rps*L(*str*A) Δ*Pho*A(*Pvu*II)^{25} HM125: KS272 *degP4l(t:.Pstl)::*ΩKan^R *eda51::*Tn*10*(Tet^R) *rpoH15* mini-Tn*10*(Tet^R) The strain KS171 HM125 is deficient in the periplasmic protease *degP* and a double mutation of *eda*/*rpoH15,* and is therefore optimised for high-level expression of secreted proteins, periplasmic accumulation and facilitated extraction.	Kanamycin, (50 µg/mL), tetracycline (25 µg/mL) cultivation at 30 °C	Georgiou et al.[482], [b]
Origami B[a]	F$^-$ *omp*T *hsd*S_B($r_B^-m_B^-$ *gal dcm lac*Y1 *ahp*C (DE3) *gor*522::Tn*10 trx*B (Kan^R, Tet^R)	Kanamycin, tetracycline	n. d.
DH5α λpir pOSIP-KL-*mcherry*	*end*A1 *hsd*R17 *gln*V44 (= *sup*E44) *thi*-1 *rec*A1 *gyr*A96 *rel*A1 ϕ80*dlac*Δ(lacZ)M15 Δ(*lac*ZYA-*arg*F)U169 *zdg*-232::Tn*10 uid*A:: λ*pir*$^+$ The strain was used for storage and the *pir*-dependent replication of pOSIP-KL-*mcherry*. Pir is the replication initiation protein for R6K-Plasmids.	Kanamycin (50 µg/mL), cultivation at 30 °C	Addgene (no. 116947)
E811 pE-Flp	The strain was used for storage and replication of pE-Flp.	Ampicillin (50 µg/mL), cultivation at 30 °C	Addgene (no. 45978)
HB101-mCherry	HB101 λ *att*B::[$P_{Antoine\ R1}$-*mcherry*- *rrn*B T2-λ*cI*^{ts}-λ*int*-Kan^R-R6K$_\gamma$ ori]	Kanamycin (50 µg/mL), cultivation at 30 °C	This thesis
HB101-mCherry^{Flp}	HB101 λ attB::[$P_{Antoine\ R1}$-*mcherry*- rrnB T2]	Streptomycin (20 µg/mL), Cultivation at 37 °C	This thesis

a) The strains were kindly provided by the group of Prof. Dr. Axel Scheidig (Kiel University). **b)** The strain HM125 and the plasmid pFimHs-NT-trc were kindly provided by Prof. Dr. Rudi Glockshuber (ETH Zürich, Switzerland).

Table 5.7-3: *E. coli* strains used in bacterial adhesion and adhesion inhibitions experiments.

***E. coli* strain**	**Genotype and features**	**Selection**	**Literature/ supplier**
SAR18 pKL174[a] (PKL1162)	CSH26: *ara*, Δ(*lac-pro*), *thi*; SAR18: CSH26 *att*B::[*bla*-$P_{A1/04/03}$-*gfpmut3* *-T0] K12 CSH26 was transduced with the *gfpmut*3 gene derived from *Aequorea victoria* at the chromosomal *att*B site, mediated by a λ integrase. *E. coli*, derived from CSH26 SAR18 therefore constitutively expresses *av*GFP under the control of the strong artificial $P_{A1/04/03}$ promoter and a T0 terminator.[465] The strain PKL1162 further expresses type 1 fimbriae encoded on pPKL174.	Ampicillin or carbenicillin (100 µg/mL) and chloramphenicol (50 µg/mL)	Reisner et al.[465],[a] Lindhorst et al. [467]
HB101 pPKL4[a]	The strain expresses type 1 fimbriae encoded on pPKL4.[296]	Ampicillin or carbenicillin (100 µg/mL)	Klemm et al. [296],[a]
HB101-mCherry^{Flp} pPKL4	HB101 λ attB::[$P_{Antoine\ R1}$-mcherry- rrnB T2] The strain expresses *mcherry* under the control of a constitutive Antoine R1 promoter and the rrnB T2 terminator. The strain further expresses type 1 fimbriae encoded on pPKL4.[296]	Ampicillin or carbenicillin (100 µg/mL)	This thesis
HB101-mCherry^{Flp} pPKL174	HB101 λ attB::[$P_{Antoine\ R1}$-mcherry- rrnB T2] The strain expresses *mcherry* under the control of a constitutive Antoine R1 promoter and the rrnB T2 terminator. The strain further expresses type 1 fimbriae encoded on pPKL174.[467]	Chloramphenicol (50 µg/mL)	This thesis
HB101-mCherry^{Flp} pPAP5	HB101 λ attB::[$P_{Antoine\ R1}$-mcherry- rrnB T2] The strain expresses *mcherry* under the control of a constitutive Antoine R1 promoter and the rrnB T2 terminator. The strain further expresses P fimbriae encoded on pPAP5.[588]	Ampicillin or carbenicillin (100 µg/mL)	This thesis
HB101 pPAP5 pETM-G (VR123-G)	The strain expresses P fimbriae encoded on pPAP5.[588] The inducible expression of a GFP-like variant, encoded on plasmid pETM-G, proceeds under the control of a T7 and lac promoter.	Kanamycin (50 mg/L) , Ampicillin or carbenicillin (100 µg/mL)	Claudia Fessele, Dissertation [576]

a) The strain was formerly provided by the group of Prof. Per Klemm (Technical University of Denmark). **b)** The strains were kindly provided by the group of Prof. Dr. Axel Scheidig (Kiel University).

Table 5.7-4: *E. coli* strains used for the expression of FimH.

***E. coli* strain**	**Features**	**Selection**	**Literature/ supplier**
HB101 pPKL241[a, b]	Expression system for the expression of $fimH_L$ with C-terminal His_6 tag under the control of an IPTG-inducible T5/*lac*O promoter and λT0 terminator.	Ampicillin or carbenicillin (100 µg/mL)	Schembri et al.[470]
SG13009[c] pPKL241[a, b]	Expression system for the expression of $fimH_L$ with C-terminal His_6 tag under the control of an IPTG-inducible T5/*lac*O promoter and λT0 terminator.	Ampicillin or carbenicillin (100 µg/mL)	This thesis
C41 (DE3)[c] pPKL241[a, b]	Expression system for the expression of $fimH_L$ with C-terminal His_6 tag under the control of an IPTG-inducible T5/*lac*O promoter and λT0 terminator. The strain C41 (DE3) is deficient in *lon* and *ompT* proteases in order to reduce proteolytic activity during protein purification. It is also suited for expression of toxic target proteins.	Ampicillin or carbenicillin (100 µg/mL)	This thesis

HM125[d] pPKL241[a, b]	Expression system for the expression and periplasmic accumulation of $FimH_L$ with C-terminal His_6 tag, under the control of an IPTG-inducible T5/*lac*O promoter and λT0 terminator.	Ampicillin or carbenicillin (100 µg/mL)	This thesis
Origami B[c] pPKL241 ΔN-sigpep[a]	Expression system for the expression of *fimH*$_L$ with C-terminal His_6 tag under the control of an IPTG-inducible *trc* promoter and *rrnB* T1 terminator. The natural N-terminal signal peptide of *fimH*$_L$ was deleted in order to circumvent secretion in favour of cytoplasmic accumulation.	Ampicillin or carbenicillin (100 µg/mL)	This thesis
HM125[d] pFimHs-NT-trc[d]	Expression system for the expression and periplasmic accumulation of FimH (without C-terminal His_6 tag) under the control of an IPTG-inducible *trc* promoter and *rrnB* T1 terminator.	Ampicillin or carbenicillin (100 µg/mL), cultivation at 30 °C	Georgiou et al. [482] Sauer et al.[337]
HM125[d] pFimH-His_6-NT-trc[a]	Expression system for the expression and periplasmic accumulation of $FimH_{trc}$ (with C-terminal His_6 tag), under the control of an IPTG-inducible *trc* promoter and *rrnB* T1 terminator.	Ampicillin or carbenicillin (100 µg/mL), cultivation at 30 °C	This thesis

a) Point mutations for the expression of N136C, R60P and S139C FimH variants were each introduced by site-directed mutagenesis. **b)** The plasmid was kindly provided by Prof. Stefan D. Knight (Uppsala University, Sweden). **c)** The strains were kindly provided by the group of Prof. Dr. Axel Scheidig (Kiel University). **d)** The strain HM125 and the plasmid pFimHs-NT-trc were kindly provided by Prof. Dr. Rudi Glockshuber (ETH Zürich, Switzerland).

5.7.4 Plasmids, vectors and expression constructs

Table 5.7-5: Plasmids and constructs used for FimH-expressing and fimbriae-expressing *E. coli* bacteria.

Plasmid/ construct	Features	Selection/ replication	Literature/ supplier
pPKL174[a]	Encodes the entire *fim* gene cluster (*fimA*, *I*, *C*, *D*, *F*, *G*, *H* and regulatory elements, including *fimB*, *E*, *S*) from *E. coli* K12 PC31, essential for the assembly of type 1 fimbriae. The gene cluster was excised from pPKL4. The cloning vector is currently unknown.	Chloramphenicol (50 µg/mL)	Lindhorst et al. [467]
pPKL4[a]	Encodes the entire *fim* gene cluster (*fimA*, *I*, *C*, *D*, *F*, *G*, *H* and regulatory elements, including *fimB*, *E*, *S*) essential for the assembly of type 1 fimbriae. The 8 kb cluster originates from the *E. coli* K12 PC31 chromosome and was subcloned at the *Hind*III and *Kpn*I sites into pBR322.	Ampicillin or carbenicillin (100 µg/mL)	Klemm et al.[296,297]
pPAP5[b]	Encodes the entire *pap* gene cluster (*papI, B*, *A*, *H*, *C*, *D, J,* K *E*, *F*, *G*) for the regulation and assembly of P fimbriae, terminated by the class I allele of the adhesin *papG.*[305] The 8.5 kb cluster originates from the chromosome of uropathogenic *E. coli* J96 and was subcloned at the *Eco*RI and *Bam*HI sites into pBR322.	Ampicillin or carbenicillin (100 µg/mL)	Lindberg et al. [210]
pETM-G	Expression plasmid for the inducible expression of a modified, GFP-like protein from *Aequorea coerulescens.* The *$_{ac}$gfp* gene was inserted in a pETM-14 vector at the *Xho*I and *Hind*III sites and is under the control of a T7-*lacO* expression system.	Kanamycin (50 µg/mL)	Claudia Fessele, Dissertation[576]
pQE30	Expression vector for the expression of (usually) N-terminally His_6-tagged target proteins under the control of the T5/*lac*O promoter. The vector is prone to (partial) deletion within the promoter sequence[472] and should actually be only used in combination with plasmids	Ampicillin or carbenicillin (100 µg/mL)	Qiagen

	encoding the lac repressor gene *lacI* (such as pREP4) or in expression strains harbouring *lacI* on the F-factor episome.[691,692]		
pKL241[c, d]	Expression plasmid that encodes a $fimH_L$ gene (encoding residue 1-156 and the N-terminal signal peptide) from pPKL4, attached by a C-terminal His_6 tag. The codon of Arg3 within the N-terminal signal sequence was codon-optimised (CGA → CGT). The $fimH_L$ gene was subcloned at the *Eco*RI and *Bam*HI sites into pQE30 and is controlled by a IPTG-inducible T5/*lac*O promoter and λT0 terminator.	Ampicillin or carbenicillin (100 µg/mL)	Schembri et al.[470]
pPKL241 ΔN-sigpep[d]	A 60 bp segment encoding the N-terminal signal peptide of $fimH_L$ was deleted from pKL241 by site-directed mutagenesis in order to circumvent secretion in favour of cytoplasmic accumulation.	Ampicillin or carbenicillin (100 µg/mL)	Sofie Todt (Bachelor thesis)[487]
pFimHs-NT-trc[e]	Expression plasmid that encodes a truncated $fimH_{trc}$ gene (encoding the natural N-terminal signal peptide and $FimH_{1-155}$) from unknown origin. The sequence on pFimHs-NT-trc differs from the K12-derived sequence in 6 silent mutations, as well as in a single mutation within the N-signal sequence, when compared to pPKL241. The $fimH_{trc}$ gene was subcloned at the *Xba*I and *Hind*III into pTrc99A and is controlled by a hybrid trp/lac (*trc*) promoter and *rrnB* T1 terminator. The lac repressor *lacI* is present on the plasmid in order to circumvent toxic effects by target protein overexpression and/or promoter deletion.	Ampicillin or carbenicillin (100 µg/mL)	Sauer et al.[337]
pFimH-His_6-NT-trc[d]	Insertion of Gly156 and His_6 tag at the C-terminal end of the $fimH_{trc}$ gene in pFimHs-NT-trc. The expression product is identical to the protein expressed with pKL241.	Ampicillin or carbenicillin (100 µg/mL)	This thesis
pOSIP-KL-*mcherry*[649, 652]	Transduction plasmid for chromosomal integration of *mcherry* at the λ *att*B site of *E. coli*. The *mcherry* gene was subcloned at the *Eco*RI and *Pst*I sites into the hybrid pOSIP-KL vector and is controlled by a constitutive Antoine-R1 promoter and the rrnB T_2 terminator. After transformation into a recipient strain, the plasmid becomes integrated into the chromosome at the λ *att*B site by λ integrase-catalysed λ *att*B-*att*P recombination. The inserted sequence contains the integration module, comprising genes coding for the λ integrase (*λint*), its temperature-sensitive λ repressor (λcI^{ts}), the kanamycin resistance marker (Kan^R) and the $R6K_\gamma$ ori, flanked by two FRT sites to enable subsequent Flp-mediated excision by FLP-FRT recombination.	Kanamycin (50 µg/mL), replication at 30 °C only in pir^+ (i.e. Pi protein-expressing) strains. The λ integrase becomes expressed at 37 °C due to inactivation of the λ repressor λcI^{ts}.	Addgene (no. 116947)
pE-FLP[649]	Plasmid for the excision of the pOSIP-integration module, catalysed by the Flp recombinase (Flippase). The *flp* gene from pCP20 was inserted into pKD46[693] (contains a temperature-sensitive replicon) under the control of the strong constitutive pE promoter from phage P2.[694] The plasmid can be removed after successful FLP-FRT recombination by cultivation at 37 °C (the pSC101 ori becomes inactive).	Ampicillin (50 µg/mL), replication at 30 °C only in pE-repressing strains.	Addgene (no. 45978)

a) The strain carrying the plasmid was formerly provided by the group of Prof. Per Klemm (Technical University of Denmark). **b)** The plasmid was provided by the group of Prof. Dr. Frank M. Aarestrup (formerly the group of Prof. Per Klemm, Technical University of Denmark). **c)** The plasmid was kindly provided by Prof. Stefan D. Knight (Uppsala University, Sweden). **d)** Point mutations for the expression of N136C, R60P and S139C FimH variants were each introduced by site-directed mutagenesis. **e)** The strain HM125 and the plasmid pFimHs-NT-trc were kindly provided by Prof. Dr. Rudi Glockshuber (ETH Zürich, Switzerland).

5.7.5 Oligonucleotides

Table 5.7-6: Oligonucleotides used for plasmid detection in PCR.

Name	Sequence (5' → 3')	T_m [°C]	T_A [°C]	Targets	Amplicon length [bp]
pQE for	GTA TCA CGA GGC CCT TTC GTC T	62.1	51.8 – 55	pPKL241,	786,
pQE rev	CAT TAC TGG ATC TAT CAA CAG GAG	59.3	51.8 – 55	pPKL241 ΔN-sigpep	725
pTrc99a for	GAG CGG ATA ACA ATT TCA CAC AGG	61	51	pFimHs-NT-trc,	685,
pTrc99a rev	GAT TTA ATC TGT ATC AGG	46.9	51	pFimH-His_6-NT-trc	706
Fimb-FimH for	CAT CAC CTA TAC CTA CAG C	54.5	50	pPKL4,	973,
Fimb-FimH rev	ACC AGC ATT AGC AAT GTC C	54.5	50	pPKL174	973
PAPG-I J96 for[695]	TCG TGC TCA GGT CCG GAA TTT	59.8	54.5	pPAP5	461
PAPG-I J96 rev[695]	TGG CAT CCC CCA ACA TTA TCG	59.8	54.5		
pE-Flp for	GGA ATA AGG GCG ACA CGG AAA TGT	64.8	59	pE-Flp	1652
pE-Flp rev	CCG TTA CAT ATC AAA GGG AAA ACT	63.7	59		

Table 5.7-7: Oligonucleotides used for chromosomal sequence detection in duplex PCR.

Name	Sequence (5' → 3')	T_m [°C]	T_A [°C]	Elongation time
P1_pOSIP-KL-*mcherry*[649]	GGC ATC ACG GCA ATA TAC	53.7	55	2 min
P2_pOSIP-KL-*mcherry*[649]	ACT TAA CGG CTG ACA TGG	53.7	55	2 min
P3_pOSIP-KL-*mcherry*[649]	GGG AAT TAA TTC TTG AAG	54	55	2 min
P4_pOSIP-KL-*mcherry*[649]	TCT GGT CTG GTA GCA ATG	53.7	55	2 min
*mcherry*20-orf_for[652]	CGT ATG AAG GCA CCC AGA	56	55	2 min

Table 5.7-8: Oligonucleotides used for site-directed mutagenesis.

Name	Sequence (5' → 3')	T_m [°C]	T_A [°C]	Elongation time
FimH N136C for[a]	CGA CAG ACC AAC TGC TAT AAC AGC GAT GAT TTC	68.2	60[e]/	3 min 40 s[e]/
FimH N136C rev[a]	GAA ATC ATC GCT GTT ATA GCA GTT GGT CTG TCG	68.2	58[f]	2 min 50 s[f]
FimH S139C for[a]	GAC CAA CAA CTA TAA CTG CGA TGA TTT CCA GTT TG	67.1	60[e]	3 min 40 s[e]
FimH S139C rev[a]	CAA ACT GGA AAT CAT CGC AGT TAT AGT TGT TGG TC	67.1		
FimH R60P for[a]	ATG TCA CAC TGC AAC CAG GCT CGG CTT ATG G	70.8	60[e]	3 min 40 s[e]
FimH R60P rev[a]	CCA TAA GCC GAG CCT GGT TGC AGT GTG ACA T	70.8		
FimH C136N C139S for[a]	GGA AAT CAT CGG AGT TAT AGT TGT TGG TCT GTC GC	69.5	60[e]	3 min 40 s[e]
FimH C136N C139S rev[a]	GCG ACA GAC CAA CAA CTA TAA CTC CGA TGA TTT CC	69.5		

FimH His Tag for[b]	CAC CAT CAC TAA GCT TGG CTG TTT TGG CGG ATG	70.7	63[e]	5 min 30 s [e]
FimH G+His Tag rev[b]	ATG GTG ATG GCC GCC AGT AGG CAC CAC CAC ATC AT	74.2		
C-Insert His Tag for[c]	TAC TGG CGG CCA TCA CCA TCA CCA TCA CTA AGC	72	65[e]	5 min 30 s[e]
C-Insert His Tag rev[c]	GCT TAG TGA TGG TGA TGG TGA TGG CCG CCA GTA	72		
FimH ΔN-sigpep for[d]	ATG TTC GCC TGT AAA ACC GCC AAT GG	64.8	53[e]	3 min 40 s[e]
FimH ΔN-sigpep rev[d]	TAC AAT CTC CTC CTG AAT TCT G	56.5		

a) Primers used for the introduction of point mutations in pPKL241 and pFimHs-NT-trc according to the QickChange™ method (Stratagene). **b)** Primers used for the C-terminal insertion of a Glycine and a His_6 tag into the *fimH*$_L$ sequence on pFimHs-NT-trc. (Insertion was proven to be incomplete). **c)** Primers used for an insertion encoding a cytidine in the incomplete His_6 tag sequence described in footnote **b)**. **d)** Primers used for the 60-bp deletion of the N-terminal signal peptide sequence in pPKL241. **e)** PCR parameters used in combination with Q5® HF DNA polymerase. **f)** PCR parameters used in combination with Phusion® HF DNA polymerase.

Table 5.7-9: Oligonucleotides used for sequencing.

Name	**Sequence (5' → 3')**	**Target**
pQE for	GTA TCA CGA GGC CCT TTC GTC T	pPKL241, pPKL241 ΔN-sigpep
M13rev-49	GAG CGG ATA ACA ATT TCA CAC AGG	pFimHs-NT-trc, pFimH-His_6-NT-trc

Both oligonucleotides are offered as standard sequencing DNA primers on the Eurofins website: https://eurofinsgenomics.eu/de/dna-rna-oligonucleotides/optimised-application-oligos/seqprimer/standard-primer/.

5.7.6 General molecular biological methods

In silico-cloning, construct design and sequence analysis

Design and analysis of constructs, alignments, *in silico*-PCR and cloning strategies as well as evaluation of sequencing results were performed with the free bioinformatics tools Unipro UGENE[696] 50.0 (http://ugene.net/) and SerialCloner 2.6.1 (http://serialbasics.free.fr/Serial_Cloner.html). Sequence graphics were also created with SnapGene 7.2 (https://www.snapgene.com/). Primer design for site-directed mutagenesis was performed by aid of the Agilent QuikChange® Primer Design tool (https://www.agilent.com/store/primerDesignProgram.jsp). Genomic DNA sequences of *Escherichia coli* were received from the BioCyc databases (Encyclopaedia of *Escherichia coli* genes and metabolism), more precisely from the *Escherichia coli* K12 substrain MG1655 reference genome database and the *Escherichia coli* J96 genome database (EcoCyc, https://ecocyc.org/ as well as BioCyc, https://biocyc.org/ GCF_000295775/ organism-summary).

Oligonucleotide synthesis and Sanger sequencing

Eurofins Genomics (https://eurofinsgenomics.eu) was commissioned for Sanger sequencing[697] of plasmids and synthesis of DNA oligonucleotides. Primer characteristics, such as melting temperature T_m and GC content that needed to be considered for PCR and cloning purposes, were also given by the supplier.

Polymerase Chain Reaction (PCR)

Polymerase chain reactions were either performed on a MiniAmp™ Plus Thermocycler (Applied Biosystems™) or on an iCycler Thermal Cycler (Bio-Rad). The ready-to-use 2 x Master Mix (Promega) was utilised for conventional detection and amplification purposes (colony PCR, plasmid detection) according to the manufacturer's recommendations. The respective primers, annealing temperatures T_A and expected lengths of PCR products are compiled in Table 5.7-6. The elongation time was adjusted according to the expected amplicon length under consideration of the turnover rate of the *Taq* DNA polymerase (~ 1 kbp/min). In general, 35 amplification cycles and a reaction volume of 15 µL or 25 µL were chosen. The PCR result was commonly evaluated by agarose gel electrophoresis.

Site-directed mutagenesis

Introduction of point mutations into plasmids, resulting in amino acid exchange, deletions and insertions, was performed in a PCR-based site-directed mutagenesis setup. Point mutations were usually inserted according to the well-known QuikChange™ procedure (Stratagene).[478] The conventional QuikChange™ approach involves the amplification of the entire target plasmid. Thereby, the mutation is introduced by use of a complementary primer-pair thus resulting in a nicked, circular plasmid. In the case of deletions or insertions, however, the primers do not overlap and bind contiguously to the target sequence, therefore producing blunt-ended dsDNA. All primers that were used for this purpose are compiled in Table 5.7-8, together with the respective annealing temperatures T_A and elongation times employed. The reaction was performed by the Q5® High-Fidelity DNA Polymerase (cat. no. M0491, New England Biolabs®) in consideration of the manufacturer's specifications. In general, the Q5 High GC Enhancer was added to the reaction mixture in order to improve the reaction performance and 25 amplification cycles were chosen. Alternatively, point mutations were introduced using the Phusion™ High-Fidelity DNA Polymerase (Thermo Scientific™, cat. no. F530) as specified by the manufacturer's recommendations, as well as under the addition of DMSO (3 % v/v) and the provided Phusion GC buffer (Thermo Scientific™) in a hot-start PCR protocol with 20 amplification cycles.

Subsequent to PCR, the parental m^6A-methylated DNA was digested with FastDigest *Dpn*I (Thermo Scientific™, cat. no. FD1703) for 20 min at 37 °C, following the manufacturer's recommendations for unpurified PCR products. The enzyme was thereafter heat-inactivated for 20 min at 80 °C. Blunt-ended dsDNA was furthermore phosphorylated at the 5'-hydroxyl termini using the T4 Polynucleotide Kinase[698] (New England Biolabs®, cat. no. M0201S, 6 Richardson units in 30 µL volume) under the addition of 1 x T4 DNA Ligase buffer (Promega, cat. no. C126B, Table 5.7-1) for 20 min at 37 °C. The enzyme was likewise inactivated for 20 min at 65 °C. The phosphorylation step enabled the ensuing self-ligation of the resulting 5'-phosphate and 3'-hydroxyl functions by conversion of the T4 DNA ligase[699] (Promega, cat. no. M180A, 3 Weiss units in 31 µL volume) for 3 h at room temperature in 1 x T4 DNA Ligase buffer. Alternatively, the KLD Enzyme Mix (New England Biolabs®, cat. no. M0554S) was used according to the recommended protocol in order to combine *Dpn*I digestion, 5'-phosphorylation and ligation in a single reaction step at room temperature for 15 min. The reaction mixtures (~ 10 µL) were transformed into *E. coli* XL-1 blue (see protocol: Transformation of DNA plasmids in *E. coli*), and the presence of plasmids was proven by colony PCR. PCR-positive clones were isolated and evaluated by Sanger sequencing. If the desired mutation could successfully be verified, the plasmids were transformed into the respective target strains.

Agarose gel electrophoresis

Analytical or preparative segregation of DNA fragments was conducted by agarose gel electrophoresis using a horizontal Sub-Cell® GT Cell system (Bio-Rad) in either a Mini-Sub® Cell GT Cell (12 x 26 x 6.5 cm) or a Wide Mini-Sub Cell® GT Cell (20 x 26 x 2.5 cm) chamber, depending on the number of samples. In general, 1 – 2 % w/v agarose in TAE buffer (Table 5.7-1) was used and the solution was carefully heated in the microwave to achieve the dissolution of the agarose. Addition of 1 – 5 µL MIDORI Green Advance stain (NIPPON Genetics) after cooling of the agarose solution facilitated the subsequent visualisation of nucleic acids under a blue light transilluminator (FastGene® FAS-DIGI Imaging System, NIPPON Genetics; Dual LED Blue/White Light Transilluminator, Invitrogen™). Prior to application, the DNA containing samples were mixed with 1 x DNA Loading Dye (Thermo Scientific™, cat. no. R0611, Table 5.7-1). A suited DNA Ladder (GeneRuler 1 kb DNA Ladder, Thermo Scientific™, cat. no. SM0311; GeneRuler 1 kb Plus DNA Ladder, Thermo Scientific™, cat. no. SM1331) was used as a size reference according to the manufacturer's recommendations. Electrophoresis was performed at constant voltage (U = 100 V) until the desired fragment separation was achieved.

Plasmid preparation

Bacterial suspensions were cultured in sterilised test tubes with LB medium (Table 5.7-1; volume between 5 – 15 mL, depending on the copy number per cell), under vigorous shaking (175 – 200 rpm) in an orbital shaker (Shaking Incubator 3031, GFL). The selection conditions and respective cultivation temperatures for the replication of each plasmid were chosen according to Table 5.7-5. The bacterial suspensions were inoculated preferably from a single colony after growth on LB agar or from a glycerol stock. Isolation of plasmids from the bacterial suspensions was performed using either the NucleoSpin® Plasmid Mini Kit (Macherey-Nagel™) or the GeneJET™ Plasmid Miniprep Kit (Thermo Scientific™) following the manufacturers' recommendations for high-copy (pOSIP-KL-*mcherry*, pE-Flp, pQE30, pPKL24, pFimHs-NT-trc and their derivatives) and low-copy (pPKL4, pPKL174, pPAP5) plasmids.

Ethanol precipitation of plasmids

Plasmids that were designated for a subsequent electroporation were eluted in nuclease-free water instead of the provided elution buffer, as the last recovery-step of the plasmid preparation protocol. Alternatively, the plasmids were precipitated in order to reduce residual salts.

For this purpose, the plasmid solution was mixed with Na-acetate buffer (Table 5.7-1) in a 10:1 ratio. 2.5 volumes of ice-cold, non-denatured 100 % ethanol were added and the mixture was frozen in liquid nitrogen. The mixture was centrifuged at maximal speed (5000 rpm) at 4 °C for 5 min (Universal 320R centrifuge, Hettich®) and the supernatant was carefully decanted. 1 mL of ice-cold, non-denatured 80 % ethanol was added, the tube was inverted and again centrifuged at maximal speed, 4 °C for 5 min and the supernatant was removed. The precipitate was once more centrifuged, decanted and dried by incubation with open lid at 300 rpm, 37 °C for 10 min (MATRIX orbital thermoshaker, IKA). The precipitate was dissolved in a desired volume of nuclease-free water and frozen at -20 °C.

Preparation of competent *E. coli*

A pre-culture in LB medium (10 mL, Table 5.7-1) with selective antibiotics added, was inoculated from an isolated, well-defined colony of the desired *E. coli* strain (which in turn was taken from a beforehand cultured LB agar plate). The pre-cultures were then incubated overnight under vigorous shaking (175 rpm, Shaking Incubator 3031, GFL) in sterilised test tubes. The respective growth temperatures and selection conditions are provided in Table 5.7-2. Fresh LB medium (50 mL, with antibiotics) was inoculated with the pre-culture and the optical density was adjusted to OD_{600} = 0.05 – 0.1. The suspension was cultured in a 100 mL culture bottle at the respective temperature under moderate agitation (100 – 120 rpm) until the OD_{600} = 0.4 – 0.6 was reached. The following steps were performed on ice and with ice-cold solutions of either sterilised $CaCl_2$ (60 mM), glycerol (15 % v/v) to obtain $CaCl_2$-competent cells, or alternatively, of sterilised glycerol (10 % v/v) for the preparation of electro-competent cells. The suspensions were cooled on ice and centrifuged at 3000 rpm, 4 °C for 15 min (Universal 320R centrifuge, Hettich®). The supernatant was removed, the residue was suspended in 10 mL of the respective glycerol solution and centrifuged once again at 3000 rpm, 4 °C for 15 min. The washing step was repeated twice and the residue was resuspended in 1 mL of the respective glycerol solution. The suspensions were aliquoted in 100 µL volumes, frozen in liquid nitrogen and stored at -20 °C.

Transformation of DNA plasmids in *E. coli*

Transformation of plasmids was either achieved chemically, combined with a short heat-shock ($CaCl_2$ method[700], TSS method[690]) or by electroporation[701] using a Gene Pulser® II Electroporation System (Bio-Rad). To recover plasmids after cloning or site-directed mutagenesis, the endonuclease-deficient *E. coli* XL-1 Blue was transformed following the $CaCl_2$ transformation protocol. If the desired cloning result or mutation could be verified, the plasmids were transformed into the respective target strains (Table 5.7-2). $CaCl_2$ competent cells were thawed on ice for ∼ 10 min. The isolated plasmid DNA (1 – 2 µL) or the reaction mixture after enzymatic ligation using T4 DNA ligase subsequent to cloning or site-directed mutagenesis (10 – 20 µL) was carefully added and incubated on ice for 20 – 30 min. The bacterial suspension was shortly heated at 42 °C (60 – 90 s) in an orbital shaker and cooled on ice for 2 min. Regular LB medium or SOC outgrowth medium optimised for transformation (500 µL, see Table 5.7-1) was added and the suspension was incubated for 1 h at 37 °C (or 30 °C) under vigorous shaking at 600 rpm (MATRIX orbital thermoshaker, IKA or ThermoMixer®, Eppendorf®). The suspension was centrifuged at ∼ 10 000 rpm for 5 min (Universal 320R centrifuge, Hettich®) and 500 µL of the supernatant was removed. The pelleted cells were resuspended in the remaining supernatant, applied onto 1 or 2 LB agar plates (with the selective antibiotics appended) and cultured overnight at the respective temperature given in Table 5.7-5. Some colonies were selected and restreaked on a fresh LB agar plate, in order to ensure that only single colonies had been isolated.

If the chemical transformation attempt was unsuccessful or the DNA uptake was supposedly more sophisticated due to large plasmid sizes (pPKL174, pPKL4 ~ 11.7 kb, pPAP5 ~ 13.6 kb) the more efficient electroporation was chosen.[701] Electro-competent cells were carefully thawed on ice, plasmid DNA (2 – 6 µL) was added and the mixture (45 µL) was transferred into an ice-cold Gene Pulser Electroporation Cuvette (0.1 cm gap, Bio-Rad). The cells were electroshocked at 1800 V, 200 Ω, 25 µF for 3 – 5 ms. LB medium or SOC outgrowth medium (300 µL, Table 5.7-1) was added and the cells were cultured at 37 °C for 1 – 2.5 h under consideration of the plasmid size (1 h/10 kb). As described for the $CaCl_2$ method, the cells were centrifuged, resuspended in residual supernatant (100 µL) and applied onto LB 1 – 2 agar plates (with the selective antibiotics appended). The cells were cultured overnight at the respective temperature given in Table 5.7-5. Eventually, some selected colonies were restreaked out onto a fresh LB agar plate and incubated once again overnight. The preparation of TSS-competent cells and the TSS transformation method is described in Chapter 5.9.

Determination of DNA concentration

Quantification of DNA samples was performed on a NanoPhotometer® P300 (Implen) by measurement of the characteristic UV/vis absorbance at 260 nm. The purity of the DNA samples was assessed based on the absorbance ratios at 260/230 and 260/280 nm, respectively.

5.7.7 General protein-biochemical methods

Bioinformatics tools

Protein parameters, such as molecular weight (M_w), theoretical isoelectric points (pI) and molar extinction coefficients (ε), were calculated by using the web-based ExPaSy ProtParam tool (https://web.expasy.org/protparam/). Monoisotopic masses were computed with the aid of the Peptide Synthetics Peptide Mass Calculator (https://www.peptidesynthetics.co.uk/tools/). Spectral data and further properties of fluorescent proteins including GFP_{mut3}, mOrange and mCherry were obtained from the open-source fluorescent protein database FPbase (https://www.fpbase.org/).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)[702]

Reducing and non-reducing SDS-polyacrylamide gel electrophoreses were performed using a Mini-PROTEAN® Tetra Cell System (Bio-Rad) in 15 % SDS-acrylamide handcast mini gels. The gels were either prepared with 1 mm spacer plates in combination with 10 well-combs or alternatively, 0.75 mm thick gels with 15 wells were used. The composition for 15 % SDS-polyacrylamide gels is given in Table 5.7-10:

Table 5.7-10: Pipetting scheme for the preparation of 15 % SDS-polyacrylamide gels.[a]

Component	SDS-Resolving gel (15 %)	SDS-Stacking Gel (5 %)
ROTIPHORESE® NF-Acrylamide/Bis solution[b]	2. 5 mL	170 µL
SDS buffer	SDS resolving gel buffer[b]: 1.2 mL	SDS stacking gel buffer[b]: 130 µL
APS (10 %)	100 µL	20 µL
TEMED	5 µL	2 µL
ddH_2O	1.2 mL	690 µL

a) The composition is calculated for the preparation of a single SDS-polyacrylamide gel (83 x 73 x 1 mm). **b)** See Table 5.7-1 for more detailed compositions of buffers and solutions.

Reducing SDS-PAGE was performed as follows: Samples with 5 – 20 µL volume (depending on the protein concentration) in 1.5 mL tubes were mixed with reducing Laemmli buffer (Table 5.7-1) approximately in a 4:1 ratio and denatured at 95 °C for 5 min. The tubes were shortly centrifuged (Galaxy Mini centrifuge, VWR™) and the protein samples were subsequently applied onto the stacking gel. A suitable protein ladder (e.g. Pierce™ Unstained Protein M_W Marker, Thermo Scientific™, cat. no. 26610; PageRuler™ Prestained Protein Ladder, Thermo Scientific™, cat. no. 26619) was used for subsequent size assignment of the separated protein species. For non-reducing SDS-PAGE, the protein samples were not denatured by heating and were mixed with non-reducing Laemmli buffer (Table 5.7-1). In general, SDS-PAGE was performed in SDS running buffer (Table 5.7-1) at constant current (I = 25 mA/gel) for approx. 1 h and/or until the migration of the bromophenol blue stain appeared sufficiently advanced.

Coomassie staining[703,704]

Protein species in SDS-polyacrylamide gels were stained in a solution of Coomassie Brilliant Blue R-250 (Table 5.7-1, ~ 50 mL/gel), generally overnight at room temperature and under gentle agitation. The gels were washed and subsequently incubated for several hours in destaining solution (Table 5.7-1) at room temperature under gentle agitation. A regular exchange of destaining solution or short boiling of the latter in the microwave has been occasionally applied in order to accelerate the destaining process.

Determination of protein concentration

After expression and purification of FimH variants, the protein concentration was generally measured on a NanoPhotometer® P300 (Implen) at 280 nm. The required molar extinction coefficients (Table 5.7-11) were calculated for each FimH variant using the ExPaSy ProtParam tool by specifying the corresponding amino acid sequences (provided in the appendix). In some cases, in particular after chemical labelling (see Chapter 5.7.7) or if the protein samples were tested in the FITC-*p*AP mannoside binding assay (Chapter 5.7.9), the protein concentrations were determined using a Pierce™ bicinchoninic acid (BCA)[705] Protein Assay Kit (Thermo Scientific™, cat. no. 23225) according to the supplier's specifications.

Table 5.7-11: Molar extinction coefficients used for the determination of protein concentration.

Variant/Construct	**Calculated molar extinction coefficient [e] [M^{-1} cm^{-1}]**
$FimH_L$-His_6[a] / $FimH_{trc}$-His_6[b]	1.458[f] / 1.451[g]
$FimH_L$-His_6 N136C[a] / $FimH_{trc}$-His_6 N136C[b]	1.459[f] / 1.452[g]
$FimH_L$-His_6 S139C[a] / $FimH_{trc}$-His_6 S139C[b]	1.457[f] / 1.402[g]
$FimH_L$-His_6 N136C ΔN-sigpep[a,c]	1.449[f] / 1.442[g]
$FimH_{trc}$-His_6 R60P[b]	1.463[f] / 1.456[g]
$FimH_{trc}$[d]	1.535[f] / 1.528[g]

a) $FimH_L$-His_6 variants encoded on pPKL241; **b)** $FimH_{trc}$-His_6 variants encoded on pFimH-His_6-NT-trc; **c)** the target protein exhibits an additional methionine as first amino acid; **d)** $FimH_{trc}$ encoded on pFimHs-NT-trc, **e)** molar extinction coefficients at 280 nm in water with a protein concentration of 1 g/L, calculated with the ExPaSy ProtParam tool (https://web.expasy.org/protparam/); **f)** assuming all pairs of Cys residues are oxidised; **g)** assuming all Cys residues are reduced.

5.7.8 General protocol of adhesion experiments under competitive and non-competitive conditions

Cultivation and preparation of *E. coli* in bacterial adhesion and adhesion inhibition experiments

LB medium (10 – 20 mL, Table 5.7-1) with appropriate antibiotics was inoculated with fimbriae-expressing *E. coli* strains from glycerol stocks and cultured in test tubes overnight at 37 °C, and under vigorous agitation (175 – 200 rpm) in an orbital shaker (Shaking Incubator 3031, GFL). Adhesive *E. coli* strains and the respective selection conditions are compiled in Table 5.7-3. If the bacterial suspension appeared dense and in late stationary or early decline phase on the following day, the same volume of fresh LB medium (with antibiotics) was added and cultured at 37 °C under moderate agitation (100 rpm) until harvest. The cells were centrifuged at 5000 rpm for 10 min at room temperature (Universal 320R centrifuge, Hettich®) and the supernatant was decanted. The pelleted cells were resuspended in 2 mL PBS (Table 5.7-1), centrifuged at 5000 rpm for 10 min and the supernatant was removed. The bacterial density was adjusted with PBS according to the experimental setup, the bacterial fluorescence intensity (if present) and/or the respective plasmids encoding the fimbrial cluster. The green-fluorescent strain SAR18 PKL1162 (harbouring the type 1 fimbriae-encoding plasmid pPKL174), which is most commonly used in standard adhesion inhibition experiments, was adjusted to OD_{600} = 0.4 (2 mg/mL). Strains carrying the plasmid pPKL4, likewise coding for type 1 fimbriae, were generally used at OD_{600} = 0.8 (4 mg/mL) unless otherwise stated.

Polysaccharide coating of microwell plates for adhesion inhibition experiments

Bacterial adhesion inhibition experiments were usually performed in black 96-microwell plates (MaxiSorp™ F96, Nunc™ Thermo Scientific™), which were coated with mannan from *Saccharomyces cerevisiae*. The mannan coating, originally introduced by Sokurenko et al.,[508] was performed according to a modified procedure by the Lindhorst group.[467] Mannan from *S. cerevisia*e (Sigma-Aldrich, 1.2 mg/mL) in carbonate buffer I (Table 5.7-1) was applied to black microwell plates (120 µL/well) and the coating solution was allowed to air-dry by incubation overnight at 37 °C, 175 rpm in an orbital shaker (Shaking Incubator 3031, GFL). The plates were washed with PBST (Table 5.7-1; 3 x 400 µL/well) in a microplate washer (Tecan HydroFlex™) and blocked with 1 % w/v polyvinyl alcohol (PVA) in PBS (Table 5.7-1; 120 µL/well) at 37 °C, 100 rpm for 2 h with the lid closed in an orbital shaker. The blocking solution was removed and the plates were washed again with PBS (3 x 400 µL/well) using a microplate washer.

Competitive adhesion inhibition experiments

The inhibitory potency of α-D-mannoside derivatives was tested in competitive adhesion inhibition experiments, following a modified procedure from Lindhorst and colleagues.[467] The inhibitor solutions in PBS were applied onto mannan-coated, PVA-blocked microwell plates (50 µL/well) in duplicate, triplicate or quadruplicate serial 1:2-dilutions. Two columns of the plate contained only PBS (2 × 8 wells) to allow to allow subsequent baseline fluorescence determination (negative control) as well as the maximum bacterial adhesion level in the absence of the inhibitor.

Suspensions of fluorescent, type 1 fimbriae-expressing *E. coli* in PBS (PKL1162, OD_{600} = 0.4; HB101-mCherry^{Flp} pPKL4, OD_{600} = 0.8) were added (50 µL/well) and the plates were incubated at 37 °C, 100 rpm for 45 min in an orbital shaker (Shaking Incubator 3031, GFL). The plates were washed with PBS in a microplate washer (3 x 400 µL/well, Tecan HydroFlex™) and refilled with PBS (100 µL/well) for the subsequent quantification of adhering bacteria via fluorescence readout.

Annotation: The *fimH* gene originating from *E. coli* K12 PC31 on the type-1 fimbriae encoding plasmids pPKL4 and pPKL241 contains the pathoadaptive A27V mutation, which tends to undergo activation upon receptor binding in the absence of shear.[393,425] It can therefore be assumed that, under the conditions used in the adhesion inhibition experiment, a significant population of the fimbriae-associated adhesins presumably adopts the high-affinity conformation, even at the moderate agitation employed (100 rpm).

Polysaccharide coating of microwell plates for adhesion experiments

In adhesion experiments under non-competitive conditions (i.e. without inhibitors), usually the concentration of the polysaccharide adsorbed on the polystyrene surface of the microplates was varied instead of the inhibitor concentration. For that purpose, black 96 microwell plates (MaxiSorp™ F96, Nunc™ Thermo Scientific™) were coated with serial dilutions of a suited polysaccharide (e.g. mannan, galactomannan) dissolved in carbonate buffer I (120 µL/well, Table 5.7-1). Wells in which only carbonate buffer was added later served as controls for nonspecific bacterial binding and to determine the baseline fluorescence of the microplate. The plates were incubated uncovered overnight at 37 °C, 175 rpm in an orbital shaker (Shaking Incubator 3031, GFL). As described above, the plates were washed with PBST in a microplate washer (3 x 400 µL/well, Tecan HydroFlex™), blocked with 1 % w/v PVA in PBS (120 µL/well, incubation for 2 h at 37 °C, 100 rpm) and were washed again with PBS (3 x 400 µL/well, see protocol: Polysaccharide coating of microwell plates for adhesion inhibition experiments).

Adhesion experiments under non-competitive conditions

Suspensions of type 1- or P-fimbriated bacteria in PBS (100 µL/well) were applied onto polysaccharide-coated and PVA-blocked black 96 microwell plates. The optical densities at 600 nm (OD_{600}) of the respective strains were adjusted according to Table 5.7-12. Wells in which only PBS (100 µL/well) was added served as controls for the calculation of the baseline fluorescence. The plates were incubated at 37 °C, 100 rpm for 45 min (Shaking Incubator 3031, GFL) and washed with PBS in a microplate washer (3 x 400 µL/well, Tecan HydroFlex™).

Table 5.7-12: Optical densities at 600 nm (OD_{600}) used in adhesion experiments under non-competitive conditions.

Fimbriae-expressing *E. coli* strain	**Optical density**
PKL1162[a]	0.2
HB101-mCherry^{Flp} pPKL4[a]	0.4
HB101-mCherry^{Flp} pPAP5[a]	0.4
VR123-G[a,b]	0.3

a) The suspensions were selected, cultured and prepared according to Table 5.7-3 (see protocol: Cultivation and preparation of *E. coli* in bacterial adhesion and adhesion inhibition experiments). **b)** The expression of GFP in *E. coli* VR123-G was induced with 1 mM IPTG at OD_{600} = 0.5 and cultured at 37 °C for 4 h, 200 rpm,[576] prior to adjustion to OD_{600} = 0.2 and use in bacterial adhesion experiments.

Fluorescence readout

The bacterial fluorescence was quantified using an Infinite® 200 Pro M Nano+ dual mode microplate reader (Tecan) at λ = 483/535 nm (excitation/emission) for the detection of GFP-expressing strains (PKL1162 and VR123-G). Red-fluorescent, fimbriae-expressing HB101-mCherry^{Flp} were detected at $\lambda_{ex/em}$ = 562/640 nm, unless otherwise stated.

Evaluation of adhesion inhibition experiments

The mean fluorescence signal of the negative control (PBS, 50 µL/well in octuplicates, no fluorescent bacteria were added) was subtracted from all other fluorescence signals (f). The mean fluorescence signal in the absence of inhibitors (bacterial suspension, 50 µL/well and PBS, 50 µL/well in octuplicates) served as fluorescence signal corresponding to maximal adhesion ($f_{max.\ adhesion}$) for the calculation of percentage inhibition (I [%]), according to the following formula:

$$I\,[\%] = 100 \cdot \left(\frac{f_{max.\ adhesion} - f_{measured}}{f_{max.\ adhesion}}\right)$$

The resulting percentage inhibition values were semi-logarithmically plotted as a function of the inhibitor concentration and fitted according to the Hill1-equation, with the asymptotes preferentially fixed at 0 and 100 %, respectively. The half-maximal inhibitory concentration (IC_{50}) is given as the inflection point of the sigmoidal Hill1 curve. In order to improve the comparability of individual adhesion inhibition experiments, the IC_{50} of all FimH inhibitors was referenced against methyl α-D-mannoside (MeMan) on the same microwell plate. The relative inhibitory potencies (RIP) were calculated from IC_{50} values as follows:

$$RIP = \frac{IC_{50}\ (\text{MeMan})}{IC_{50}\ (\text{Inhibitor})}$$

Toxicity test of FimH inhibitors

In order to preclude that an observed inhibitory behaviour of a given FimH inhibitor may be attributed to toxic effects, the viability of *E. coli* was tested after incubation with different concentrations of the inhibitor. Therefore, after 45 min incubation at 37 °C and under agitation at 100 rpm (Shaking Incubator 3031, GFL) of the respective mixtures of bacteria and inhibitors, the latter were removed from the microplate and centrifuged for 5 min at 10 000 rpm (Universal 320R centrifuge, Hettich®). The supernatant was removed and the pelleted cells were resuspended in 50 µL PBS. The suspensions were plated on LB agar with antibiotics (see Table 5.7-3) and cultured at 37 °C overnight. On the following day, the colony-forming units (CFU) were counted and compared to a control experiment, in which no inhibitor was added. A significant reduction in CFU number, which would strongly indicate a toxic effect, was not observed among the monovalent inhibitors presented in this thesis.

5.8 Supplemental information to Chapter 3: "Photocontrol of the bacterial lectin FimH"

5.8.1 Repair of pPKL241: Renewed subcloning of the *fimH* gene in pQE30

The expression plasmid pPKL241,[470] which was formerly kindly provided by Prof. Stefan Knight (Uppsala University, Sweden), unintentionally contained an extensive deletion in the hybrid T5/*lac*O promoter region. This deletion prevents recognition by the RNA polymerase and thus, the expression of the target gene. The deleted sequence, which is frequently observed in pQE30-based constructs,[472] thus had to be replaced by reintegration of the *fimH* gene from the deficient plasmid into the intact pQE30 vector.

Enzymatic digestion of DNA with restriction enzymes

The undamaged target vector pQE30 was obtained from a storage strain (*E. coli* XL-1 pQE30) and prepared as described in Chapter 5.7.6 (see details in protocol: Plasmid preparation). About 10 µg of the isolated target vector as well as the T5/*lac*O promoter-deficient pPKL241 (and pPKL241 N136C, pPKL241 S139C, coding for the respective cysteine variants) were digested with the restriction enzymes BamHI High-Fidelity (New England Biolabs®, cat. no. R3136, 40 U) and *Eco*RI High-Fidelity (New England Biolabs®, cat. no. R3101, 40 U) in 1 x CutSmart® buffer (New England Biolabs, Table 5.7-1), 50 µL reaction volume at 37 °C, 600 rpm for 1 h (MATRIX orbital thermoshaker, IKA or ThermoMixer®, Eppendorf®). The restriction enzymes were subsequently inactivated at 70 °C for 5 min. The success of the double digestion was evaluated by agarose gel electrophoresis (Figures 5.7-1 – 5.7-2) to confirm the expected fragment sizes given in Table 5.8-1. The 584-bp *fimH*$_L$-His$_6$ fragments, encoding the K12-derived FimH$_L$ as well as the cysteine variants N136C and S139C that resulted from the digestion of the respective pPKL241-constructs, and the unimpaired 3404-bp pQE30 backbone fragment were excised from the agarose gel. The linearised DNA was extracted from the gel pieces using the NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel™) according to the manufacturer's recommendations.

Table 5.8-1: Expected DNA fragments after double digestion of pQE30 and pPKL241 with restriction enzymes EcoRI and BamHI.

Plasmid/vector	**Expected fragment lengths [bp]**	
	***Eco*RI → *Bam*HI**	***Bam*HI → *Eco*RI**
pPKL241[a] (intact)	584 (*fimH*$_L$-His$_6$ fragments)	3404 (pQE30 backbone)
pPKL241[a] (T5/*lac*O-deficient)[b]	584 (*fimH*$_L$-His$_6$ fragments)	~ 3372
pQE30	57	3404

a) Plasmid variants of pPKL241, encoding the *E. coli* K12 derived FimH$_L$-His$_6$, including the cysteine variants N136C and S139C. **b)** Calculated fragment sizes for pPKL241 with deletion of a 32-bp segment in the T5/*lac*O promoter region.

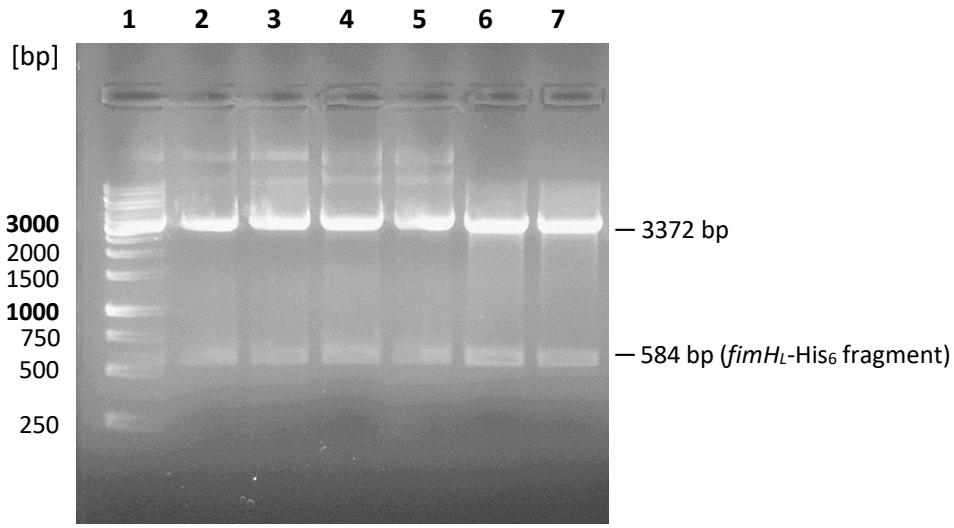


Figure 5.8-1: Agarose gel to evaluate the success of the double digestion of T5/*lac*O promoter-deficient pPKL241-constructs with the restriction enzymes *Eco*RI and *Bam*HI.

1: GeneRuler 1 kb DNA Ladder (cat. no. no. SM0311, Thermo Scientific™); **2** – **7**: digestion attempts with *Eco*RI and *Bam*HI; **2** – **3**: T5/*lac*O-deficient pPKL241 N136C; **4** – **5**: T5/*lac*O-deficient pPKL241 S139C; **6** – **7**: T5/*lac*O-deficient pPKL241. Agarose gel (1.5 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 30 min. The expected fragment lengths for a *Eco*RI and *Bam*HI double-digestion are given in

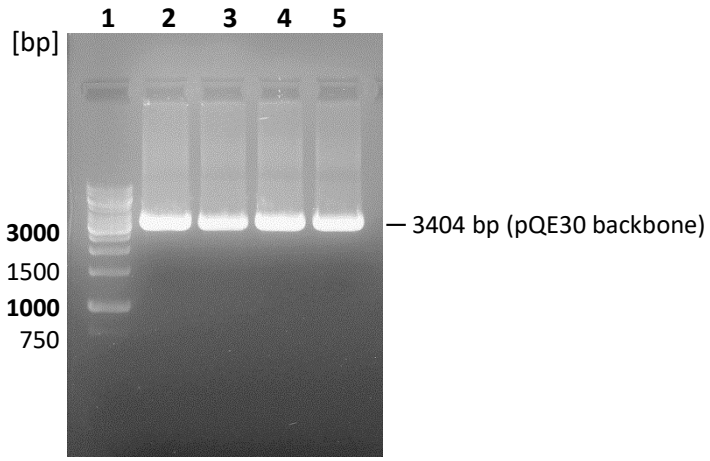


Figure 5.8-2: Agarose gel to evaluate the success of the double digestion of target vector pQE30 with the restriction enzymes *Eco*RI and *Bam*HI.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **5**: digestion attempts of pQE30 with *Eco*RI and *Bam*HI. Agarose gel (1.5 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 35 min. The expected fragment lengths for a *Eco*RI and *Bam*HI double-digestion are given in Table 5.8-1.

Dephosphorylation

The terminal 3´and 5'-phosphate groups were removed from the digested and purified vector pQE30 in order to prevent self-ligation. Dephosphorylation was achieved using FastAP™ Thermosensitive Alkaline Phosphatase (Thermo Scientific™, cat. no. EF0651, 1 U) in 1 x FastAP™ Reaction buffer (Thermo Scientific™, Table 5.7-1), 20 µL reaction volume at 37 °C for 15 min. The enzyme was subsequently inactivated at 75 °C for 5 min.

DNA Ligation

The ligation of the dephosphorylated vector and the different *fimH*$_L$-His$_6$ fragments (100 ng each) was catalysed by T4 DNA Ligase[699] (Promega, cat. no. M180A, 1 Weiss unit) in 1 x T4 DNA Ligase buffer (Promega, cat. no. C126B, Table 5.7-1) with 10 µL reaction volume. The ligation mixture was incubated at room temperature overnight under moderate agitation (300 rpm) in a thermal shaker (ThermoMixer®, Eppendorf®). The ligation mixture was finally transformed into *E. coli* XL-1 according to the $CaCl_2$ method described in Chapter 5.7.6 (see protocol: Transformation of DNA plasmids in *E. coli*).

Evaluation by colony PCR and Sanger sequencing

The obtained clones were tested for transformation with successfully re-assembled pPKL241 plasmids by colony PCR as described in Chapter 5.7.6 (see protocol: Polymerase Chain Reaction (PCR)), using the pQE30 for/rev primer pair (Table 5.7-6). The result of the colony PCR was evaluated by agarose gel electrophoresis (Figure 5.8-3). PCR-positive plasmids were further analysed by Sanger sequencing (Eurofins Genomics), which confirmed successful repair of the plasmids pPKL241 and pPKL241 S139C. In the respective plasmid for the expression of the variant FimH N136C, however, the partial deletion of the T5/*lac*O promoter had reoccurred. The corresponding plasmid was thus produced by site-directed mutagenesis from the reconstructed precursor plasmid pPKL241 as described in the following Chapter 5.8.2.

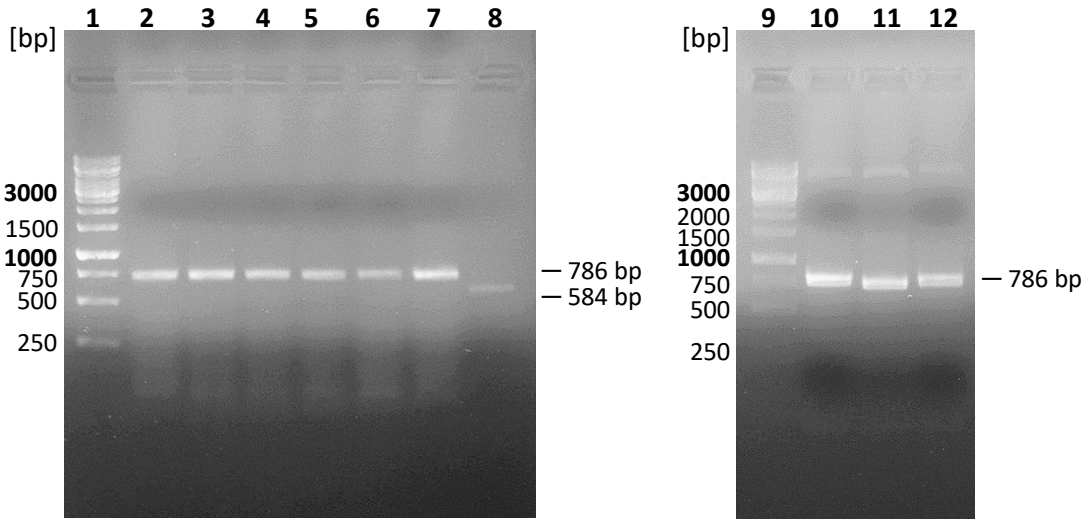


Figure 5.8-3: Agarose gels after colony PCR to demonstrate successful re-integration of the *fimH*$_L$-His$_6$ fragments in the pQE30 backbone.

1, 9: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **4**: attempts for colony PCR after transformation with pPKL241 N136C; **5** – **7**: attempts for colony PCR after transformation with pPKL241 S139C; **8:** Unligated *fimH*$_L$-His$_6$ fragment (control, 584 bp); **10** – **12**: attempts for colony PCR after transformation with pPKL241. Agarose gels (1.5 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 30 min. The presence of a PCR fragment with a length of 786 bp indicates a successful re-assembly of the pPKL241 plasmids.

5.8.2 Site-directed mutagenesis of pPKL241 and pFimH-His$_6$-NT-trc

All further expression plasmids, such as the missing plasmids for the expression of the cysteine variant N136C and the proline variant R60P of FimH, were obtained by site-directed mutagenesis of pPKL241 as described in Chapter 5.7.6 (see protocol: Site-directed mutagenesis). For this purpose, the corresponding primer pairs, annealing temperatures, and elongation times, provided in Table 5.7-8, were employed. The deletion of the natural N-terminal signal peptide within the *fimH*$_L$ gene on pPKL241 N136C was performed by Sofie Todt as part of her Bachelor thesis in order to obtain plasmid pPKL241 N136C ΔN-sigpep.[487] Deletion of the N-terminal signal peptide was achieved with primer pair FimH ΔN-signal for/rev. The approach also included the insertion, as well as the conservation of an N-terminal start codon (AUG, coding for *N*-formylmethionine, ^{f}Met). Due to the recurring problem with the reappearance of the T5/lacO promoter deletion in all pQE30-based plasmids, the expression plasmid was changed in the later course of this research. The literature-unknown plasmid pFimHs-NT-trc was kindly provided by Prof. Rudi Glockshuber (ETH Zürich, Switzerland) and is equipped with the repressor *lacI* and a less strong bacterial trc-promoter (trp/lac hybrid) compared to the viral T5/lacO promoter, making a recurrence of a promoter deletion unlikely. To avoid confusion between the respective truncated *fimH*(-His$_6$) constructs on both plasmids and their derivatives in this thesis, the corresponding K12-derived gene (and its expression product) on pPKL241 is referred to as *fimH*$_L$ (FimH$_L$), whereas the respective gene on pFimHs-NT-trc is designated as *fimH*$_{trc}$ (FimH$_{trc}$).

An attempt was made to insert the sequence coding for a C-terminal His$_6$ tag and Gly156 on the plasmid pFimHs-NT-trc using the primer pair FimH G+His Tag for/rev. A single cytidine was missing in the resulting sequence, which was subsequently inserted using primer pair C-Insert His Tag for/rev. The cysteine and proline point mutations were introduced in the *fimH*$_{trc}$ gene on the resulting plasmid pFimH-His$_6$-NT-trc in the same manner as described for pPKL241. The respective expression products from pPKL241 and pFimH-His$_6$-NT-trc are identical since all mutations within the *fimH*$_{trc}$ gene from the precursor plasmid pFimHs-NT-trc remain silent. A sequence alignment of the *fimH* genes on pPKL241 and pFimH-His$_6$-NT-trc as well as DNA sequences of miscellaneous plasmids used in this thesis are provided in the appendix. The completed and sequenced expression plasmids were transformed into the corresponding target strains as described in Chapter 5.7.6 (see protocol: Transformation of DNA plasmids in *E. coli*). The respective expression systems for FimH expression used in the course of the research are compiled in Table 5.7-4.

5.8.3 Bacterial adhesion experiments with *N*-acetyl-protected azobenzene mannosides

Adhesion inhibition experiments with UV exposure

The influence of *trans* → *cis* photoisomerisation of the *N*-acetyl-protected azobenzene mannosides on bacterial binding was investigated in adhesion inhibition experiments with previous UV irradiation. The coating of black 96-well microplates with mannan from *S. cerevisiae*, the subsequent blocking with 1% w/v PVA in PBS and the cultivation and preparation of GFP- and type 1 fimbriae-expressing *E. coli* PKL1162 was performed as described in Chapter 5.7.8 (see protocols Cultivation and preparation of *E. coli* in bacterial adhesion and adhesion inhibition experiments and Polysaccharide coating of microwell plates for adhesion inhibition experiments).

Solutions of the azobenzene mannoside derivatives used in Chapter 3.4 were relaxed overnight at 45 °C under exclusion of light. To improve the solubility of the moderately water-soluble inhibitors, these were dissolved in PBS/MeOH, 9:1 and equilibrated at 37 °C for 5 min. The inhibitors were applied onto the microplates in two triplicate 1:2 dilution series (50 µL/well in PBS/MeOH, 9:1). The reference inhibitor methyl *α*-D-mannopyranoside was applied as duplicate dilution series. A triplicate of the photoswitchable inhibitor as well as one row with the reference inhibitor was wrapped in aluminium foil to prevent UV exposure. The rest of the microplate was irradiated with a UV lamp at 365 nm (LED NC4U133AT, Nichia Corporation) at 50 % intensity for 10 min. The bacterial suspension of SAR18 PKL1162 (OD_{600} = 0.4 in PBS, 50 µL/well) was added immediately after UV exposure. The microplates were covered with aluminium foil and incubated at 37 °C, 100 rpm for typically 15 min (Shaking Incubator 3031, GFL). Longer incubation times were also tested (30 and 45 min) and frequently resulted in higher quality of the dose-response curves, but this was partly compromised by progression of the relaxation process. The microplates were washed with PBS in a microplate washer (3 x 400 µL/well, Tecan HydroFlex™), subsequently filled with PBS (100 µL/well) and fluorescence readout was performed at $\lambda_{ex/em}$ = 483/535 nm. The calculation of IC_{50} and RIP was performed according to Chapter 5.7.8 (see protocol: Evaluation of adhesion inhibition experiments). In addition, the quotient of the IC_{50} values after/without UV exposure was calculated:

$$IC_{50}\,\text{ratio} = \frac{IC_{50}\ \text{not irradiated}}{IC_{50}\ \text{irradiated}}$$

An IC_{50} ratio < 1 indicates a decrease of the antiadhesive effect upon UV exposure and/or *trans* → *cis* photoisomerisation, as it is expected for *cis*-azobenzene derivatives.[410]

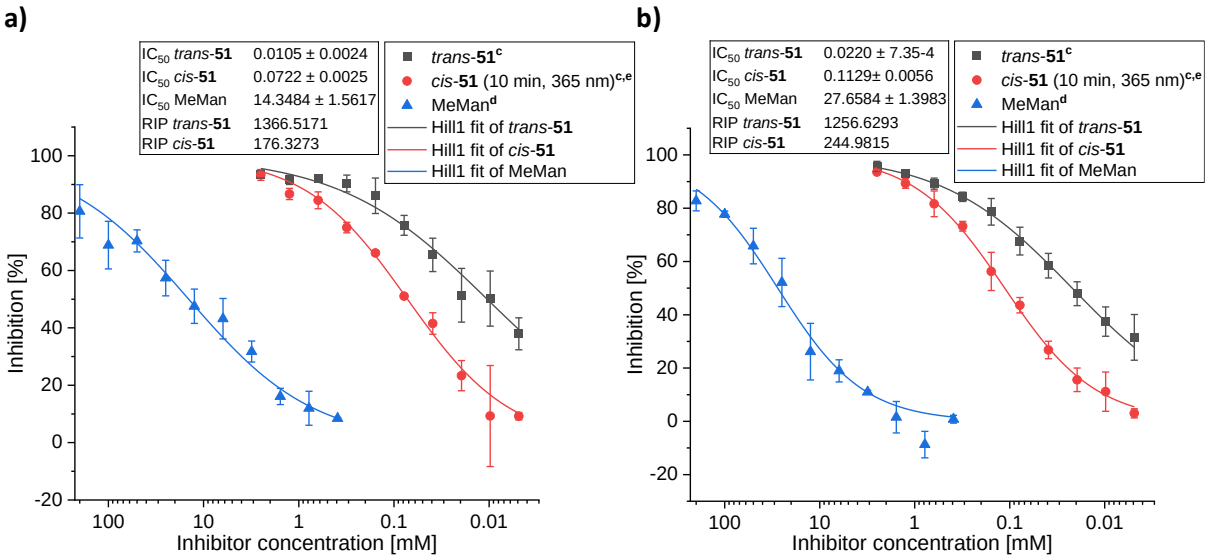


Figure 5.8-4: Representative results from bacterial adhesion inhibition experiments with diethylene glycol-linked fluoroazobenzene mannoside derivative **51** after photoirradiation at 365 nm.

a) Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). **b)** Bacterial fluorescence was measured after 30 min of relaxation at 37 °C. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **c)** triplicate or **d)** duplicate measurements. **e)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

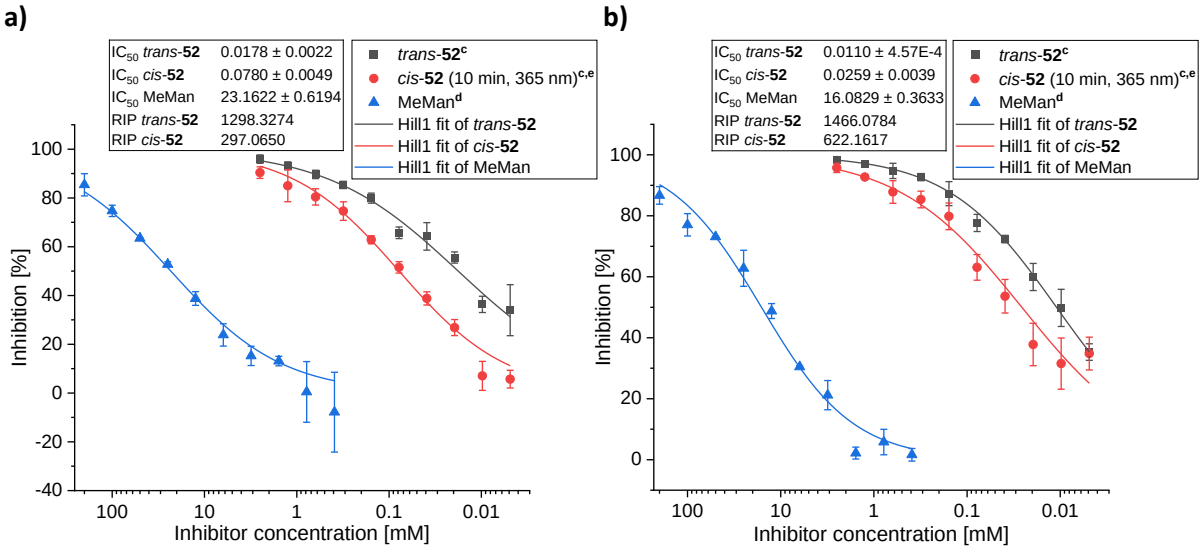


Figure 5.8-5: Representative results from bacterial adhesion inhibition experiments with triethylene glycol-linked fluoroazobenzene mannoside derivative **52** after photoirradiation at 365 nm.

a) Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). **b)** Bacterial fluorescence was measured after 30 min of relaxation at 37 °C. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **c)** triplicate or **d)** duplicate measurements. **e)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

Table 5.8-2: Antiadhesive properties of tri- and diethylene glycol-linked fluoroazobenzene mannosides **51** and **52** after photoirradiation.

Value	***trans*-51**[a,d] **Not irradiated**	***cis*-51**[a,d] **Irradiated at 365 nm**	***trans*-52**[a] **Not irradiated**	***cis*-52**[a] **Irradiated at 365 nm**
IC_{50} [mM]	$1.68 \cdot 10^{-2}$ ($\pm 8.45 \cdot 10^{-3}$)	$7.53 \cdot 10^{-2}$ ($\pm 2.3 \cdot 10^{-2}$)	$1.90 \cdot 10^{-2}$ ($\pm 6.10 \cdot 10^{-3}$)	$8.78 \cdot 10^{-2}$ ($\pm 5.11 \cdot 10^{-2}$)
RIP[b]	1484.93 (± 445.53)	279.63 (± 116.26)	1668.26 (± 588.63)	413.70 (± 149.89)
IC_{50} ratio[c]	0.23 ($\pm 1.45 \cdot 10^{-1}$)		0.26 ($\pm 1.12 \cdot 10^{-1}$)	

a) Average IC_{50}, RIP and IC_{50} ratio from 4 individual experiments. The respective ± standard deviations are given in brackets. **b)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **c)** Mean ratio of IC_{50} values of the respective inhibitors without UV exposure and after 10 min irradiation at 365 nm $\left(IC_{50}\text{ ratio} = \frac{IC_{50}\text{ not irradiated}}{IC_{50}\text{ irradiated}}\right)$. **d)** The IC_{50} and RIP values are probably elevated due to poor water solubility of the diethylene glycol-linked fluoroazobenzene mannoside **51**.

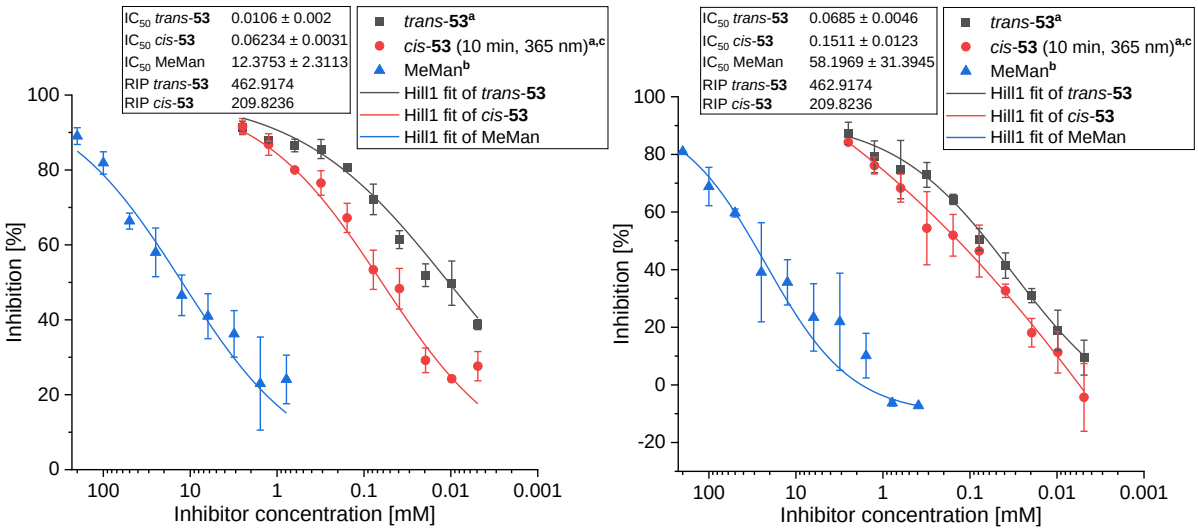


Figure 5.8-6: Results from bacterial adhesion inhibition experiments with diethylene glycol-linked azobenzene mannoside derivative **53** after photoirradiation at 365 nm.

Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **a)** triplicate or **b)** duplicate measurements. **c)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension. The inhibition curves are affected by the poor water solubility of the diethylene glycol-linked azobenzene mannoside **53**.

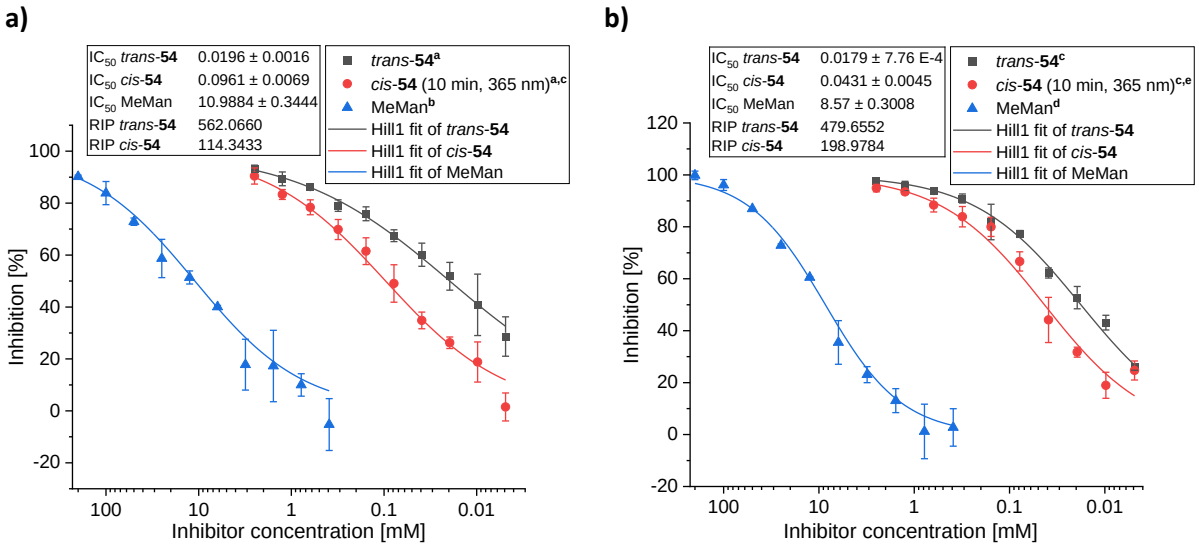


Figure 5.8-7: Representative results from bacterial adhesion inhibition experiments with triethylene glycol-linked azobenzene mannoside derivative **54** after photoirradiation at 365 nm.

a) Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). **b)** Bacterial fluorescence was measured after 45 min of relaxation at 37 °C. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **c)** triplicate or **d)** duplicate measurements. **e)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

Table 5.8-3: Antiadhesive properties of tri- and diethylene glycol-linked azobenzene mannosides **53** and **54** after photoirradiation.

Value	*trans*-53[a,e] Not irradiated	*cis*-53[a,e] Irradiated at 365 nm	*trans*-54[b] Not irradiated	*cis*-54[b] Irradiated at 365 nm
IC_{50} [mM]	$3.95 \cdot 10^{-2}$ ($\pm 4.09 \cdot 10^{-2}$)	$1.07 \cdot 10^{-1}$ ($\pm 6.27 \cdot 10^{-2}$)	$2.48 \cdot 10^{-2}$ ($\pm 5.33 \cdot 10^{-3}$)	$9.65 \cdot 10^{-2}$ ($\pm 2.96 \cdot 10^{-2}$)
RIP[c]	815.75 (± 498.98)	204.17 (± 8.0)	499.82 (± 133.40)	127.59 (± 15.24)
IC_{50} ratio[d]	0.31 ($\pm 2.00 \cdot 10^{-1}$)		0.27 ($\pm 7.39 \cdot 10^{-2}$)	

a) Average IC_{50}, RIP and IC_{50} ratio from 2 individual experiments. **b)** Average IC_{50}, RIP and IC_{50} ratio from 4 individual experiments. The respective ± standard deviations are given in brackets. **c)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **d)** Mean ratio of IC_{50} values of the respective inhibitors without UV exposure and after 10 min irradiation at 365 nm $\left(IC_{50}\ \text{ratio} = \frac{IC_{50}\ \text{not irradiated}}{IC_{50}\ \text{irradiated}}\right)$. **e)** The IC_{50} and RIP values are affected by the poor water solubility of the diethylene glycol-linked azobenzene mannoside **53**.

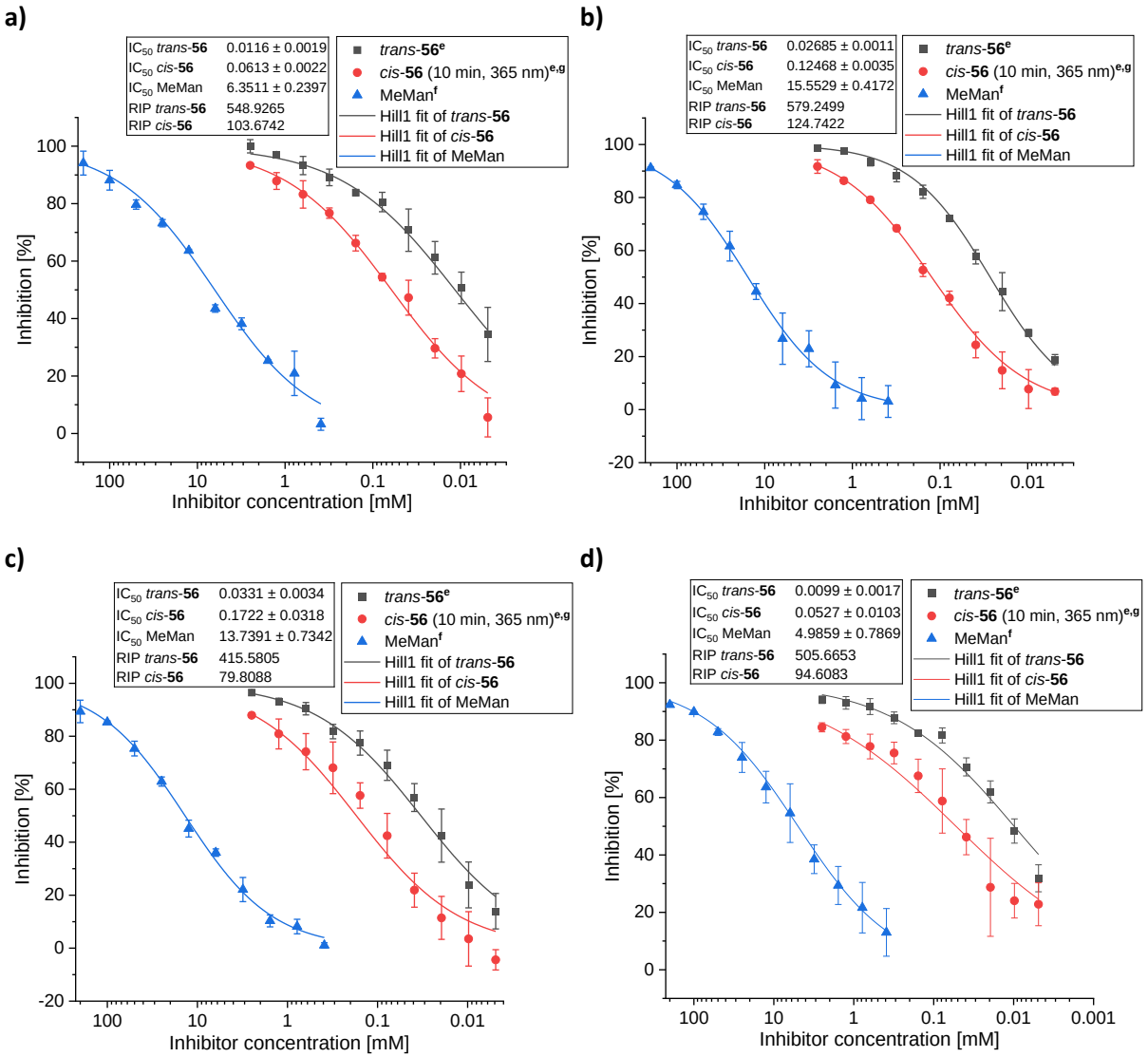


Figure 5.8-8: Representative results from bacterial adhesion inhibition experiments with azobenzene mannoside derivative **56** after photoirradiation at 365 nm.

Bacterial fluorescence was measured after **a)** 15 min, **b)** 30 min, **c)** 45 min at 37 °C (incubation step with bacterial suspension) or **d)** 20 h of relaxation at room temperature. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **e)** triplicate or **f)** duplicate measurements. **g)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

Table 5.8-4: Antiadhesive properties of azobenzene mannoside derivative **56** after photoirradiation.

Value	***trans*-56**[a] **Not irradiated**	***cis*-56**[a] **Irradiated 10 min at 365 nm**
IC_{50} [mM]	$2.64 \cdot 10^{-2}$ ($\pm 1.57 \cdot 10^{-2}$)	$1.17 \cdot 10^{-1}$ ($\pm 5.45 \cdot 10^{-2}$)
RIP[b]	493.76 (± 126.26)	104.26 (± 23.28)
IC_{50} ratio[c]	0.23 ($\pm 1.28 \cdot 10^{-1}$)	

a) Average IC_{50}, RIP and IC_{50} ratio from 6 individual experiments. The respective ± standard deviations are given in brackets. **b)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **c)** Mean ratio of IC_{50} of the respective inhibitors without UV exposure and after 10 min irradiation at 365 nm $\left(IC_{50}\text{ ratio} = \frac{IC_{50}\text{ not irradiated}}{IC_{50}\text{ irradiated}}\right)$.

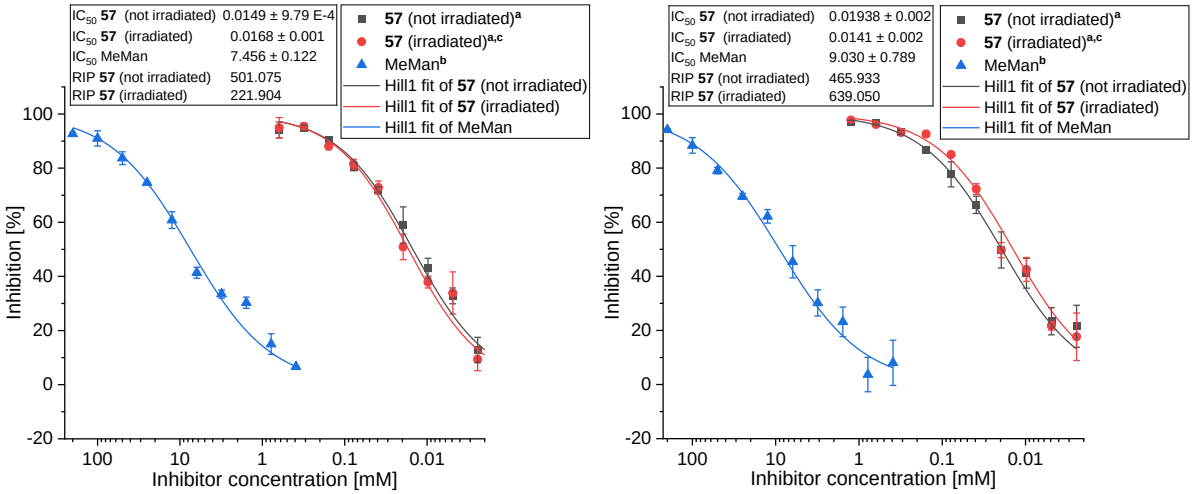


Figure 5.8-9: Representative results from bacterial adhesion inhibition experiments with control compound 2-Fluoro-4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **57** with and without previous UV exposure.

Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **a)** triplicate or **b)** duplicate measurements. [c)]The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addion of the bacterial suspension.

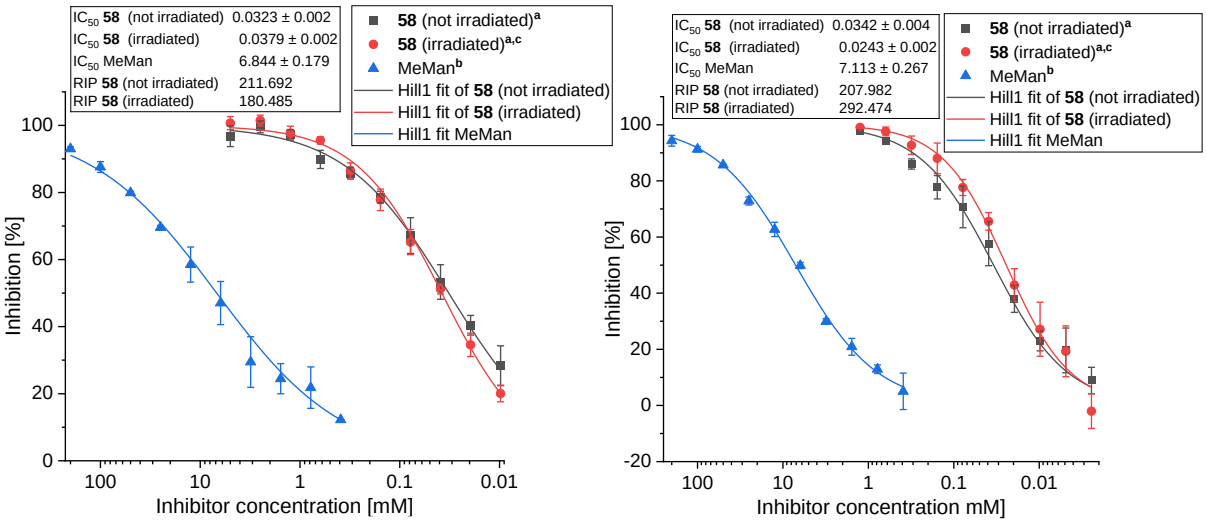


Figure 5.8-10: Representative results from bacterial adhesion inhibition experiments with control compound 4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **58** with and without UV exposure.

Bacterial fluorescence was measured after 15 min of relaxation at 37 °C (incubation step with bacterial suspension). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of **a)** triplicate or **b)** duplicate measurements. **c)** The inhibitor solutions were applied as 1:2 dilution series on the mannan-coated microwell plate and irradiated at 365 nm for 10 min, prior to the addition of the bacterial suspension.

Table 5.8-5: Influence of UV irradiation (10 min at 365 nm) on the antiadhesive properties of the control compounds 2-Fluoro-4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **57** and 4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **58**.

Value	57[a]		58[a]	
	Not irradiated	**Irradiated at 365 nm**	**Not irradiated**	**Irradiated at 365 nm**
IC_{50} [mM]	$4.56 \cdot 10^{-2}$ ($\pm 4.93 \cdot 10^{-2}$)	$2.73 \cdot 10^{-2}$ ($\pm 2.05 \cdot 10^{-2}$)	$3.26 \cdot 10^{-2}$ ($\pm 1.54 \cdot 10^{-2}$)	$3.03 \cdot 10^{-2}$ ($\pm 6.96 \cdot 10^{-3}$)
RIP[b]	437.87 (± 80.97)	593.65 (± 133.61)	207.36 (± 4.16)	231.65 (± 56.62)
IC_{50} ratio[c]	1.42 ($\pm 5.67 \cdot 10^{-1}$)		1.12 ($\pm 2.78 \cdot 10^{-1}$)	

a) Average IC_{50} and RIP values, as well as IC_{50} ratios from 3 individual experiments. The respective ± standard deviations are given in brackets. **b)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl *α*-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **c)** Mean ratio of IC_{50} of the respective inhibitors without UV exposure and after 10 min irradiation at 365 nm $\left(IC_{50}\text{ ratio} = \frac{IC_{50}\text{ not irradiated}}{IC_{50}\text{ irradiated}}\right)$.

Adhesion inhibition experiments without UV-exposition

Due to the unavoidable temperature increase during and after UV irradiation of the microplate, the IC_{50} and RIP values of the inhibitors, used in Chapter 3.4, were also determined under standard conditions, as described in Chapter 5.7.8. To improve the solubility of the moderately water-soluble azobenzene mannoside derivatives, the inhibitors were dissolved in PBS/MeOH 9:1 and equilibrated at 37 °C for 5 min prior to use.

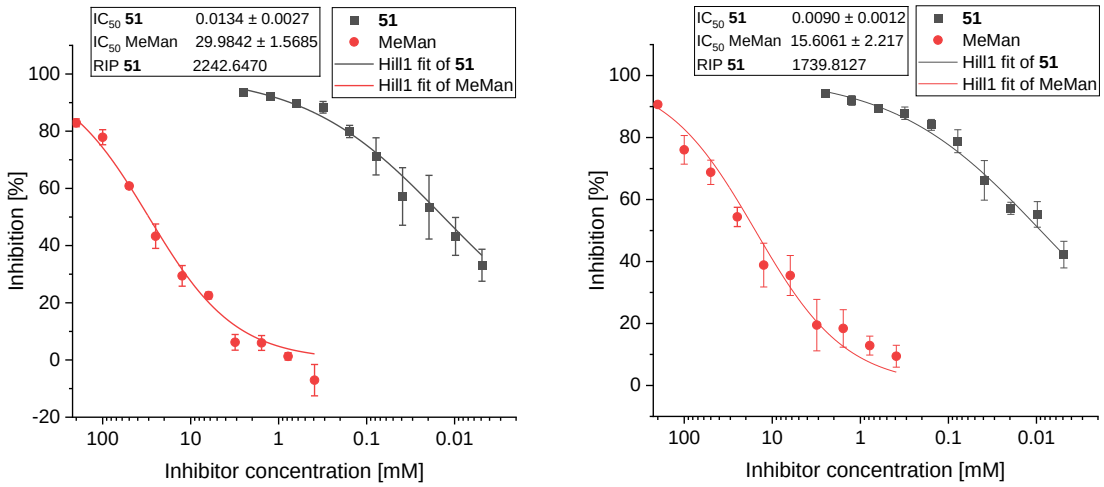


Figure 5.8-11: Two Representative results from bacterial adhesion inhibition experiments with diethylene glycol-linked fluoroazobenzene mannoside derivative **51**.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

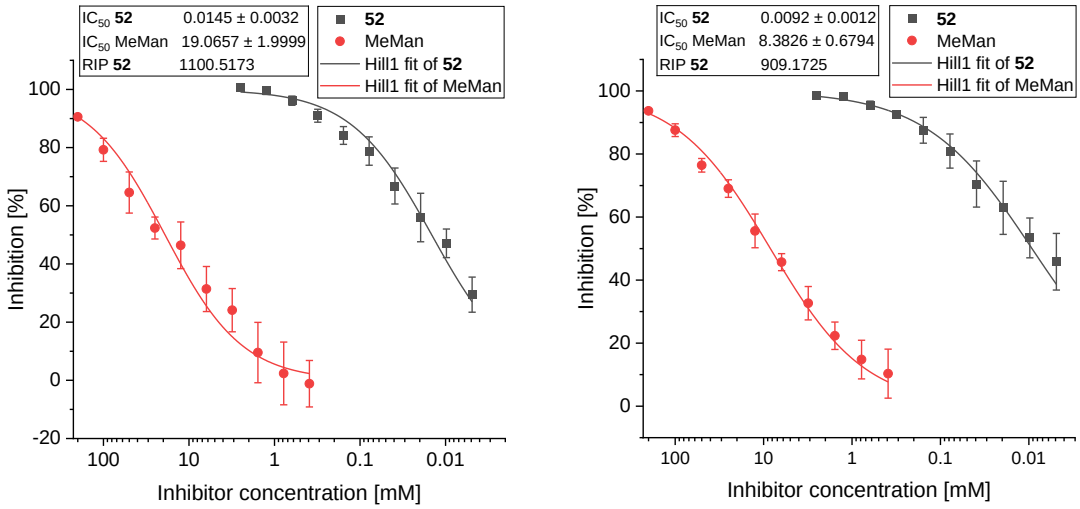


Figure 5.8-12: Two representative results from bacterial adhesion inhibition experiments with triethylene glycol-linked fluoroazobenzene mannoside derivative **52**.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

Table 5.8-6: Antiadhesive properties of tri- and diethylene glycol-linked fluoroazobenzene mannosides **51** and **52**.

Value	MeMan[a]	51[b,e]	52[c]
IC_{50} [mM]	15.10 (± 7.62)	$9.11 \cdot 10^{-3}$ (± 4.20 $\cdot 10^{-3}$)	$1.02 \cdot 10^{-2}$ (± $2.88 \cdot 10^{-3}$)
RIP[d]	≡ 1	2240.58 (± 499.74)	1110.61 (± 127.51)

a) Average IC_{50} from 7 individual experiments. **b)** Average IC_{50} and RIP from 3 individual experiments. **c)** Average IC_{50} and RIP from 4 individual experiments. The respective ± standard deviations are given in brackets. **d)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **e)** The IC_{50} and RIP values are probably significantly elevated due to poor water solubility of the diethylene glycol-linked fluoroazobenzene mannoside **51**.

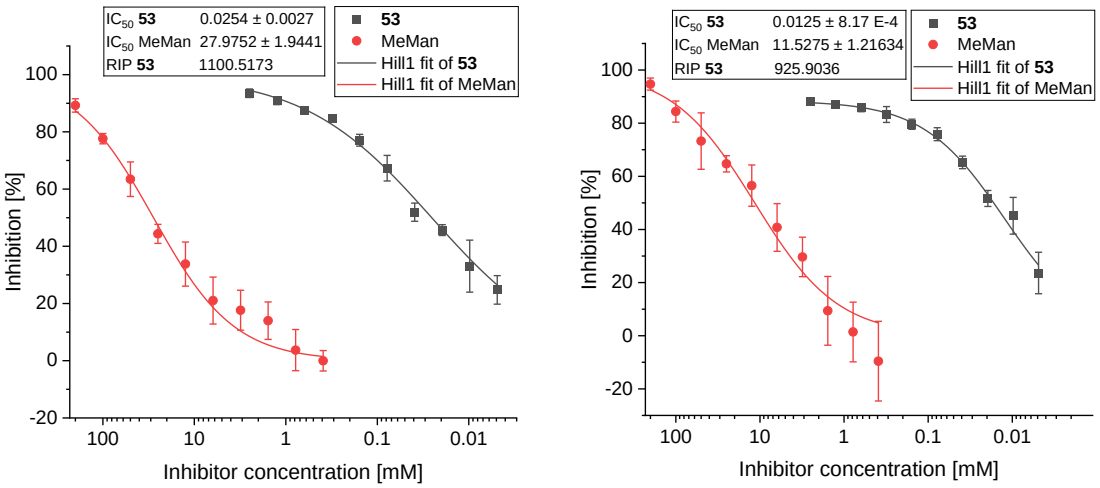


Figure 5.8-13: Two representative results from bacterial adhesion inhibition experiments with diethylene glycol-linked azobenzene mannoside derivative **53**.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

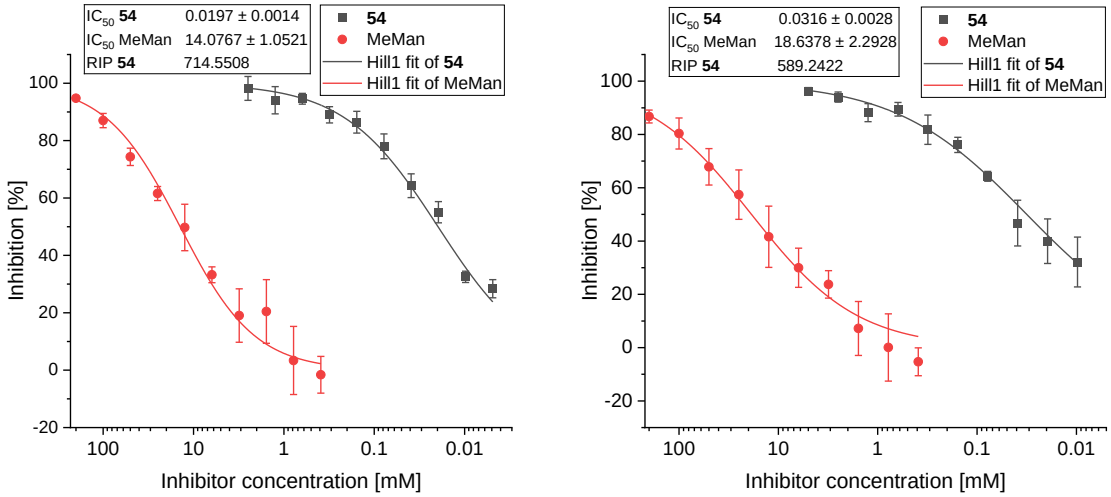


Figure 5.8-14: Two representative results from bacterial adhesion inhibition experiments with triethylene glycol-linked azobenzene mannoside derivative **54**.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

Table 5.8-7: Antiadhesive properties of tri- and diethylene glycol-linked azobenzene mannosides **53** and **54**.

Value	MeMan[a]	53[b,e]	54[c]
IC_{50} [mM]	14.32 (± 6.32)	$1.39 \cdot 10^{-2}$ (± $6.64 \cdot 10^{-3}$)	$2.32 \cdot 10^{-2}$ (± $1.96 \cdot 10^{-2}$)
RIP[d]	≡ 1	1095.80 (± 129.0)	614.90 (± 89.27)

a) Average IC_{50} from 8 individual experiments. **b)** Average IC_{50} and RIP from 5 individual experiments. **c)** Average IC_{50} and RIP from 3 individual experiments. The respective ± standard deviations are given in brackets. **d)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$). **e)** The IC_{50} and RIP values are probably significantly elevated due to poor water solubility of the diethylene glycol-linked azobenzene mannoside **53**.

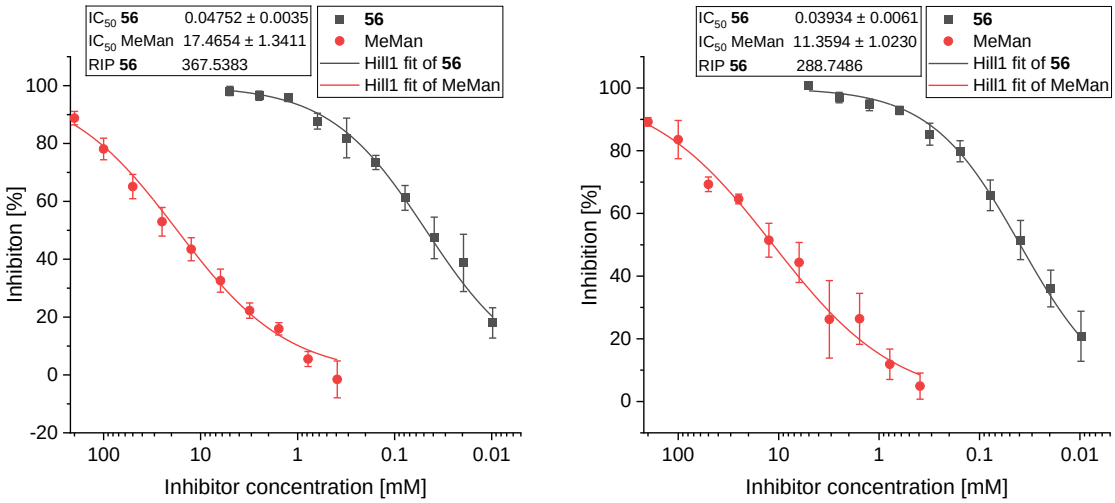


Figure 5.8-15: Two representative results from bacterial adhesion inhibition experiments with azobenzene mannoside derivative **56**.

Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

Table 5.8-8: Antiadhesive properties of azobenzene mannoside derivative **56**.

Value	MeMan[a]	56[a]
IC_{50} [mM]	18.61 (± 1.88 · 10^{-1})	5.53 · 10^{-2} (± 2.09 · 10^{-2})
RIP[b]	≡ 1	332.75 (± 40.19)

a) Average IC_{50} and RIP from 3 individual experiments. The respective ± standard deviations are given in brackets.
b) RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl α-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$).

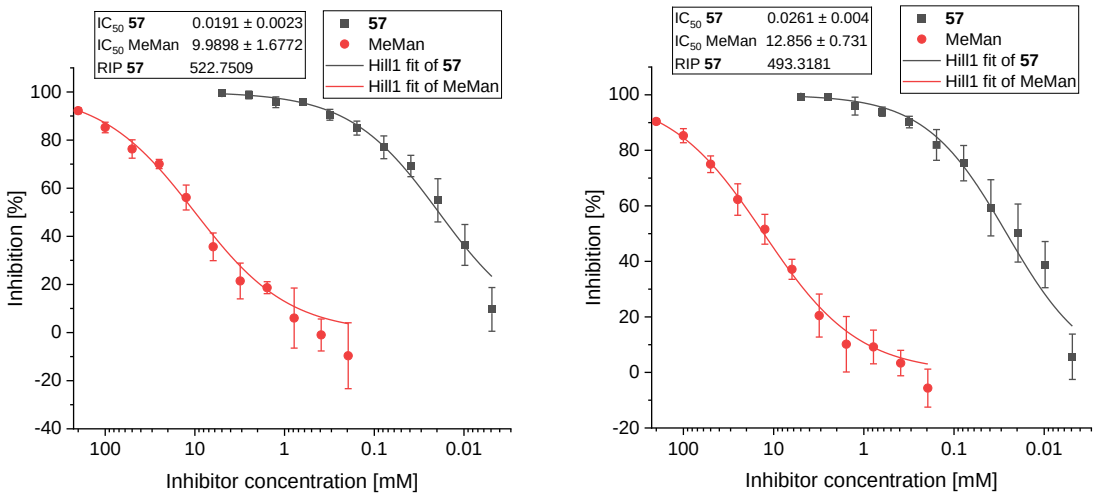


Figure 5.8-16: Two representative results from bacterial adhesion inhibition experiments with control compound 2-Fluoro-4-(*N*-acetylamino)phenyl α-D-mannopyranoside **57**.
Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

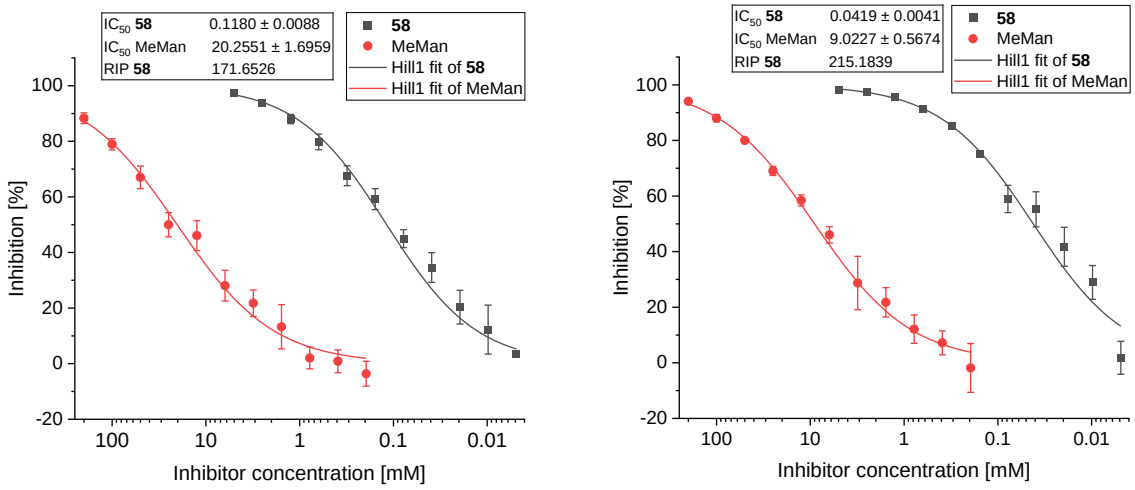


Figure 5.8-17: Two representative results from bacterial adhesion inhibition experiments with control compound 4-(*N*-acetylamino)phenyl α-D-mannopyranoside **58**.
Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

Table 5.8-9: Antiadhesive properties of the control compounds 2-Fluoro-4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **57** and 4-(*N*-acetylamino)phenyl *α*-D-mannopyranoside **58**.

Value	MeMan[a]	57[b]	58[c]
IC_{50} [mM]	14.06 (± 6.10)	$2.18 \cdot 10^{-2}$ (± $3.73 \cdot 10^{-3}$)	$6.50 \cdot 10^{-2}$ (± $3.91 \cdot 10^{-2}$)
RIP[d]	≡ 1	522.72 (± 29.39)	270.78 (± 91.30)

a) Average IC_{50} from 7 individual experiments. **b)** Average IC_{50} and RIP from 3 individual experiments. **c)** Average IC_{50} and RIP from 4 individual experiments. The respective ± standard deviations are given in brackets. **d)** RIP values represent the inhibitory potency of each inhibitor compared to the reference inhibitor methyl *α*-D-mannopyranoside (MeMan) tested on the same microplate ($IP_{MeMan} \equiv 1$).

Control experiments

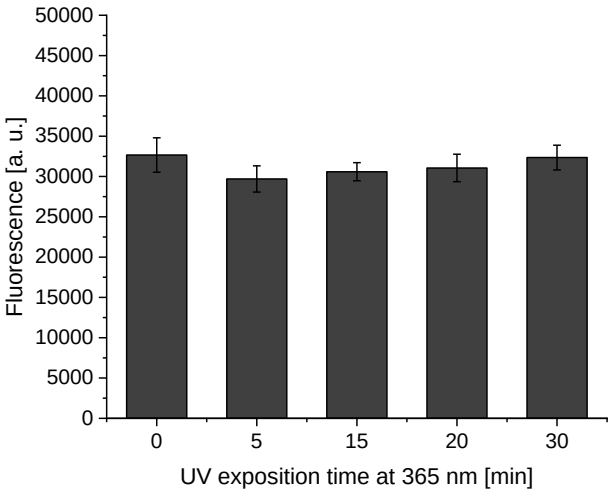
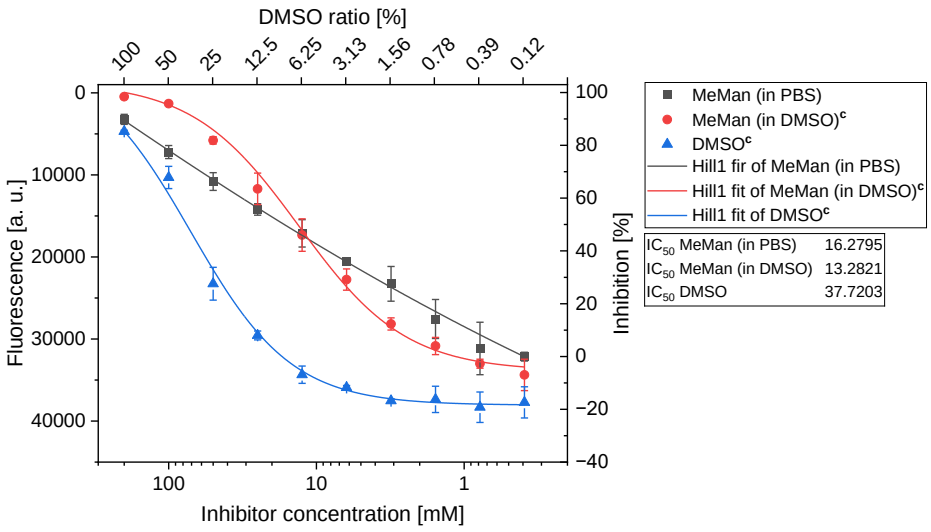


Figure 5.8-18: Adhesion of PKL1162 to a mannan-coated polystyrene surface after exposition to UV light (365 nm).

A mannan-coated black 96 microwell plate was filled with PBS (50 µL/well), irradiated with a UV lamp at 365 nm (LED NC4U133AT, Nichia Corporation) and PKL1162 was allowed to adhere at 37°C, 100 rpm for 45 min (Shaking Incubator 3031, GFL). Unexposed wells were wrapped in aluminium foil prior to irradiation. Error bars in represent the ± standard deviation of 18 measurements.

a)



b)

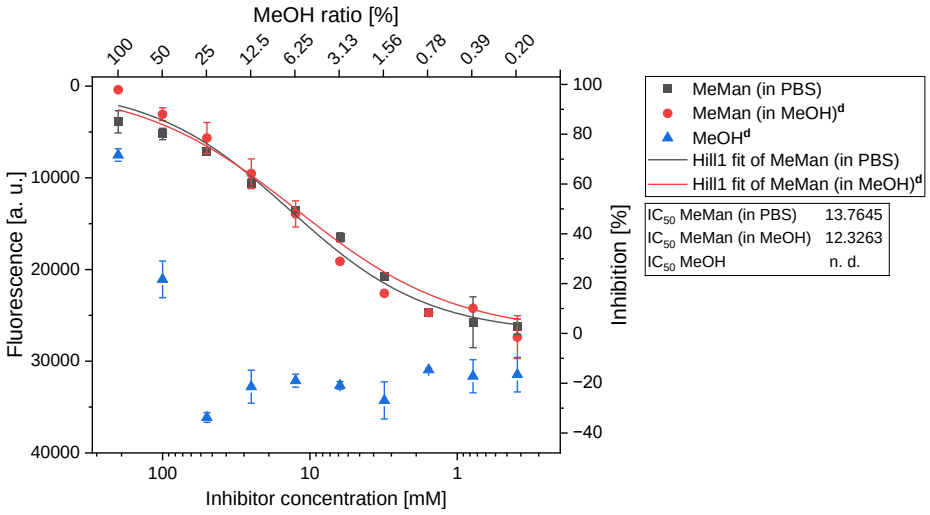


Figure 5.8-19: Control inhibition experiments with methyl *α*-D-mannopyranoside (MeMan) in the presence of **a)** DMSO and **b)** MeOH, to improve the solubility of azobenzene mannoside derivatives. **c)** A solution of MeMan (200 mM) in DMSO, or only DMSO (control without inhibitor) was serially 1:2-diluted with PBS. **d)** A solution of MeMan (200 mM) in MeOH, or only MeOH (control without inhibitor) was serially 1:2-diluted with PBS. **e)** The dose-response curves of solvent controls (without inhibitor) refer only to the upper X-axis. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of duplicate measurements.

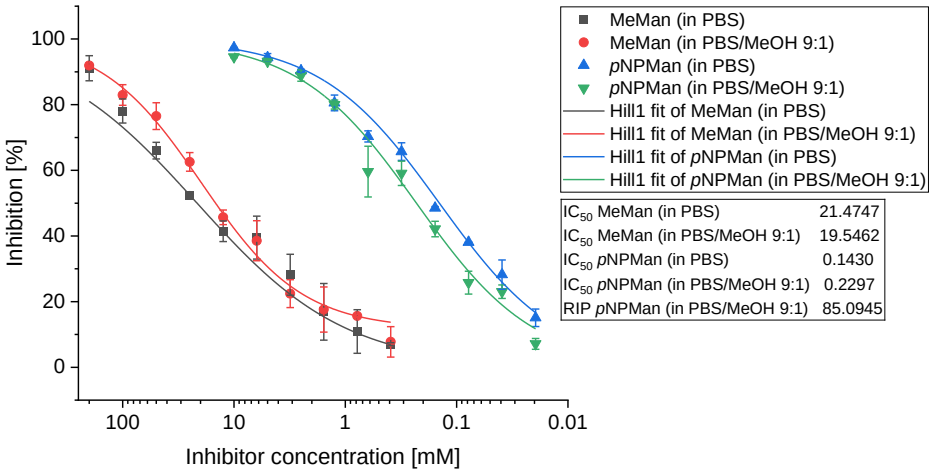


Figure 5.8-20: Control experiment to demonstrate the influence of 10 % v/v MeOH in adhesion inhibition experiments.

Inhibitors were dissolved and serially diluted either in PBS or PBS/MeOH 9:1, with dilution factor 1:2. The result suggests no or little impact of small amounts of MeOH in the buffer used in biological assays (PBS) on the inhibition effect. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of duplicate measurements.

5.8.4 Expression of FimH-His_6 variants encoded on pPKL241- and pFimH-His_6-NT-trc-based constructs

Cultivation of *E. coli* for protein expression

Expression strains were pre-cultured in LB medium (10 – 20 mL, Table 5.7-1) with antibiotics (ampicillin or carbenicillin, 100 µg/mL) under vigorous shaking (175 – 200 rpm) in an orbital shaker (Shaking Incubator 3031, GFL) overnight. The pre-cultures were generally inoculated with single, well-defined colonies that were previously cultured under selective conditions on LB agar. The definite growth conditions of the respective strains are compiled in Table 5.7-4. In general, pPKL241-based expression systems were cultured at 37 °C according to the procedure of Schembri et al.,[470] whereas expression in HM125 and/or strains with pFimHs-NT-trc-derived plasmids were grown at lower temperatures (≤ 30 °C).[337] The next day, the pre-cultures were centrifuged at room temperature, 5000 rpm for 10 min (Universal 320R centrifuge, Hettich®) and the pelleted cells were resuspended in fresh medium (1 – 2 mL). The expression media, LB medium (3 – 5 L) or alternatively, in order to achieve higher cell densities, more nutritious Terrific broth (TB) medium (1.5 – 3 L, Table 5.7-1) were inoculated with the bacterial suspensions and the optical density was adjusted to OD_{600} = 0.01 – 0.05. Antibiotics (ampicillin or carbenicillin, 100 µg/mL) and a few drops of Antifoam A concentrate (Sigma Aldrich) were added and the expression cultures were shaken in baffled 5 L Erlenmeyer flasks (maximum filling volume 1 L/flask) at 80 – 100 rpm in an orbital shaker (Multitron, Infors HT). Small-scale test expressions, on the other hand, were performed in 100 mL culture flasks with LB medium (50 mL) with antibiotics added (ampicillin or carbenicillin, 100 µg/mL) under agitation at 150 rpm (Shaking Incubator 3031, GFL), at 37 °C or 30 °C, respectively. When the intended OD_{600} of ∼ 0.5 was reached, the expression was induced by addition of 1 mM isopropyl *β*-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich). In the case of pPKL241-based expression systems, the temperature was initially reduced to 30 °C for the first 4 – 5 h after induction. Expression overnight was generally performed at 25 – 28 °C. If appropriate, small amounts of culture (100 µL) were removed at regular intervals in order to retrospectively monitor the course of expression by SDS-PAGE (see protocol: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in Chapter 5.7.7) and centrifuged 5 min at 12 000 rpm (Universal 320R centrifuge, Hettich®). The bacterial cells from preparative expression volumes were harvested the next day in multiple centrifugation steps, each 10 min at 10000 rpm, 4 °C (Avanti J-26 XP with JA-10 fixed angle rotor, Beckman Coulter). If necessary, the cell pellets were frozen in liquid nitrogen and stored at −20 °C. However, for a subsequent periplasmic extraction it proved advantageous to avoid prior freezing in order to prevent complete cell lysis and thus release of cytoplasmic proteins.

Periplasmic extraction

If the N-terminal signal sequence was present on the expression plasmids, periplasmic extraction was preferably performed according to a procedure introduced by Glockshuber and co-workers for the purification of FimH.[480] The method is based on the strong permeabilising effect of polymyxin B sulfate on the outer membrane of Gram-negative bacteria,[706,707] which is further enhanced by the addition of EDTA as a chelating agent.[708,709] The pelleted cells were completely resuspended in the smallest possible volume of lysis buffer I (Table 5.7-1) and incubated at 4 °C for 2 – 3 h under gentle rotation (Intelli-Mixer, neoLab®). A protease inhibitor cocktail (SigmaFast™, Sigma-Aldrich, cat. no. S8830) was added to the lysis buffer I according to the supplier's specifications, especially if expression strains were used that produce periplasmic proteases. Subsequently, the insoluble cell components were removed by centrifugation at 12 000 – 14 000 rcf, 4 °C, for 20 min (Avanti J-26 XP with JA-25.50 fixed angle rotor, Beckman Coulter).

The periplasmic extract was dialysed (MWCO = 10 kDa) against binding buffer (≥ 5 L, Table 5.7-1) overnight at 5 °C with several buffer changes to largely remove EDTA from the protein extract. Finally, the periplasmic extract was thoroughly cleared by another centrifugation step (12 000 – 14 000 rcf) at 4 °C, 20 min (Avanti J-26 XP with JA-25.50 fixed angle rotor, Beckman Coulter). The cell pellets of samples, which were intended to be subjected to SDS-PAGE, were also treated with lysis buffer I (100 µL, 4 °C, 2 h) and centrifuged at 10 000 rpm, 5 min, 4 °C (Universal 320R centrifuge, Hettich®). If centrifugation was not sufficient, the protein extract was sterile-filtered through a bottle top filter (Filtropur BT 50, 0.1 µm, Sarstedt, cat. no. 83.3941.102) under reduced pressure. The use of the cyclic peptide antibiotic polymyxin B sulfate (1.3 kDa) is preferable over a lysozyme-mediated protein extraction due to the possible confusion of the target protein (17.8 kDa) with lysozyme (14.3 kDa) on the SDS-polyacrylamide gel.

Mechanical cell disruption

Optionally, in particular for the expression of FimH variants lacking the N-terminal signal peptide (N136C ΔN-sigpep), a mechanical cell disruption was chosen. For this purpose, both high-pressure homogenisation and a sonication treatment were available as options. The pelleted cells were resuspended in the smallest possible volume of lysis buffer II (Table 5.7-1) and either disrupted by one or two passages through a high-pressure homogeniser (Emulsiflex C3, Avestin GmbH) at 1200 bar, 5 °C. Alternatively, the cells were disrupted by ultrasound treatment (Sonoplus ultrasonic homogeniser HD 2200 with a VS70 T sonotrode, Bandelin electronic GmbH) in 3 runs for one minute each (cycle 6 x 10 %, Power 30 %), interrupted by 1 min incubation on ice. Finally, the cell debris was carefully separated by centrifugation at 75 600 rcf (Avanti J-26 XP with JA-25.50 fixed angle rotor, Beckman Coulter). In the course of the research, periplasmic extraction was favoured over mechanical cell disruption, if possible. This was mainly due to the lower contamination by cytoplasmic proteins as shown in Figure 5.8-21. Samples that were collected for later analysis by SDS-PAGE were also resuspended in lysis buffer II (100 µL, Table 5.7-1) and disrupted by several cycles of flash-freezing in liquid nitrogen and subsequent thawing at 37 °C. The cells were harvested by centrifugation at 10 000 rpm, 5 min, 4 °C (Universal 320R centrifuge, Hettich®).

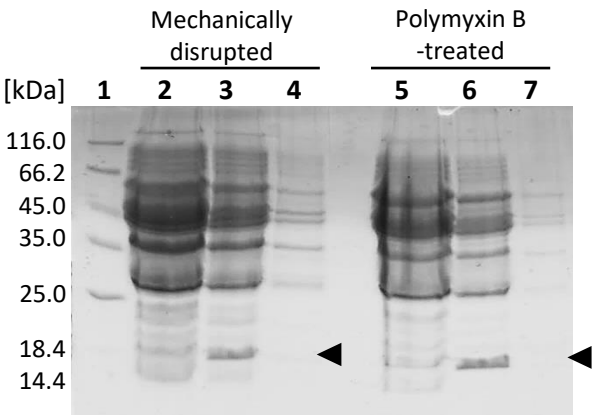


Figure 5.8-21: Comparison of mechanical cell disruption and periplasmic extraction after overexpression of $FimH_L$ in HM125 pPKL241 investigated by SDS-PAGE.

1: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **2** – **4**: different volumes of protein extract after mechanical cell disruption by sonication and subsequent centrifugation of the cell debris; **2**: 15 µL; **3**: 5 µL; **4**: 1 µL; **5** – **7**: different volumes of protein extract obtained by periplasmic extraction with polymyxin B sulfate and subsequent centrifugation; **5**: 15 µL; **6**: 5 µL; **7**: 1 µL. SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 1 h at constant current (I = 25 mA). The harvested cells were frozen at -20 °C prior to protein extraction. Black arrowheads indicate the expression bands of the overexpressed target protein $FimH_L$ (17.8 kDa).

5.8.5 Purification strategy for $FimH_L$-His_6 variants encoded on pPKL241-based constructs

Immobilised Metal Affinity Chromatography (IMAC) of $FimH_L$-His_6 variants

The purification of C-terminally His_6-tagged $FimH_L$ variants was performed using ÄKTA™ pure and ÄKTA™ purifier chromatography systems (GE Healthcare) equipped with a UV/vis and conductivity detector. For this purpose, the cleared protein extracts containing the impure target proteins (Chapter 5.8.3) were applied onto a Ni^{2+}-charged IMAC column (5 mL HisTrap™ HP column, GE Healthcare, cat. no. 17-5248; or 5 mL HisPur™ Ni^{2+}-NTA Chromatography Cartridges, cat. no. 90099) that had previously been equilibrated with 5 – 10 column volumes (CV) of binding buffer supplemented with 5 % elution buffer I (Table 5.7-1, ≙ 35 mM imidazole). After sample application, the column was washed with 5 – 10 CV binding buffer supplemented with 6 – 10 % elution buffer I (Wash I, ≙ 39 mM imidazole). The ratio of the elution buffer I was subsequently increased in a linear gradient up to 100 % (500 mM imidazole) throughout 10 – 15 CV. Elution of the target proteins occurred reaching approximately 12 % elution buffer I (~ 69 mM imidazole). The fractions were analysed for the presence of the target protein by SDS-PAGE (see Chapter 5.7.7). Eligible fractions were pooled and concentrated in a centrifugal filter (MWCO = 10 kDa, Amicon® Ultra, Merck Millipore, cat. no. UFC9010 or Vivaspin® 20, Cytiva, cat. no. 28-9323-60) to a volume of 1 – 2 mL by several centrifugation steps at 3500 rpm, 4 °C (Centrifuge 5417 R Eppendorf®).

Preparative size exclusion chromatography (SEC) of $FimH_L$-His_6 variants

Preparative size exclusion chromatography was also performed by employing ÄKTA™ pure or ÄKTA™ purifier chromatography systems (GE Healthcare). The pooled and concentrated fractions resulting from the first IMAC purification step were subjected to size exclusion chromatography on a 120 mL HiLoad® 16/600 Superdex® prep-grade (pg) column (GE Healthcare, cat. no. 28-9893-33) that had been beforehand equilibrated with PBS (Table 5.7-1). Elution of the target proteins occurred after a passage of approx. 0.58 – 0.7 CV PBS. Appropriate fractions were once again analysed by SDS-PAGE (Chapter 5.7.7). If sufficiently purified target protein was proven, the fractions were unified and concentrated in a centrifugal filter (MWCO = 10 kDa, Amicon® Ultra, Millipore®, cat. no. UFC9010 or Vivaspin® 20, Cytiva, cat. no. 28-9323-60) to 0.5 to 2 mg/mL according to the manufacturers' recommendations. The concentrated proteins were aliquoted in 100 µL volumes, flash-frozen in liquid nitrogen and stored at −80 °C as suggested by Bouckaert and co-workers.[376]

Exemplary results from the expression and purification of $FimH_L$-His_6 variants

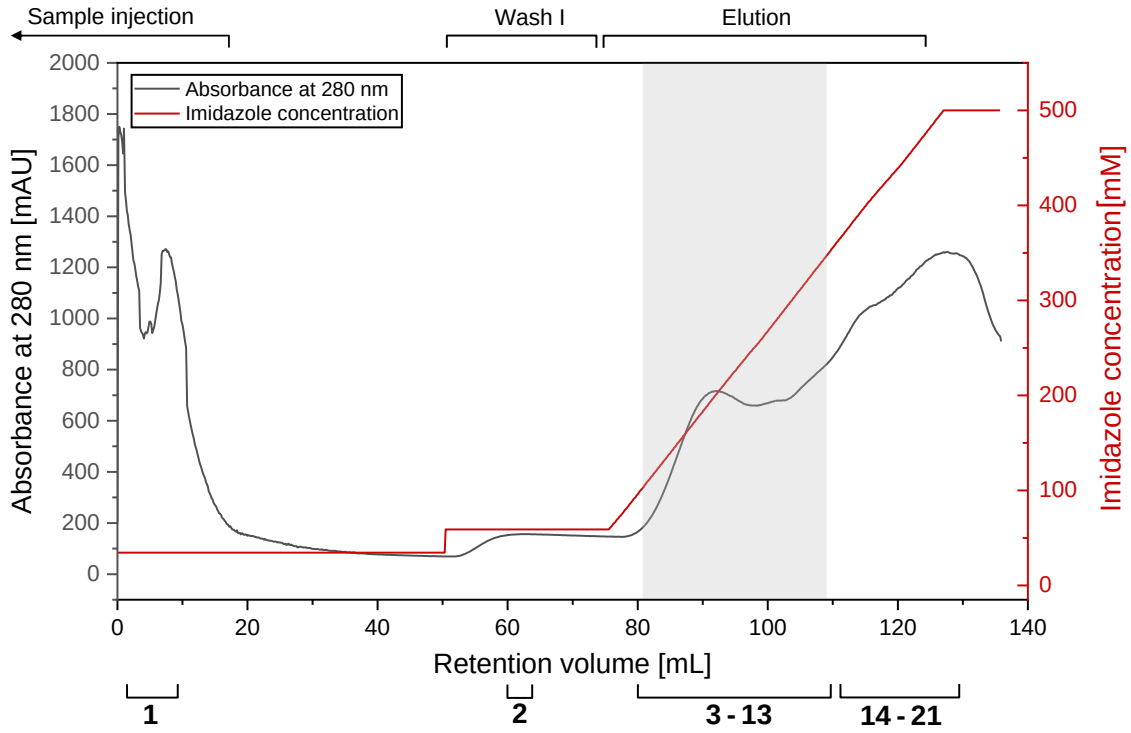


Figure 5.8-22: Chromatogram from the IMAC purification of $FimH_L$ using a linear imidazole gradient.
$FimH_L$ was expressed in *E. coli* HM125 pPKL241 in 4 L LB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare) equilibrated with binding buffer. The column was washed with 59 mM imidazole (Wash I, 5 CV) prior to elution of the target protein via a linear imidazole gradient from 60 to 500 mM imidazole throughout 10 CV. Fractions **1** – **21** were analysed by SDS-PAGE (Figure 5.8-23). The main peak (fractions **3** – **13**), containing the target protein, is highlighted in grey.

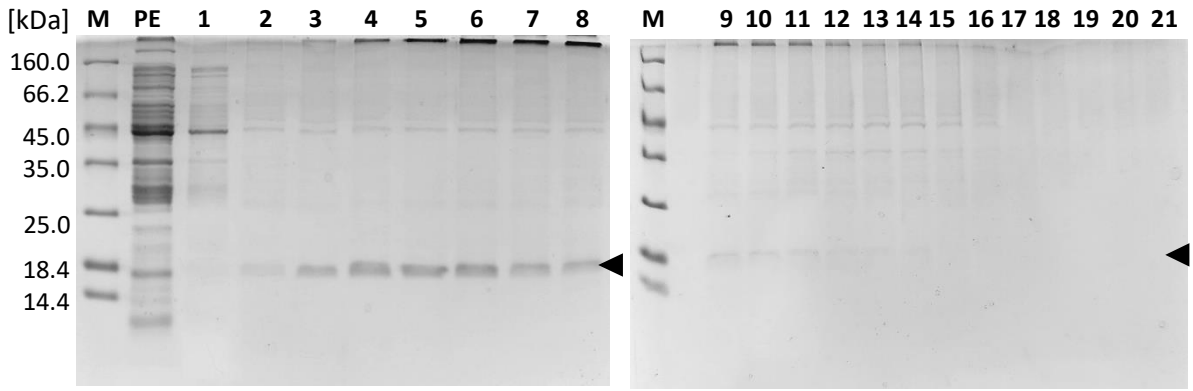


Figure 5.8-23: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_L$.
M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **21**: collected fractions from the IMAC purification shown in Figure 5.8-22. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 25 mA/gel). Black arrowheads indicate the migration bands corresponding to the overexpressed target protein $FimH_L$ (17.8 kDa).

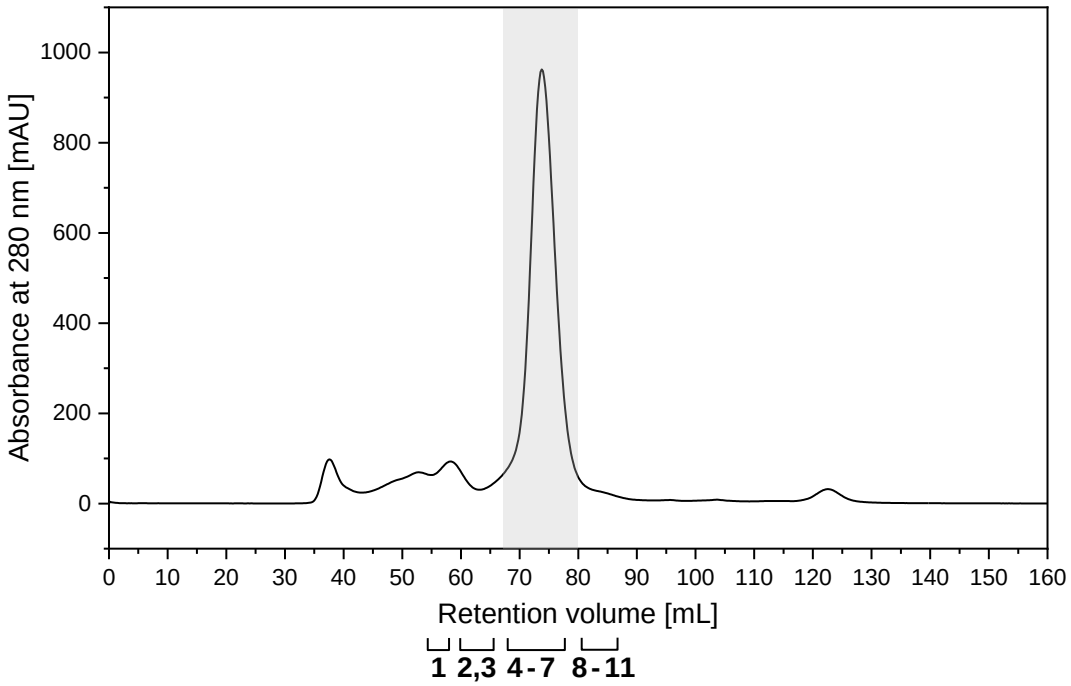


Figure 5.8-24: Chromatogram from the SEC purification of $FimH_L$.

The collected eluate fractions from the IMAC purification of $FimH_L$, shown in Figure 5.8-22, were concentrated to a final volume of 1.6 mL in a centrifugal filter (Vivaspin® Turbo, MWCO = 3 kDa), and applied onto a 120 mL SEC column (HiLoad® 16/600 Superdex® prep-grade column, GE Healthcare, cat. no. 28-9893-33) equilibrated with PBS. The indicated fractions **1** – **11** were further evaluated by SDS-PAGE (Figure 5.8-25). All fractions of the main peak (highlighted in grey), which contained the target protein, were pooled, concentrated to a final volume of 1.08 mL. Individual aliquots were used for chemical modification (Chapter 3.6) and mass spectrometric analysis (Chapter 3.7).

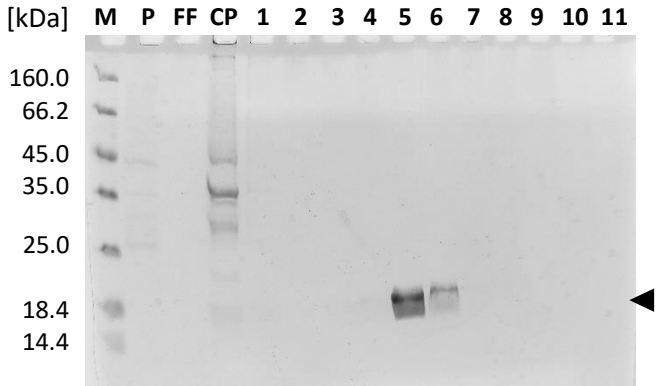


Figure 5.8-25: SDS-PAGE analysis to evaluate the SEC purification of $FimH_L$.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **P**: pooled fractions from the IMAC purification shown in Figure 7.7-22; **FF**, **CP**: flowthrough and concentrated pool after concentration with a centrifugal filter (MWCO = 3 kDa); **1** – **11**: collected fractions from the SEC purification shown in Figure 5.8-24. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 40 min at constant current (I = 30 mA). The black arrowhead indicates the migration bands corresponding to the overexpressed target protein $FimH_L$ (17.8 kDa).

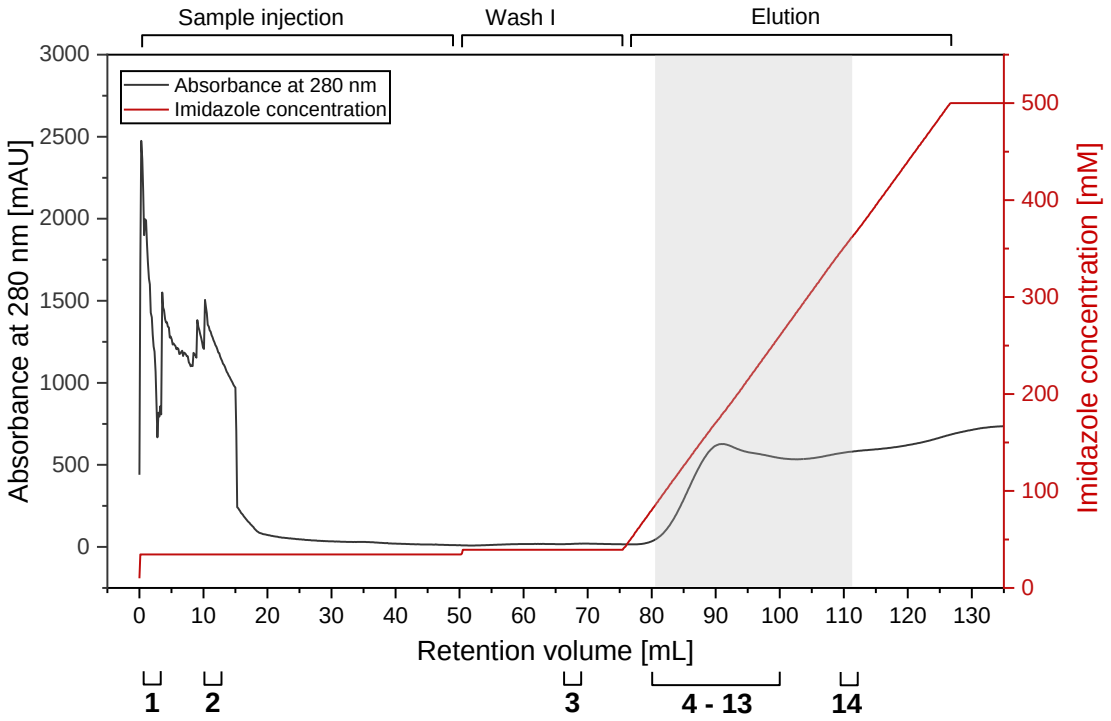


Figure 5.8-26: Chromatogram from the IMAC purification of FimH$_L$ S139C.

FimH$_L$ S139C was expressed in *E. coli* HM125 pPKL241 4 L LB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare). The column was washed with 39 mM imidazole (Wash I, 5 CV) prior to elution of the target protein via a linear imidazole gradient from 39 to 500 mM imidazole throughout 10 CV. Fractions **1** – **14** were analysed by SDS-PAGE (Figure 5.8-27). The main peak (fractions **4**– **14**), containing the target protein, is highlighted in grey.

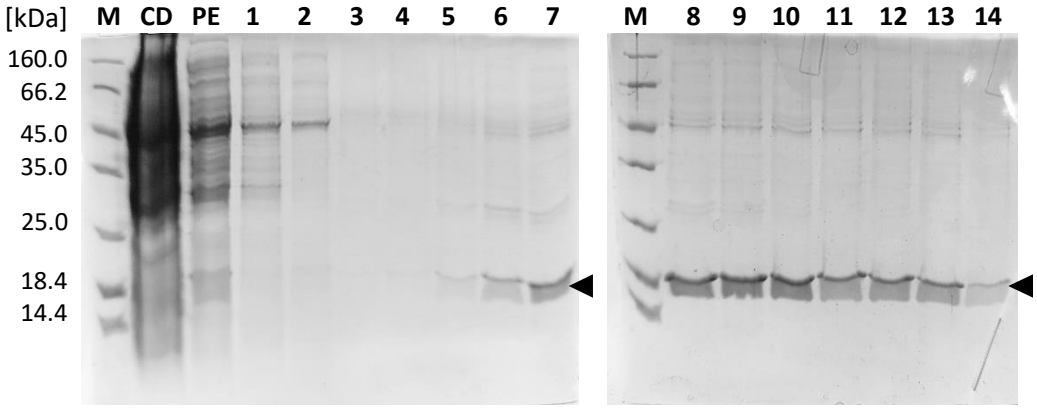


Figure 5.8-27: SDS-PAGE analysis to evaluate the IMAC purification of FimH$_L$.S139C.

M: Pierce™ Unstained Protein M$_W$ Marker (cat. no. 26610, Thermo Scientific™); **CD**: cell debris; **PE**: periplasmic extract; **1** – **14**: collected fractions from the IMAC purification shown in Figure 5.8-26. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 50 min at constant current (I = 25 mA/gel). Black arrowheads indicate the expression bands of the overexpressed target protein FimH$_L$ S139C (17.9 kDa).

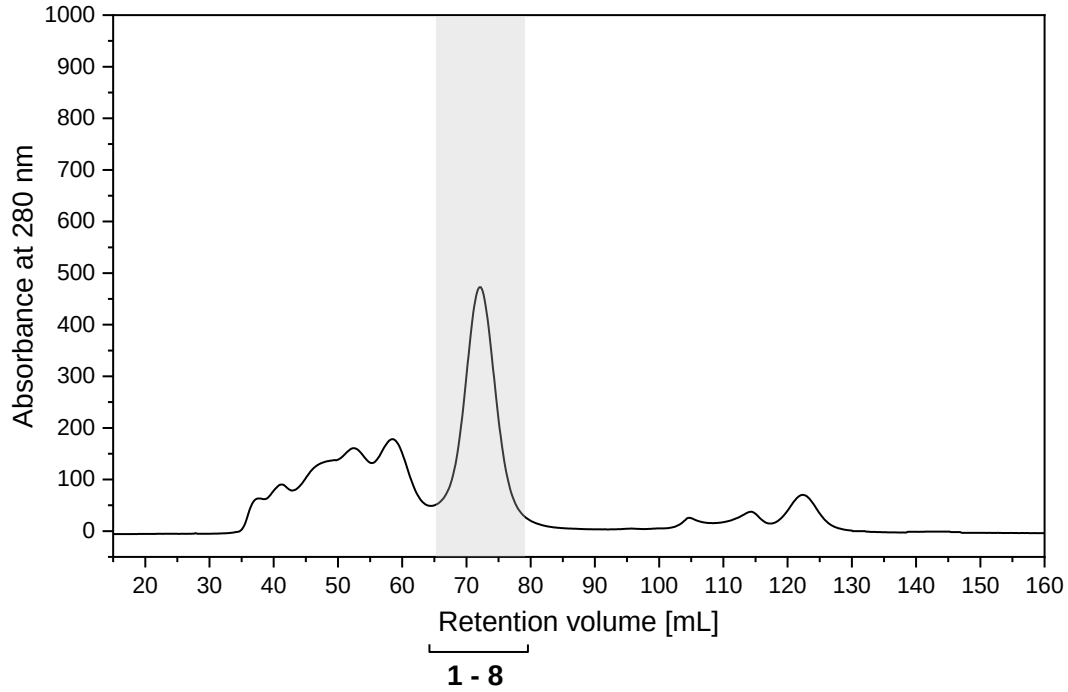


Figure 5.8-28: Chromatogram from the SEC purification of $FimH_L$ S139C.

The collected eluate fractions from the IMAC purification of $FimH_L$ S139C shown in Figure 5.8-26 were concentrated to a final volume of 2 mL in a centrifugal filter (Vivaspin® Turbo, MWCO = 3 kDa), and applied onto a 120 mL SEC column (HiLoad® 16/600 Superdex® prep-grade column, GE Healthcare, cat. no. 28-9893-33) equilibrated with PBS. The indicated fractions **1** – **8** were further evaluated by SDS-PAGE (Figure 5.8-27). All fractions of the main peak (highlighted in grey), which contained the target protein, were pooled and concentrated to a final volume of 1.08 mL. Individual aliquots were used for chemical modification (Chapter 3.6) and mass spectrometric analysis (Chapter 3.7).

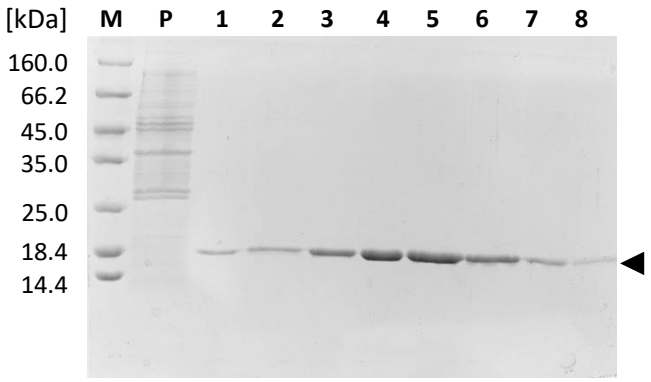


Figure 5.8-29: SDS-PAGE analysis to evaluate the SEC purification of $FimH_L$.S139C.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **P**: pooled fractions from the IMAC purification shown in Figure 5.8-26; **1** – **8**: collected fractions from the SEC purification shown in Figure 5.8-28. SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 40 min at constant current (I = 25 mA/gel). Black arrowheads indicate the migration bands corresponding to the overexpressed target protein $FimH_L$ S139C (17.9 kDa).

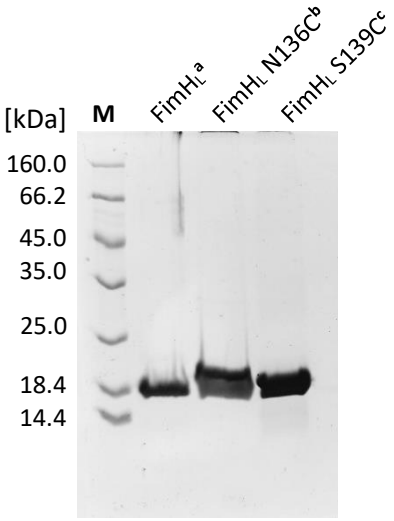


Figure 5.8-30: SDS-PAGE analysis to evaluate the purity of FimH$_L$ variants used in this thesis.

M: Pierce™ Unstained Protein M$_W$ Marker (cat. no. 26610, Thermo Scientific™). Protein concentrations: **a)** FimH$_L$, 2.59 mg/mL (PBS); **b)** FimH$_L$ N136C, 1.61 mg/mL (PBS); **c)** FimH$_L$ S139C 2.09 mg/mL (PBS). Expression and purification of the here visualised target proteins is documented in Chapter 3.5.1 (Figure 3.5-5 – 3.5-8) and in Chapter 5.8.5 (Figure 5.7-22 – 5.7-29). Individual aliquots were used for chemical modification (Chapters 3.6 and 5.7.7) and mass spectrometric analysis (Chapters 3.7 and 5.8.8). SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 40 min at constant current (I = 30 mA).

Table 5.8-10: Yields of FimH$_L$ variants obtained after expression in pPKL241-based expression systems and purified by IMAC (linear gradient).

Protein variant	**FimH$_L$**	**FimH$_L$ N136C ΔN-sigpep**	**FimH$_L$ N136C**	**FimH$_L$ S139C**
Total yield[a] [mg]	2.05[b,g] (PBS) 2.80[f,i] (PBS)	0.12[c,h] (PBS)	0.54[d,g] (PBS) 2.01[f,i] (PBS)	1.22[e,i] (PBS) 2.51[f,i] (PBS)
Yield[a]/cell mass [mg/g]	0.41[b,g] 0.21[f,i]	$6.73 \cdot 10^{-3}$ [c,h]	0.11[d,g] 0.07[f,i]	0.16[e,i] 0.18[f,i]
Yield[a]/expression volume [mg/L]	0.69[b,g] 0.70[f,i]	$7.85 \cdot 10^{-3}$ [c,h]	0.18[d,g] 0.40[f,i]	0.24[e,i] 0.50[f,i]

a) The protein concentration was determined on a NanoPhotometer® P300 (Implen). The variants were expressed in: **b)** HB101; **c)** Origami B pPKL241 ΔN-sigpep in 1.5 L TB medium (coning, expression and IMAC was performed by Sofie Todt as part of her Bachelor Thesis[487]); **d)** SG13009; **e)** C41 (DE3); **f)** HM125. Cell disruption was achieved by: **g)** high-pressure homogenisation (Emulsiflex C3, Avestin GmbH); **h)** sonication (Sonoplus, Bandelin electronic GmbH). **i)** The protein extract was obtained by periplasmic extraction. All FimH$_L$ variants were expressed in pPKL241-based expression systems, in 2 – 5 L LB medium (unless otherwise stated) and purified by IMAC (5 mL HisTrap HP), followed by concentration in a spin column (MWCO = 3 kDa) and SEC (HiLoad® 16/600 Superdex® pg, GE Healthcare) in PBS. The fractions/pools were again concentrated using a centrifugal filter (MWCO = 3 kDa).

5.8.6 Purification strategy for $FimH_{trc}$-His_6 variants encoded on pFimH-His_6-NT-trc-based constructs

Immobilised Metal Affinity Chromatography (IMAC) of $FimH_{trc}$-His_6 variants

The purification of C-terminally His_6-tagged $FimH_{trc}$ variants was likewise performed using the ÄKTA™ pure and ÄKTA™ purifier chromatography systems (GE Healthcare). The cleared periplasmic extracts containing the impure target proteins (Chapter 5.8.3) were applied onto a 1 mL or 5 mL Ni^{2+}-sepharose column (HisTrap™ HP column, GE Healthcare, cat. no. 17-5247, 17-5248), depending on the extracted cell mass and expected expression level of the respective target proteins. The column had been beforehand equilibrated with binding buffer (Table 5.7-1). If necessary, 10 % v/v glycerol was added to both binding buffer and elution buffer in order to reduce protein aggregation. Subsequent to sample application, the column was washed with 10 CV binding buffer under admixtion of 10 % elution buffer II (Table 5.7-1, Wash I, ≙ 25 mM imidazole), followed by another washing step with 20 % elution buffer II over 5 – 10 CV (Wash II, ≙ 50 mM imidazole). The target proteins were eluted with 100 % elution buffer II (250 mM imidazole) throughout 2 CV. Eligible fractions were analysed by SDS-PAGE (Chapter 5.7.7) and possibly unified or separately dialysed (MWCO = 10 kDa) overnight at 5 °C against assay buffer[376] or PBS. Finally, the individual fractions were aliquoted in 100 µL volumes and stored at 4 °C for early application. Alternatively, the aliquots were flash-frozen in liquid nitrogen for long-term storage at −80 °C. If the SDS-PAGE result revealed that a large part of the elution fractions still contained other, undesired proteins, the fractions were subjected to a further chromatographic purification step (SEC or AEC).

Preparative and analytical size exclusion chromatography (SEC) of $FimH_{trc}$-His_6 variants

The cysteine variant $FimH_{trc}$ N136C was required in the highest possible purity and concentration for subsequent chemical modification (Chapter 5.7.7) and analysis of the lectin activity in a FITC-*p*AP-Man binding assay (Chapter 5.7.9). Therefore, contaminated fractions from the IMAC purification step were carefully concentrated to a volume of ∼ 2 mL in a centrifugal filter (MWCO = 3 kDa, Amicon® Ultra, Millipore®, cat. no. UFC9003) following the supplier's specifications and under regular addition of 10 % v/v glycerol to the buffer (PBS). A solution of DTT (∼ 5 mM) was added and the protein concentrate was incubated for 5 min on ice. The reduced protein concentrate was loaded onto a 120 mL HiLoad® 16/600 Superdex® prep-grade column (GE Healthcare, cat. no. 28-9893-33) equilibrated with PBS and 10 % v/v glycerol. The obtained elution fractions with target protein were separately stored at 4 °C/ -80°C. Protein aliquots that showed no observable contamination with undesired proteins after SDS-PAGE (Chapter 5.7.7) were possibly subjected to analytical SEC in order to roughly test for possible oligomerisation states and if possible, to assign the monomeric state. For this purpose, about 150 µL of the protein samples were analysed using a 24 mL SEC column (Superdex® 200 Increase 10/300 GL, Cytiva, cat. no. 28-9909-44) with PBS and 10 % v/v glycerol in an ÄKTA go™ protein purification system (GE Healthcare). Fractions that contained only the pure, predominantly monomeric target protein were used for chemical labelling (Chapter 5.7.7) with subsequent testing for lectin activity (Chapter 5.7.9).

Anion exchange chromatography (AEC) of $FimH_{trc}$-His_6 N136C

As an alternative to SEC, the purity of the IMAC fractions was increased by anion exchange chromatography (AEC) in accordance with a frequently used approach, introduced by Glockhuber and co-workers, for the expression and purification of *fimH*-constructs without affinity tag.[480] For this purpose, contaminated fractions from the IMAC purification step were dialysed (MWCO = 10 kDa) overnight at 5 °C against AEC low-salt buffer (5 L, Table 5.7-1, 10 mM NaCl). The emerging white precipitate of aggregated protein was removed by centrifugation at 4200 rpm, 4 °C for 10 min (Centrifuge 5417 R, Eppendorf®). The supernatant was sterile-filtered through a bottle top filter (Filtropur BT 50, 0.1 µm, Sarstedt, cat. no. 83.3941.102) under reduced pressure. The cleared protein solution was loaded onto an AEC column (HiTrap™ Q FF, 1 mL, Cytiva, cat. no. 17505301; MonoQ®, 1 mL, Cytiva, cat. no. 17-5166-01) that had been beforehand equilibrated with AEC low-salt buffer in an ÄKTA go™ protein purification system (GE Healthcare). After sample application and washing with 10 CV of AEC low-salt buffer, the loaded protein species were separated according to the isoelectric point in a linear gradient from 0 % to 100 % AEC high-salt buffer (Table 5.7-1, ≙ 10 mM – 1 M NaCl) throughout 30 – 50 CV (Theoretical $pI_{FimH\ N136C}$: 6.7, as calculated by the web-based Protparam tool, https://web.expasy.org/protparam/). The purity of resulting fractions was evaluated by SDS-PAGE (Chapter 5.7.7) and carefully re-buffered in PBS with 10 % v/v glycerol using a centrifugal filter (MWCO = 10 kDa, Amicon® Ultra, Millipore®, cat. no. UFC9010).

Exemplary results from the expression and purification of $FimH_{trc}$-His_6 variants

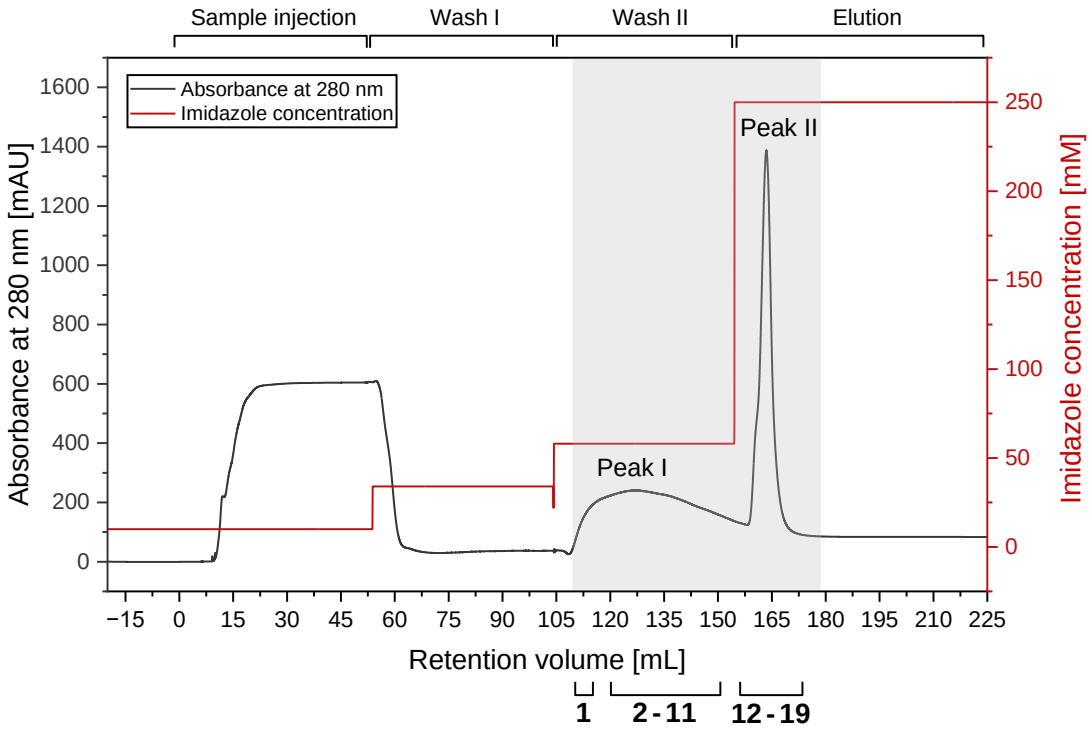


Figure 5.8-31: Chromatogram from the IMAC purification of $FimH_{trc}$.

$FimH_{trc}$ was expressed in *E. coli* HM125 pFimH-His_6-NT-trc in 2 L TB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare). The column was washed with 25 mM imidazole (Wash I, 10 CV) and 50 mM imidazole (Wash II, 10 CV) prior to elution with 250 mM imidazole (step gradient). The fractions **1** – **19** were analysed by SDS-PAGE (Figure 5.8-32). The washing step II (fractions **1** – **11**) and the main elution peak II (fractions **12** – **19**) containing the purified target protein are highlighted in grey.

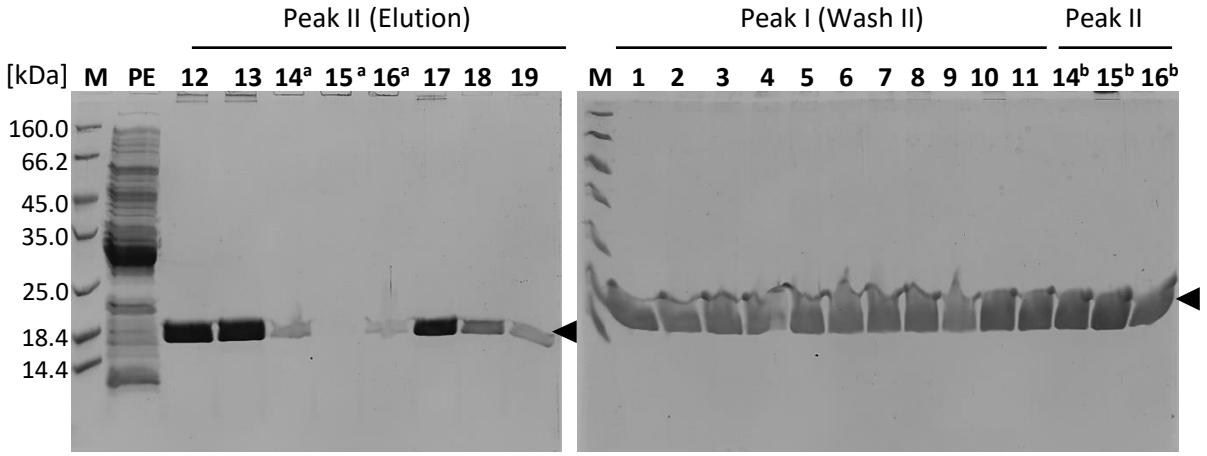


Figure 5.8-32: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_{trc}$.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **19**: collected fractions from the IMAC purification shown in Figure 5.8-31. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 25 mA/gel). Black arrowheads indicate the expression bands of the target protein $FimH_{trc}$ (17.8 kDa). **a)** Samples aggregated and did not migrate through the polyacrylamide gel, due to high concentration of target protein. **b)** Samples were 1:3-diluted with water prior to denaturation (95 °C) and subsequent sample application.

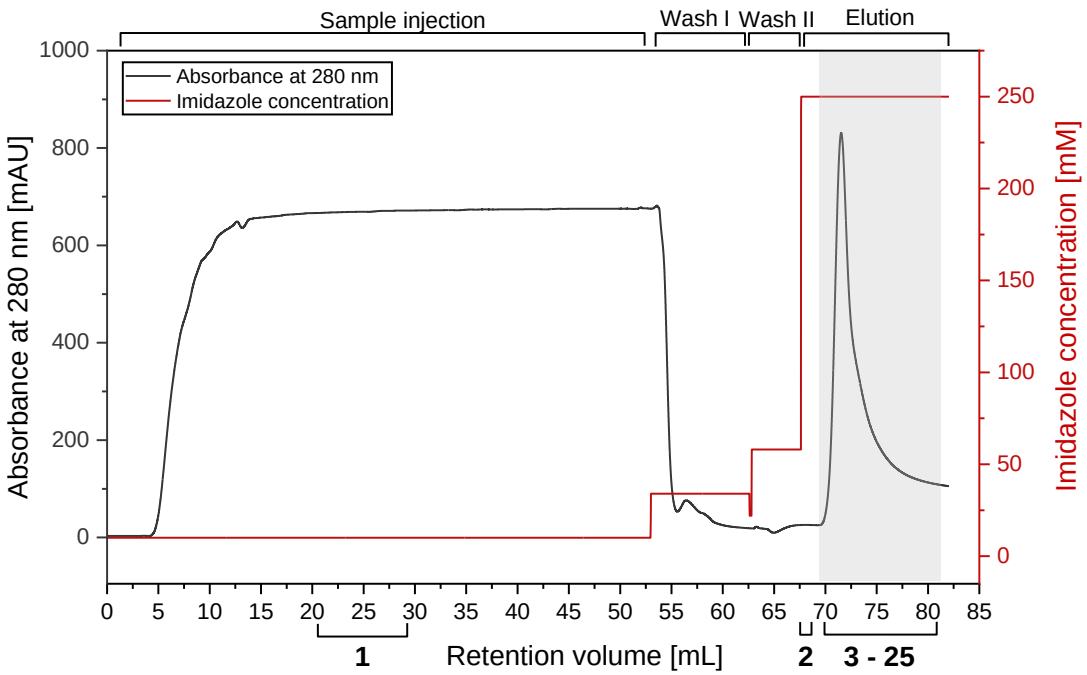


Figure 5.8-33: Chromatogram from the IMAC purification of $FimH_{trc}$ R60P.

$FimH_{trc}$ R60P was expressed in *E. coli* HM125 pFimH-His_6-NT-trc in 1.5 L TB medium. The protein extract was obtained by periplasmic extraction and purified using a 1 mL HisTrap™ HP column (GE Healthcare). The column was washed with 25 mM imidazole (Wash I, 10 CV) and 50 mM imidazole (Wash II, 5 CV) prior to elution with 250 mM imidazole (step gradient). Fractions **1** – **25** were analysed by SDS-PAGE (Figure 5.8-34). The main peak (fractions **3** – **25**), containing the purified target protein, is highlighted in grey.

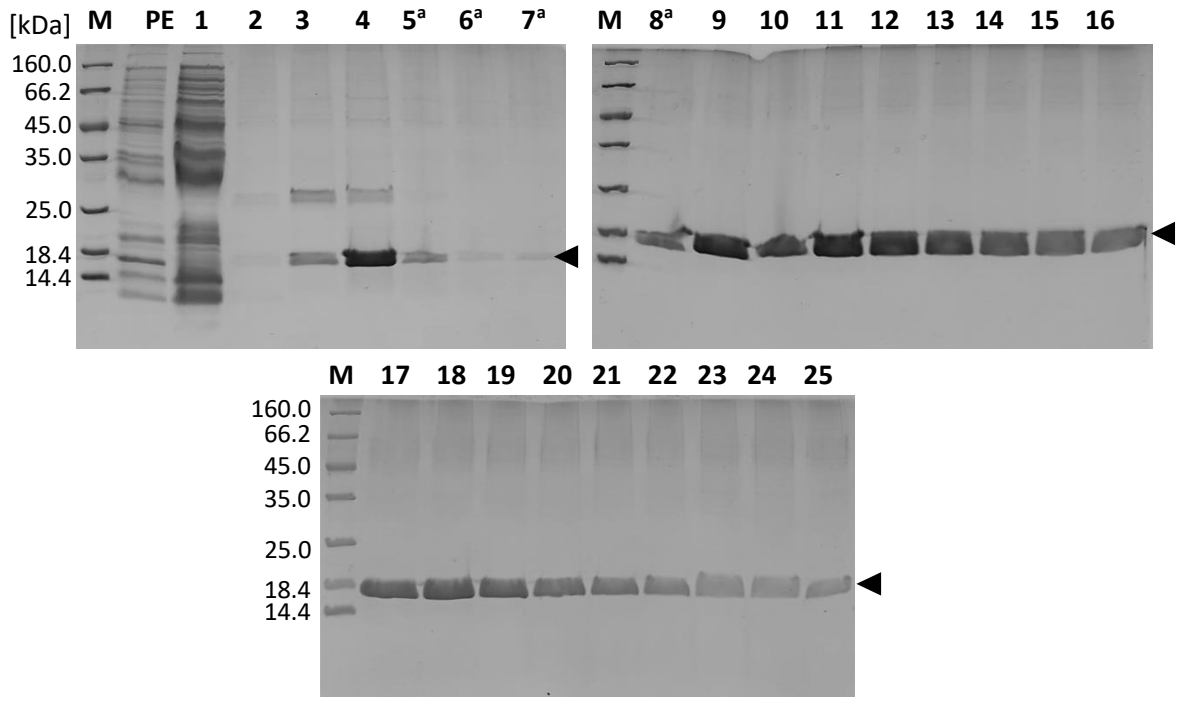


Figure 5.8-34: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_{trc}$ R60P.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **25**: collected fractions from the IMAC purification shown in Figure 5.8-34. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 50 min at constant current (I = 30 mA/gel). Black arrowheads indicate the expression bands of the overexpressed target protein $FimH_{trc}$ R60P (17.8 kDa). **a)** Samples aggregated and did not migrate through the polyacrylamide gel, due to high concentration of target protein.

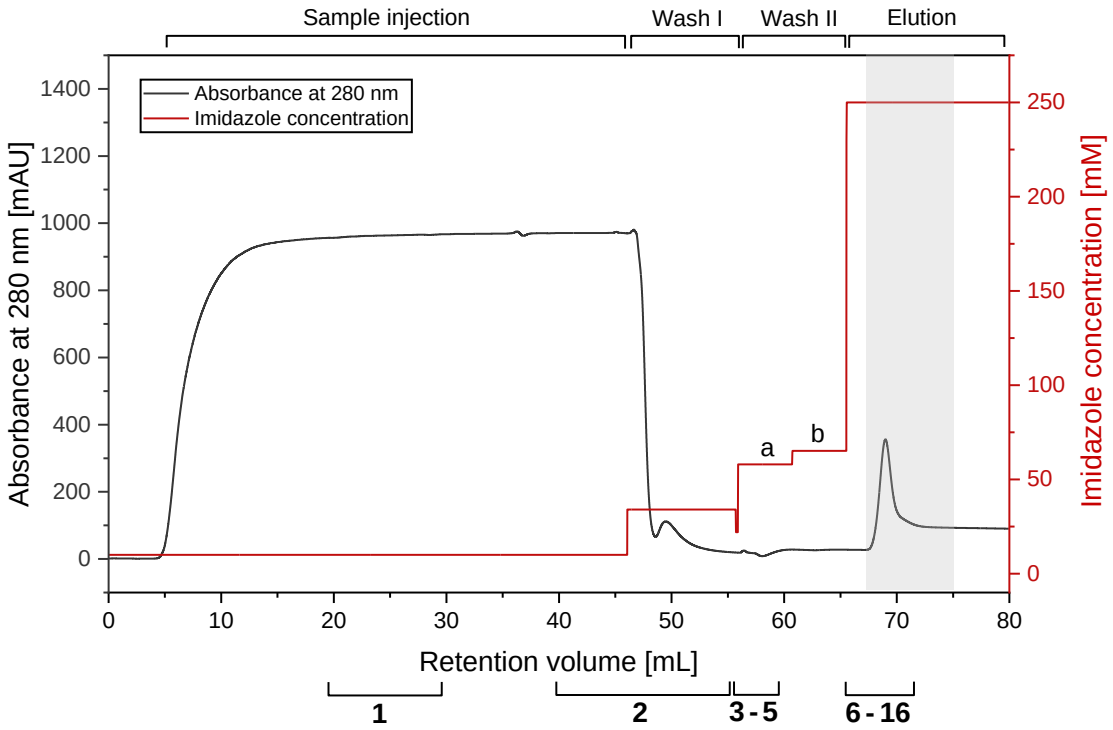


Figure 5.8-35: Chromatogram from the IMAC purification of $FimH_{trc}$ S139C.

$FimH_{trc}$ S139C was expressed in *E. coli* HM125 pFimH-His_6-NT-trc in 1.6 L TB medium. The protein extract was obtained by periplasmic extraction and purified using a 1 mL HisTrap™ HP column (GE Healthcare). The column was washed with 25 mM imidazole (Wash I, 10 CV), 50 mM imidazole (Wash IIa, 5 CV) and 57 mM imidazole (Wash IIb, 5 CV) prior to elution with 250 mM imidazole (step gradient). Fractions **1** – **16** were analysed by SDS-PAGE (Figure 5.8-36). The main peak (fractions **11** – **16**), containing the purified target protein, is highlighted in grey.

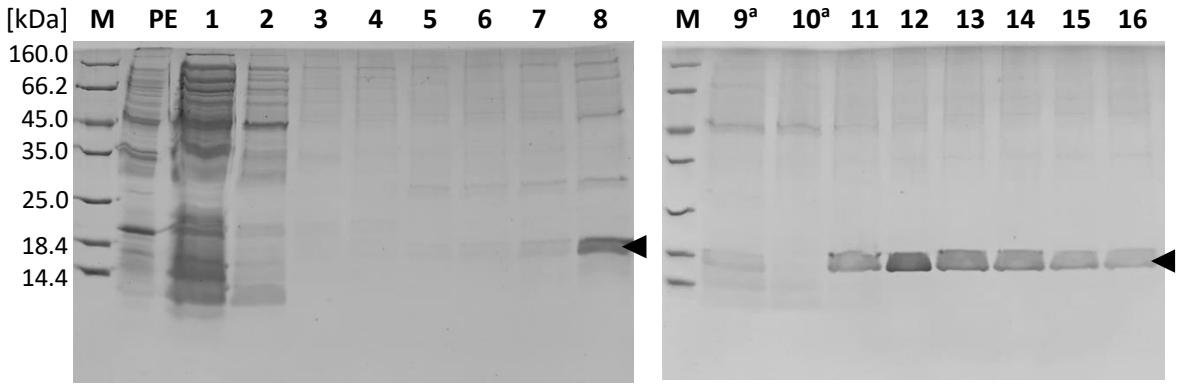


Figure 5.8-36: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_{trc}$ S139C.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **16**: collected fractions from the IMAC purification shown in Figure 5.8-35. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 50 min at constant current (I = 25 mA/gel). Black arrowheads indicate the expression bands of the overexpressed target protein $FimH_{trc}$ S139C (17.9 kDa). **a)** Samples aggregated and did not migrate through the polyacrylamide gel, due to high concentration of target protein.

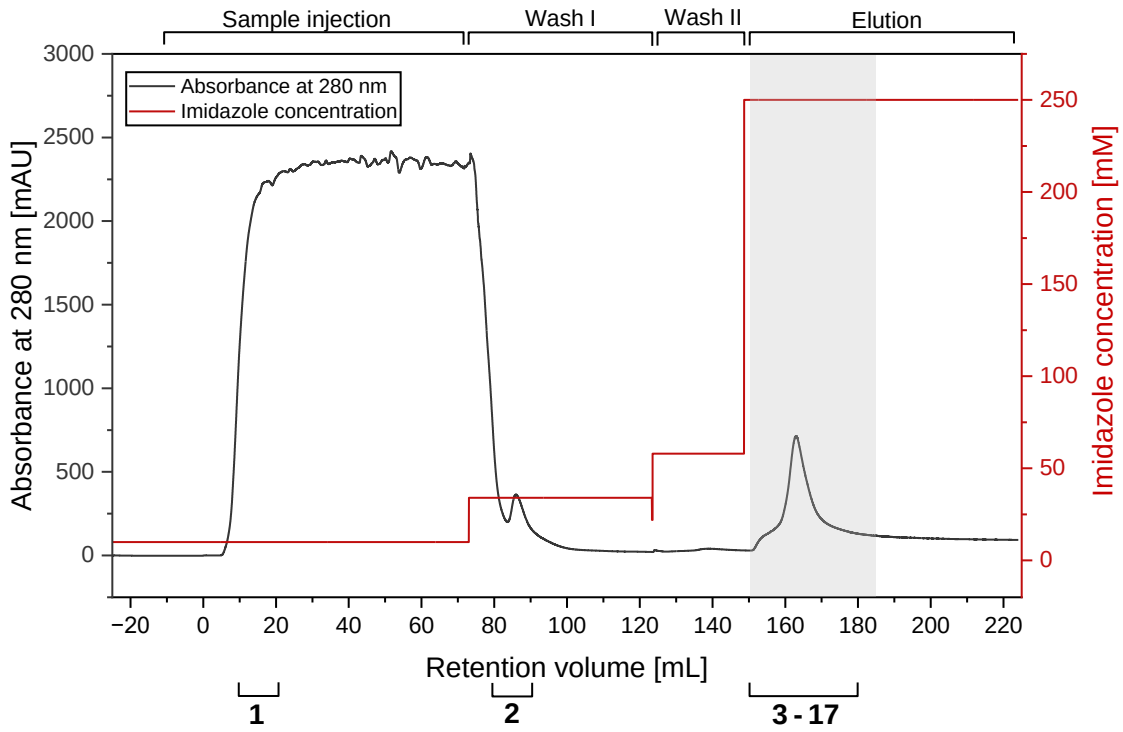


Figure 5.8-37: Chromatogram from the IMAC purification of $FimH_{trc}$ N136C.

$FimH_{trc}$ N136C was expressed in *E. coli* HM125 pFimH-His_6-NT-trc in 2 L TB medium. The protein extract was obtained by periplasmic extraction and purified using a 5 mL HisTrap™ HP column (GE Healthcare). The column was washed with 25 mM imidazole (Wash I, 10 CV) and 50 mM imidazole (Wash II, 5 CV) prior to elution with 250 mM imidazole (step gradient). Fractions **1** – **17** were analysed by SDS-PAGE (Figure 5.8-38). The main peak (fractions **3** – **17**), containing the target protein, is highlighted in grey.

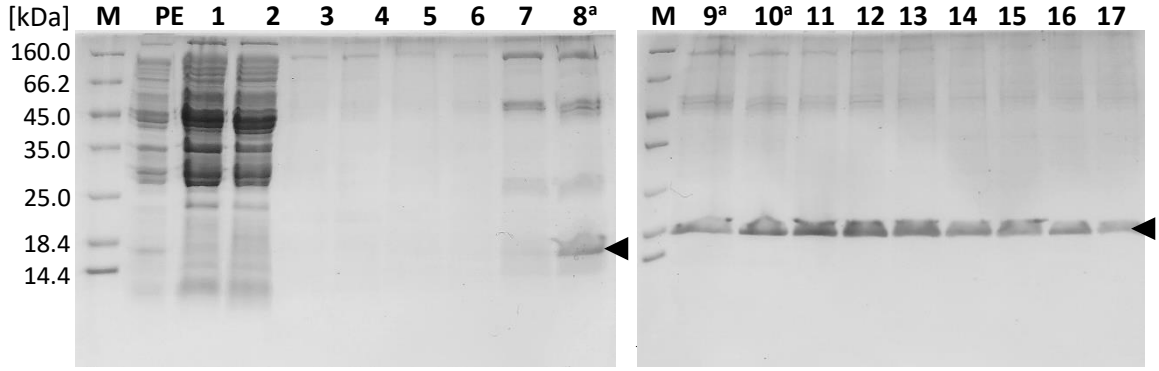


Figure 5.8-38: SDS-PAGE analysis to evaluate the IMAC purification of $FimH_{trc}$ N136C.

M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); **PE:** periplasmic extract; **1** – **17**: collected fractions from the IMAC purification shown in Figure 5.8-37. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 27.5 mA/gel). Black arrowheads indicate the expression bands of the target protein $FimH_{trc}$ N136C (17.8 kDa). **a)** The smeary migration behaviour results from the gel-like consistency of the SDS-PAGE samples, due to the high concentration of target protein.

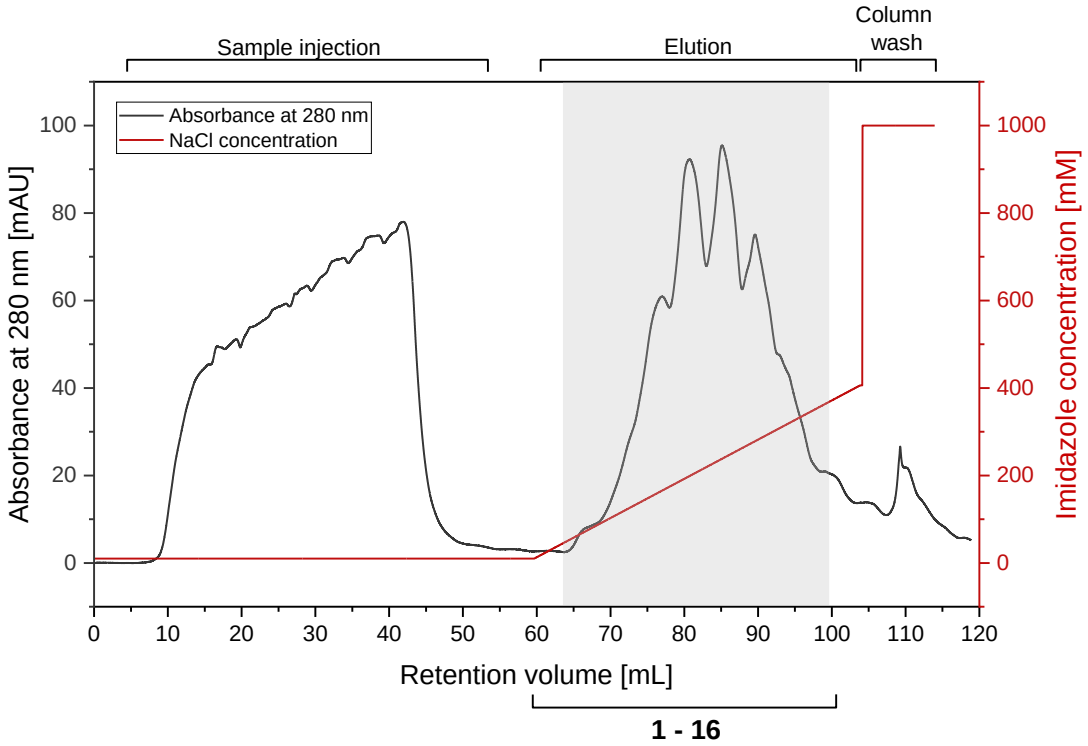


Figure 5.8-39: Chromatogram from the AEC purification of FimH$_{trc}$ N136C.

The collected eluate fractions from the IMAC purification of FimH$_{trc}$ N136C shown in Figures 5.7-37 – 5.7-38 were dialysed against AEC low-salt buffer and the emerging protein aggregate was removed by centrifugation and sterile-filtration. The target protein was purified using an AEC column (MonoQ®, 1 mL Cytiva cat. no. 17-5166-01) in a linear NaCl gradient from 10 mM to 400 mM throughout 45 CV. The fractions **1** – **16** (highlighted in grey), which contained the purified target protein, were subjected to SDS-PAGE (Figure 5.8-40).

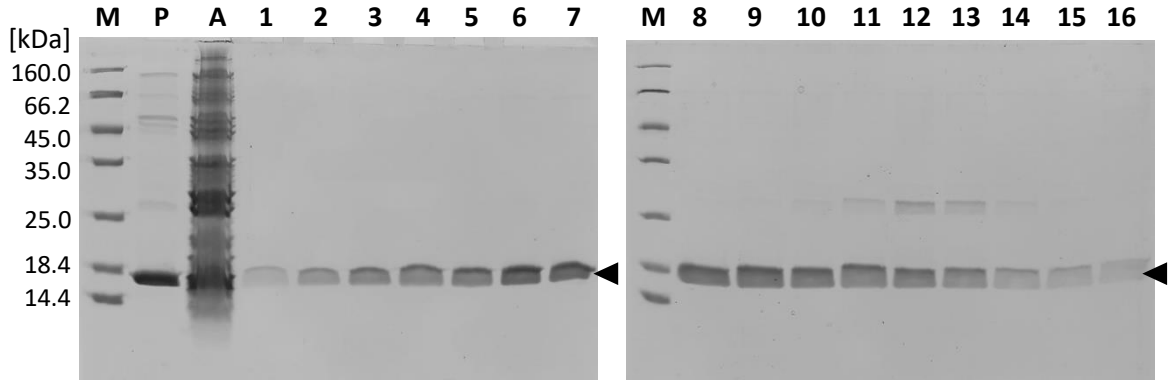


Figure 5.8-40: SDS-PAGE analysis to evaluate the AEC purification of FimH$_{trc}$ N136C.

M: Pierce™ Unstained Protein M$_W$ Marker (cat. no. 26610, Thermo Scientific™); **P:** pooled fractions resulting from the IMAC purification shown in Figure 5.8-39; **A**: protein aggregate, which emerged during dialysis in AEC low-salt buffer; **1** – **16**: collected fractions from the AEC purification shown in Figure 5.8-39. SDS-polyacrylamide gels (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 27.5 mA/gel). Black arrowheads indicate the overexpressed target protein FimH$_{trc}$ N136C (17.8 kDa).

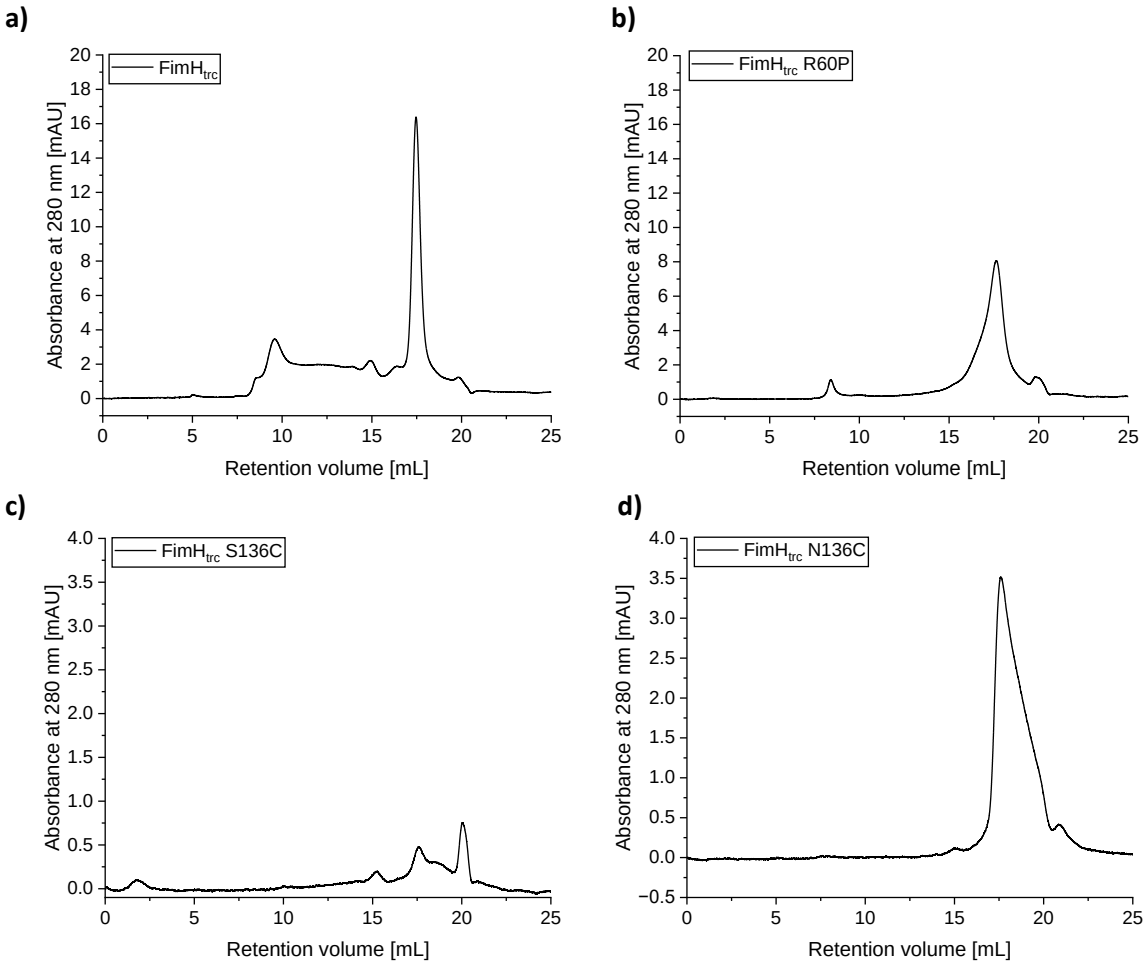


Figure 5.8-41: Chromatograms resulting from the analytical SEC of FmH$_{trc}$ variants.

a) SEC result of the pooled eluate fractions obtained by IMAC purification of FimH$_{trc}$ (Figure 5.8-31, c = 1 mg/mL); **b)** SEC result of the pooled eluate fractions obtained by IMAC purification of FimH$_{trc}$ R60P (Figure 5.8-33, c = 0.45 mg/mL); **c)** SEC result of the pooled eluate fractions obtained by IMAC purification of FimH$_{trc}$ S139C (Figure 5.8-35, c = 0.2 mg/mL); **d)** fraction peak from the SEC purification of FimH$_{trc}$ N136C (Figures 3.5-7 – 3.5-8, c = 0.46 mg/mL) containing the monomeric (and covalent-dimeric) target protein. The protein solutions (150 µL) were applied onto an analytical SEC column (Superdex® 200 Increase 10/300 GL, Cytiva, cat. no. 28-9909-44) equilibrated with PBS and 10 % v/v.

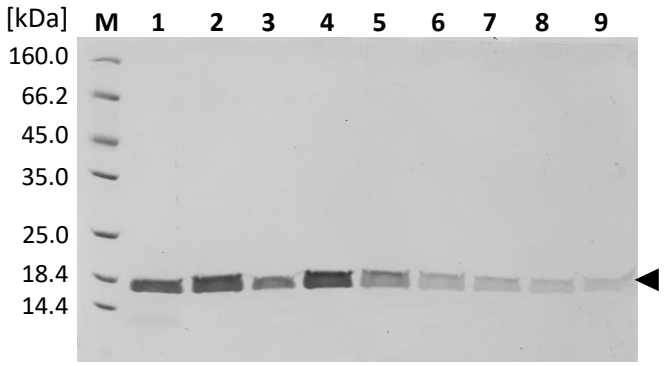


Figure 5.8-42: SDS-PAGE analysis to evaluate the SEC purification of FimH$_{trc}$ N136C, discussed in Chapter 3.5.2.

M: Pierce™ Unstained Protein M$_W$ Marker (cat. no. 26610, Thermo Scientific™); **1** – **9**: fractions of Peak IV resulting from SEC purification, illustrated in Figure 3.5-7. SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 30 mA). The black arrowhead indicates the purified target protein FimH$_{trc}$ N136C (17.8 kDa).

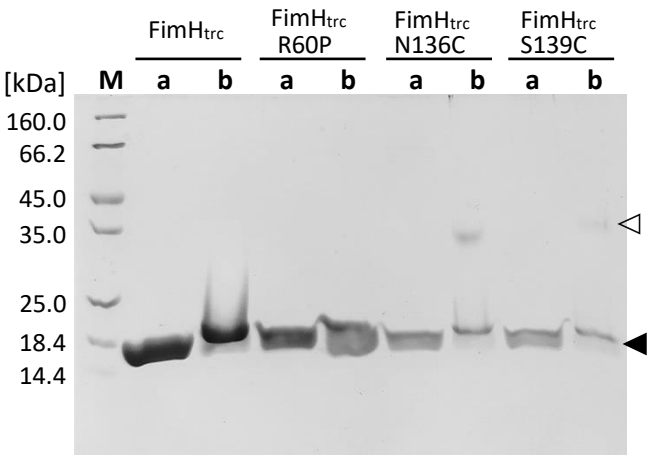


Figure 5.8-43: Results of reducing and non-reducing SDS-PAGE of $FimH_{trc}$ variants used in this thesis.
M: Pierce™ Unstained Protein M_W Marker (cat. no. 26610, Thermo Scientific™); Protein concentrations: $FimH_{trc}$: 1 mg/mL (assay buffer), $FimH_{trc}$ R60P: 0.45 mg/mL (PBS), $FimH_{trc}$ N136C: 0.3 mg/mL (PBS); $FimH_{trc}$ S139C: 0.2 mg/mL (PBS). The black arrowhead indicates the migration bands corresponding to the monomeric target proteins (17.8 -17.9 kDa). The white arrowhead indicates the migration bands corresponding to the disulfide-linked dimers of the cysteine variants (35.6- 35.8 kDa). Expression and purification of the here visualised target proteins is documented in Chapter 3.5.1 (Figures 3.5-5 – 3.5-8) and in Chapter 5.8.6, (Figures 5.7-31 – 5.7-40). Individual aliquots were used for chemical modification (Chapter 3.6) and in FITC-*p*AP-Man binding assays (Chapter 3.8). **a)** Samples were denatured at 95 °C for 5 min in the presence of β-mercaptoethanol (10 % v/v, reducing SDS-PAGE). **b)** Samples were not heated and dissolved in non-reducing Laemmli buffer (non-reducing SDS-PAGE). SDS-polyacrylamide gel (15 %) stained with Coomassie Brilliant Blue R-250. Electrophoresis 55 min at constant current (I = 25 mA).

Table 5.8-11: Yields of $FimH_{trc}$ variants obtained after expression in HM125 pFimH-His_6-NT-trc and purified according to the optimised IMAC protocol (step gradient).

Protein variant	**$FimH_{trc}$**	**$FimH_{trc}$ R60P**	**$FimH_{trc}$ N136C**	**$FimH_{trc}$ S139C**
Total yield[a] [mg]	30.4 [c] (assay buffer)[h]	1.98 [d] (PBS)[h]	3.6 [c] (impure) 2.3 [e] (PBS/glycerol 9:1)[h] 3.8 [f] (impure) 2.42 [g] (PBS/glycerol 9:1)[h]	0.70 [d] (PBS)[h]
Yield[a]/cell mass [mg/g]	0.79 [c]	0.06[d]	0.04 [e] 0.062 [g]	0.03[d]
Yield[a]/expression volume [mg/L]	15.2 [c] (Lit. 11 mg in M9 minimal medium)[480]	1.26[d]	0.58 [e] 0.81 [g]	0.43[d]
Highest concentrations in individual fractions/ pools [mg/L]	1.0 [a,c], (1.52 [b,c])	0.45 [d], (1.0 [b,d])	0.49 [a,e], (0.85 [b,e]) 0.41 [a,g], (0.46 [b,g])	0.2 [a,d], (0.2 [b,d])

a) The protein concentration was in general determined on a NanoPhotometer® P300 (Implen). **b)** The protein concentration of the same pool/fraction was also determined in a BCA assay (Pierce™ Thermo Scientific™). The $FimH_{trc}$ variant was purified by **c)** IMAC (5 mL HisTrap HP); **d)** IMAC (1 mL HisTrap HP); **e)** IMAC (5 mL HisTrap HP), followed by AEC (MonoQ®, 1 mL, Cytiva) and concentrated in a spin column (MWCO = 10 kDa); **f)** IMAC (5 mL HisTrap HP), followed by AEC (HiTrap™ Q FF, 1 mL, Cytiva); **g)** IMAC (5 mL HisTrap HP) in the presence of 10 % v/v glycerol, followed by concentration in a spin column (MWCO = 10 kDa) and SEC (HiLoad® 16/600 Superdex® pg, GE Healthcare) in PBS with 10 % v/v glycerol. **h)** The fractions/pools were dialysed or rebuffered in a spin column (MWCO = 10 kDa) against the respective buffer. All FimH variants were expressed in HM125 pFimH-His_6-NT-trc in 1.5 – 3 L TB medium at 30 °C and the protein extract was obtained by periplasmic extraction.

5.8.7 Chemical modification of FimH with photoswitchable iodoacetamide derivatives

The FimH cysteine variants N136C and S139C were site-selectively labelled with IAA-functionalised azobenzene and 3-fluoroazobenzene mannosides **H1**, **H2**, **F1**, and **F2**. For this purpose, the following procedure was pursued: "Basic protocol 1: alkylation of a protein of known size and composition with haloacyl reagents or *N*-ethylmaleimide", which was introduced by Crankshaw and Grant.[218] For the calculation of the molar amount of protein disulfide that was intended for alkylation, only the free cysteine residues Cys136 and Cys139, introduced by site-directed mutagenesis, but not the intrachain disulfide bridge between Cys3 and Cys44 were considered:

$$Protein\ disulfide\ [\mathrm{nmol}] = \ Protein\ mass\ [\ \mu\mathrm{g}] \cdot \left(\frac{Protein\ amount\ [\mathrm{nmol}]}{Protein\ mass\ [\ \mu\mathrm{g}]}\right) \cdot \ 0.5\ Disulfides \cdot$$

Aliquots of the cysteine variants (100 µL or 500 µL, respectively; protein concentrations: 0.49 – 2.6 mg/mL) in PBS (Table 5.7-1) were transferred to 1.5 mL Eppendorf tubes and reduced with 10 equivalents of dithiothreitol (DTT) per expected protein disulfide.

$$DTT\ [\mathrm{nmol}] = \ Protein\ disulfide\ [\mathrm{nmol}] \ \cdot \ 10$$

The tubes were flushed for 2 min with gaseous nitrogen through a cannula that was inserted through the lid of the Eppendorf tube. The latter was then carefully sealed with Parafilm® and incubated for 1 h at 37 °C (Accu Block™ Digital Dry Bath, Labnet Inc.). The iodoacetamide-derivatives were added in a fivefold excess over the total number of sulfhydryl groups, corresponding to both DTT- and cysteine thiols:

$$Alkylating\ agent\ [nmol] = \big((DTT\ [nmol] \cdot 2) + (Protein\ disulfide\ [\mathrm{nmol}] \cdot 2)\big) \ \cdot \ 5$$

The tubes were wrapped in aluminium foil in order to exclude light exposure and thus iodine formation. The samples were once again flushed with nitrogen for 2 min, sealed with Parafilm® and incubated for 1 h at 37 °C. The alkylation reaction was quenched with a tenfold molar excess of 2-mercaptoethanol over the molar amount of iodoacetamide derivatives. Finally, the alkylated proteins were desalted using Zeba™ spin columns (MWCO = 7 kDa, 500 µL, cat. no. 89882; 5 mL, cat. no. 89891, Thermo Scientific™) according to the manufacturer's recommendations and by centrifugation at 1500 rpm, 4 °C, 2 min (Universal 320R centrifuge, Hettich®). If the alkylation attempt was performed for a subsequent top-down or bottom-up LC-MS/MS analysis (Chapter 5.8.8), quenching and desalting of the alkylation reaction was not necessary. Alkylated FimH aliquots were stored at 4 °C for early application or at –80 °C for long-term storage.

5.8.8 Mass spectrometric analysis of FimH variants

Evaluation of chemically modified FimH variants by LC-MS/MS

The cysteine variants of $FimH_L$ investigated in this work were expressed in *E. coli* HM125 pPKL241 unless otherwise stated. The expression and purification of $FimH_L$ N136C and S139C are described in Chapter 5.8.5. Subsequent chemical modification with the iodoacetamide-functionalised azobenzene and 3-fluoroazobenzene mannoside derivatives (**H1**, **H2**, **F1**, and **F2**) was performed as described in Chapter 5.7.7. All proteomics measurements and the resulting experimental data, presented in this chapter, were performed, analysed, and compiled by Dr. Philipp T. Kaulich under the supervision of Prof. Dr. Andreas Tholey from the Institute of Experimental Medicine (Kiel University).

For top-down proteomics, the protein samples (30 µg) were precipitated in a mixture of methanol, chloroform and water[710] and resuspended in a solution containing 3 % v/v acetonitrile and 0.1 % v/v TFA (300 µL). In some cases, unmodified cysteine variants were reduced in the presence of 20 mM TCEP, for 10 min at 60 °C[497] prior to precipitation and resuspension. The sample solutions (1 µL) were injected into a Dionex UltiMate™ 3000 UHPLC system (Thermo Scientific™) equipped with a C4 reversed-phase analytical column (50 cm × 75 µm, 2.6 µm, 150 Å, Thermo Scientific™) and a C4 pre-column (C4 PepMap™ 300, 5 µm, 300 Å, Thermo Scientific™) for proteoform separation and sample-loading, respectively. The proteoforms were analysed in an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Scientific™) using the multi-compensation voltage (CV) field asymmetric ion mobility spectrometry (FAIMS) method which was introduced for top-down proteomics approaches by Philipp Kaulich and co-workers.[489] The method takes the advantage of improved proteoform fractionation in an electric field with asymmetric waveform and with high and low electric fields at radio frequency (FAIMS), combined with multiple CVs per run in order to optimise the signal-to-noise ratio and increase the sensitivity of the measurement. The CVs correct the trajectory of the ions, which become fractionated in the asymmetric electric field and thus enable the entry into the mass spectrometer with strong dependence between proteoform mass and selected voltage. Therefore, multiple measurements with different CVs are required to cover a full mass range. Intact-mass spectra were recorded with the following settings: mass range 400 – 1800 *m*/*z*, resolution 120 000, normalised AGC target 200 %, maximum injection time 246 ns, 4 microscans. Fragment spectra were recorded with a cycle time of 3 s, charge states 4 – 50, dynamic exclusion 60 s, n = 2, 150 – 2000 *m*/*z*, isolation window 5 *m*/*z*, resolution 60 000, normalised AGC target 1000 %, maximum injection time 250 ms, 4 microscans. The following CVs were applied: −40, −30, −20, 0 V, collision-induced dissociation (CID, 25 %); −40 V, higher-energy-collision induced dissociation (HCD), stepped (15, 20, 25 %) and fixed (25 %), electron-transfer/higher-energy collision dissociation (EThcD, 10 ms, 23 %), CID-assisted (15, 30, 45 %). TCEP-reduced samples were measured without interposed fractionation by FAIMS. The following software was used for data analysis: FreeStyle™ 1.8, Flexera, Thermo Scientific™ (Inspection of the raw files); FLASHDeconv 2.0,[711] with default settings (deconvolution of raw files); pyOpenMS[712] (theoretical isotope distribution); MSTopDiff 1.1.0[490] (detection of over- or underlabelling); ProSightPD, Thermo Scientific™ (Customised proteoform database including all possible modifications, Deconvolution with Xtract default settings, annotated and subsequence search, filtered PrSMs and proteoforms at a false discovery rate of 1 %).

In the bottom-up approach, the protein samples were electrophoretically separated on a 16 % SDS-PAGE and the bands, containing the target proteins, were excised from the gel. The proteins within the shredded SDS-PAA gel pieces were reduced with DTT (10 mM, 45 min, 56 °C), alkylated with 2-iodoacetamide (55 mM, 30 min at room temperature under exclusion of light) and trypsinised (100 ng trypsin, 37 °C, overnight). The peptides were extracted from the SDS-PAA gel and analysed using a Q Exactive™ Plus Orbitrap mass spectrometer (Thermo Scientific™, 2 h gradient, top 15 method, HCD, 27.5 normalised collision energy). Data analysis was performed using the ProteomeDiscoverer 2.2 (Thermo Scientific™, SequestHT precursor tolerance 10 ppm, fragment tolerance 0.02 Da; fixed-value PSM validator).

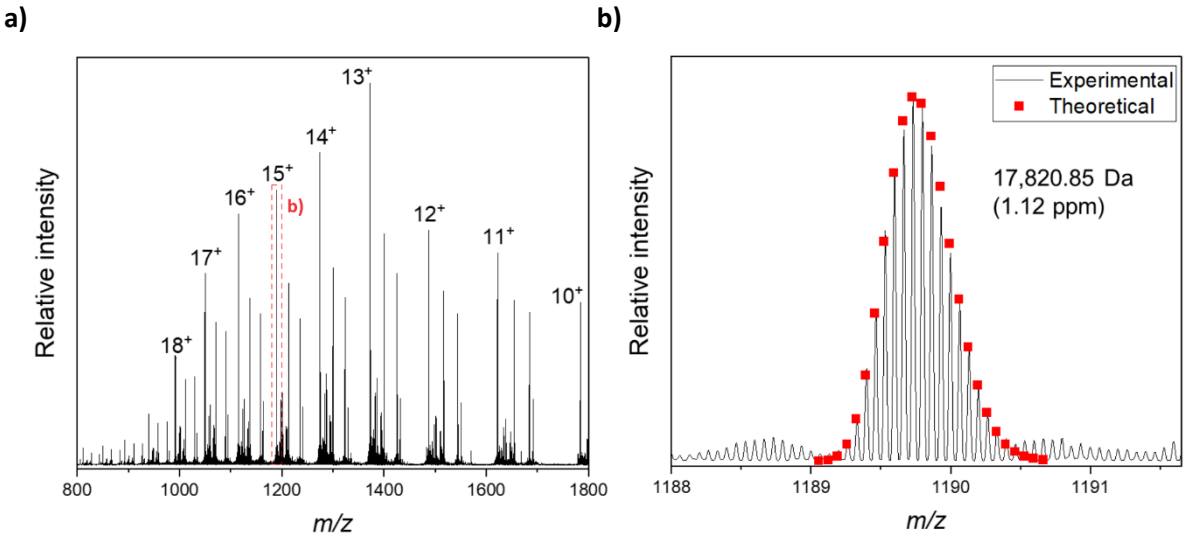


Figure 5.8-44: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of unmodified $FimH_L$ N136C without prior reduction.

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

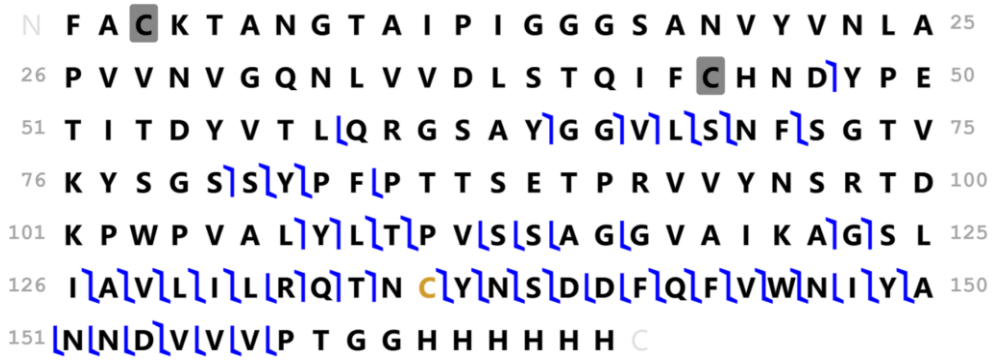


Figure 5.8-45: Fragment map of the identified proteoform of unmodified $FimH_L$ N136C without prior reduction.

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis.

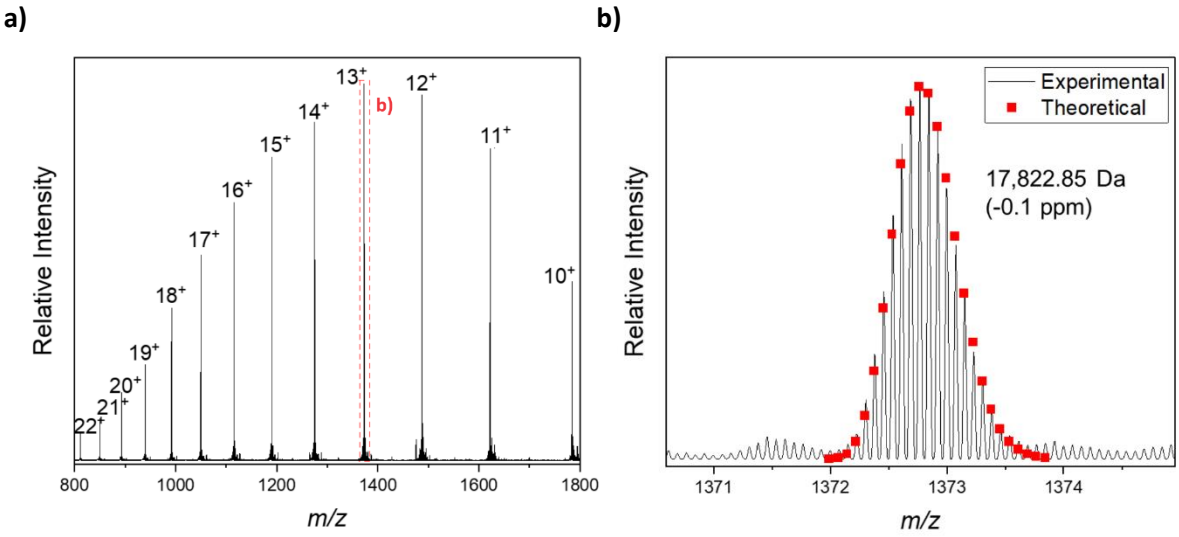


Figure 5.8-46: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of unmodified $FimH_L$ N136C after reduction with 20 mM TCEP, 10 min at 60 °C.[497]

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

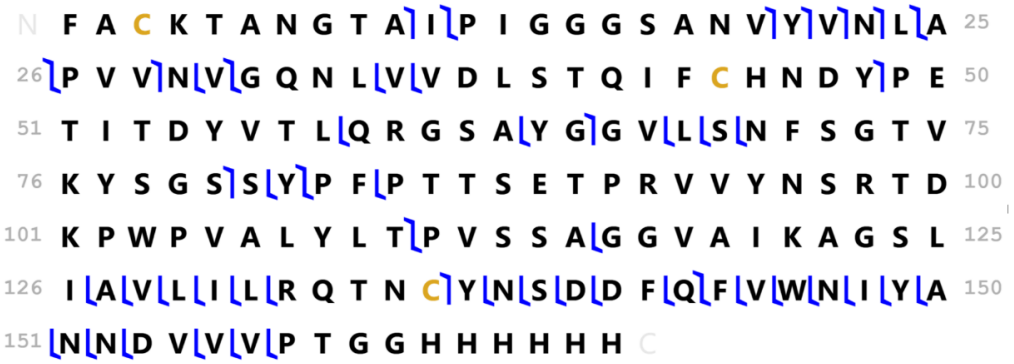


Figure 5.8-47: Fragment map of the identified proteoform of unmodified $FimH_L$ N136C after reduction with 20 mM TCEP, 10 min at 60 °C.[497]

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); yellow: reduced cysteine residues.

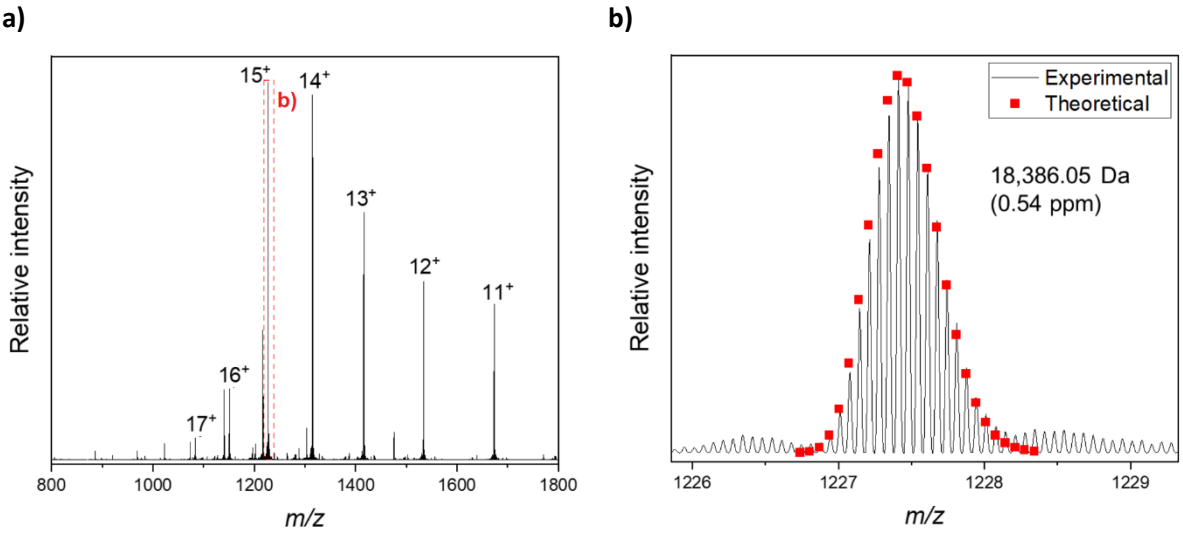


Figure 5.8-48: **a)** Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ N136C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F2**.
The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

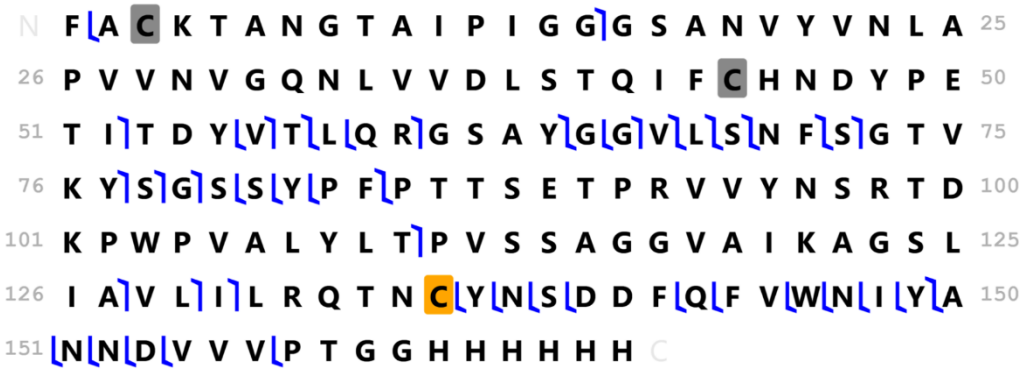


Figure 5.8-49: Fragment map of the identified proteoform of $FimH_L$ N136C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F2**.
The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis and alkylated with **F2**.

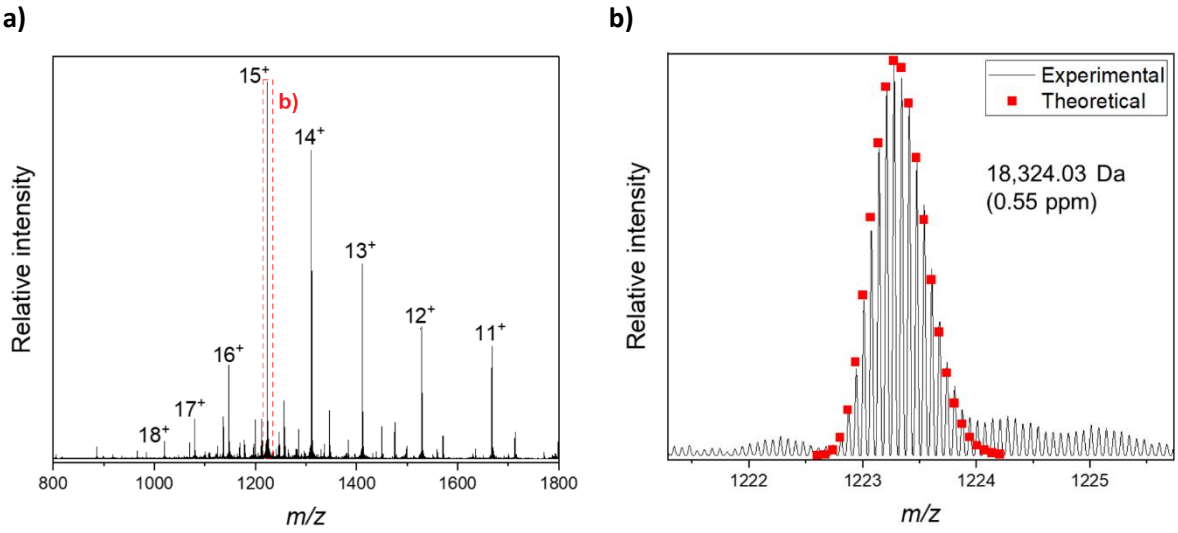


Figure 5.8-50: **a)** Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ N136C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H1**.
The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

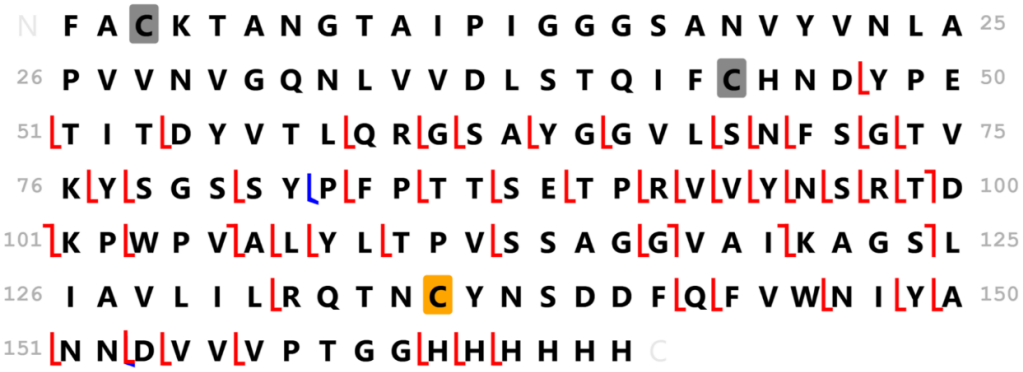


Figure 5.8-51: Fragment map of the identified proteoform of $FimH_L$ N136C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H1**.
The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); red: fragmentation pattern (EthcD); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis and alkylated with **H1**.

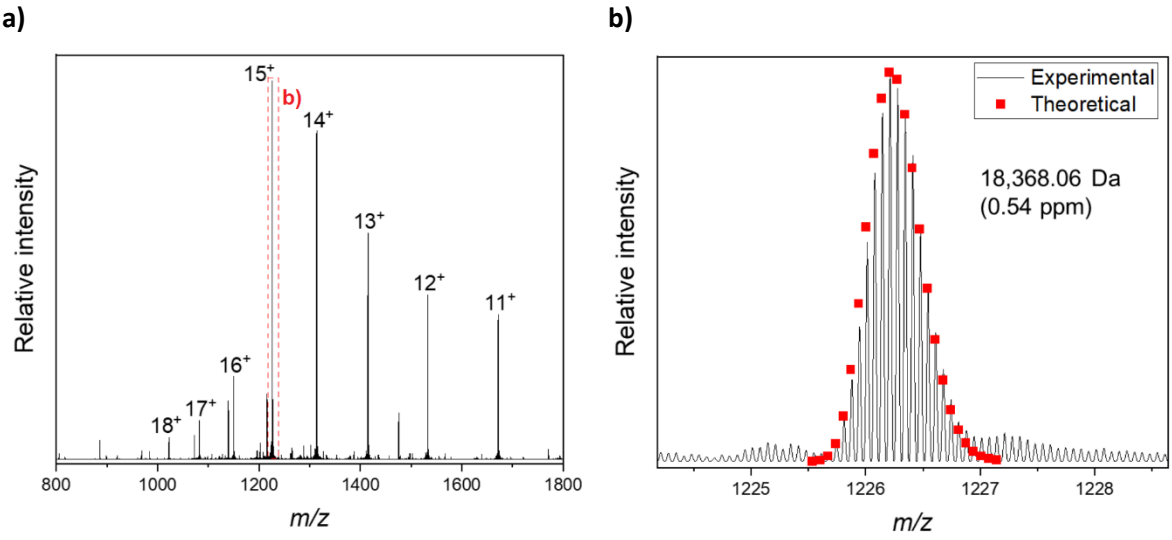


Figure 5.8-52: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of FimH$_L$ N136C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H2**.



Figure 5.8-53: Fragment map of the identified proteoform of FimH$_L$ N136C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H2**.

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis and alkylated with **H2**.

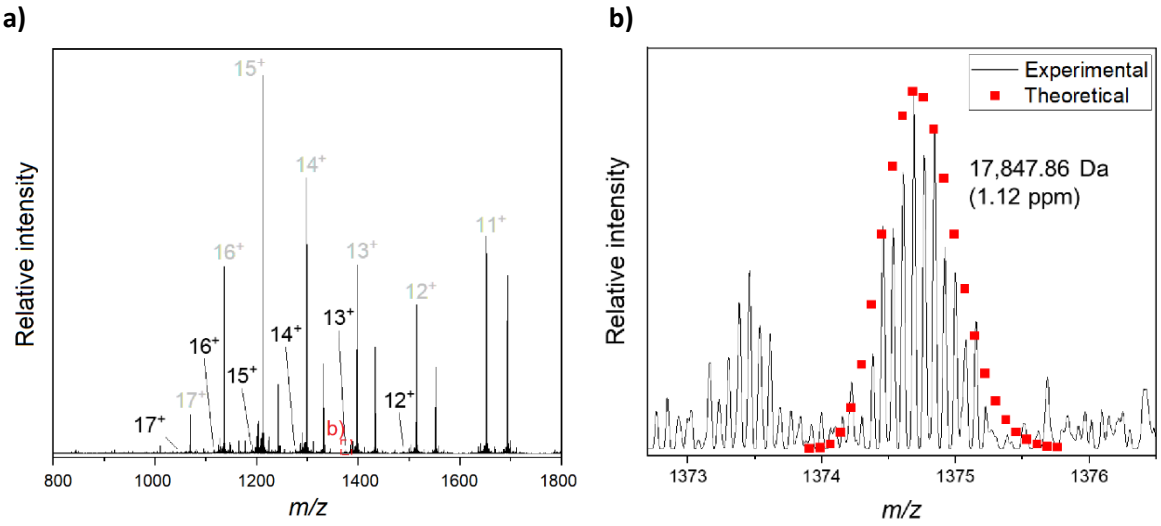


Figure 5.8-54: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of unmodified $FimH_L$ S139C without prior reduction.

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).



Figure 5.8-55: Fragment map of the identified proteoform of unmodified $FimH_L$ S139C without prior reduction.

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys136 introduced by site-directed mutagenesis.

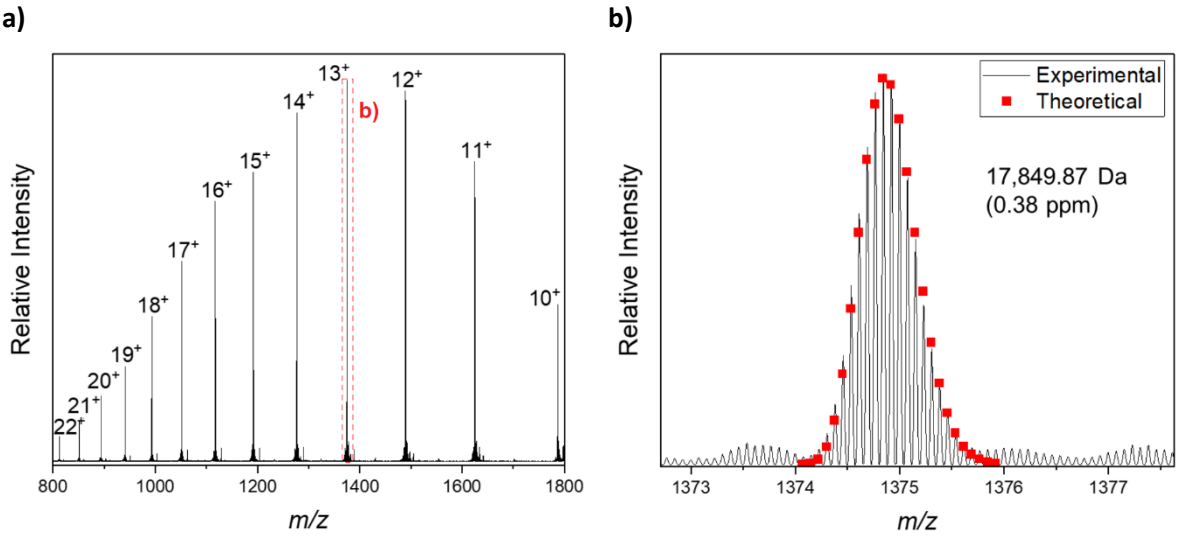


Figure 5.8-56: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of unmodified FimH$_L$ S139C after reduction with 20 mM TCEP, 10 min at 60 °C.[497]

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

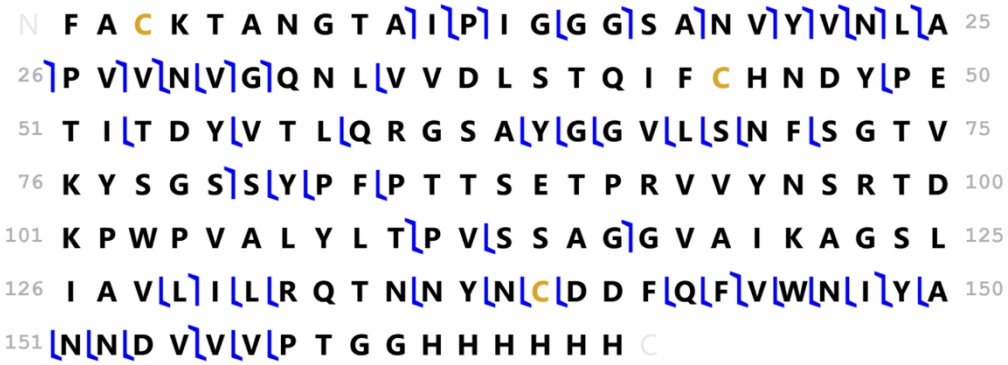


Figure 5.8-57: Fragment map of the identified proteoform of unmodified FimH$_L$ S139C after reduction with 20 mM TCEP, 10 min at 60 °C.[497]

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID); yellow: reduced cysteine residues.

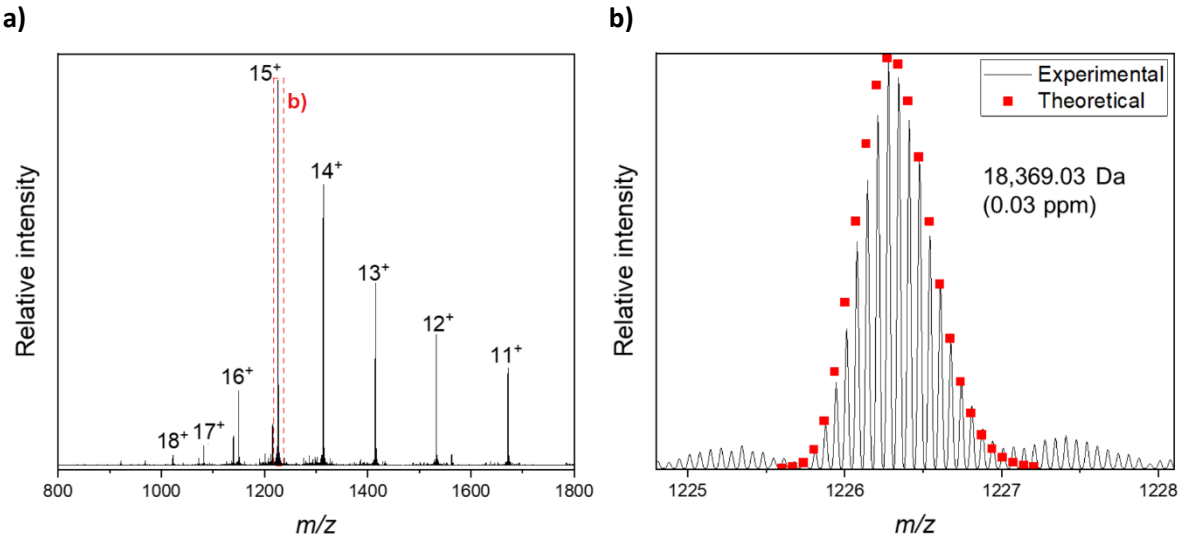


Figure 5.8-58: a) Overview intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F1**.
The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).

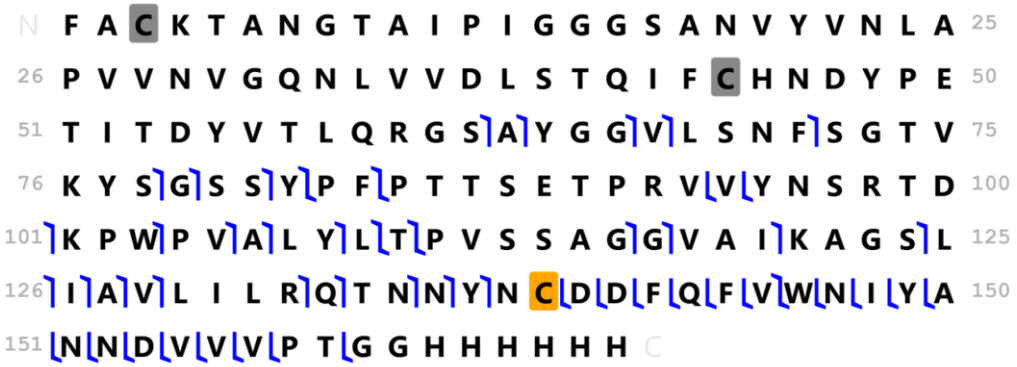


Figure 5.8-59: Fragment map of the identified proteoform of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F1**.
The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID fragmentation); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys139 introduced by site-directed mutagenesis and alkylated with **F1**.

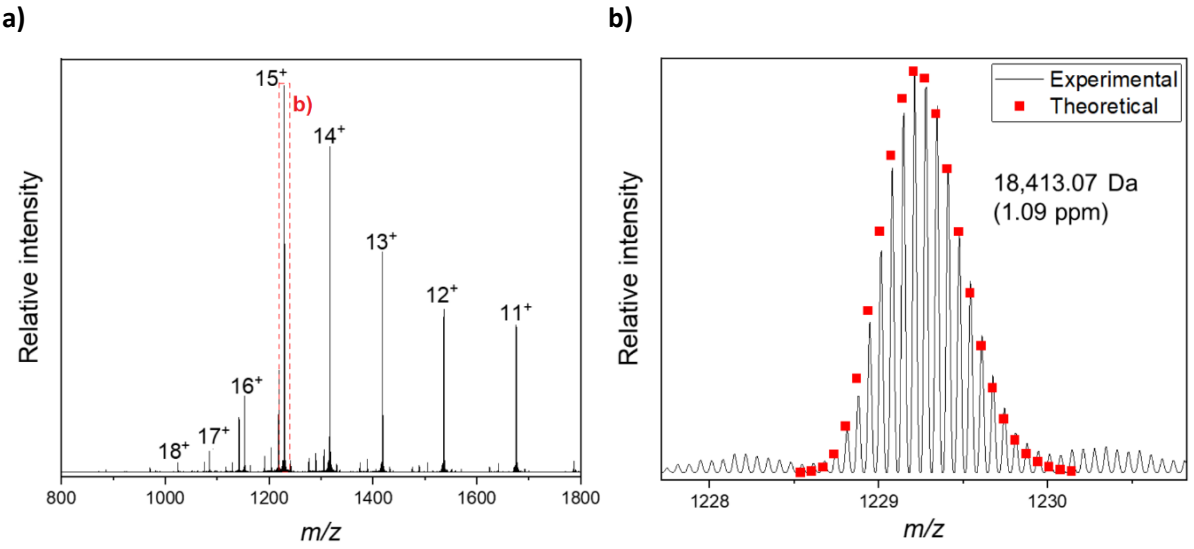


Figure 5.8-60 a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F2**.
The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).



Figure 5.8-61: Fragment map of the identified proteoform of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised 3-fluoroazobenzene mannoside **F2**.
The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID fragmentation); Rred: fragmentation pattern (EthcD); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys139 introduced by site-directed mutagenesis and alkylated with **F2**.

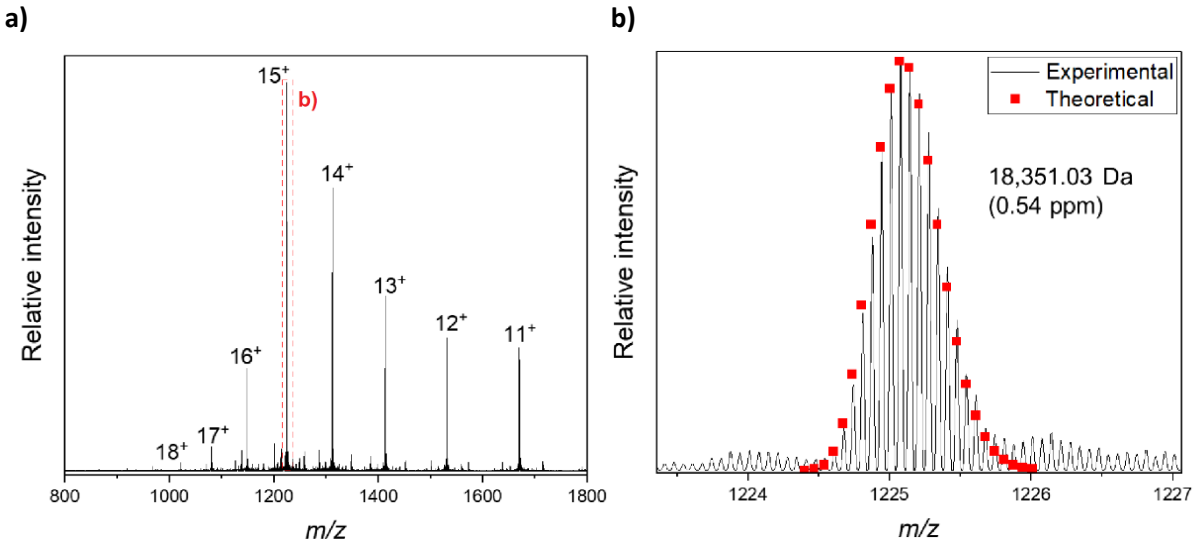


Figure 5.8-62: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H1**.
The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).



Figure 5.8-63: Fragment map of the identified proteoform of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H1**.
The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID fragmentation); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys139 introduced by site-directed mutagenesis and alkylated with **H1**.

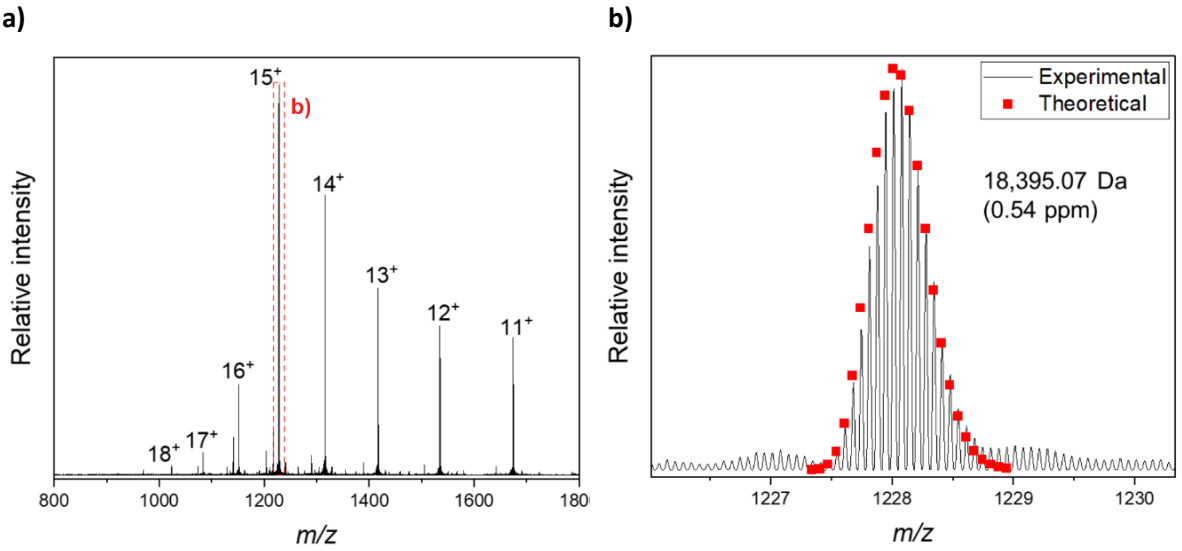


Figure 5.8-64: a) Overview of the intact-mass spectrum and **b)** zoomed-in section showing the isotope distribution of theoretical versus experimental proteoforms obtained by top-down LC-MS/MS analysis of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H2**.

The figures were created and kindly provided by Dr. Philipp Kaulich (Kiel University).



Figure 5.8-65: Fragment map of the identified proteoform of $FimH_L$ S139C after alkylation with iodoacetamide-functionalised azobenzene mannoside **H2**.

The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University). Blue: fragmentation pattern (CID fragmentation); red: fragmentation pattern (EthcD); grey: disulfide bridge formed between Cys3 and Cys44; yellow: Cys139 introduced by site-directed mutagenesis and alkylated with **H2**.

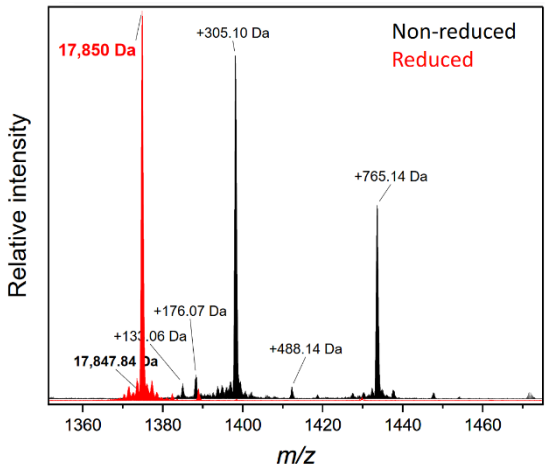


Figure 5.8-66: Overview of the intact-mass spectrum of reduced (red) versus non-reduced (black) $FimH_L$ S139C obtained by top-down LC-MS/MS analysis of the unmodified variant prior to alkylation with photoswitchable ligands.

The additional, mass-shifted proteoforms of the non-reduced S139C variant presumably result from the following biological modifications: + 133.06 Da, *S*-homocysteine; + 176.07 Da, hexuronic acid; + 305.10 Da glutathione disulfide; + 765.14 Da, coenzyme A. These modifications were efficiently removed after reduction with 20 mM TCEP for 10 min at 60 °C.[497] The figure was created and kindly provided by Dr. Philipp Kaulich (Kiel University).

Table 5.8-12: Monoisotopic intact masses identified by top-down LC-MS/MS analysis of unmodified and labelled $FimH_L$ N136C expressed in non-*degP*-deficient SG13009 pPKL241.

	Unmodified $FimH_L$ N136C	**$FimH_L$ N136C alkylated with F1**	**$FimH_L$ N136C alkylated with F2**
Exp. mass[a] [Da]	17 820.85	18 342.03	18 386.06
Obs. mass [Da]	17 409.63	17 930.77	17 974.82
Obs. Mass difference[b] (M_{obs}-M_{exp}) [Da]	- 411.22	- 411.27	- 411.24
Exp. Mass difference (unmodified-alkylated) [Da]	–	- 521.18	- 565.21
Obs. mass difference[c] (unmodified-alkylated) [Da]	–	- 521.14	- 565.19

a) Monoisotopic masses calculated for $FimH_L$ N136C with disulfide bridge formed by Cys3 and Cys44. **b)** The observed mass shifts presumably result from the absence of three histidine residues within the C-terminal His_6 tag (M_{mono} = 411.17673 g/mol). **c)** The observed mass difference of unmodified and alkylated $FimH_L$ N136C indicates a successful cysteine modification with the respective ligands. All data was obtained and compiled by Dr. Philipp T. Kaulich (Kiel University).

Table 5.8-13: Non-tryptic peptides identified by bottom-up LC-MS/MS analysis of unmodified and labelled $FimH_L$ N136C expressed in non-*degP*-deficient SG13009 pPKL241.

Sequence section	RGSAYGGVLSNFSGTVKYS (aa 60 – 78)	KYSGSSYPFPTTSETPRV (aa 76 – 93)	RTDKPWPVALYLTPVSSAGGVAIKA (aa 98 – 122)	YANNDVVVPTGGHHHHHH (aa 149 – 162)
Identified peptides	GSAYGGVLSN GSAYGGVLSNFSG GSAYGGVLSNFSGTVK GSAYGGVLSNFSGTVKY	YSGSSYPFPT YSGSSYPFPTT YSGSSYPFPTTS YSGSSYPFPTTSE YSGSSYPFPTTSET YSGSSYPFPTTSETP YSGSSYPFPTTSETPR	TDKPWPVA TDKPWPVAL TDKPWPVALY TDKPWPVALYLTPV TDKPWPVALYLTPVSS TDKPWPVALYLTPVSSA TDKPWPVALYLTPVSSAG TDKPWPVALYLTPVSSAGG TDKPWPVALYLTPVSSAGGV TDKPWPVALYLTPVSSAGGVAIK	ANNDVVVPTGGH ANNDVVVPTGGHH ANNDVVVPGly1GHHH

The observed non-tryptic peptides indicate (exo-)peptidase activity, probably during the protein purification process and/or accumulation of the target protein in the periplasm of the DegP expressing *E. coli* strain.

MALDI-TOF-MS analysis of FimH variants

The protein solutions (1 – 1.5 µL) were either premixed with the matrix solution (1:9) or directly mixed on the MALDI target (~ 1:2), which was then submitted for measurement on a MALDI-TOF AutoFlex® mass spectrometer (Bruker Corporation) directly after solvent evaporation. The following matrices were used as saturated solutions in acetonitrile/H_2O/TFA, 500:500:1: 2,5-Dihydroxybenzoic acid (DHB), sinapinic acid (SA), α-cyano-4-hydroxycinnamic acid (CCA) or 4-chloro-α-cyanocinnamic acid (ClCCA).

Table 5.8-14: Observed *m/z* ratios from MALDI mass spectra of genetically and chemically modified as well as of unmodified FimH variants.

FimH variant	Average molecular mass[a] [Da]	Monoisotopic mass[a] [Da]	Obs. m/z ratios
$FimH_{trc}$-His_6	17842.858	17831.877	17835.8 17845.3 17866.8
$FimH_{trc}$-His_6 N136C	17831.897	17820.843	17840 17853
$FimH_{trc}$-His_6 N136C + **H1**[b]	18335.417	18324.0412	18327.4 18341.0 18359.8 9170.8 (z = 2)
$FimH_{trc}$-His_6 N136C + **F1**[b]	18353.407	18342.0318	18342.0 18358.2 18362 9171.7 (z = 2)
$FimH_{trc}$-His_6 N136C + **F2**[b]	18397.457	18386.058	18393 19396 18411 18428
$FimH_{trc}$-His_6 S139C	17858.922	17847.854	17789.9 17841.1
$FimH_L$-His_6 N136C ΔN-sigpep	17963.093	17951.884	17955
$FimH_{trc}$-His_6 R60P	17783.79	17772.829	17777.0 17792.1

a) Average molecular masses and monoisotopic masses were calculated based on the protein sequence, under consideration of the disulfide bridge between Cys3 and Cys44 and assuming unmodified N- and C-termini using the PeptideSynthetics Peptide Mass Calculator (https://www.peptidesynthetics.co.uk/tools/). **b)** Monoisotopic masses of alkylated ligands were calculated with the Exact Mass Calculator from Scientific Instrument Services (https://www.sisweb.com/referenc/tools/exactmass.htm). MALDI mass spectra were recorded on a MALDI-TOF AutoFlex® mass spectrometer (Bruker Corporation). Exemplary MALDI mass spectra of the variants ($FimH_{trc}$, $FimH_L$ N136C ΔN-sigpep, and $FimH_{trc}$ R60P), which were not further analysed by LC-MS/MS, are provided in the appendix.

5.8.9 FITC-*p*AP mannopyranoside binding assay

The expression and purification of the here investigated $FimH_{trc}$ variants and $FimH_L$ N136C is described in Chapter 3.5 and Chapter 5.8.6 . The alkylation of the FimH variants N136C and S139C was performed according to Chapter 5.7.7. The alkylation success of the $FimH_{trc}$ cysteine variants was evaluated by SDS-PAGE (Figure 3.7-4) and MALDI mass spectrometry (Table 5.8-14). It was ensured that unmodified FimH cysteine variants were reduced with 5 mM DTT for 5 min on ice in the course of the purification, in order to remove potentially present biological cysteine modifications. Alternatively, the cysteine variants were treated with 1 mM DTT, and incubated for 5 min at room temperature and desalted using a Zeba™ Spin desalting column with a suitable volume capacity and MWCO prior to use in the FITC-*p*AP-Man binding assay.

Standard FITC-*p*AP mannopyranoside binding assays

All FITC-*p*AP mannopyranoside binding assays were performed on Ni^{2+}-chelate-coated black 96-well microplates for the immobilisation of hexahistidine-tagged proteins (Pierce™, Thermo Scientific™, cat. no. 15342). The supplied plates are already blocked with bovine serum albumin (BSA) and were washed with PBST (Table 5.7-1; 3 x 400 µL/well) in a microplate washer (Tecan HydroFlex™) prior to use. The FimH solutions were applied in duplicates or triplicates (100 µL/well) and immobilised for 2 h at 20 °C under moderate agitation at 300 rpm (MATRIX orbital thermoshaker, IKA). Usually, protein concentrations between 100 and 200 µg/mL (5.6 – 11.26 µM) were used. Dilution series (1 :10) were prepared with the respective buffers (Table 5.7-1) in separate Eppendorf tubes and subsequently pipetted on the microplates in order to prevent premature immobilisation and thus inaccurate protein concentrations. The immobilisation step was optionally inhibited by the addition of imidazole (100 mM) during the initial incubation step with the lectin. BSA and PBS (100 µL/well) served as controls for nonspecific binding of FITC-*p*AP mannopyranoside and to determine the baseline fluorescence signal. After immobilisation, unbound protein was removed by washing with PBST (3 x 400 µL/well) in a microplate washer. A solution of FITC-labelled *p*-aminophenyl *α*-D-mannopyranoside (FITC-*p*AP-Man) **59** (1 – 2 eq., 11.2 µM, in PBS, 100 µL/well) was added and incubated at 20 °C for 1 h under agitation (300 rpm). The binding step was competitively inhibited by addition of methyl *α*-D-mannopyranoside (MeMan, 50 mM) in order to demonstrate the mannoside-specificity of the binding event. Eventually, the plates were washed with PBST (2 x 400 µL/well) in a microplate washer, re-filled with PBS (100 µL/well) and the fluorescence of the residual FITC-*p*AP-Man derivative was measured on an Infinite® 200 Pro M Nano+ dual mode microplate reader (Tecan) at $\lambda_{ex/em}$ = 483/535 nm.

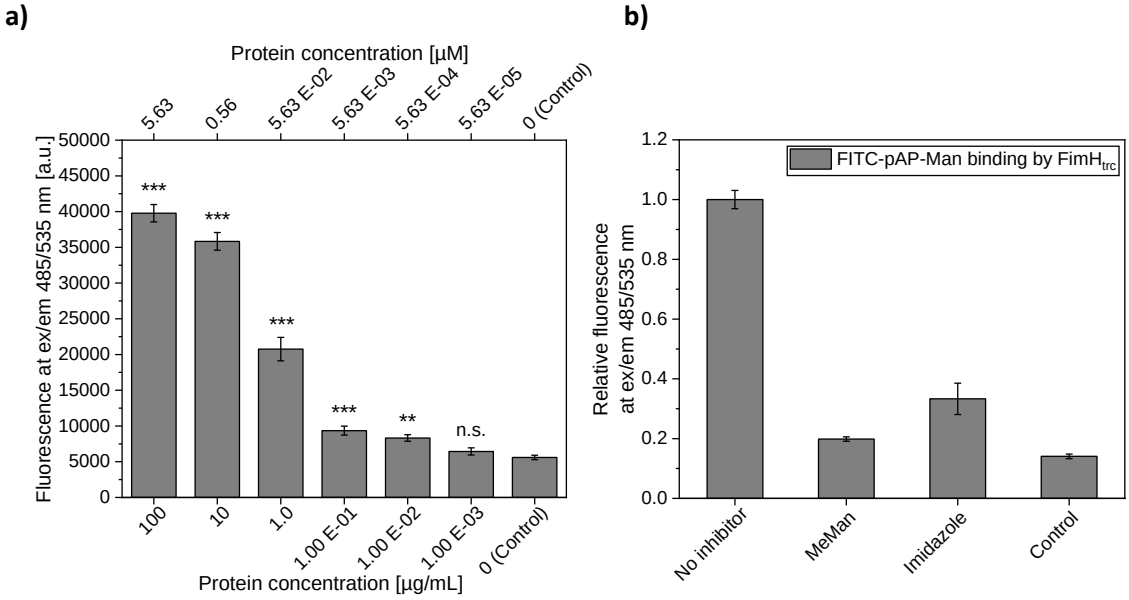


Figure 5.8-67: a) Concentration dependence and **b)** inhibition of FITC-*p*AP mannopyranoside binding activity of $FimH_{trc}$.

$FimH_{trc}$ solutions in assay buffer (100 µg/mL; 5.63 µM as well as a premixed 1:10 dilution series, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and the lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well). The immobilisation was blocked by addition of imidazole (100 mM) and the lectin activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). Error bars represent the ± standard deviation of triplicate measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (0 µg/mL $FimH_{trc}$).

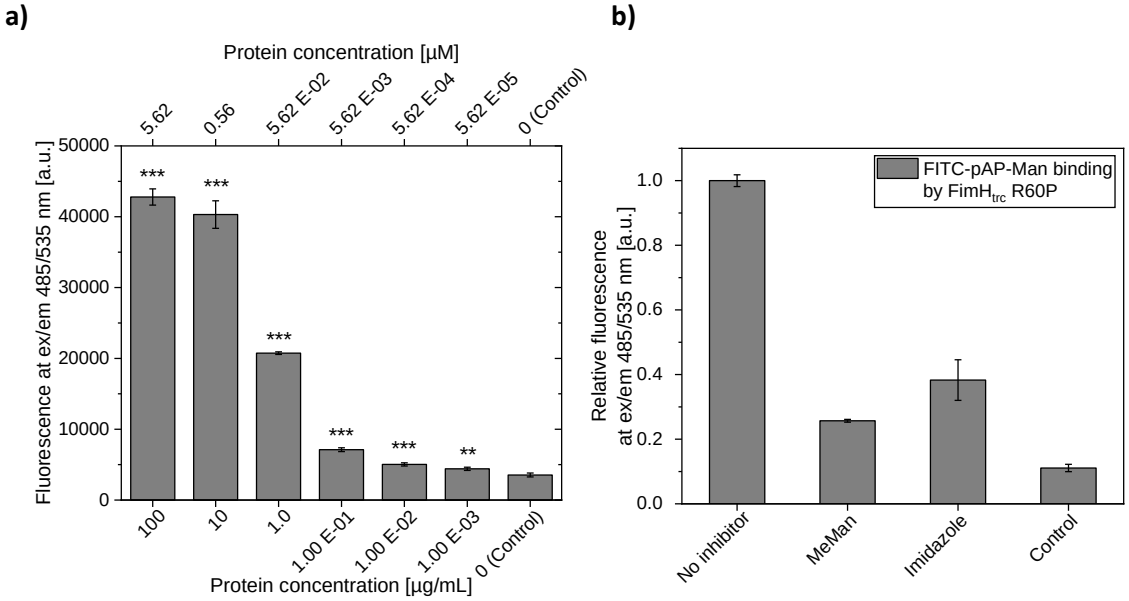


Figure 5.8-68: a) Concentration dependence and **b)** inhibition of FITC-*p*AP mannopyranoside binding activity of $FimH_{trc}$ R60P.

$FimH_{trc}$ R60P solutions in PBS (100 µg/mL; 5.62 µM as well as a premixed 1:10 dilution series, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and the lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well). The immobilisation was blocked by addition of imidazole (100 mM) and the lectin activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). Error bars represent the ± standard deviation of triplicate measurements and the asterisks (*) denote statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (0 µg/mL $FimH_{trc}$ R60P).

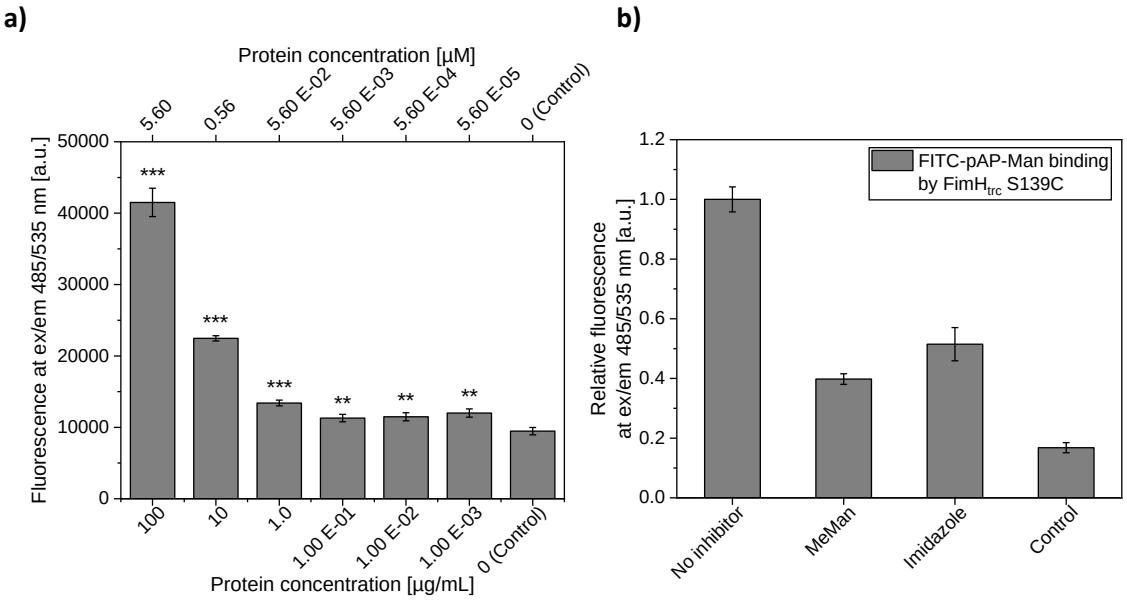


Figure 5.8-69: a) Concentration dependence and **b)** inhibition of FITC-*p*AP mannopyranoside binding activity of unmodified $FimH_{trc}$ S139C.

$FimH_{trc}$ S139C solutions in PBS (100 µg/mL; 5.60 µM as well as a premixed 1:10 dilution series, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and the lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well). The immobilisation was blocked by addition of imidazole (100 mM) and the lectin activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). Error bars represent the ± standard deviation of triplicate measurements and the asterisks (*) denote statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (0 µg/mL $FimH_{trc}$ S139C).

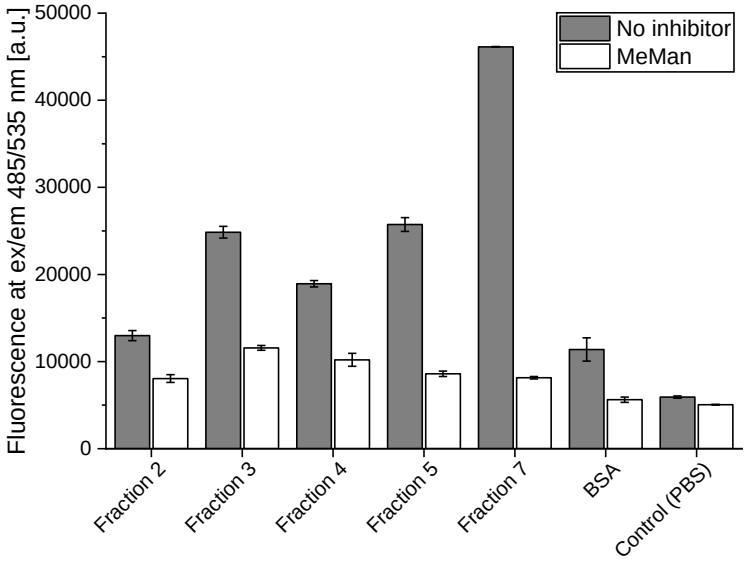


Figure 5.8-70: Investigation of SEC fractions from the purification of $FimH_{trc}$ N136C in the FITC-*p*AP mannopyranoside bindig assay.

Fractions from the purification shown in Figure 3.5-8a: F2, F3 (main signal and shoulder of SEC peak I), F4 (SEC peak II), F5 (SEC peak III) and F7 (SEC peak IV). The lectin (100 µg/mL, 100 µL/well) was immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and the activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well). The binding activity was inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). BSA (100 µg/mL) and PBS (each 100 µL/well) were used as controls. The protein concentrations were beforehand determined in a BCA assay (Pierce™ Thermo Scientific). Error bars represent the ± standard deviation of duplicate measurements.

Control experiments for the optimisation of the FITC-*p*AP mannopyranoside binding assay with UV exposure

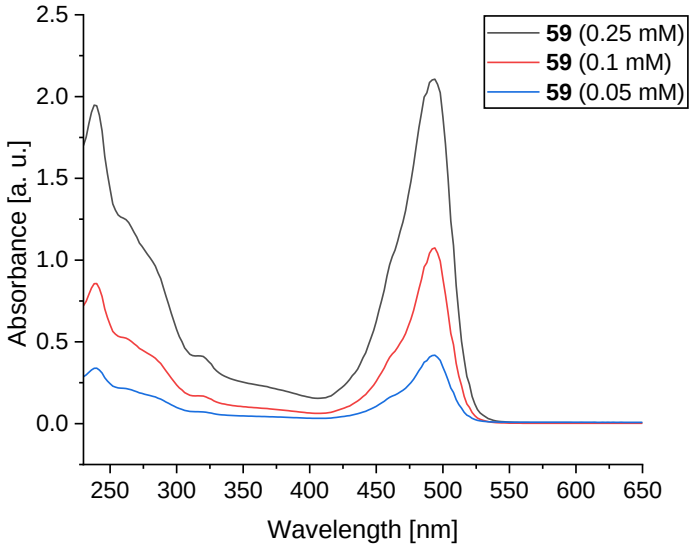


Figure 5.8-71: UV/vis absorbance spectrum of FITC-*p*AP mannoside **59**.

The fluorescein-labelled mannoside derivative **59** in different concentrations (PBS) was applied onto a transparent UV-Star® F96-well plate (100 µL) and the absorbance was measured on a Tecan plate reader at 37 °C.

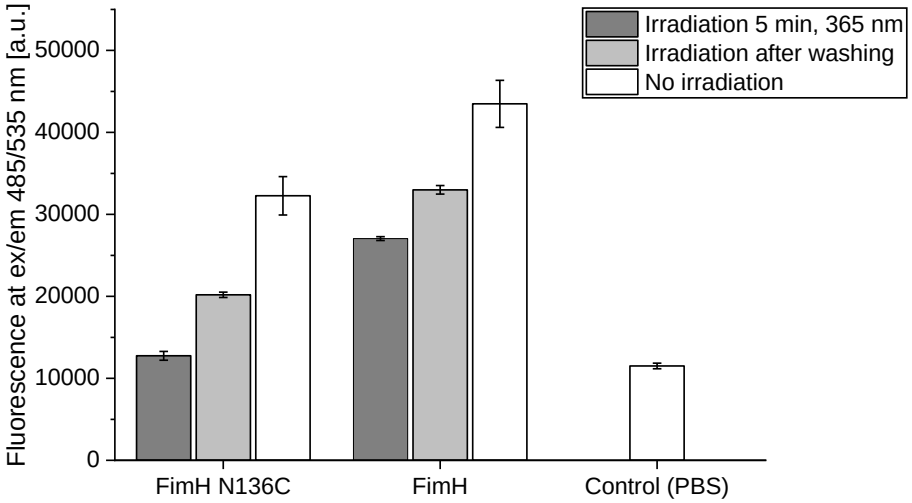


Figure 5.8-72: FITC-*p*AP mannopyranoside binding experiment with on-plate irradiation at 365 nm.

Unlabelled $FimH_{trc}$ N136C (100 µg/mL, 5.62 µM, in PBS/glycerol, 9:1) and $FimH_{trc}$ (100 µg/mL, 5.63 µM in assay buffer) were applied onto on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) and irradiated at 365 nm, 5 min after lectin immobilisation (2 h, at 300 rpm, 20 °C) or directly after washing with PBST (3 x 400 µL/well). PBS (100 µL/well) was used as control without irradiation instead of the lectins. Unexposed wells were wrapped in aluminium foil prior to irradiation. Error bars represent the ± standard deviation of duplicate measurements.

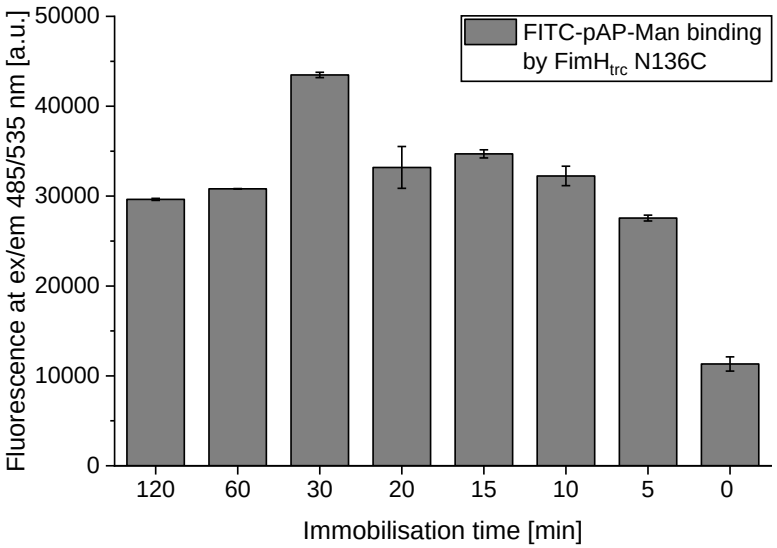


Figure 5.8-73: Influence of the lectin immobilisation time in the FITC-*p*AP mannopyranoside binding assay.

Unlabelled FimH$_{trc}$ N136C (100 µg/mL, 5.62 µM in PBS/glycerol, 9:1; 100 µL/well) was immobilised onto a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™) with varying incubation times at 300 rpm, 20 °C. The lectin activity was tested with 2 eq. of FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well, 1 h at 300 rpm, 20 °C). Error bars represent the ± standard deviation of duplicate measurements. The fluctuations, in particular the outlier after 30 min immobilisation, could result from the multiple reuse of the plate, which may have affected the quality of the measurement.

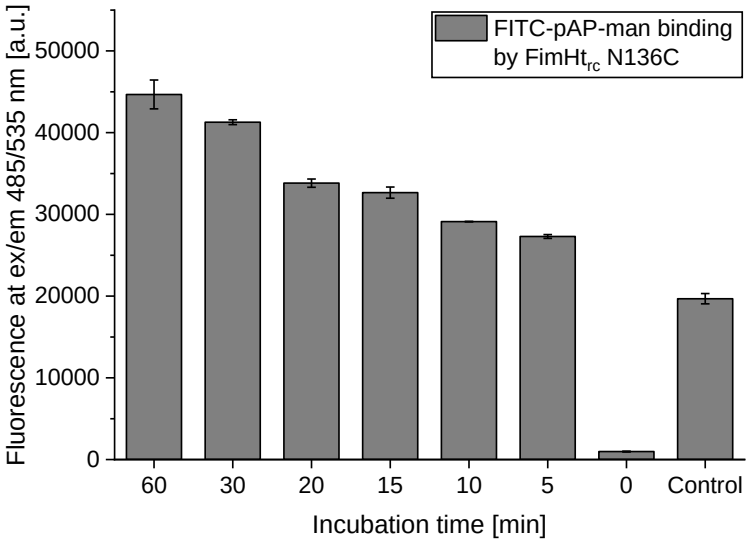


Figure 5.8-74: Influence of the incubation time in the FITC-*p*AP mannopyranoside binding assay.

Unlabelled FimH$_{trc}$ N136C (100 µg/mL, 5.62 µM in PBS/glycerol 9:1; 100 µL/well) was immobilised onto a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™, 2 h, at 300 rpm, 20 °C) and the lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well) with varying incubation times (0 – 60 min) at 300 rpm, 20 °C. PBS (100 µL/well) was applied as control, a solution of **59** (11.2 µM in PBS, 100 µL/well) was added and incubated for 60 min at 300 rpm, 20 °C. Error bars represent the ± standard deviation of duplicate measurements.

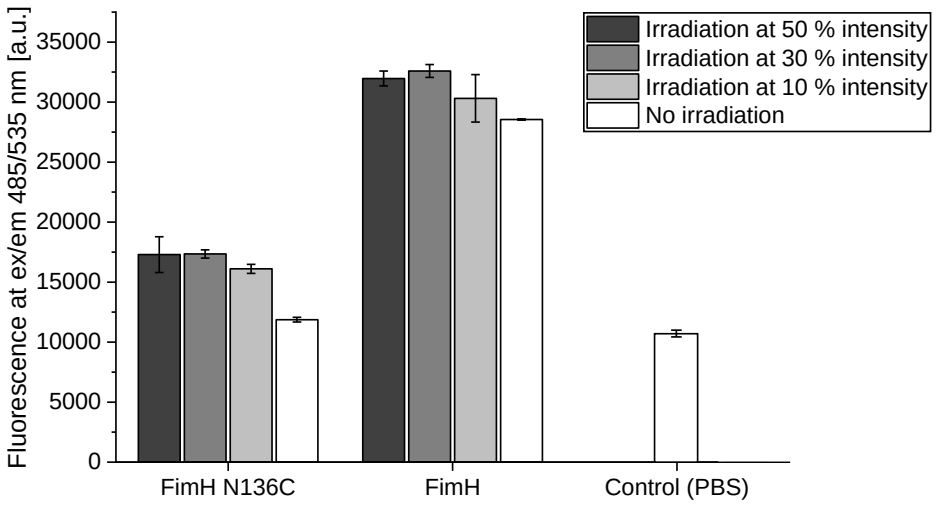


Figure 5.8-75: FITC-*p*AP mannopyranoside binding experiment with pre-irradiation of the lectin at 365 nm prior to immobilisation.

Unlabelled $FimH_{trc}$ N136C (49 µg/mL, 2.75 µM in PBS) and $FimH_{trc}$ (100 µg/mL, 5.63 µM in assay buffer) were irradiated 5 min at 365 nm (LED NC4U133AT, Nichia Corporation) at different intensities prior to immobilisation on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™, 2 h at 300 rpm, 20 °C). PBS (100 µL/well) was used as control without irradiation instead of the lectins. Error bars represent the ± standard deviation of duplicate measurements. The experiment indicates no negative impact of the pre-irradiation on the FITC-*p*AP-Man binding activity of the lectins.

FITC-*p*AP mannopyranoside binding assay of labelled $FimH_{trc}$ N136C variants and UV exposure

In FITC-*p*AP-Man binding experiments with modified FimH cysteine derivatives that were beforehand alkylated with azobenzene mannoside derivatives **H1** and **F2**, all steps were performed under exclusion of light as far as possible. In experiments with UV exposure, the lectin solutions were irradiated at 365 nm for 5 min at 30 to 40 % intensity (LED NC4U133AT, Nichia Corporation) in 2 mL glass vials on ice prior to immobilisation onto the Ni^{2+}-coated 96-well plate. The incubation steps at 20 °C and 300 rpm, for the protein immobilisation and binding of FITC-*p*AP-Man, were reduced to 10 min in order to minimise *trans*/*cis* photoisomerisation of the azobenzene moiety.

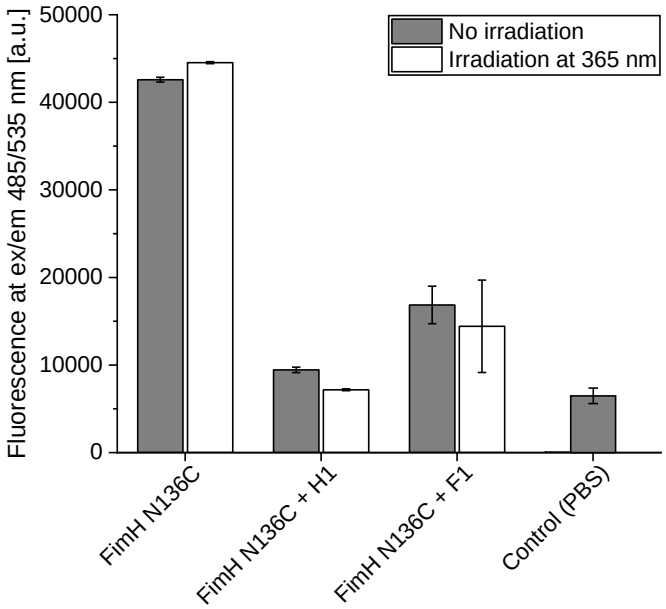


Figure 5.8-76: Binding of FITC-*p*AP mannopyranoside by modified and unlabelled $FimH_{trc}$ N136C with/without prior irradiation at 365 nm.

Unmodified $FimH_{trc}$ N136C and the lectin labelled with azobenzene mannoside derivatives **H1** and **F1** (162 µg/mL; 9.1 µM; in PBS/glycerol, 9 :1; 100 µL/well) were exposed to UV light at 365 nm, 5 min (30 % intensity, LED NC4U133AT, Nichia Corporation) and immobilised onto a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Sscientific™; 10 min, at 300 rpm, 20 °C). The lectin activity was tested with FITC-*p*AP mannopyranoside (**59**, 11.2 µM, in PBS, 100 µL/well, 10 min at 300 rpm, 20 °C). PBS (100 µL/well) was used as a control. The protein concentrations were determined in a BCA assay (Pierce™ Thermo Scientific™). Error bars represent the ± standard deviation of duplicate measurements.

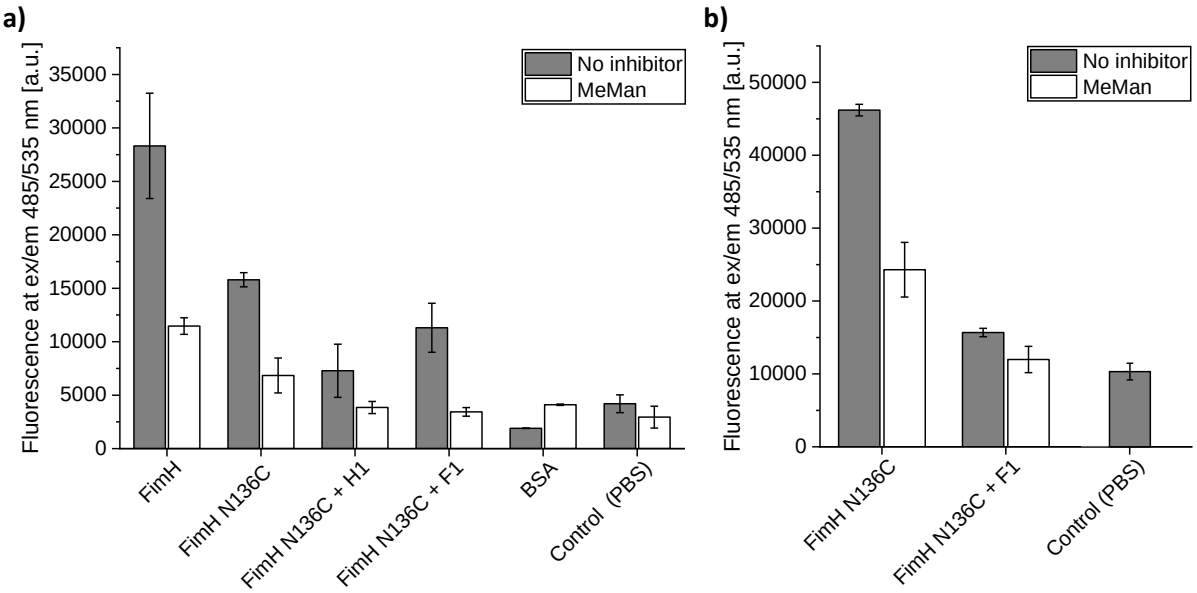


Figure 5.8-77: Binding of FITC-*p*AP mannopyranoside by modified and unlabelled FimH N136C.

a) Unmodified $FimH_{trc}$, $FimH_{trc}$ N136C and the cysteine variant labelled with azobenzene mannoside derivatives **H1** and **F1** (200 µg/mL; 11.24 µM, in PBS/glycerol, 9 :1; 100 µL/well) and **b)** $FimH_{trc}$ N136C and the variant labelled with **F1** (100 µg/mL, 5.62 µM in PBS, 100 µL/well) were immobilised on a black Ni^{2+}-coated 96-well plate (Pierce™, Thermo Scientific™, 25 min, at 300 rpm, 20 °C). The lectin activity was tested with FITC-*p*AP mannoside (**59**, 11.2 µM in PBS, 100 µL/well, 1 h at 300 rpm, 20 °C) and inhibited with methyl *α*-D-mannopyranoside (MeMan, 50 mM). BSA (100 µg/mL) and PBS (each 100 µL/well) were used as controls. The protein concentrations were determined in a BCA assay (Pierce™ Thermo Scientific™). Error bars represent the ± standard deviation of duplicate measurements.

5.9 Supplemental information to Chapter 4: "Red-fluorescent bacteria for adhesion studies: Chromosomal integration of *mcherry* in *E. coli*"

The site-specific chromosomal insertion of the gene encoding mCherry, a red-fluorescent protein derived from DsRed (which in turn originates from *Discosoma sp.*), was achieved by following the "clonetegration" strategy implemented by Shearwin and co-workers.[662] The procedure provides chromosomal integration of DNA sequences only relying on a single transformation event. Hence, the clonetegration approach represents a faster, simpler alternative to conventional techniques that are usually more time-consuming, such as the "conditional-replication, integration, and modular" (CRIM) system.[639]

The "One-Step Integration Plasmids" (pOSIPs) combine all elements necessary for a site-specific integration of a gene of interest into the *E. coli* chromosome. We used pOSIP-KL-*mcherry* equipped with a gene encoding the λ Integrase, a recombinase originating from the *Escherichia* virus λ, also known as lambda phage. The integrase catalyses recombination at the λ *att*B site on the chromosome with the λ *att*P site on the plasmid, resulting in chromosomal integration of the entire plasmid sequence. The insertion can be selected via the co-inserted kanamycin acetyltransferase within the so-called integration module. The integration module, which also includes the λ integrase gene (λ*int*) and a temperature-sensitive λ repressor gene (λcI^{ts}), is flanked by two recombination sites (FRT) that are recognised by another recombinase, the flippase (Flp).[713] This provides the opportunity to remove the integration module by a further transformation event with the flippase-encoding plasmid pE-Flp. The plasmid can be selected with ampicillin, which is mediated by the encoded *β*-lactamase. The temperature-sensitive operon on pE-Flp enables subsequent removal of the plasmid after successful flipping of the integration module. The subcloning of the mcherry gene into pOSIP-KL was performed by Bikard and colleagues.[652] Both plasmids (pOSIP-KL-*mcherry* and pE-Flp) are available from Addgene (cat. no. 116947, cat. no. 45978).

5.9.1 Chromosomal integration of pOSIP-KL-*mcherry*

Preparation of pOSIP-KL-*mcherry* and pE-Flp

The plasmids were ordered from Addgene and stored in appropriate bacterial host strains DH5α λpir (pOSIP-KL-*mcherry*) and E811 (pE-Flp). The storage strains were cultured according to the supplier's recommendations (Table 5.7-2), and the plasmids were isolated with a kit suited for plasmid preparation, following the standard protocol for high-copy plasmids (see protocol: Plasmid preparation in Chapter 5.7.6).

Transformation and Storage Solution (TSS) method [690]

According to the pOSIP protocol,[662] transformation of pOSIP-KL-*mcherry* and pE-Flp can be accomplished using "Transformation and Storage Solution" (TSS, see Table 5.7-1). The TSS method is based on the same permeabilisation principle (depolarisation of the cell membrane's electric potential) as the well-known $CaCl_2$ transformation method,[700] but the preparation of TSS-competent cells requires no thorough washing procedure as is necessary in conventional approaches. To obtain TSS-competent cells for the subsequent transformation with pOSIP-KL-*mcherry*, the respective strains (*E. coli* HB101; *E. coli* HB101 pPKL4; *E. coli* HB101 pPAP5, also referred to as VR123) were cultured in a 100 mL culture flask with 50 mL LB medium (Table 5.7-1) and streptomycin (20 µg/mL) at 37 °C, and agitation at 180 rpm.

When the optical density had risen from approximately OD_{600} = 0.05 to OD_{600} = 0.5, the suspensions were cooled on ice for 10 min and centrifuged at 5000 rpm, 4 °C, 10 min (Universal 320R centrifuge, Hettich®). The cell pellets were resuspended in 5 mL (10:1 ratio) ice-cold TSS (Table 5.7-1) and aliquoted in 100 µL volumes. The aliquots were either directly transformed with the intended plasmids or frozen in liquid nitrogen and stored at –80 °C until use.

TSS transformation with pOSIP-KL-*mcherry*

For this purpose, the aliquots were thawed on ice and 1.5 µL (10 ng/µL) of the plasmid pOSIP-KL-*mcherry* was added. After incubation on ice for 20 min, the suspensions were heat-shocked at 42 °C for 45 s and subsequently incubated on ice for 2 min. LB medium (1 mL, Table 5.7-1) was added and the suspensions were incubated at 37 °C, 1 h under agitation at 600 rpm. The incubation at 37 °C leads to the inactivation of the temperature-sensitive repressor λ*cI*^{ts} and accordingly to the expression of the λ integrase. The cells were pelleted by centrifugation at 12 000 rpm, 5 min and 1 mL of the supernatant was removed. The residues were resuspended, plated on LB agar plates with kanamycin (20 mg/mL) and incubated at 30 °C overnight in order to re-activate the repression of the λ integrase. Selected colonies were re-streaked on LB agar with kanamycin (20 µg/mL) and cultured once again at 30 °C overnight. It is important to maintain the specified kanamycin concentration, which is lower than usual, for instance, for multicopy plasmids, since the resistance gene is only present as a single copy within the genome after successful clonetegration.

Chromosomal DNA preparation

Subsequent to transformation, the pOSIP protocol actually proposes a verification of chromosomal integration by colony PCR using four different primers (duplex PCR). However, the duplex PCR using bacterial colonies for the provision of DNA templates frequently led to poorly reproducible or undesired results, presumably due to insufficient cell disruption. An interposed extraction of the bacterial chromosomal DNA significantly improved the quality and reproducibility of the subsequent PCR result. For this purpose, the transformants were picked and cultured in liquid LB medium (Table 5.7-1) with kanamycin (20 mg/mL) at 30 °C, 180 rpm overnight. The chromosomal DNA was extracted using the DNeasy® UltraClean® Microbial Kit (Qiagen) with the corresponding vortex adapter for 24 x 1.5 – 2 mL tubes (Qiagen) according to the manufacturer's specifications.

Duplex PCR to demonstrate chromosomal integration

Transformants with successful integration of a single pOSIP insert were screened by duplex PCR with four different primers (P1, P2, P3, P4; the sequences are provided in Table 5.6-7). The inversely oriented primers P1 and P4 bind on the bacterial chromosome on both sides next to the *att*B site, whereas the also inversely oriented P3 and P2 bind next to the *att*P and adjacent FRT site on the plasmid. The primers were designed by Shearwin and co-workers[662,662] and the expected product sizes for duplex PCR are given in Table 5.9-1.

The duplex PCR was performed with the Q5® High-Fidelity DNA Polymerase (cat. no. M0491, New England Biolabs®) using isolated, well-defined colonies or the extracted chromosomal DNA as template. Beyond the protocol recommended by the manufacturer, a reaction volume of 15 µL, an annealing temperature of 55 °C, an elongation time of 2 min, a final extension time of 10 min and 30 amplification cycles were successfully chosen. The Q5 High GC Enhancer was added to the reaction mixture in order to improve the reaction performance. The primers were usually added as a mix including all primers in each reaction (P1 – P4; 100 ng/µL).

Table 5.9-1: Expected product sizes generated in a duplex PCR to demonstrate site-specific integration of pOSIP-KL-*mcherry* at the chromosomal *att*B site in *E. coli*.

Integration result	**Single integration**		**No integration at *att*B site**	**Tanden integration/episomal pOSIP-KL-*mcherry***
Primer pair	P1 + P2	P3 + P4	P1 + P4	P2 + P3
Expected length [bp]	631	1156	741	1046/1045

The primer sequences of P1, P2, P3, and P4 were designed by Shearwin et al.[662,662] and are compiled in Table 5.6-7. A successful integration of pOSIP is indicated by two PCR fragments (631 bp and 1156 bp). A third PCR product (1046 bp) would appear if multiple, consecutive pOSIP inserts were integrated (tandem integration). Theoretically, a PCR fragment of approx. the same size and a PCR fragment with a length of 741 bp would indicate the presence of the intact plasmid without chromosomal integration.

The result of the duplex PCR was evaluated by agarose electrophoresis as shown in Figure 5.9-1. Individual colonies of *E. coli* HB101 (lane **2**) and HB101 pPAP5 (lane **3**) showed a successful site-specific integration of a single pOSIP at the *att*B site, as indicated by two PCR products with a migration behaviour approximate to 631 bp (P1 + P2 product) and 1156 bp (P3 + P4 product). The absence of an additional PCR amplicon with a length of 1045 bp precludes the integration of multiple pOSIP sections (tandem integration). By contrast, the transformation and/or integration of pOSIP-KL-*mcherry* in HB101 pPKL4 (lane **4**) was unsuccessful, as shown by a 741-bp PCR product, indicating an unaltered *att*B sequence. The same amplicon was observed with a non-transformed *E. coli* HB101 serving as a negative control (lane **5**). The successful integration of the *mcherry* gene within the pOSIP insert in HB101 and HB101 pPAP5 (hereafter termed HB101-mCherry and HB101-mCherry pPAP5) was further confirmed by a visible pink discoloration of the bacterial colonies after a few days of storing the petri dishes in the refrigerator at 5 °C.

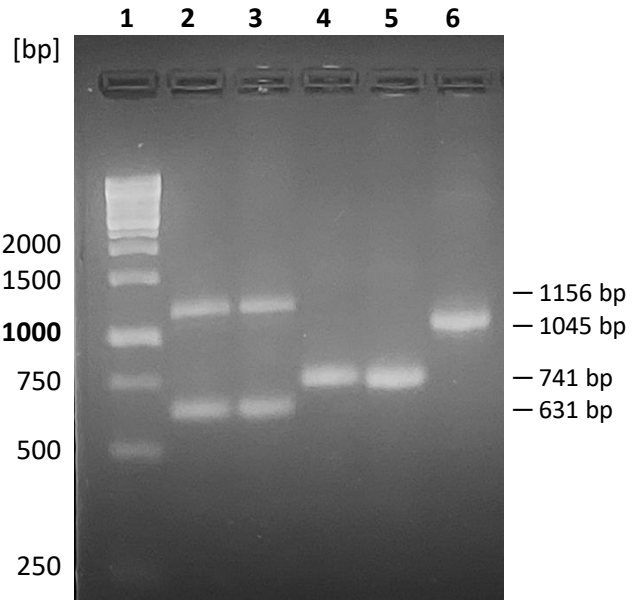


Figure 5.9-1: Agarose gel after duplex PCR to demonstrate site-specific integration of pOSIP-KL-*mcherry* at the chromosomal *att*B site in *E. coli*.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **5**: duplex PCR with extracted chromosomal DNA after transformation with pOSIP-KL-*mcherry* of the following *E. coli*: **2**: HB101-mCherry; **3**: HB101-mCherry pPAP5; **4**: HB101 pPKL4 (transformation and/or integration was unsuccessful); **5**: HB101 (negative control, not transformed); **6**: pOSIP-KL-*mcherry* (PCR with primers P2 + P3). Agarose gel (2 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 75 min. The expected amplicon lengths of each primer pair are given in Table 5.9-1.

5.9.2 Removal of the integration module from the pOSIP-KL-*mcherry* insert

The resulting mCherry-expressing strains should always be cultured and used at 30 °C since otherwise, the λ integrase becomes expressed due to inactivation of the temperature-sensitive transcription factor λcI^{ts}. In usual adhesion and adhesion inhibition experiments in the Lindhorst group, the central incubation step of the bacterial suspensions on the polysaccharide surface is performed at 37 °C. Therefore, the removal of the λ integrase gene from the pOSIP insert is desirable in order to ensure the applicability and comparability of the red-fluorescent strains in adhesion experiments. The integration module within the pOSIP insert is flanked by two flippase recognition targets (FRTs), allowing convenient removal of the module by the flippase-mediated Flp-FRT recombination. The antibiotic marker (Kan^R) is also removed in this recombination step, enabling a subsequent negative selection via resistance loss, if desired. The flippase was introduced by TSS transformation with the plasmid pE-Flp, which is selectable utilising its encoded *β*-lactamase. Due to the selection irreconcilability in the presence of the likewise Amp^R selectable, fimbriae-encoding plasmids pPKL4, pPKL174 and pPAP5, only HB101-mCherry was eligible for transformation with pE-Flp.

TSS transformation of pE-Flp into HB101-mCherry

The flippase-encoding plasmid pE-Flp and TSS-competent HB101-mCherry were prepared as described above (see protocols: Preparation of pOSIP-KL-*mcherry* and pE-Flp and Transformation and Storage Solution method). The cells were thawed on ice and 4 µL (10 ng/µL) of the plasmid pE-Flp were added. The mixture was incubated on ice for 20 min, the suspensions were heat-shocked at 42 °C for 60 s and subsequently cooled on ice for 2 min. SOC outgrowth medium (1 mL, Table 5.7-1) was added and the suspensions were incubated at 30 °C under agitation (600 rpm). The cells were pelleted by centrifugation at 11 000 rpm, 5 min and 1 mL of the supernatant was removed. The residues were resuspended, plated on LB agar plates with ampicillin (100 µg/mL) and incubated at 30 °C overnight, allowing both plasmid replication and Flp-FRT recombination to occur.

PCR to demonstrate successful removal of the integration module

Isolated transformants were picked from the LB agar plate and cultured in liquid LB medium (Table 5.7-1) with ampicillin (100 µg/mL) at 30 °C, 180 rpm overnight. The chromosomal DNA was extracted using the DNeasy® UltraClean® Microbial Kit (Qiagen) according to the manufacturer's recommendations. The success of the flippase reaction was once more evaluated by PCR using the primers P1, P2, P3, P4 and the target-specific primer *mcherry*20-orf_for (Table 5.6-7), followed by agarose gel electrophoresis (Figure 5.9-2). The correspondingly expected PCR product sizes are given in Table 5.9-2.

Table 5.9-2: Expected product sizes generated in a PCR after successful removal of the integration module from the pOSIP insert, mediated by Flp-FRT recombination.

Recombination result	**Removal of the integration module**			**Presence of pE-Flp**
Primer pair	P1 + P2	P3 + P4	P1 + *mcherry*20-orf_for	pE-Flp for + pE-Flp rev
Expected length [bp]	–	1156	1422	1652

The primer sequences of P1, P2, P3, P4 and *mcherry*20-orf_for were designed by Shearwin et al.[662,662] and are provided in Table 5.6-7. A successful removal of the integration module is indicated by a PCR product generated with primer P1 and the insert-specific primer *mcherry*20-orf_for (1422 bp). In addition, the amplicon generated by primer pair P1 + P2 is not observable any more. The presence of the plasmid pE-Flp can be demonstrated by PCR with primer pair pE-Flp for/rev (1652 bp).

The binding site of primer P2 is located within the integration module. As a consequence, the respective PCR product of primer pair P1 + P2 is no longer generated a after a successful Flp-FRT recombination. However, upon removal of the integration module, the *mcherry* gene becomes in sufficient proximity to allow a 1422-bp fragment to be generated with primer pair P1 + *mcherry*20-orf_for. The amplicon resulting from primer pair P3 + P4 (1156 bp) is retained.

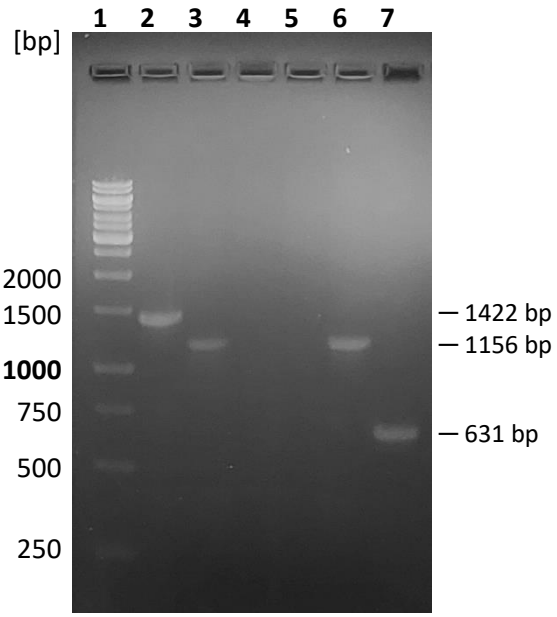


Figure 5.9-2: Agarose gel after PCR to demonstrate successful flippase-catalysed removal of the integration module in *E. coli* HB101-mCherry.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **4**: PCR with extracted chromosomal DNA after transformation of HB101-mCherry with pE-Flp (HB101-mCherry^{Flp}); **2**: PCR with primer pair P1 + *mcherry*20-orf_for; **3**: PCR with primer pair P3 + P4; **4**: PCR with primer pair P1 + P2 (negative control); **5** – **7**: PCR with extracted chromosomal DNA of HB101-mCherry (negative control, not transformed with pE-Flp); **5**: PCR with primer pair P1 + *mcherry*20-orf_for; **6**: PCR with primer pair P3 + P4; **7**: PCR with primer pair P1 + P2. Agarose gel (1.5 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 75 min. The expected amplicon lengths of each primer pair are given in Table 5.9-2.

To enable a subsequent transformation of the resulting HB101-mCherry^{Flp} with fimbriae-encoding plasmids, the removal of the flippase-encoding plasmid pE-Flp was required. This was achieved by exploiting the temperature-sensitivity of the replicon originating from pCP20,[714] which allows for plasmid replication only at low temperatures (30 °C) but not at 37 °C. Colonies that were confirmed PCR-positive (Figure 5.9-2), were cultured in liquid LB medium (Table 5.7-1) without the addition of antibiotics at 37 °C, 180 rpm overnight. On the following day, the suspension was streaked on three LB agar plates (one with ampicillin, 100 µg/mL; one with kanamycin, 20 µg/mL; one without antibiotics) and cultured initially for 2 h at 42 °C, followed by overnight incubation at 37 °C. The streaking and incubation at 37 °C of some selected colonies was repeated twice until no growth on LB agar in the presence of antibiotics was observed any more. The loss of pE-Flp was evidenced by a PCR reaction using the primer pair pE-Flp for/rev (see Table 5.7-6). The absence of the respective PCR product (1652 bp), in both colony PCR and with purified plasmids, suggested successful plasmid removal, as observed in Figure 5.9-3.

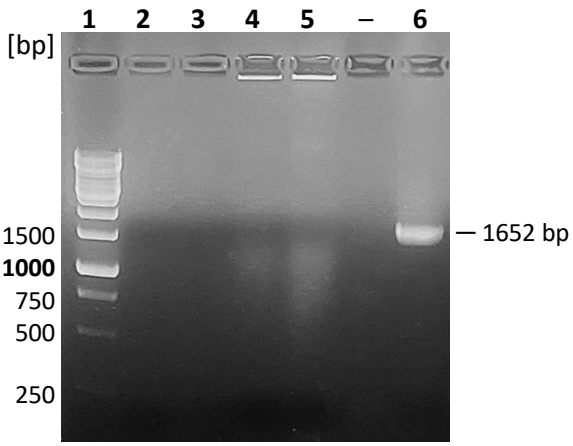


Figure 5.9-3: Agarose gel after PCR to demonstrate successful loss of pE-Flp in *E. coli* HB101-mCherry^{Flp} after 3 d incubation at 37 °C.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **3**: PCR after plasmid preparation from clone 1 and clone 2; **4** – **5**: colony PCR of the same clones 1 and 2; **6**: PCR with purified plasmid pE-Flp (positive control). Agarose gel (1.2 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 30 min. Used primer pair: pE-Flp for/rev (exp. amplicon length 1652 bp).

5.9.3 Transformation of HB101-mCherry^{Flp} with fimbriae-coding plasmids

The fimbriae-encoding plasmids pPKL4, pPKL174 and pPAP5 were transformed into HB101-mCherry^{Flp} via electroporation as described in Chapter 5.7.6 (see protocols: Preparation of competent *E. coli* and Transformation of DNA plasmids in *E. coli*). Due to the comparatively large plasmid sizes (pPKL4: ~ 11.7 kbp, pPAP5: ~ 13.60 kbp), longer recovery times about 2 – 2.5 h were chosen. The transformation result was assessed by PCR using the primer pairs Fimb-FimH for/rev for the detection of the *fimH* gene associated with the *fim* cluster and PAPG-I J96 for/rev for the detection of the class I *papG* gene within the *pap* cluster[695] (see Table 5.7-6 for sequences, annealing temperatures and amplicon lengths). The PCR result was visualised by agarose gel electrophoresis as illustrated in Figure 5.8-4 – 5.8-6).

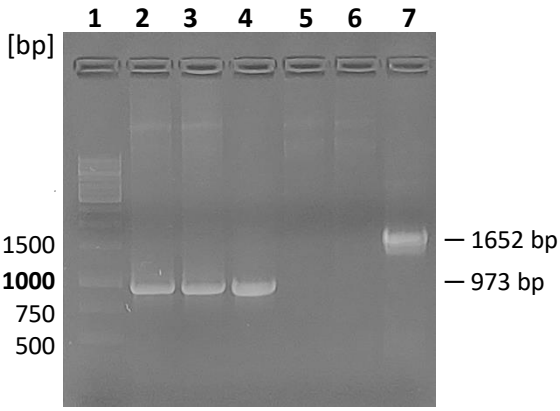


Figure 5.9-4: Agarose gel after PCR to verify the transformation of pPKL4 in HB101-mCherry^{Flp}.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **4**: PCR using primer pair Fimb-FimH for/rev (exp. amplicon length 973 bp); **2**: PCR after plasmid preparation of clone 1, **3**: PCR after plasmid preparation of clone 2; **4**: PCR with pPKL4 (positive control); **5** – **7**: PCR using primer pair pE-Flp for/rev (exp. amplicon length 1652 bp); **5**: PCR after plasmid preparation of clone 1; **6**: PCR after plasmid preparation of clone 2; **7**: PCR with pE-Flp (positive control). Agarose gel (1.2 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 50 min.

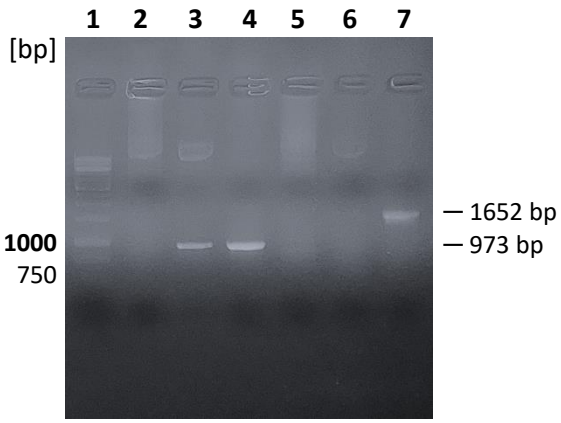


Figure 5.9-5: Agarose gel after PCR to verify the transformation of pPKL174 in HB101-mCherry^{Flp}.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **4**: PCR using primer pair Fimb-FimH for/rev (exp. amplicon length 973 bp); **2**: PCR after plasmid preparation of clone 1 (PCR-negative), **3**: PCR after plasmid preparation of clone 2; **4**: PCR with pPKL174 (positive control); **5** – **7**: PCR using primer pair pE-Flp for/rev (exp. amplicon length 1652 bp); **5**: PCR after plasmid preparation of clone 1; **6**: PCR after plasmid preparation of clone 2; **7**: PCR with pE-Flp (positive control). Agarose gel (1.2 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 35 min.

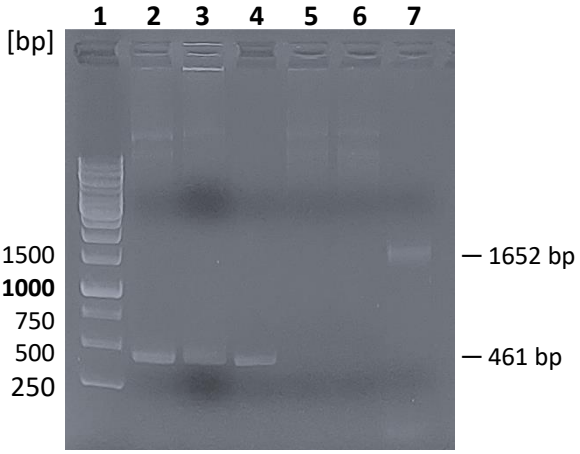


Figure 5.9-6: Agarose gel after PCR to verify the transformation of pPAP5 in HB10-mCherry^{Flp}.

1: GeneRuler 1 kb DNA Ladder (cat. no. SM0311, Thermo Scientific™); **2** – **4**: PCR using primer pair PAPG-I J96 for/rev[136] (exp. amplicon length 461 bp); **2**: colony PCR with clone 1, **3**: colony PCR with clone 2; **4**: PCR with pPAP (positive control); **5** – **7**: PCR using primer pair pE-Flp for/rev (exp. amplicon length 1652 bp); **5**: colony PCR with clone 1; **6**: colony PCR with clone 2; **7**: PCR with pE-Flp (positive control). Agarose gel (1.2 % w/v), stained with MIDORI green and visualised under blue light; electrophoresis performed at constant voltage (U = 100 V) for 50 min.

5.9.4 Fluorescence measurements

The fluorescence signal intensity of red-fluorescent HB101-mCherry^{Flp}, transformed with fimbriae-encoding plasmids (pPKL4, pPKL174: *fim* cluster; pPAP5: *pap* cluster) was evaluated at different bacterial suspension densities (OD_{600}) in order to demonstrate a linear correlation. The results were compared to the green fluorescence signal intensity of SAR18 PKL1162.

Suspensions of fluorescent bacteria were cultured and prepared in PBS, according to Chapter 5.7.8 (see protocol: Cultivation and preparation of *E. coli* in bacterial adhesion experiments). The suspensions were applied onto black 96 microwell plates (MaxiSorp™ F96, Nunc™ Thermo Scientific™) in serial 1:2-dilutions (100 µL/well) and the red fluorescence was measured at $\lambda_{ex/em}$ = 568/630 nm on an Infinite® 200 Pro M Nano+ dual mode microplate reader (Tecan). The green fluorescence of SAR18 PKL1162 was measured at $\lambda_{ex/em}$ = 485/535 nm, as usual.

The resulting relationships between bacterial density and fluorescence intensity of HB101-mCherry^{Flp} and its derivatives (Figure 4.4-1, and Figures 5.8-7 – 5.8-9) demonstrate both a linear correlation and adequate signal intensity at the chosen OD_{600} on a level comparable to the GFP-expressing SAR18 PKL1162 (Figure 5.9-10). These findings suggest a good suitability for the use of red-fluorescent HB101-mCherry^{Flp} strains in bacterial adhesion and adhesion inhibition assays.

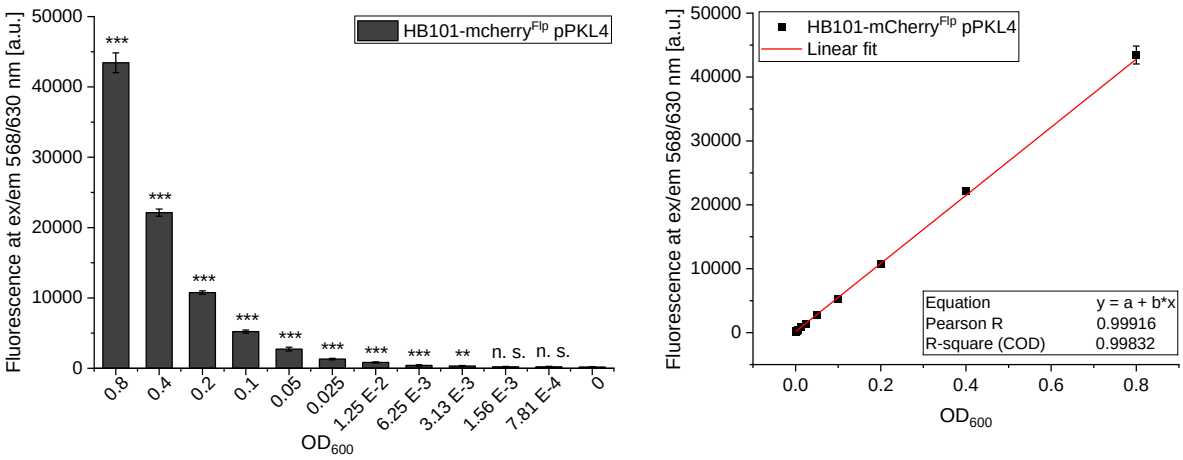


Figure 5.9-7: Red fluorescence of *E. coli* HB101-mCherry^{Flp} pPKL4.

Error bars represent the ± standard deviation of five measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (PBS, OD_{600} = 0).

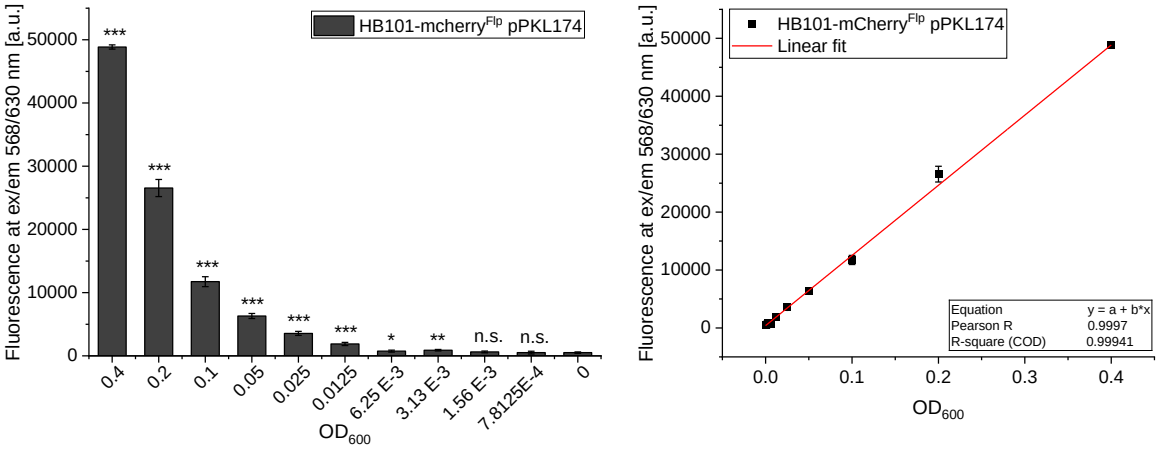


Figure 5.9-8: Red fluorescence of *E. coli* HB101-mCherry^{Flp} pPKL174.

Error bars represent the ± standard deviation of five measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (*p < 0.05, **p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (PBS, OD_{600} = 0).

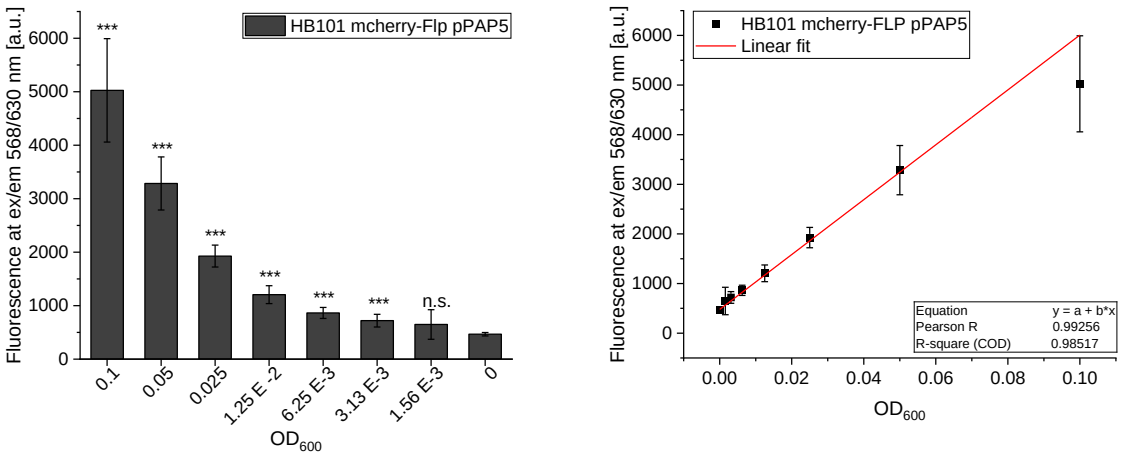


Figure 5.9-9: Red fluorescence of *E. coli* HB101-mCherry^{Flp} pPAP5.

Error bars represent the ± standard deviation of five measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (***p < 0.001, two-sample t-test) compared to the control (PBS, OD_{600} = 0).

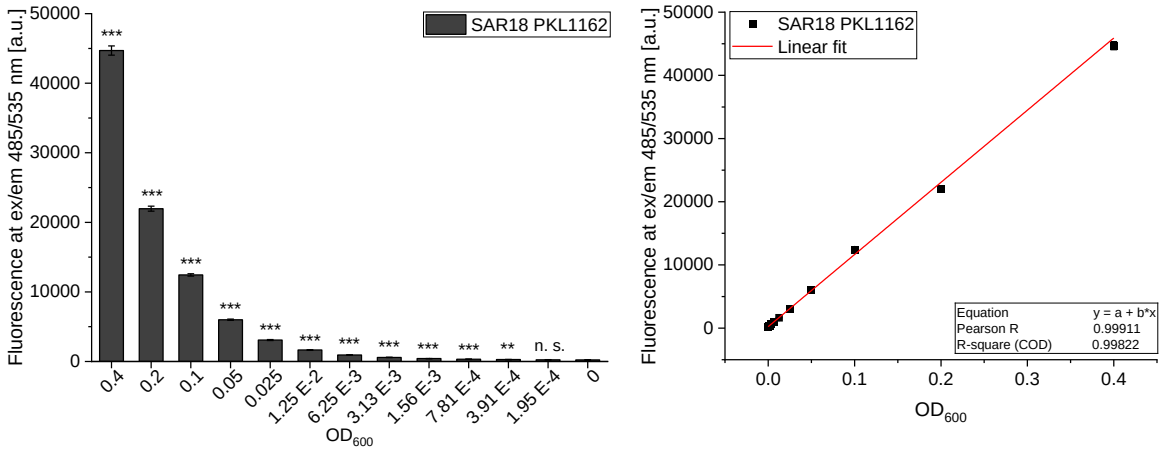


Figure 5.9-10: Green fluorescence of SAR18 PKL1162.

Error bars represent the ± standard deviation of five measurements. n.s. denotes statistically non-significant signal differences and the asterisks (*) denote statistically significant signal differences (**p < 0.01, ***p < 0.001, two-sample t-test) compared to the control (PBS, OD_{600} = 0).

5.9.5 Adhesion and adhesion inhibition experiments with HB101-mCherry^{Flp}

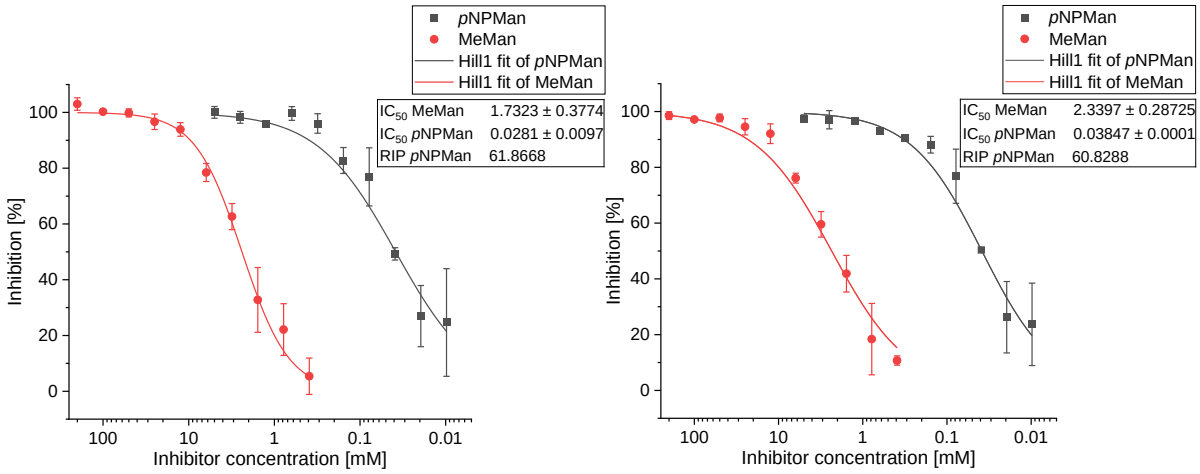


Figure 5.9-11: Representative results from adhesion inhibition experiments with HB101-mCherry^{Flp} pPKL4 using *p*NPMan as an inhibitor.

Inhibition experiments with *p*-nitrophenyl *α*-D-mannopyranoside (*p*NPMan). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

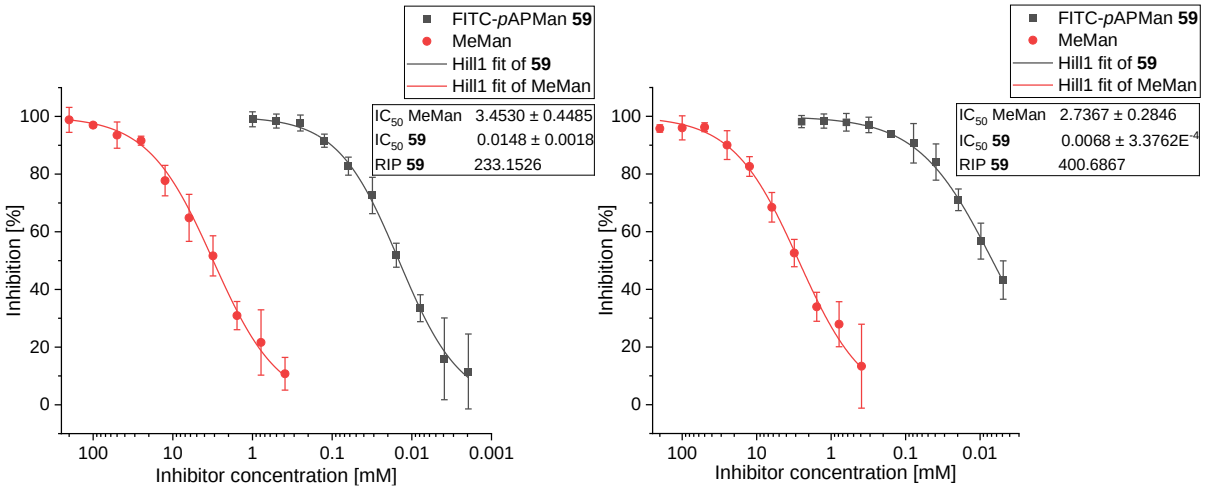


Figure 5.9-12: Representative results from adhesion inhibition experiments with HB101-mCherry^{Flp} pPKL4 using **59** (FITC-*p*APMan) as an inhibitor.

Inhibition experiments with fluorescein-labelled *α*-D-mannopyranoside **59** (FITC-*p*AP-Man). Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curves represent the ± standard deviation of quadruplicate measurements.

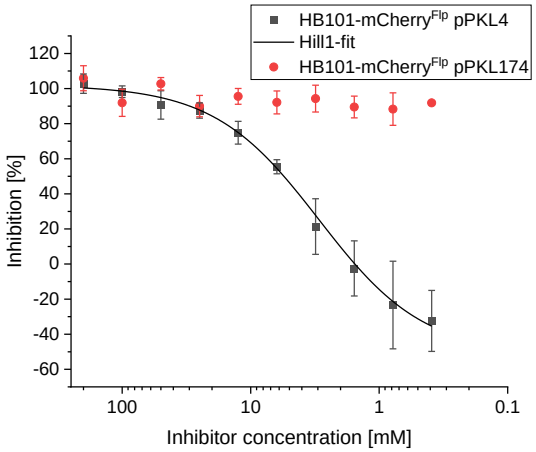


Figure 5.9-13: Bacterial adhesion inhibition experiment with HB101-mCherry^{Flp}, transformed with the type-1 fimbriae encoding plasmids pPKL4 (black) and pPKL174 (red).

The bacterial adhesion was inhibited with serial dilutions of methyl *α*-D-mannopyranoside (MeMan). In contrast to the red-fluorescent strain, transformed with the plasmid pPKL4, no inhibition effect of HB101-mCherry^{Flp} pPKL174 was observed. Sigmoidal fitting was performed according to the Hill1 equation. Error bars in the dose-response curve represent the ± standard deviation of quadruplicate measurements.

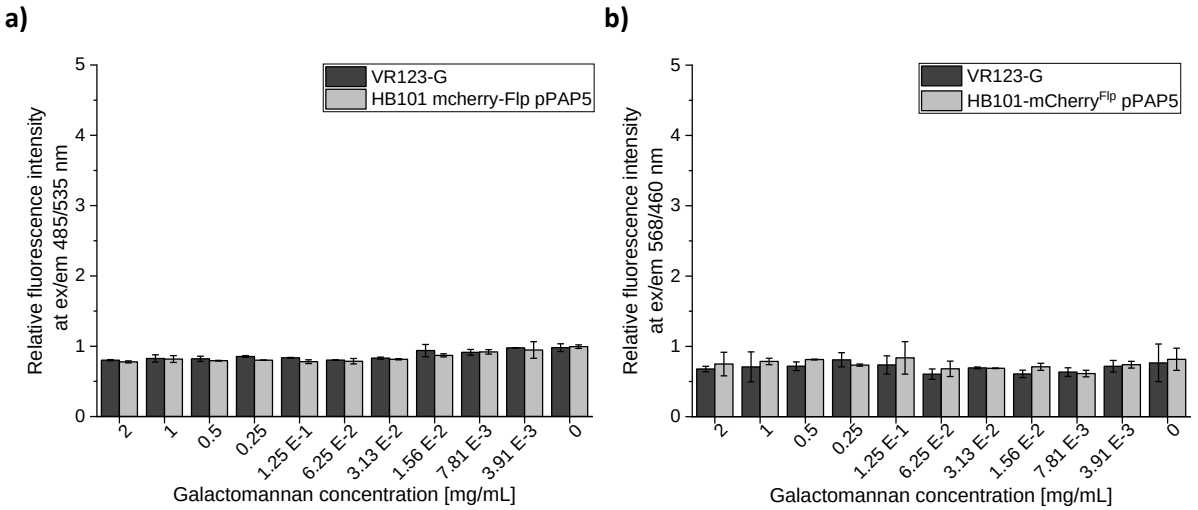


Figure 5.9-14: Adhesion experiment with P-fimbriated *E. coli* (green-fluorescent VR123-G and red-fluorescent HB101-mCherry^{Flp} pPAP5) on a polystyrene-surface, coated with serial 1:2 dilutions of galactomannan.

a) Fluorescence at $\lambda_{ex/em}$ = 485/353 nm. **b)** Fluorescence at $\lambda_{ex/em}$ = 568/640 nm. Error bars in the dose-response relationships represent the ± standard deviation of triplicate experiments.

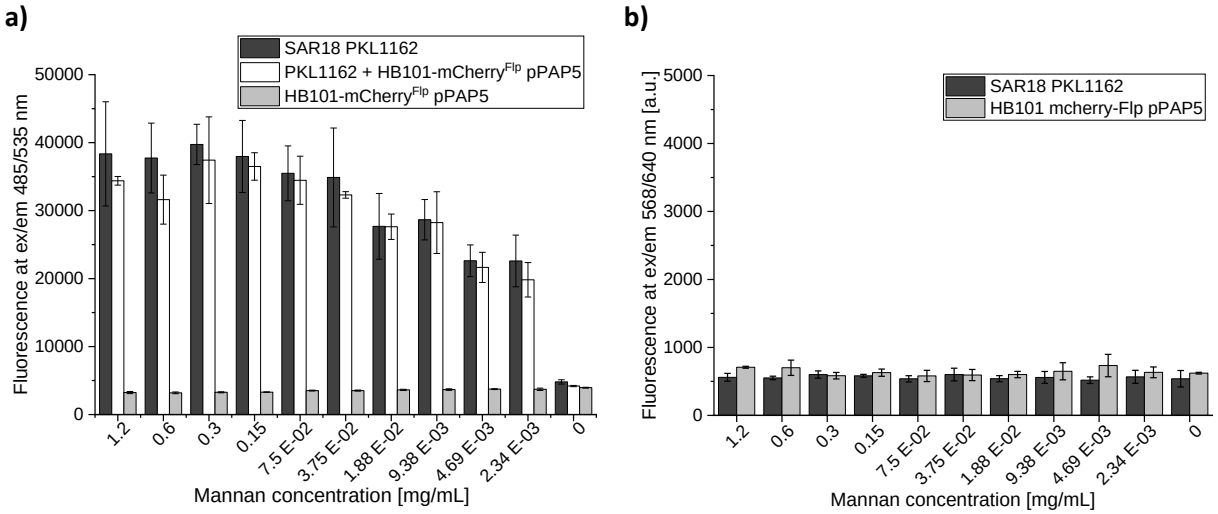


Figure 5.9-15: Selective adhesion experiment with green-fluorescent, type 1-fimbriated SAR18 PKL1162 and red-fluorescent, P-fimbriated HB101-mCherry^{Flp} pPAP5 on a polystyrene-surface, coated with serial 1:2 dilutions of mannan from *Saccharomyces cerevisiae*.

a) Fluorescence at $\lambda_{ex/em}$ = 485/353 nm. **b)** Fluorescence at $\lambda_{ex/em}$ = 568/640 nm. Error bars in the dose-response relationships represent the ± standard deviation of triplicate experiments.

6. Additional tools

The following additional tools were employed in this dissertation: In general, figures and illustrations were generated using PowerPoint (Office 2016), unless otherwise stated. Chemical structures were created using the ChemDraw Prime and ChemDraw Professional software (PerkinElmer), which also provided several templates for biological macromolecules and cellular organelles, taken from the BioDraw, Advanced BioDraw, and Anatomy toolbars. For data analysis and graphical visualisation, the software Origin 2021 (OriginLab) was employed. The latest version (7) of the citation software Zotero (https://www.zotero.org/) in combination with the corresponding Word Processor plugin was utilised for generating the reference list. Where necessary, translation support from German to English was provided by DeepL (https://www.deepl.com/ translator) as well as by online dictionaries such as dict.cc (https://www.dict.cc/) and LEO (https://dict.leo.org/). Final proofreading with regard to correct grammar, spelling, and idiomatic language use was performed using ChatGPT (https://chatgpt.com, OpenAI). No artificial intelligence tools were used for generating any of the scientific content, images, and graphics or for data analysis presented in this work.

7. References

[1] K. J. Hellingwerf, *J. Photochem. Photobiol. B* **2000**, *54*, 94–102.
[2] M. A. van der Horst, K. J. Hellingwerf, *Acc. Chem. Res.* **2004**, *37*, 13–20.
[3] X. Li, H. Ren, M. Kundu, Z. Liu, F. W. Zhong, L. Wang, J. Gao, D. Zhong, *Nat. Commun.* **2020**, *11*, 4316.
[4] R. Podolec, R. Ulm, *Curr. Opin. Plant Biol.* **2018**, *45*, 18–25.
[5] A.-Y. Shin, Y.-J. Han, A. Baek, T. Ahn, S. Y. Kim, T. S. Nguyen, M. Son, K. W. Lee, Y. Shen, P.-S. Song, J.-I. Kim, *Nat. Commun.* **2016**, *7*, 11545.
[6] A. Takemiya, S. Inoue, M. Doi, T. Kinoshita, K. Shimazaki, *Plant Cell* **2005**, *17*, 1120–1127.
[7] T. Kinoshita, M. Doi, N. Suetsugu, T. Kagawa, M. Wada, K. Shimazaki, *Nature* **2001**, *414*, 656–660.
[8] Q. Wang, C. Lin, *Annu. Rev. Plant Biol.* **2020**, *71*, 103–129.
[9] Z. Liu, L. Wang, D. Zhong, *Phys. Chem. Chem. Phys.* **2015**, *17*, 11933–11949.
[10] J. A. Jarillo, H. Gabrys, J. Capel, J. M. Alonso, J. R. Ecker, A. R. Cashmore, *Nature* **2001**, *410*, 952–954.
[11] T. Kagawa, T. Sakai, N. Suetsugu, K. Oikawa, S. Ishiguro, T. Kato, S. Tabata, K. Okada, M. Wada, *Science* **2001**, *291*, 2138–2141.
[12] G. Kreimer, in *Int. Rev. Cytol.* (Eds.: K.W. Jeon, J. Jarvik), Academic Press, **1994**, pp. 229–310.
[13] P. Schopfer, A. Brennicke, in *Pflanzenphysiologie*, Springer Spektrum, Berlin, **2010**, pp. 445–470.
[14] W. L. Butler, K. H. Norris, H. W. Siegelman, S. B. Hendricks, *Proc. Natl. Acad. Sci.* **1959**, *45*, 1703–1708.
[15] F.-W. Li, M. Melkonian, C. J. Rothfels, J. C. Villarreal, D. W. Stevenson, S. W. Graham, G. K.-S. Wong, K. M. Pryer, S. Mathews, *Nat. Commun.* **2015**, *6*, 7852.
[16] F. Nagy, E. Schäfer, *Annu. Rev. Plant Biol.* **2002**, *53*, 329–355.
[17] J. Li, G. Li, H. Wang, X. Wang Deng, *Arab. Book Am. Soc. Plant Biol.* **2011**, *9*, e0148.
[18] S. N. Warnasooriya, B. L. Montgomery, *Plant Signal. Behav.* **2011**, *6*, 2047–2050.
[19] Y. Y. Levy, C. Dean, *Plant Cell* **1998**, *10*, 1973–1989.
[20] R. A. Sharrock, *Genome Biol.* **2008**, *9*, 230.
[21] M. Legris, Y. Ç. Ince, C. Fankhauser, *Nat. Commun.* **2019**, *10*, 5219.
[22] V. N. Pham, P. K. Kathare, E. Huq, *Plant Physiol.* **2018**, *176*, 1025–1038.
[23] Y. Hirshberg, *Compt. Rend. Acad. Sci.* **1950**, 903.
[24] J. C. Lagarias, H. Rapoport, *J. Am. Chem. Soc.* **1980**, *102*, 4821–4828.
[25] N. C. Rockwell, J. C. Lagarias, *Plant Cell* **2006**, *18*, 4–14.
[26] F. Frascogna, B. Ledermann, J. Hartmann, E. Pérez Patallo, F. Zeqiri, E. Hofmann, N. Frankenberg-Dinkel, *Plant Physiol.* **2023**, *193*, 246–258.
[27] Y. Yang, T. Stensitzki, L. Sauthof, A. Schmidt, P. Piwowarski, F. Velazquez Escobar, N. Michael, A. D. Nguyen, M. Szczepek, F. N. Brünig, R. R. Netz, M. A. Mroginski, S. Adam, F. Bartl, I. Schapiro, P. Hildebrandt, P. Scheerer, K. Heyne, *Nat. Chem.* **2022**, *14*, 823–830.
[28] C. J. P. Spruit, R. E. Kendrick, *Photochem. Photobiol.* **1977**, *26*, 133–138.
[29] F. Velazquez Escobar, C. Kneip, N. Michael, T. Hildebrandt, N. Tavraz, W. Gärtner, J. Hughes, T. Friedrich, P. Scheerer, M. A. Mroginski, P. Hildebrandt, *J. Phys. Chem. B* **2020**, *124*, 4044–4055.
[30] M. Bischoff, G. Hermann, S. Rentsch, D. Strehlow, *Biochemistry* **2001**, *40*, 181–186.
[31] M. Carrillo, S. Pandey, J. Sanchez, M. Noda, I. Poudyal, L. Aldama, T. N. Malla, E. Claesson, W. Y. Wahlgren, D. Feliz, V. Šrajer, M. Maj, L. Castillon, S. Iwata, E. Nango, R. Tanaka, T. Tanaka, L. Fangjia, K. Tono, S. Owada, S. Westenhoff, E. A. Stojković, M. Schmidt, *Struct. Lond. Engl. 1993* **2021**, *29*, 743-754.e4.
[32] E. S. Burgie, J. Zhang, R. D. Vierstra, *Structure* **2016**, *24*, 448–457.
[33] E. S. Burgie, A. N. Bussell, J. M. Walker, K. Dubiel, R. D. Vierstra, *Proc. Natl. Acad. Sci.* **2014**, *111*, 10179–10184.

[34] A. Schmidt, L. Sauthof, M. Szczepek, M. F. Lopez, F. V. Escobar, B. M. Qureshi, N. Michael, D. Buhrke, T. Stevens, D. Kwiatkowski, D. von Stetten, M. A. Mroginski, N. Krauß, T. Lamparter, P. Hildebrandt, P. Scheerer, *Nat. Commun.* **2018**, *9*, 4912.
[35] J. Moon, Y. Lee, H. Ihee, *Chem. Commun.* **2024**, *60*, 9472–9482.
[36] A. M. Orville, *Curr. Opin. Struct. Biol.* **2020**, *65*, 193–208.
[37] B. Durbeej, O. A. Borg, L. A. Eriksson, *Phys. Chem. Chem. Phys.* **2004**, *6*, 5066–5073.
[38] C. Klose, F. Nagy, E. Schäfer, *Mol. Plant* **2020**, *13*, 386–397.
[39] V. Gordeliy, K. Kovalev, E. Bamberg, F. Rodriguez-Valera, E. Zinovev, D. Zabelskii, A. Alekseev, R. Rosselli, I. Gushchin, I. Okhrimenko, in *Rhodopsin Methods Protoc.* (Ed.: V. Gordeliy), Springer US, New York, NY, **2022**, pp. 1–52.
[40] E. Kühne, in *Untersuchungen Über Den Sehpurpur*, C. Winter Universitätsbuchhandlung, Heidelberg, **1878**, pp. 139–218.
[41] R. A. Bogomolni, J. L. Spudich, *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 6250–6254.
[42] H. Irieda, T. Morita, K. Maki, M. Homma, H. Aiba, Y. Sudo, *J. Biol. Chem.* **2012**, *287*, 32485–32493.
[43] G. M. Avelar, R. I. Schumacher, P. A. Zaini, G. Leonard, T. A. Richards, S. L. Gomes, *Curr. Biol. CB* **2014**, *24*, 1234–1240.
[44] H. Kandori, *Front. Mol. Biosci.* **2015**, *2*, 52.
[45] N. D. Larusso, B. E. Ruttenberg, A. K. Singh, T. H. Oakley, *J. Mol. Evol.* **2008**, *66*, 417–423.
[46] K. Kojima, Y. Sudo, *RSC Adv.* **2023**, *13*, 5367–5381.
[47] T. E. Meyer, *Biochim. Biophys. Acta BBA - Bioenerg.* **1985**, *806*, 175–183.
[48] T. E. Meyer, E. Yakali, M. A. Cusanovich, G. Tollin, *Biochemistry* **1987**, *26*, 418–423.
[49] W. D. Hoff, P. Dux, K. Hard, B. Devreese, I. M. Nugteren-Roodzant, W. Crielaard, R. Boelens, R. Kaptein, J. Van Beeumen, K. J. Hellingwerf, *Biochemistry* **1994**, *33*, 13959–13962.
[50] W. W. Sprenger, W. D. Hoff, J. P. Armitage, K. J. Hellingwerf, *J. Bacteriol.* **1993**, *175*, 3096–3104.
[51] D. Goodsell, PDB101: Molecule of the Month: Photoactive Yellow Protein. http://doi.org/10.2210/rcsb_pdb/mom_2017_3 [13 March 2025].
[52] G. E. Borgstahl, D. R. Williams, E. D. Getzoff, *Biochemistry* **1995**, *34*, 6278–6287.
[53] A. Xie, L. Kelemen, J. Hendriks, B. J. White, K. J. Hellingwerf, W. D. Hoff, *Biochemistry* **2001**, *40*, 1510–1517.
[54] P. L. Ramachandran, J. E. Lovett, P. J. Carl, M. Cammarata, J. H. Lee, Y. O. Jung, H. Ihee, C. R. Timmel, J. J. van Thor, *J. Am. Chem. Soc.* **2011**, *133*, 9395–9404.
[55] R. S. Molday, O. L. Moritz, *J. Cell Sci.* **2015**, *128*, 4039–4045.
[56] P. A. Hargrave, J. H. McDowell, *FASEB J.* **1992**, *6*, 2323–2331.
[57] G. Wald, P. K. Brown, P. H. Smith, *J. Gen. Physiol.* **1955**, *38*, 623–681.
[58] S. Hattar, H.-W. Liao, M. Takao, D. M. Berson, K.-W. Yau, *Science* **2002**, *295*, 1065–1070.
[59] M. T. H. Do, K.-W. Yau, *Physiol. Rev.* **2010**, *90*, 1547–1581.
[60] J. J. Gooley, S. M. W. Rajaratnam, G. C. Brainard, R. E. Kronauer, C. A. Czeisler, S. W. Lockley, *Sci. Transl. Med.* **2010**, *2*, 31ra33-31ra33.
[61] S. N. Peirson, S. Halford, R. G. Foster, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2009**, *364*, 2849–2865.
[62] Y. Shichida, T. Matsuyama, *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 2881–2895.
[63] T. Okada, K. Palczewski, *Curr. Opin. Struct. Biol.* **2001**, *11*, 420–426.
[64] D. Bownds, *Nature* **1967**, *216*, 1178–1181.
[65] M. F. Brown, *Chem. Phys. Lipids* **1994**, *73*, 159–180.
[66] G. Wald, *Science* **1968**, *162*, 230–239.
[67] A. K. Kusnetzow, A. Dukkipati, K. R. Babu, L. Ramos, B. E. Knox, R. R. Birge, *Proc. Natl. Acad. Sci.* **2004**, *101*, 941–946.
[68] S. Yokoyama, *Prog. Retin. Eye Res.* **2000**, *19*, 385–419.
[69] S. Deeb, *Clin. Genet.* **2005**, *67*, 369–377.
[70] H. Horiguchi, J. Winawer, R. F. Dougherty, B. A. Wandell, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E260–E269.
[71] L. Zhu, Y. Imanishi, S. Filipek, A. Alekseev, B. Jastrzebska, W. Sun, D. A. Saperstein, K. Palczewski, *J. Biol. Chem.* **2006**, *281*, 22289–22298.

[72] J. C. Gardner, G. Liew, Y.-H. Quan, B. Ermetal, H. Ueyama, A. E. Davidson, N. Schwarz, N. Kanuga, R. Chana, E. R. Maher, A. R. Webster, G. E. Holder, A. G. Robson, M. E. Cheetham, J. Liebelt, J. B. Ruddle, A. T. Moore, M. Michaelides, A. J. Hardcastle, *Hum. Mutat.* **2014**, *35*, 1354–1362.
[73] L. Stryer, *J. Biol. Chem.* **1991**, *266*, 10711–10714.
[74] R. Hubbard, A. Kropf, *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 130–139.
[75] X. E. Zhou, K. Melcher, H. E. Xu, *Acta Pharmacol. Sin.* **2012**, *33*, 291–299.
[76] R. W. Schoenlein, L. A. Peteanu, R. A. Mathies, C. V. Shank, *Science* **1991**, *254*, 412–415.
[77] J. E. Kim, M. J. Tauber, R. A. Mathies, *Biochemistry* **2001**, *40*, 13774–13778.
[78] A. Cooper, *Nature* **1979**, *282*, 531–533.
[79] U. M. Ganter, W. Gärtner, F. Siebert, *Biochemistry* **1988**, *27*, 7480–7488.
[80] O. P. Ernst, F. J. Bartl, *ChemBioChem* **2002**, *3*, 968–974.
[81] D. Salom, D. T. Lodowski, R. E. Stenkamp, I. L. Trong, M. Golczak, B. Jastrzebska, T. Harris, J. A. Ballesteros, K. Palczewski, *Proc. Natl. Acad. Sci.* **2006**, *103*, 16123–16128.
[82] H. Nakamichi, T. Okada, *Proc. Natl. Acad. Sci.* **2006**, *103*, 12729–12734.
[83] R. Vogel, M. Mahalingam, S. Lüdeke, T. Huber, F. Siebert, T. P. Sakmar, *J. Mol. Biol.* **2008**, *380*, 648–655.
[84] H.-W. Choe, Y. J. Kim, J. H. Park, T. Morizumi, E. F. Pai, N. Krauß, K. P. Hofmann, P. Scheerer, O. P. Ernst, *Nature* **2011**, *471*, 651–655.
[85] P. Scheerer, J. H. Park, P. W. Hildebrand, Y. J. Kim, N. Krauß, H.-W. Choe, K. P. Hofmann, O. P. Ernst, *Nature* **2008**, *455*, 497–502.
[86] H. M. Rarick, N. O. Artemyev, H. E. Hamm, *Science* **1992**, *256*, 1031–1033.
[87] E. H. Choi, A. Daruwalla, S. Suh, H. Leinonen, K. Palczewski, *J. Lipid Res.* **2021**, *62*, 100040.
[88] P. D. Kiser, M. Golczak, A. Maeda, K. Palczewski, *Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids* **2012**, *1821*, 137–151.
[89] K. J. Hellingwerf, J. Hendriks, T. Gensch, *J. Phys. Chem. A* **2003**, *107*, 1082–1094.
[90] L. J. G. W. van Wilderen, M. A. van der Horst, I. H. M. van Stokkum, K. J. Hellingwerf, R. van Grondelle, M. L. Groot, *Proc. Natl. Acad. Sci.* **2006**, *103*, 15050–15055.
[91] J. D. Hong, K. Palczewski, *BioEssays* **2023**, *45*, 2300068.
[92] O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown, H. Kandori, *Chem. Rev.* **2014**, *114*, 126–163.
[93] R. S. H. Liu, A. E. Asato, *Tetrahedron* **1984**, *40*, 1931–1969.
[94] S. He, A. M. Linz, S. L. R. Stevens, P. Q. Tran, F. Moya-Flores, B. O. Oyserman, J. R. Dwulit-Smith, K. T. Forest, K. D. McMahon, *Mol. Ecol.* **2023**, *32*, 2798–2817.
[95] N. Yutin, E. V. Koonin, *Biol. Direct* **2012**, *7*, 34.
[96] D. Oesterhelt, W. Stoeckenius, *Nature. New Biol.* **1971**, *233*, 149–152.
[97] D. Oesterhelt, W. Stoeckenius, *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70*, 2853–2857.
[98] Y. Mukohata, Y. Sugiyama, K. Ihara, M. Yoshida, *Biochem. Biophys. Res. Commun.* **1988**, *151*, 1339–1345.
[99] O. Béjà, L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, E. F. DeLong, *Science* **2000**, *289*, 1902–1906.
[100] J. K. Lanyi, *J. Biol. Chem.* **1997**, *272*, 31209–31212.
[101] J. Pinhassi, E. F. DeLong, O. Béjà, J. M. González, C. Pedrós-Alió, *Microbiol. Mol. Biol. Rev. MMBR* **2016**, *80*, 929–954.
[102] R. Henderson, P. N. T. Unwin, *Nature* **1975**, *257*, 28–32.
[103] S. Grzesiek, N. A. Dencher, *Proc. Natl. Acad. Sci.* **1988**, *85*, 9509–9513.
[104] Y. Mukai, N. Kamo, S. Mitaku, *Protein Eng. Des. Sel.* **1999**, *12*, 755–759.
[105] E. M. Landau, J. P. Rosenbusch, *Proc. Natl. Acad. Sci.* **1996**, *93*, 14532–14535.
[106] E. Pebay-Peyroula, G. Rummel, J. P. Rosenbusch, E. M. Landau, *Science* **1997**, *277*, 1676–1681.
[107] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J. K. Lanyi, *J. Mol. Biol.* **1999**, *291*, 899–911.
[108] J. Tittor, D. Oesterhelt, *FEBS Lett.* **1990**, *263*, 269–273.
[109] C. Wickstrand, P. Nogly, E. Nango, S. Iwata, J. Standfuss, R. Neutze, *Annu. Rev. Biochem.* **2019**, *88*, 59–83.

[110] A. Shigeta, Y. Otani, R. Miyasa, Y. Makino, I. Kawamura, T. Okitsu, A. Wada, A. Naito, *Membranes* **2022**, *12*, 279.
[111] T. Takahashi, Y. Mochizuki, N. Kamo, Y. Kobatake, *Biochem. Biophys. Res. Commun.* **1985**, *127*, 99–105.
[112] A. Matsuno-Yagi, Y. Mukohata, *Arch. Biochem. Biophys.* **1980**, *199*, 297–303.
[113] B. Schobert, J. K. Lanyi, *J. Biol. Chem.* **1982**, *257*, 10306–10313.
[114] A. K. Sharma, D. A. Walsh, E. Bapteste, F. Rodriguez-Valera, W. Ford Doolittle, R. T. Papke, *BMC Evol. Biol.* **2007**, *7*, 79.
[115] S. Yoshizawa, Y. Kumagai, H. Kim, Y. Ogura, T. Hayashi, W. Iwasaki, E. F. DeLong, K. Kogure, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6732–6737.
[116] T. Hasemi, T. Kikukawa, N. Kamo, M. Demura, *J. Biol. Chem.* **2016**, *291*, 355–362.
[117] D. Oesterhelt, *Isr. J. Chem.* **1995**, *35*, 475–494.
[118] M. Kolbe, H. Besir, L. O. Essen, D. Oesterhelt, *Science* **2000**, *288*, 1390–1396.
[119] M. Schreiner, R. Schlesinger, J. Heberle, H. H. Niemann, *J. Struct. Biol.* **2015**, *190*, 373–378.
[120] L.-O. Essen, *Curr. Opin. Struct. Biol.* **2002**, *12*, 516–522.
[121] S. Mous, G. Gotthard, D. Ehrenberg, S. Sen, T. Weinert, P. J. M. Johnson, D. James, K. Nass, A. Furrer, D. Kekilli, P. Ma, S. Brünle, C. M. Casadei, I. Martiel, F. Dworkowski, D. Gashi, P. Skopintsev, M. Wranik, G. Knopp, E. Panepucci, V. Panneels, C. Cirelli, D. Ozerov, G. F. X. Schertler, M. Wang, C. Milne, J. Standfuss, I. Schapiro, J. Heberle, P. Nogly, *Science* **2022**, *375*, 845–851.
[122] G. Nagel, D. Ollig, M. Fuhrmann, S. Kateriya, A. M. Musti, E. Bamberg, P. Hegemann, *Science* **2002**, *296*, 2395–2398.
[123] S. O. Mast, *J. Exp. Zool.* **1916**, *20*, 1–17.
[124] K. W. Foster, R. D. Smyth, *Microbiol. Rev.* **1980**, *44*, 572–630.
[125] O. A. Sineshchekov, K.-H. Jung, J. L. Spudich, *Proc. Natl. Acad. Sci.* **2002**, *99*, 8689–8694.
[126] G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann, E. Bamberg, *Proc. Natl. Acad. Sci.* **2003**, *100*, 13940–13945.
[127] P. Hegemann, *Planta* **1997**, *203*, 265–274.
[128] S. Kateriya, G. Nagel, E. Bamberg, P. Hegemann, *Physiology* **2004**, *19*, 133–137.
[129] M. Bessen, R. Fay, G. Witman, *J. Cell Biol.* **1980**, *86*, 446–455.
[130] S. K. Sanyal, M. Awasthi, P. Ranjan, S. Sharma, G. K. Pandey, S. Kateriya, *Int. J. Biol. Macromol.* **2023**, *245*, 125492.
[131] K. Seth, G. Kumawat, P. Vyas, Harish, *J. Basic Microbiol.* **2022**, *62*, 1169–1178.
[132] B. HyperBook, *Bioeng. Hyperbook* **2024**.
[133] E. G. Govorunova, K.-H. Jung, O. A. Sineshchekov, J. L. Spudich, *Biophys. J.* **2004**, *86*, 2342–2349.
[134] T. M. Mittelmeier, J. S. Boyd, M. R. Lamb, C. L. Dieckmann, *J. Cell Biol.* **2011**, *193*, 741–753.
[135] H. C. Watanabe, K. Welke, F. Schneider, S. Tsunoda, F. Zhang, K. Deisseroth, P. Hegemann, M. Elstner, *J. Biol. Chem.* **2012**, *287*, 7456–7466.
[136] H. E. Kato, F. Zhang, O. Yizhar, C. Ramakrishnan, T. Nishizawa, K. Hirata, J. Ito, Y. Aita, T. Tsukazaki, S. Hayashi, P. Hegemann, A. D. Maturana, R. Ishitani, K. Deisseroth, O. Nureki, *Nature* **2012**, *482*, 369–374.
[137] Y. Hontani, M. Marazzi, K. Stehfest, T. Mathes, I. H. M. van Stokkum, M. Elstner, P. Hegemann, J. T. M. Kennis, *Sci. Rep.* **2017**, *7*, 7217.
[138] M.-K. Verhoefen, C. Bamann, R. Blöcher, U. Förster, E. Bamberg, J. Wachtveitl, *ChemPhysChem* **2010**, *11*, 3113–3122.
[139] K. Oda, T. Nomura, T. Nakane, K. Yamashita, K. Inoue, S. Ito, J. Vierock, K. Hirata, A. D. Maturana, K. Katayama, T. Ikuta, I. Ishigami, T. Izume, R. Umeda, R. Eguma, S. Oishi, G. Kasuya, T. Kato, T. Kusakizako, W. Shihoya, H. Shimada, T. Takatsuji, M. Takemoto, R. Taniguchi, A. Tomita, R. Nakamura, M. Fukuda, H. Miyauchi, Y. Lee, E. Nango, R. Tanaka, T. Tanaka, M. Sugahara, T. Kimura, T. Shimamura, T. Fujiwara, Y. Yamanaka, S. Owada, Y. Joti, K. Tono, R. Ishitani, S. Hayashi, H. Kandori, P. Hegemann, S. Iwata, M. Kubo, T. Nishizawa, O. Nureki, *eLife* **2021**, *10*, e62389.
[140] M. Nack, I. Radu, M. Gossing, C. Bamann, E. Bamberg, G. F. von Mollard, J. Heberle, *Photochem. Photobiol. Sci.* **2010**, *9*, 194–198.

[141] Q. Xin, W. Zhang, S. Yuan, *Int. J. Mol. Sci.* **2023**, *24*, 5667.
[142] M. Müller, C. Bamann, E. Bamberg, W. Kühlbrandt, *J. Mol. Biol.* **2015**, *427*, 341–349.
[143] O. Volkov, K. Kovalev, V. Polovinkin, V. Borshchevskiy, C. Bamann, R. Astashkin, E. Marin, A. Popov, T. Balandin, D. Willbold, G. Büldt, E. Bamberg, V. Gordeliy, *Science* **2017**, *358*, eaan8862.
[144] Y. Shan, L. Zhao, M. Chen, X. Li, M. Zhang, D. Pei, *Nat. Commun.* **2024**, *15*, 7292.
[145] K. Deisseroth, G. Feng, A. K. Majewska, G. Miesenböck, A. Ting, M. J. Schnitzer, *J. Neurosci. Off. J. Soc. Neurosci.* **2006**, *26*, 10380–10386.
[146] E. S. Boyden, *F1000 Biol. Rep.* **2011**, *3*, 11.
[147] W. Chen, C. Li, W. Liang, Y. Li, Z. Zou, Y. Xie, Y. Liao, L. Yu, Q. Lin, M. Huang, Z. Li, X. Zhu, *Front. Aging Neurosci.* **2022**, *14*, 867863.
[148] B. V. Zemelman, G. A. Lee, M. Ng, G. Miesenböck, *Neuron* **2002**, *33*, 15–22.
[149] C. Lang-Hinrichs, I. Queck, G. Büldt, U. Stahl, V. Hildebrandt, *Mol. Gen. Genet. MGG* **1994**, *244*, 183–188.
[150] G. Nagel, B. Möckel, G. Büldt, E. Bamberg, *FEBS Lett.* **1995**, *377*, 263–266.
[151] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel, K. Deisseroth, *Nat. Neurosci.* **2005**, *8*, 1263–1268.
[152] Y. K. Cho, D. Li, in *Optogenetics Basic Concepts Their Dev.* (Ed.: A. Kianianmomeni), Springer, New York, NY, **2016**, pp. 1–17.
[153] V. Gradinaru, K. R. Thompson, K. Deisseroth, *Brain Cell Biol.* **2008**, *36*, 129–139.
[154] A. Kianianmomeni, A. Hallmann, *Trends Biochem. Sci.* **2015**, *40*, 624–627.
[155] X. Li, D. V. Gutierrez, M. G. Hanson, J. Han, M. D. Mark, H. Chiel, P. Hegemann, L. T. Landmesser, S. Herlitze, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17816–17821.
[156] G. Nagel, M. Brauner, J. F. Liewald, N. Adeishvili, E. Bamberg, A. Gottschalk, *Curr. Biol.* **2005**, *15*, 2279–2284.
[157] B. R. Arenkiel, J. Peca, I. G. Davison, C. Feliciano, K. Deisseroth, G. J. Augustine, M. D. Ehlers, G. Feng, *Neuron* **2007**, *54*, 205–218.
[158] A. D. Douglass, S. Kraves, K. Deisseroth, A. F. Schier, F. Engert, *Curr. Biol.* **2008**, *18*, 1133–1137.
[159] X. Han, X. Qian, J. G. Bernstein, H. Zhou, G. T. Franzesi, P. Stern, R. T. Bronson, A. M. Graybiel, R. Desimone, E. S. Boyden, *Neuron* **2009**, *62*, 191–198.
[160] M. Stierl, P. Stumpf, D. Udwari, R. Gueta, R. Hagedorn, A. Losi, W. Gärtner, L. Petereit, M. Efetova, M. Schwarzel, T. G. Oertner, G. Nagel, P. Hegemann, *J. Biol. Chem.* **2011**, *286*, 1181–1188.
[161] A. Bi, J. Cui, Y.-P. Ma, E. Olshevskaya, M. Pu, A. M. Dizhoor, Z.-H. Pan, *Neuron* **2006**, *50*, 23–33.
[162] J.-A. Sahel, E. Boulanger-Scemama, C. Pagot, A. Arleo, F. Galluppi, J. N. Martel, S. D. Esposti, A. Delaux, J.-B. de Saint Aubert, C. de Montleau, E. Gutman, I. Audo, J. Duebel, S. Picaud, D. Dalkara, L. Blouin, M. Taiel, B. Roska, *Nat. Med.* **2021**, *27*, 1223–1229.
[163] P. Klán, J. Wirz, in *Mol. Photoswitches - Chem. Prop. Appl.*, Wiley-VCH, Weinheim, Germany, **2022**, pp. 4–19.
[164] J. Volarić, W. Szymanski, N. A. Simeth, B. L. Feringa, *Chem. Soc. Rev.* **2021**, *50*, 12377–12449.
[165] M. Zhou, S. Mathew, B. de Bruin, *J. Am. Chem. Soc.* **2023**, *145*, 645–657.
[166] H. Bouas-Laurent, H. Dürr, *Pure Appl. Chem.* **2001**, *73*, 639–665.
[167] H. M. D. Bandara, S. C. Burdette, *Chem. Soc. Rev.* **2012**, *41*, 1809–1825.
[168] A. Santiago, R. S. Becker, *J. Am. Chem. Soc.* **1968**, *90*, 3654–3658.
[169] R. Navrátil, S. Wiedbrauk, J. Jašík, H. Dube, J. Roithová, *Phys. Chem. Chem. Phys.* **2018**, *20*, 6868–6876.
[170] E. A. Abourashed, *J. Nat. Prod.* **2017**, *80*, 577–577.
[171] L. Kortekaas, W. R. Browne, *Chem. Soc. Rev.* **2019**, *48*, 3406–3424.
[172] S.-Z. Pu, Q. Sun, C.-B. Fan, R.-J. Wang, G. Liu, *J. Mater. Chem. C* **2016**, *4*, 3075–3093.
[173] W. Szymański, J. M. Beierle, H. A. V. Kistemaker, W. A. Velema, B. L. Feringa, *Chem. Rev.* **2013**, *113*, 6114–6178.
[174] G. Despras, V. Poonthiyil, T. K. Lindhorst, in *Mol. Photoswitches - Chem. Prop. Appl.*, Wiley-VCH, Weinheim, Germany, **2022**, pp. 1015–1045.
[175] J. Morstein, A. C. Impastato, D. Trauner, *Chembiochem Eur. J. Chem. Biol.* **2021**, *22*, 73–83.
[176] W. A. Velema, W. Szymanski, B. L. Feringa, *J. Am. Chem. Soc.* **2014**, *136*, 2178–2191.

[177] G. S. Hartley, *Nature* **1937**, *140*, 281–281.
[178] J. Harada, K. Ogawa, S. Tomoda, *Acta Crystallogr. B* **1997**, *53*, 662–672.
[179] H. Rau, *Angew. Chem. Int. Ed. Engl.* **1973**, *12*, 224–235.
[180] M. Asbjørn, R. Chr., H. Sten, R. J. J. C. Lousberg, W. U., *Acta Chem. Scand.* **1971**, *25*, 3561–3568.
[181] E. R. Talaty, J. C. Fargo, *Chem. Commun. Lond.* **1967**, 65–66.
[182] J. Maciejewski, A. Sobczuk, A. Claveau, A. Nicolai, R. Petraglia, L. Cervini, E. Baudat, P. Miéville, D. Fazzi, C. Corminboeuf, G. Sforazzini, *Chem. Sci.* **2016**, *8*, 361–365.
[183] I. K. Lednev, T.-Q. Ye, P. Matousek, M. Towrie, P. Foggi, F. V. R. Neuwahl, S. Umapathy, R. E. Hester, J. N. Moore, *Chem. Phys. Lett.* **1998**, *290*, 68–74.
[184] C. L. Forber, E. C. Kelusky, N. J. Bunce, M. C. Zerner, *J. Am. Chem. Soc.* **1985**, *107*, 5884–5890.
[185] P. Hamm, S. M. Ohline, W. Zinth, *J. Chem. Phys.* **1997**, *106*, 519–529.
[186] H. Fliegl, A. Köhn, C. Hättig, R. Ahlrichs, *J. Am. Chem. Soc.* **2003**, *125*, 9821–9827.
[187] E. Merino, M. Ribagorda, *Beilstein J. Org. Chem.* **2012**, *8*, 1071–1090.
[188] J. Sun, F. Wang, H. Zhang, K. Liu, *Adv. NanoBiomed Res.* **2021**, *1*, 2100020.
[189] J. Broichhagen, J. A. Frank, D. Trauner, *Acc. Chem. Res.* **2015**, *48*, 1947–1960.
[190] E. Merino, *Chem. Soc. Rev.* **2011**, *40*, 3835–3853.
[191] G. C. Mueller, J. A. Miller, *J. Biol. Chem.* **1949**, *180*, 1125–1136.
[192] H. Mori, Y. Mori, S. Sugie, N. Yoshimi, M. Takahashi, H. Ni-i, H. Yamazaki, K. Toyoshi, G. M. Williams, *Cancer Res.* **1986**, *46*, 1654–1658.
[193] M. Londoño-Berrío, S. Pérez-Buitrago, I. C. Ortiz-Trujillo, L. M. Hoyos-Palacio, L. Y. Orozco, L. López, D. G. Zárate-Triviño, J. A. Capobianco, P. Mena-Giraldo, *Polymers* **2022**, *14*, 3119.
[194] S. L. Jacques, *Phys. Med. Biol.* **2013**, *58*, R37-61.
[195] J. Keyvan Rad, Z. Balzade, A. R. Mahdavian, *J. Photochem. Photobiol. C Photochem. Rev.* **2022**, *51*, 100487.
[196] M. Dong, A. Babalhavaeji, S. Samanta, A. A. Beharry, G. A. Woolley, *Acc. Chem. Res.* **2015**, *48*, 2662–2670.
[197] A. A. Blevins, G. J. Blanchard, *J. Phys. Chem. B* **2004**, *108*, 4962–4968.
[198] J. García-Amorós, D. Velasco, *Beilstein J. Org. Chem.* **2012**, *8*, 1003–1017.
[199] A. A. Beharry, O. Sadovski, G. A. Woolley, *J. Am. Chem. Soc.* **2011**, *133*, 19684–19687.
[200] S. Samanta, A. A. Beharry, O. Sadovski, T. M. McCormick, A. Babalhavaeji, V. Tropepe, G. A. Woolley, *J. Am. Chem. Soc.* **2013**, *135*, 9777–9784.
[201] D. Bléger, J. Schwarz, A. M. Brouwer, S. Hecht, *J. Am. Chem. Soc.* **2012**, *134*, 20597–20600.
[202] S. Crespi, N. A. Simeth, B. König, *Nat. Rev. Chem.* **2019**, *3*, 133–146.
[203] R. Siewertsen, H. Neumann, B. Buchheim-Stehn, R. Herges, C. Näther, F. Renth, F. Temps, *J. Am. Chem. Soc.* **2009**, *131*, 15594–15595.
[204] D. Das, M. K. Yadav, L. Singla, A. Kumar, M. Karanam, S. Dev, A. R. Choudhury, *ChemistrySelect* **2020**, *5*, 13957–13962.
[205] C. Di Berardino, M. A. Strauss, D. Schatz, H. A. Wegner, *Chem. – Eur. J.* **2022**, *28*, e202104284.
[206] A. L. Sanna, T. Pachova, A. Catellani, A. Calzolari, G. Sforazzini, *Mol. Basel Switz.* **2024**, *29*, 1929.
[207] T. J. Dougherty, C. J. Gomer, B. W. Henderson, G. Jori, D. Kessel, M. Korbelik, J. Moan, Q. Peng, *JNCI J. Natl. Cancer Inst.* **1998**, *90*, 889–905.
[208] K. Hüll, J. Morstein, D. Trauner, *Chem. Rev.* **2018**, *118*, 10710–10747.
[209] J. E. Warias, F. Reise, S. C. Hövelmann, R. P. Giri, M. Röhrl, J. Kuhn, M. Jacobsen, K. Chatterjee, T. Arnold, C. Shen, S. Festersen, A. Sartori, P. Jordt, O. M. Magnussen, T. K. Lindhorst, B. M. Murphy, *Sci. Rep.* **2023**, *13*, 11480.
[210] I. Linsmeier, S. Banerjee, P. W. Oakes, W. Jung, T. Kim, M. P. Murrell, *Nat. Commun.* **2016**, *7*, 12615.
[211] D. D. Young, A. Deiters, *Org. Biomol. Chem.* **2007**, *5*, 999–1005.
[212] L. Josa-Culleré, A. Llebaria, *ChemPhotoChem* **2021**, *5*, 296–314.
[213] A. Reiner, E. Y. Isacoff, *Methods Mol. Biol. Clifton NJ* **2014**, *1148*, 45–68.
[214] V. M. Krishnamurthy, V. Semetey, P. J. Bracher, N. Shen, G. M. Whitesides, *J. Am. Chem. Soc.* **2007**, *129*, 1312–1320.

[215] J. Broichhagen, A. Damijonaitis, J. Levitz, K. R. Sokol, P. Leippe, D. Konrad, E. Y. Isacoff, D. Trauner, *ACS Cent. Sci.* **2015**, *1*, 383–393.
[216] R. A. Bednar, *Biochemistry* **1990**, *29*, 3684–3690.
[217] P. Leippe, J. Koehler Leman, D. Trauner, *Biochemistry* **2017**, *56*, 5214–5220.
[218] M. W. Crankshaw, G. A. Grant, *Curr. Protoc. Protein Sci.* **2001**, *Chapter 15*, Unit15.1.
[219] F.-J. Chen, J. Gao, *Chem. Weinh. Bergstr. Ger.* **2022**, *28*, e202201843.
[220] E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira, G. J. L. Bernardes, *Nat. Rev. Chem.* **2019**, *3*, 147–171.
[221] D. D. Roberts, S. D. Lewis, D. P. Ballou, S. T. Olson, J. A. Shafer, *Biochemistry* **1986**, *25*, 5595–5601.
[222] T. P. King, Y. Li, L. Kochoumian, *Biochemistry* **1978**, *17*, 1499–1506.
[223] A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y.-S. Lin, B. L. Pentelute, *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.
[224] L. Xu, M. J. S. A. Silva, P. M. P. Gois, S. L. Kuan, T. Weil, *Chem. Sci.* **2021**, *12*, 13321–13330.
[225] I. N. Gober, A. J. Riemen, M. Villain, *J. Pept. Sci.* **2021**, *27*, e3323.
[226] A. D. Baldwin, K. L. Kiick, *Bioconjug. Chem.* **2011**, *22*, 1946–1953.
[227] E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007–2010.
[228] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org. Lett.* **2000**, *2*, 2141–2143.
[229] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.
[230] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16793–16797.
[231] M. L. Blackman, M. Royzen, J. M. Fox, *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.
[232] Y.-H. Tsai, S. Essig, J. R. James, K. Lang, J. W. Chin, *Nat. Chem.* **2015**, *7*, 554–561.
[233] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, *ACS Chem. Biol.* **2008**, *3*, 373–382.
[234] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86–89.
[235] B. V. Zemelman, N. Nesnas, G. A. Lee, G. Miesenböck, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1352–1357.
[236] M. J. Caterina, M. A. Schumacher, M. Tominaga, T. A. Rosen, J. D. Levine, D. Julius, *Nature* **1997**, *389*, 816–824.
[237] H. Knotkova, M. Pappagallo, A. Szallasi, *Clin. J. Pain* **2008**, *24*, 142.
[238] S. F. Traynelis, L. P. Wollmuth, C. J. McBain, F. S. Menniti, K. M. Vance, K. K. Ogden, K. B. Hansen, H. Yuan, S. J. Myers, R. Dingledine, *Pharmacol. Rev.* **2010**, *62*, 405–496.
[239] G. Riedel, B. Platt, J. Micheau, *Behav. Brain Res.* **2003**, *140*, 1–47.
[240] M. Volgraf, P. Gorostiza, S. Szobota, M. R. Helix, E. Y. Isacoff, D. Trauner, *J. Am. Chem. Soc.* **2007**, *129*, 260–261.
[241] A. Reiter, A. Skerra, D. Trauner, A. Schiefner, *Biochemistry* **2013**, *52*, 8972–8974.
[242] M. Volgraf, P. Gorostiza, R. Numano, R. Kramer, E. Isacoff, D. Trauner, *Nat. Chem. Biol.* **2006**, *2*, 47–52.
[243] R. Numano, S. Szobota, A. Y. Lau, P. Gorostiza, M. Volgraf, B. Roux, D. Trauner, E. Y. Isacoff, *Proc. Natl. Acad. Sci.* **2009**, *106*, 6814–6819.
[244] J. M. Berg, J. L. Tymoczko, L. Stryer, in *Stryer Biochem.*, Springer Spektrum, Berlin, **2017**, pp. 529–567.
[245] D. Klemm, B. Heublein, H.-P. Fink, A. Bohn, *Angew. Chem. Int. Ed.* **2005**, *44*, 3358–3393.
[246] U. Lüttge, M. Kluge, G. Thiel, in *Bot. - Umfassende Biol. Pflanz.*, Wiley-VCH, Weinheim, **2010**, pp. 279–305.
[247] W. K. Purves, D. Sadava, in *Biologie*, Spektrum Akademischer Verlag, München, **2006**, pp. 735–754.
[248] W. K. Purves, D. Sadava, in *Biologie*, Spektrum Akademischer Verlag, München, **2006**, pp. 783–801.
[249] C. Whitfield, C. M. Szymanski, A. L. Lewis, M. Aebi, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H.

Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[250] H. S. BENNETT, *J. Histochem. Cytochem.* **1963**, *11*, 14–23.
[251] S. Reitsma, D. W. Slaaf, H. Vink, M. A. M. J. van Zandvoort, M. G. A. oude Egbrink, *Pflugers Arch.* **2007**, *454*, 345–359.
[252] J. M. Rini, K. W. Moremen, B. G. Davis, J. D. Esko, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[253] P. Stanley, K. W. Moremen, N. E. Lewis, N. Taniguchi, M. Aebi, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[254] I. Brockhausen, H. H. Wandall, K. G. T. Hagen, P. Stanley, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[255] C. L. R. Merry, U. Lindahl, J. Couchman, J. D. Esko, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[256] R. S. Haltiwanger, L. Wells, H. H. Freeze, H. Jafar-Nejad, T. Okajima, P. Stanley, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[257] R. L. Schnaar, R. Sandhoff, M. Tiemeyer, T. Kinoshita, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[258] P. Gagneux, T. Hennet, A. Varki, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[259] S. Schröter, C. Osterhoff, W. McArdle, R. Ivell, *Hum. Reprod. Update* **1999**, *5*, 302–313.
[260] M. E. Taylor, K. Drickamer, A. Imberty, Y. van Kooyk, R. L. Schnaar, M. E. Etzler, A. Varki, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[261] H. Franz, P. Ziska, J. Mohr, *Acta Histochem.* **1982**, *71*, 19–21.
[262] P. H. Stillmark, *Über Ricin, ein giftiges Ferment aus den Samen von Ricinus comm. L. und einigen anderen Euphorbiaceen*, Dissertation, Kaiserliche Universität zu Dorpat, Estonia, **1888**.
[263] W. C. Boyd, E. Shapleigh, *Science* **1954**, *119*, 419.
[264] J. B. Sumner, S. F. Howell, *J. Bacteriol.* **1936**, *32*, 227–237.
[265] K. D. Hardman, C. F. Ainsworth, *Biochemistry* **1972**, *11*, 4910–4919.
[266] R. D. Cummings, R. L. Schnaar, J. D. Esko, R. J. Woods, K. Drickamer, M. E. Taylor, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[267] Z. Fujimoto, H. Tateno, J. Hirabayashi, *Methods Mol. Biol. Clifton NJ* **2014**, *1200*, 579–606.
[268] D. Xu, J. H. Prestegard, R. J. Linhardt, J. D. Esko, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.

[269] R. D. Cummings, E. Chiffoleau, Y. van Kooyk, R. P. McEver, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[270] T. Angata, S. von Gunten, R. L. Schnaar, A. Varki, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[271] R. D. Cummings, F.-T. Liu, G. A. Rabinovich, S. R. Stowell, G. R. Vasta, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[272] N. Dahms, T. Braulke, A. Varki, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[273] R. D. Cummings, M. E. Etzler, T. N. C. Ramya, K. Kato, G. A. Rabinovich, A. Surolia, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[274] P. C. Nowell, *Cancer Res.* **1960**, *20*, 462–466.
[275] A. L. Lewis, J. J. Kohler, M. Aebi, in *Essent. Glycobiol.* (Eds.: A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, J.H. Prestegard, R.L. Schnaar, P.H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2022**.
[276] I. Ofek, D. L. Hasty, N. Sharon, *FEMS Immunol. Med. Microbiol.* **2003**, *38*, 181–191.
[277] N. Sharon, *Biochim. Biophys. Acta BBA - Gen. Subj.* **2006**, *1760*, 527–537.
[278] A. Asadi, S. Razavi, M. Talebi, M. Gholami, *Infection* **2019**, *47*, 13–23.
[279] J. J. Skehel, D. C. Wiley, *Annu. Rev. Biochem.* **2000**, *69*, 531–569.
[280] G. N. Rogers, J. C. Paulson, *Virology* **1983**, *127*, 361–373.
[281] D. G. Pina, L. Johannes, *Toxicon* **2005**, *45*, 389–393.
[282] M. J. Walsh, J. E. Dodd, G. M. Hautbergue, *Virulence* **2013**, *4*, 774–784.
[283] A. Olsén, A. Jonsson, S. Normark, *Nature* **1989**, *338*, 652–655.
[284] D. E. Bessen, A. Kalia, *Infect. Immun.* **2002**, *70*, 1159–1167.
[285] A. Filloux, *J. Bacteriol.* **2010**, *192*, 3243–3245.
[286] W. Deng, N. C. Marshall, J. L. Rowland, J. M. McCoy, L. J. Worrall, A. S. Santos, N. C. J. Strynadka, B. B. Finlay, *Nat. Rev. Microbiol.* **2017**, *15*, 323–337.
[287] R. Fronzes, H. Remaut, G. Waksman, *EMBO J.* **2008**, *27*, 2271–2280.
[288] H. Sakellaris, J. R. Scott, *Mol. Microbiol.* **1998**, *30*, 681–687.
[289] J. S. Mattick, *Annu. Rev. Microbiol.* **2002**, *56*, 289–314.
[290] G. Waksman, S. J. Hultgren, *Nat. Rev. Microbiol.* **2009**, *7*, 765–774.
[291] D. J. Wurpel, S. A. Beatson, M. Totsika, N. K. Petty, M. A. Schembri, *PloS One* **2013**, *8*, e52835.
[292] B. Pakbin, W. M. Brück, J. W. A. Rossen, *Int. J. Mol. Sci.* **2021**, *22*, 9922.
[293] Y. Zhou, Z. Zhou, L. Zheng, Z. Gong, Y. Li, Y. Jin, Y. Huang, M. Chi, *Int. J. Mol. Sci.* **2023**, *24*, 10537.
[294] C. Struve, M. Bojer, K. A. Krogfelt, *Infect. Immun.* **2008**, *76*, 4055–4065.
[295] R. Kolenda, M. Ugorski, K. Grzymajlo, *Front. Microbiol.* **2019**, *10*, 1017.
[296] P. Klemm, B. J. Jørgensen, I. van Die, H. de Ree, H. Bergmans, *Mol. Gen. Genet.* **1985**, *199*, 410–414.
[297] K. A. Krogfelt, P. Klemm, *Microb. Pathog.* **1988**, *4*, 231–238.
[298] P. Klemm, *EMBO J.* **1986**, *5*, 1389–1393.
[299] J. M. Abraham, C. S. Freitag, J. R. Clements, B. I. Eisenstein, *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 5724–5727.
[300] S. A. Joyce, C. J. Dorman, *Mol. Microbiol.* **2002**, *45*, 1107–1117.
[301] D. L. Gally, J. Leathart, I. C. Blomfield, *Mol. Microbiol.* **1996**, *21*, 725–738.
[302] W. R. Schwan, S. Shibata, S.-I. Aizawa, A. J. Wolfe, *J. Bacteriol.* **2007**, *189*, 7159–7163.

[303] Y. Wu, F. W. Outten, *J. Bacteriol.* **2009**, *191*, 1248–1257.
[304] M. L. Valenski, S. L. Harris, P. A. Spears, J. R. Horton, P. E. Orndorff, *J. Bacteriol.* **2003**, *185*, 5007–5011.
[305] G. T. Werneburg, D. G. Thanassi, *EcoSal Plus* **2018**, *8*, 10.1128/ecosalplus.ESP-0007–2017.
[306] N. Raad, D. Tandon, S. Hapfelmeier, N. Polacek, *Nucleic Acids Res.* **2022**, *50*, 11858–11875.
[307] C. P. Corcoran, C. J. Dorman, *Mol. Microbiol.* **2009**, *74*, 1071–1082.
[308] P. Nilsson, S. Naureckiene, B. E. Uhlin, *J. Bacteriol.* **1996**, *178*, 683–690.
[309] P. Klemm, M. Schembri, *EcoSal Plus* **2004**, *1*, 10.1128/ecosalplus.8.3.2.6.
[310] M. Sarshar, P. Behzadi, C. Ambrosi, C. Zagaglia, A. T. Palamara, D. Scribano, *Antibiot. Basel Switz.* **2020**, *9*, 397.
[311] A. P. Maillard, K. Chan, F. Duong, in *Madame Curie Biosci. Database Internet*, Landes Bioscience, **2013**.
[312] M. Totsika, B. Heras, D. J. Wurpel, M. A. Schembri, *J. Bacteriol.* **2009**, *191*, 3901–3908.
[313] M. D. Crespo, C. Puorger, M. A. Schärer, O. Eidam, M. G. Grütter, G. Capitani, R. Glockshuber, *Nat. Chem. Biol.* **2012**, *8*, 707–713.
[314] C. Hal Jones, P. N. Danese, J. S. Pinkner, T. J. Silhavy, S. J. Hultgren, *EMBO J.* **1997**, *16*, 6394–6406.
[315] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S. J. Hultgren, S. D. Knight, *Science* **1999**, *285*, 1061–1066.
[316] F. G. Sauer, K. Fütterer, J. S. Pinkner, K. W. Dodson, S. J. Hultgren, G. Waksman, *Science* **1999**, *285*, 1058–1061.
[317] O. Eidam, F. S. N. Dworkowski, R. Glockshuber, M. G. Grütter, G. Capitani, *FEBS Lett.* **2008**, *582*, 651–655.
[318] C.-S. Hung, J. Bouckaert, D. Hung, J. Pinkner, C. Widberg, A. DeFusco, C. G. Auguste, R. Strouse, S. Langermann, G. Waksman, S. J. Hultgren, *Mol. Microbiol.* **2002**, *44*, 903–915.
[319] C. Giese, J. Eras, A. Kern, M. A. Schärer, G. Capitani, R. Glockshuber, *Angew. Chem. Int. Ed.* **2016**, *55*, 9350–9355.
[320] H. Remaut, R. J. Rose, T. J. Hannan, S. J. Hultgren, S. E. Radford, A. E. Ashcroft, G. Waksman, *Mol. Cell* **2006**, *22*, 831–842.
[321] F. G. Sauer, J. S. Pinkner, G. Waksman, S. J. Hultgren, *Cell* **2002**, *111*, 543–551.
[322] G. Phan, H. Remaut, T. Wang, W. J. Allen, K. F. Pirker, A. Lebedev, N. S. Henderson, S. Geibel, E. Volkan, J. Yan, M. B. A. Kunze, J. S. Pinkner, B. Ford, C. W. M. Kay, H. Li, S. J. Hultgren, D. G. Thanassi, G. Waksman, *Nature* **2011**, *474*, 49–53.
[323] T. W. Ng, L. Akman, M. Osisami, D. G. Thanassi, *J. Bacteriol.* **2004**, *186*, 5321–5331.
[324] M. Du, Z. Yuan, H. Yu, N. Henderson, S. Sarowar, G. Zhao, G. T. Werneburg, D. G. Thanassi, H. Li, *Nature* **2018**, *562*, 444–447.
[325] G. T. Werneburg, N. S. Henderson, E. B. Portnoy, S. Sarowar, S. J. Hultgren, H. Li, D. G. Thanassi, *Nat. Struct. Mol. Biol.* **2015**, *22*, 540–546.
[326] C. Puorger, M. Vetsch, G. Wider, R. Glockshuber, *J. Mol. Biol.* **2011**, *412*, 520–535.
[327] C. N. Spaulding, H. L. Schreiber IV, W. Zheng, K. W. Dodson, J. E. Hazen, M. S. Conover, F. Wang, P. Svenmarker, A. Luna-Rico, O. Francetic, M. Andersson, S. Hultgren, E. H. Egelman, *eLife* **2018**, *7*, e31662.
[328] Y. Hasegawa, K. Nagano, *Jpn. Dent. Sci. Rev.* **2021**, *57*, 190–200.
[329] K. A. Krogfelt, H. Bergmans, P. Klemm, *Infect. Immun.* **1990**, *58*, 1995–1998.
[330] I. Ofek, D. Mirelman, N. Sharon, *Nature* **1977**, *265*, 623–625.
[331] M. A. Schembri, E. V. Sokurenko, P. Klemm, *Infect. Immun.* **2000**, *68*, 2638–2646.
[332] S. G. Stahlhut, V. Tchesnokova, C. Struve, S. J. Weissman, S. Chattopadhyay, O. Yakovenko, P. Aprikian, E. V. Sokurenko, K. A. Krogfelt, *J. Bacteriol.* **2009**, *191*, 6592–6601.
[333] M. A. Schembri, P. Klemm, *Infect. Immun.* **2001**, *69*, 1322–1328.
[334] M.-M. L, H. T, J. J, *Expert Opin. Drug Discov.* **2017**, *12*, 711–731.
[335] E. V. Sokurenko, V. Chesnokova, D. E. Dykhuizen, I. Ofek, X.-R. Wu, K. A. Krogfelt, C. Struve, M. A. Schembri, D. L. Hasty, *Proc. Natl. Acad. Sci.* **1998**, *95*, 8922–8926.
[336] D. J. Schwartz, V. Kalas, J. S. Pinkner, S. L. Chen, C. N. Spaulding, K. W. Dodson, S. J. Hultgren, *Proc. Natl. Acad. Sci.* **2013**, *110*, 15530–15537.

[337] M. M. Sauer, R. P. Jakob, J. Eras, S. Baday, D. Eriş, G. Navarra, S. Bernèche, B. Ernst, T. Maier, R. Glockshuber, *Nat. Commun.* **2016**, *7*, 10738.
[338] G. Zhou, W.-J. Mo, P. Sebbel, G. Min, T. A. Neubert, R. Glockshuber, X.-R. Wu, T.-T. Sun, X.-P. Kong, *J. Cell Sci.* **2001**, *114*, 4095–4103.
[339] G. Min, H. Wang, T.-T. Sun, X.-P. Kong, *J. Cell Biol.* **2006**, *173*, 975–983.
[340] H. Yanagisawa, Y. Kita, T. Oda, M. Kikkawa, *Commun. Biol.* **2023**, *6*, 1–11.
[341] I. Kątnik-Prastowska, J. Lis, A. Matejuk, *Glycoconj. J.* **2014**, *31*, 623–636.
[342] M. M. Sauer, R. P. Jakob, T. Luber, F. Canonica, G. Navarra, B. Ernst, C. Unverzagt, T. Maier, R. Glockshuber, *J. Am. Chem. Soc.* **2019**, *141*, 936–944.
[343] J. Bouckaert, J. Mackenzie, J. L. De Paz, B. Chipwaza, D. Choudhury, A. Zavialov, K. Mannerstedt, J. Anderson, D. Piérard, L. Wyns, P. H. Seeberger, S. Oscarson, H. De Greve, S. D. Knight, *Mol. Microbiol.* **2006**, *61*, 1556–1568.
[344] J. Pak, Y. Pu, Z. T. Zhang, D. L. Hasty, X. R. Wu, *J. Biol. Chem.* **2001**, *276*, 9924–9930.
[345] A. Stsiapanava, C. Xu, M. Brunati, S. Zamora-Caballero, C. Schaeffer, M. Bokhove, L. Han, H. Hebert, M. Carroni, S. Yasumasu, L. Rampoldi, B. Wu, L. Jovine, *EMBO J.* **2020**, *39*, e106807.
[346] P. Thumbikat, R. E. Berry, G. Zhou, B. K. Billips, R. E. Yaggie, T. Zaichuk, T.-T. Sun, A. J. Schaeffer, D. J. Klumpp, *PLoS Pathog.* **2009**, *5*, e1000415.
[347] S. Langermann, R. Möllby, J. E. Burlein, S. R. Palaszynski, C. Gale Auguste, A. DeFusco, R. Strouse, M. A. Schenerman, S. J. Hultgren, J. S. Pinkner, J. Winberg, L. Guldevall, M. Söderhäll, K. Ishikawa, S. Normark, S. Koenig, *J. Infect. Dis.* **2000**, *181*, 774–778.
[348] D. S. Eto, T. A. Jones, J. L. Sundsbak, M. A. Mulvey, *PLoS Pathog.* **2007**, *3*, e100.
[349] G. Chevalier, A. Laveissière, G. Desachy, N. Barnich, A. Sivignon, M. Maresca, C. Nicoletti, E. Di Pasquale, M. Martinez-Medina, K. W. Simpson, V. Yajnik, H. Sokol, T. Adegbamigbe, T. Ahmad, I. Arnott, Y. Bouhnik, F. Carbonnel, J.-F. Colombel, G. Doherty, J. R. F. Cummings, X. Hébuterne, H. Herfarth, D. Kevans, G. P. de Chambrun, M. Nachury, S. Nancey, X. Roblin, M. A. W. Tremelling, J. Plassais, F. Strozzi, A. Cervino, R. Morra, C. Bonny, MOBIDIC Study Investigators, *Microbiome* **2021**, *9*, 176.
[350] I. Ofek, E. H. Beachey, *Infect. Immun.* **1978**, *22*, 247–254.
[351] C. S. Giampapa, S. N. Abraham, T. M. Chiang, E. H. Beachey, *J. Biol. Chem.* **1988**, *263*, 5362–5367.
[352] L. Byrd-Leotis, C. Gao, N. Jia, A. Y. Mehta, J. Trost, S. F. Cummings, J. Heimburg-Molinaro, R. D. Cummings, D. A. Steinhauer, *J. Virol.* **2019**, *93*, 10.1128/jvi.01178-19.
[353] J. E. Cutler, *Med. Mycol.* **2001**, *39 Suppl 1*, 75–86.
[354] W. E. Thomas, E. Trintchina, M. Forero, V. Vogel, E. V. Sokurenko, *Cell* **2002**, *109*, 913–923.
[355] M. Dembo, D. C. Torney, K. Saxman, H. Da, D. Hammer, **1988**, DOI 10.1098/rspb.1988.0038.
[356] W. Chen, J. Lou, C. Zhu, *J. Biol. Chem.* **2010**, *285*, 35967–35978.
[357] C. D. Buckley, J. Tan, K. L. Anderson, D. Hanein, N. Volkmann, W. I. Weis, W. J. Nelson, A. R. Dunn, *Science* **2014**, *346*, 1254211.
[358] D. L. Huang, N. A. Bax, C. D. Buckley, W. I. Weis, A. R. Dunn, *Science* **2017**, *357*, 703–706.
[359] R. P. McEver, C. Zhu, *Annu. Rev. Cell Dev. Biol.* **2010**, *26*, 363–396.
[360] A. M. Ding, R. J. Palmer, J. O. Cisar, P. E. Kolenbrander, *Appl. Environ. Microbiol.* **2010**, *76*, 1294–1297.
[361] D. I. Kisiela, J. J. Kramer, V. Tchesnokova, P. Aprikian, V. Yarov-Yarovoy, S. Clegg, E. V. Sokurenko, *J. Biol. Chem.* **2011**, *286*, 38136–38147.
[362] M. Mathelié-Guinlet, F. Viela, G. Pietrocola, P. Speziale, D. Alsteens, Y. F. Dufrêne, *Nat. Commun.* **2020**, *11*, 5431.
[363] S. Lancellotti, M. Sacco, M. Basso, R. D. Cristofaro, *Biomol. Concepts* **2019**, *10*, 194–208.
[364] V. Walhorn, A.-K. Möller, C. Bartz, T. Dierks, D. Anselmetti, *Sci. Rep.* **2018**, *8*, 16849.
[365] P. Aprikian, G. Interlandi, B. A. Kidd, I. Le Trong, V. Tchesnokova, O. Yakovenko, M. J. Whitfield, E. Bullitt, R. E. Stenkamp, W. E. Thomas, E. V. Sokurenko, *PLoS Biol.* **2011**, *9*, e1000617.
[366] L. A. Carlucci, W. E. Thomas, *Biophys. J.* **2020**, *118*, 514a.
[367] M. Forero, O. Yakovenko, E. V. Sokurenko, W. E. Thomas, V. Vogel, *PLoS Biol.* **2006**, *4*, e298.
[368] E. Miller, T. Garcia, S. Hultgren, A. F. Oberhauser, *Biophys. J.* **2006**, *91*, 3848–3856.
[369] L. A. Carlucci, K. C. Johnson, W. E. Thomas, *Biophys. J.* **2024**, *123*, 3038–3050.

[370] O. Yakovenko, S. Sharma, M. Forero, V. Tchesnokova, P. Aprikian, B. Kidd, A. Mach, V. Vogel, E. Sokurenko, W. E. Thomas, *J. Biol. Chem.* **2008**, *283*, 11596–11605.
[371] P. Magala, R. E. Klevit, W. E. Thomas, E. V. Sokurenko, R. E. Stenkamp, *Proteins Struct. Funct. Bioinforma.* **2020**, *88*, 593–603.
[372] L. M. Nilsson, W. E. Thomas, E. Trintchina, V. Vogel, E. V. Sokurenko, *J. Biol. Chem.* **2006**, *281*, 16656–16663.
[373] J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhrer, C.-S. Hung, J. Pinkner, R. Slättegård, A. Zavialov, D. Choudhury, S. Langermann, S. J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S. D. Knight, H. De Greve, *Mol. Microbiol.* **2005**, *55*, 441–455.
[374] B. Fiege, S. Rabbani, R. C. Preston, R. P. Jakob, P. Zihlmann, O. Schwardt, X. Jiang, T. Maier, B. Ernst, *ChemBioChem* **2015**, *16*, 1235–1246.
[375] A. Wellens, M. Lahmann, M. Touaibia, J. Vaucher, S. Oscarson, R. Roy, H. Remaut, J. Bouckaert, *Biochemistry* **2012**, *51*, 4790–4799.
[376] S. Rabbani, E.-M. Krammer, G. Roos, A. Zalewski, R. Preston, S. Eid, P. Zihlmann, M. Prévost, M. F. Lensink, A. Thompson, B. Ernst, J. Bouckaert, *IUCrJ* **2017**, *4*, 7–23.
[377] M. Hartmann, T. K. Lindhorst, *Eur. J. Org. Chem.* **2011**, *2011*, 3583–3609.
[378] L. Pang, S. Kleeb, K. Lemme, S. Rabbani, M. Scharenberg, A. Zalewski, F. Schädler, O. Schwardt, B. Ernst, *ChemMedChem* **2012**, *7*, 1404–1422.
[379] S. Kleeb, L. Pang, K. Mayer, D. Eris, A. Sigl, R. C. Preston, P. Zihlmann, T. Sharpe, R. P. Jakob, D. Abgottspon, A. S. Hutter, M. Scharenberg, X. Jiang, G. Navarra, S. Rabbani, M. Smiesko, N. Lüdin, J. Bezençon, O. Schwardt, T. Maier, B. Ernst, *J. Med. Chem.* **2015**, *58*, 2221–2239.
[380] W. Schönemann, J. Cramer, T. Mühlethaler, B. Fiege, M. Silbermann, S. Rabbani, P. Dätwyler, P. Zihlmann, R. P. Jakob, C. P. Sager, M. Smieško, O. Schwardt, T. Maier, B. Ernst, *ChemMedChem* **2019**, *14*, 749–757.
[381] L. Mousavifar, G. Vergoten, G. Charron, R. Roy, *Molecules* **2019**, *24*, 3566.
[382] T. Klein, D. Abgottspon, M. Wittwer, S. Rabbani, J. Herold, X. Jiang, S. Kleeb, C. Lüthi, M. Scharenberg, J. Bezençon, E. Gubler, L. Pang, M. Smiesko, B. Cutting, O. Schwardt, B. Ernst, *J. Med. Chem.* **2010**, *53*, 8627–8641.
[383] W. Reinisch, X. Hébuterne, A. Buisson, S. Schreiber, P. Desreumaux, C. Primas, J.-M. Paillarse, G. Chevalier, C. Bonny, *J. Gastroenterol. Hepatol.* **2022**, *37*, 832–840.
[384] Z. Zhou, R. Tang, Y. Fang, T. Lv, J. Liu, X. Wang, *ACS Med. Chem. Lett.* **2024**, *15*, 1724–1730.
[385] M. J. Bishop, V. J. Colandrea, Y. LI, E. L. Stewart, I. STRAMBEANU, K. L. Widdowson, J. W. Janetka, L. K. MCGRANE, *C-Mannoside Compounds Useful for the Treatment of Urinary Tract Infections*, **2020**, WO2020012336A1.
[386] M. Touaibia, A. Wellens, T. C. Shiao, Q. Wang, S. Sirois, J. Bouckaert, R. Roy, *ChemMedChem* **2007**, *2*, 1190–1201.
[387] P. Hoyos, A. Perona, O. Juanes, Á. Rumbero, M. J. Hernáiz, *Chem. – Eur. J.* **2021**, *27*, 7593–7624.
[388] J. Cramer, B. Pero, X. Jiang, C. Bosko, M. Silbermann, S. Rabbani, S. Wilke, D. Nemli, B. Ernst, M. W. Peczuh, *Eur. J. Med. Chem.* **2024**, 116225.
[389] L. Mydock-McGrane, Z. Cusumano, Z. Han, J. Binkley, M. Kostakioti, T. Hannan, J. S. Pinkner, R. Klein, V. Kalas, J. Crowley, N. P. Rath, S. J. Hultgren, J. W. Janetka, *J. Med. Chem.* **2016**, *59*, 9390–9408.
[390] P. Aprikian, V. Tchesnokova, B. Kidd, O. Yakovenko, V. Yarov-Yarovoy, E. Trinchina, V. Vogel, W. Thomas, E. Sokurenko, *J. Biol. Chem.* **2007**, *282*, 23437–23446.
[391] I. Le Trong, P. Aprikian, B. A. Kidd, M. Forero-Shelton, V. Tchesnokova, P. Rajagopal, V. Rodriguez, G. Interlandi, R. Klevit, V. Vogel, R. E. Stenkamp, E. V. Sokurenko, W. E. Thomas, *Cell* **2010**, *141*, 645–655.
[392] G. Interlandi, W. E. Thomas, *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 990–1008.
[393] W. E. Thomas, L. Carlucci, O. Yakovenko, G. Interlandi, I. Le Trong, P. Aprikian, P. Magala, L. Larson, Y. Sledneva, V. Tchesnokova, R. E. Stenkamp, E. V. Sokurenko, *J. Mol. Biol.* **2022**, *434*, 167681.
[394] O. Languin-Cattoën, F. Sterpone, G. Stirnemann, *Biophys. J.* **2023**, *122*, 2744–2756.

[395] V. B. Rodriguez, B. A. Kidd, G. Interlandi, V. Tchesnokova, E. V. Sokurenko, W. E. Thomas, *J. Biol. Chem.* **2013**, *288*, 24128–24139.
[396] D. I. Kisiela, P. Magala, G. Interlandi, L. A. Carlucci, A. Ramos, V. Tchesnokova, B. Basanta, V. Yarov-Yarovoy, H. Avagyan, A. Hovhannisyan, W. E. Thomas, R. E. Stenkamp, R. E. Klevit, E. V. Sokurenko, *PLoS Pathog.* **2021**, *17*, e1009440.
[397] S. Rabbani, B. Fiege, D. Eris, M. Silbermann, R. P. Jakob, G. Navarra, T. Maier, B. Ernst, *J. Biol. Chem.* **2018**, *293*, 1835–1849.
[398] J. Liu, L. Amaral, S. Keten, *Proteins* **2020**, DOI 10.1002/prot.26013.
[399] E. V. Sokurenko, V. Tchesnokova, G. Interlandi, R. Klevit, W. E. Thomas, *J. Mol. Biol.* **2022**, *434*, 167717.
[400] V. Tchesnokova, P. Aprikian, O. Yakovenko, C. LaRock, B. Kidd, V. Vogel, W. Thomas, E. Sokurenko, *J. Biol. Chem.* **2008**, *283*, 7823–7833.
[401] D. I. Kisiela, V. B. Rodriguez, V. Tchesnokova, H. Avagyan, P. Aprikian, Y. Liu, X.-R. Wu, W. E. Thomas, E. V. Sokurenko, *Proc. Natl. Acad. Sci.* **2013**, *110*, 19089–19094.
[402] D. Eris, R. C. Preston, M. Scharenberg, F. Hulliger, D. Abgottspon, L. Pang, X. Jiang, O. Schwardt, B. Ernst, *ChemBioChem* **2016**, *17*, 1012–1020.
[403] O. Yakovenko, V. Tchesnokova, E. V. Sokurenko, W. E. Thomas, *Proc. Natl. Acad. Sci.* **2015**, *112*, 9884–9889.
[404] V. Kalas, J. S. Pinkner, T. J. Hannan, M. E. Hibbing, K. W. Dodson, A. S. Holehouse, H. Zhang, N. H. Tolia, M. L. Gross, R. V. Pappu, J. Janetka, S. J. Hultgren, *Sci. Adv.* **2017**, *3*, e1601944.
[405] F. Hommais, S. Gouriou, C. Amorin, H. Bui, M. C. Rahimy, B. Picard, E. Denamur, *Infect. Immun.* **2003**, *71*, 3619–3622.
[406] K. Deisseroth, *Nat. Neurosci.* **2015**, *18*, 1213–1225.
[407] M. Hartmann, *Type 1 Fimbriae-Mediated Bacterial Adhesion to Glycosylated Surfaces*, Dissertation, Christiana Albertina University of Kiel, **2011**.
[408] V. Chandrasekaran, *Sweet Switches: Azobenzene Glycosides as Photosensitive Lectin Inhibitors in Solution and on Surfaces*, Dissertation, Christiana Albertina University of Kiel, **2013**.
[409] E. Fast, *Synthese von Glycoazobenzol- Konjugaten Zur Verbesserung Der Photoschaltbarkeit von Glyco-SAMs*, Dissertation, Christiana Albertina University of Kiel, **2019**.
[410] V. Chandrasekaran, K. Kolbe, F. Beiroth, T. K. Lindhorst, *Beilstein J. Org. Chem.* **2013**, *9*, 223–233.
[411] L. M. Friedrich, T. K. Lindhorst, *Beilstein J. Org. Chem.* **2025**, *21*, 736–748.
[412] E. Fast, A. Schlimm, I. Lautenschläger, K. U. Clausen, T. Strunskus, C. Spormann, T. K. Lindhorst, F. Tuczek, *Chemistry* **2020**, *26*, 485–501.
[413] T. Weber, V. Chandrasekaran, I. Stamer, M. B. Thygesen, A. Terfort, T. K. Lindhorst, *Angew. Chem. Int. Ed.* **2014**, *53*, 14583–14586.
[414] L. Möckl, A. Müller, C. Bräuchle, T. K. Lindhorst, *Chem. Commun.* **2016**, *52*, 1254–1257.
[415] G. Despras, L. Möckl, A. Heitmann, I. Stamer, C. Bräuchle, T. K. Lindhorst, *ChemBioChem* **2019**, *20*, 2373–2382.
[416] F.-I. Beiroth, *Untersuchung und Kontrolle bakterieller Adhäsion: Synthese, spezifische Markierung und photochemische Funktionalisierung des bakteriellen Lektins FimH*, Dissertation, Christiana Albertina University of Kiel, **2016**.
[417] T. Hayashi, I. Hamachi, *Acc. Chem. Res.* **2012**, *45*, 1460–1469.
[418] R. A. Homan, J. D. Lapek, C. M. Woo, S. Niessen, L. H. Jones, C. G. Parker, *Nat. Rev. Methods Primer* **2024**, *4*, 1–23.
[419] Y. Koshi, E. Nakata, M. Miyagawa, S. Tsukiji, T. Ogawa, I. Hamachi, *J. Am. Chem. Soc.* **2008**, *130*, 245–251.
[420] F. Reise, *Functional Glycomimetics to Explore Bacterial Adhesion and Membrane Behaviour: Synthesis and Applications*, Dissertation, Christiana Albertina University of Kiel, **2018**.
[421] D. William R. JR., *Introduction to Fluorine NMR*, John Wiley & Sons, Ltd, **2016**.
[422] A. Miseta, P. Csutora, *Mol. Biol. Evol.* **2000**, *17*, 1232–1239.
[423] R. J. Dutton, D. Boyd, M. Berkmen, J. Beckwith, *Proc. Natl. Acad. Sci.* **2008**, *105*, 11933–11938.
[424] Z. Yang, A. B. Attygalle, *J. Mass Spectrom.* **2007**, *42*, 233–243.

[425] R. Pouttu, T. Puustinen, R. Virkola, J. Hacker, P. Klemm, T. K. Korhonen, *Mol. Microbiol.* **1999**, *31*, 1747–1757.
[426] Schrödinger Release 2019-4: Glide, **2019**.
[427] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
[428] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, R. M. Levy, *J. Comput. Chem.* **2005**, *26*, 1752–1780.
[429] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739–1749.
[430] S. G. Ludwig, C. L. Kiyohara, L. A. Carlucci, D. Kisiela, E. V. Sokurenko, W. E. Thomas, *J. Biol. Eng.* **2021**, *15*, 3.
[431] Schrödinger Release 2019-4: MacroModel, **2019**.
[432] A. Scarpino, G. G. Ferenczy, G. M. Keserű, *J. Chem. Inf. Model.* **2018**, *58*, 1441–1458.
[433] A. Scarpino, G. G. Ferenczy, G. M. Keserű, *Curr. Pharm. Des.* **2020**, *26*, 5684–5699.
[434] A. Scarpino, G. G. Ferenczy, G. M. Keserű, *Methods Mol. Biol. Clifton NJ* **2021**, *2266*, 73–88.
[435] Schrödinger Release 2024-2: CovDock, **2024**.
[436] C. Lu, C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G. A. Ross, M. K. Dahlgren, E. Russell, C. D. Von Bargen, R. Abel, R. A. Friesner, E. D. Harder, *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300.
[437] Schrödinger Release 2024-2: Prime, **2024**.
[438] C. Spormann, *Regulation von Adhäsivität: Proteinsynthese Und Chemische Synthese Für Die Modifizierung Des Bakteriellen Lektins FimH*, Master Thesis, Christiana Albertina University of Kiel, **2017**.
[439] R. R. Schmidt, W. Kinzy, *Adv. Carbohydr. Chem. Biochem.* **1994**, *50*, 21–123.
[440] G. Blay, M. L. Cardona, M. B. Garcia, J. R. Pedro, *Synthesis* **1989**, *21*, 438–439.
[441] V. Chandrasekaran, E. Johannes, H. Kobarg, F. D. Sönnichsen, T. K. Lindhorst, *ChemistryOpen* **2014**, *3*, 99–108.
[442] B. Priewisch, K. Rück-Braun, *J. Org. Chem.* **2005**, *70*, 2350–2352.
[443] C. Mills, *J. Chem. Soc. Trans.* **1895**, *67*, 925–933.
[444] G. Zemplén, E. Pacsu, *Ber. Dtsch. Chem. Ges. (A and B)* **1929**, *62*, 1613–1614.
[445] A. W. Williamson, *Ann.* **1851**, *77*, 37.
[446] B. E.-D. M. El-Gendy, E. H. Ghazvini Zadeh, A. C. Sotuyo, G. G. Pillai, A. R. Katritzky, *Chem. Biol. Drug Des.* **2013**, *81*, 577–582.
[447] I. Lundt, A. J. Steiner, A. E. Stütz, C. A. Tarling, S. Ully, S. G. Withers, T. M. Wrodnigg, *Bioorg. Med. Chem.* **2006**, *14*, 1737–1742.
[448] J. Chen, C. Wang, E. Keliher, *Myeloperoxidase Imaging Agents*, **2018**, WO2018094005A1.
[449] G. R. Bebernitz, E. Cody, R. F. Day, T. J. Patel, M. Qian, L. Whitehead, T. Zabawa, F. Zecri, *Polycyclic Herg Activators*, **2015**, WO2015161052A1.
[450] D. W. Piotrowski, B. N. Rogers, W. W. Mcwhorter, D. P. Walker, J. W. Corbett, V. E. Groppi, D. G. Rudmann, *Positive Allosteric Modulators of the Nicotinic Acetylcholine Receptor*, **2003**, US2003236287A1.
[451] C. D. Selassie, C. Hansch, T. A. Khwaja, C. B. Dias, S. Pentecost, *J. Med. Chem.* **1984**, *27*, 347–357.
[452] S. M. Seyedi, Z. Jafari, N. Attaran, H. Sadeghian, M. R. Saberi, M. M. Riazi, *Bioorg. Med. Chem.* **2009**, *17*, 1614–1622.
[453] P. Griess, *Justus Liebigs Ann. Chem.* **1858**, *106*, 123–125.
[454] C. Jiménez-Castells, R. Stanton, S. Yan, P. Kosma, I. B. Wilson, *Glycobiology* **2016**, *26*, 1297–1307.
[455] D. R. Vutukuri, P. Bharathi, Z. Yu, K. Rajasekaran, M.-H. Tran, S. Thayumanavan, *J. Org. Chem.* **2003**, *68*, 1146–1149.
[456] H. Staudinger, J. Meyer, *Helv. Chim. Acta* **1919**, *2*, 635–646.
[457] J. Bernhofen, *Synthese Und Untersuchung von Photoschaltbaren Thioglycosylierten Macrocyclen*, Master Thesis, Christiana Albertina University of Kiel, **2022**.
[458] V. Poonthiyil, F. Reise, G. Despras, T. K. Lindhorst, *Eur. J. Org. Chem.* **2018**, *2018*, 6241–6248.

[459] A. Müller, *Synthesis of Azobenzene Glycoconjugates to Control Biological Function*, Dissertation, Christiana Albertina University of Kiel, **2016**.
[460] T. Weber, V. Chandrasekaran, I. Stamer, M. B. Thygesen, A. Terfort, T. K. Lindhorst, *Angew. Chem. Int. Ed Engl.* **2014**, *53*, 14583–14586.
[461] V. Chandrasekaran, T. K. Lindhorst, *Chem. Commun.* **2012**, *48*, 7519–7521.
[462] Y. Fan, A. E. Rhaz, S. Maisonneuve, E. Gillon, M. Fatthalla, F. L. Bideau, G. Laurent, S. Messaoudi, A. Imberty, J. Xie, *Beilstein J. Org. Chem.* **2024**, *20*, 1486–1496.
[463] C. E. Ballou, in *Adv. Enzymol. Relat. Areas Mol. Biol.*, John Wiley & Sons, Ltd, **1974**, pp. 239–270.
[464] C. Ballou, in *Adv. Microb. Physiol.* (Eds.: A.H. Rose, D.W. Tempest), Academic Press, **1976**, pp. 93–158.
[465] A. Reisner, J. A. J. Haagensen, M. A. Schembri, E. L. Zechner, S. Molin, *Mol. Microbiol.* **2003**, *48*, 933–946.
[466] L. Diederich, L. J. Rasmussen, W. Messer, *Plasmid* **1992**, *28*, 14–24.
[467] M. Hartmann, A. K. Horst, P. Klemm, T. K. Lindhorst, *Chem. Commun. Camb. Engl.* **2010**, *46*, 330–332.
[468] H. C. Ansel, W. P. Norred, I. L. Roth, *J. Pharm. Sci.* **1969**, *58*, 836–839.
[469] F. Ganske, U. T. Bornscheuer, *Biotechnol. Lett.* **2006**, *28*, 465–469.
[470] M. A. Schembri, H. Hasman, P. Klemm, *FEMS Microbiol. Lett.* **2000**, *188*, 147–151.
[471] M. S. Hanson, J. Hempel, C. C. Brinton, *J. Bacteriol.* **1988**, *170*, 3350–3358.
[472] M. Kawe, U. Horn, A. Plückthun, *Microb. Cell Factories* **2009**, *8*, 8.
[473] M. Bzymek, S. T. Lovett, *Proc. Natl. Acad. Sci.* **2001**, *98*, 8319–8325.
[474] D. Pribnow, *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 784.
[475] H. Schaller, C. Gray, K. Herrmann, *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 737.
[476] A. Pingoud, A. Jeltsch, *Nucleic Acids Res.* **2001**, *29*, 3705–3727.
[477] S. T. Estrem, W. Ross, T. Gaal, Z. W. S. Chen, W. Niu, R. H. Ebright, R. L. Gourse, *Genes Dev.* **1999**, *13*, 2134–2147.
[478] C. Papworth, A. Greener, J. Braman, D. Wright, *Strategies* **1996**, *9*, 3–4.
[479] A. Wellens, C. Garofalo, H. Nguyen, N. V. Gerven, R. Slättegård, J.-P. Hernalsteens, L. Wyns, S. Oscarson, H. D. Greve, S. Hultgren, J. Bouckaert, *PLOS ONE* **2008**, *3*, e2040.
[480] M. Vetsch, P. Sebbel, R. Glockshuber, *J. Mol. Biol.* **2002**, *322*, 827–840.
[481] C. A. Gross, D. B. Straus, J. W. Erickson, T. Yura, *Cold Spring Harb. Monogr. Arch.* **1990**, *19*, 167–189.
[482] H. J. Meerman, G. Georgiou, *Nat Biotechnol* **1994**, *12*, 1107–1110.
[483] C. Spiess, A. Beil, M. Ehrmann, *Cell* **1999**, *97*, 339–347.
[484] D. D. Isaac, J. S. Pinkner, S. J. Hultgren, T. J. Silhavy, *Proc. Natl. Acad. Sci.* **2005**, *102*, 17775–17779.
[485] J. Skórko-Glonek, A. Wawrzynów, K. Krzewski, K. Kurpierz, B. Lipińska, *Gene* **1995**, *163*, 47–52.
[486] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst, *Anal. Biochem.* **2010**, *407*, 188–195.
[487] S. Todt, *Mutagenese Und Expression Der R60P- Und ΔN-Signalpeptid-Varianten Des Bakteriellen Lektins FimH*, Bachelor Thesis, Christiana Albertina University of Kiel, **2023**.
[488] H. A. de Boer, L. J. Comstock, M. Vasser, *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 21–25.
[489] P. T. Kaulich, L. Cassidy, K. Winkels, A. Tholey, *Anal. Chem.* **2022**, *94*, 3600–3607.
[490] P. T. Kaulich, K. Winkels, T. B. Kaulich, C. Treitz, L. Cassidy, A. Tholey, *J. Proteome Res.* **2022**, *21*, 20–29.
[491] Y. Jiang, D. A. B. Rex, D. Schuster, B. A. Neely, G. L. Rosano, N. Volkmar, A. Momenzadeh, T. M. Peters-Clarke, S. B. Egbert, S. Kreimer, E. H. Doud, O. M. Crook, A. K. Yadav, M. Vanuopadath, A. D. Hegeman, M. L. Mayta, A. G. Duboff, N. M. Riley, R. L. Moritz, J. G. Meyer, *ACS Meas. Sci. Au* **2024**, *4*, 338–417.
[492] D. S. Roberts, J. A. Loo, Y. O. Tsybin, X. Liu, S. Wu, J. Chamot-Rooke, J. N. Agar, L. Paša-Tolić, L. M. Smith, Y. Ge, *Nat. Rev. Methods Primer* **2024**, *4*, 1–23.
[493] P. T. Kaulich, K. Jeong, O. Kohlbacher, A. Tholey, *Nat. Methods* **2024**, 1–11.
[494] J. Mitchell Wells, S. A. McLuckey, in *Methods Enzymol.*, Academic Press, **2005**, pp. 148–185.

[495] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, *Nat. Methods* **2007**, *4*, 709–712.
[496] C. K. Frese, A. F. M. Altelaar, H. van den Toorn, D. Nolting, J. Griep-Raming, A. J. R. Heck, S. Mohammed, *Anal. Chem.* **2012**, *84*, 9668–9673.
[497] L. Cassidy, A. O. Helbig, P. T. Kaulich, K. Weidenbach, R. A. Schmitz, A. Tholey, *J. Proteomics* **2021**, *230*, 103988.
[498] C. Wülfing, A. Plückthun, *Mol. Microbiol.* **1994**, *12*, 685–692.
[499] D. Missiakas, S. Raina, *J. Bacteriol.* **1997**, *179*, 2465–2471.
[500] H. Kadokura, J. Beckwith, *Antioxid. Redox Signal.* **2010**, *13*, 1231–1246.
[501] L. M. Nilsson, W. E. Thomas, E. V. Sokurenko, V. Vogel, *Structure* **2008**, *16*, 1047–1058.
[502] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed Engl.* **1999**, *38*, 1784–1788.
[503] H. Hemmi, *Methods Mol. Biol. Clifton NJ* **2014**, *1200*, 501–509.
[504] A. W. Overhauser, *Phys. Rev.* **1953**, *92*, 411–415.
[505] S. Gouin, G. Roos, J. Bouckaert, in *Top. Med. Chem.*, **2014**.
[506] A. Palmioli, P. Sperandeo, S. Bertuzzi, A. Polissi, C. Airoldi, *Bioorganic Chem.* **2021**, *112*, 104876.
[507] T. K. Lindhorst, C. Kieburg, U. Krallmann-Wenzel, *Glycoconj. J.* **1998**, *15*, 605–613.
[508] E. V. Sokurenko, V. A. McMackin, D. L. Hasty, in *Methods Enzymol.*, Academic Press, **1995**, pp. 519–528.
[509] M. Scharenberg, D. Abgottspon, E. Cicek, X. Jiang, O. Schwardt, S. Rabbani, B. Ernst, *Assay Drug Dev. Technol.* **2011**, *9*, 455–464.
[510] D. Abgottspon, G. Rölli, L. Hosch, A. Steinhuber, X. Jiang, O. Schwardt, B. Cutting, M. Smiesko, U. Jenal, B. Ernst, A. Trampuz, *J. Microbiol. Methods* **2010**, *82*, 249–255.
[511] A. Velázquez-Campoy, H. Ohtaka, A. Nezami, S. Muzammil, E. Freire, *Curr. Protoc. Cell Biol.* **2004**, *23*, 17.8.1-17.8.24.
[512] J.-P. E. Grolier, J. M. del Río, *J. Chem. Thermodyn.* **2012**, *55*, 193–202.
[513] U. Jönsson, L. Fägerstam, B. Ivarsson, B. Johnsson, R. Karlsson, K. Lundh, S. Löfås, B. Persson, H. Roos, I. Rönnberg, *BioTechniques* **1991**, *11*, 620–627.
[514] M. Malmqvist, *Nature* **1993**, *361*, 186–187.
[515] N. J. Moerke, *Curr. Protoc. Chem. Biol.* **2009**, *1*, 1–15.
[516] K. Kakehi, Y. Oda, M. Kinoshita, *Anal. Biochem.* **2001**, *297*, 111–116.
[517] M. Silbermann, *Kinetic and Thermodynamic Characterization of the Bacterial Lectin FimH*, Dissertation, University of Basel, **2021**.
[518] T. K. Dam, C. F. Brewer, *Chem. Rev.* **2002**, *102*, 387–430.
[519] W. Vornholt, M. Hartmann, M. Keusgen, *Biosens. Bioelectron.* **2007**, *22*, 2983–2988.
[520] D. Kisiela, A. Laskowska, A. Sapeta, M. Kuczkowski, A. Wieliczko, M. Ugorski, *Microbiol. Read. Engl.* **2006**, *152*, 1337–1346.
[521] B. Y. Yang, J. S. S. Gray, R. Montgomery, *Carbohydr. Res.* **1996**, *287*, 203–212.
[522] N. Takahashi, K. B. Lee, H. Nakagawa, Y. Tsukamoto, K. Masuda, Y. C. Lee, *Anal. Biochem.* **1998**, *255*, 183–187.
[523] J. Boratyński, R. Roy, *Glycoconj. J.* **1998**, *15*, 131–138.
[524] Z. Han, J. S. Pinkner, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano, S. J. Hultgren, J. W. Janetka, *J. Med. Chem.* **2010**, *53*, 4779–4792.
[525] M. P. Williamson, *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.
[526] S. Vanwetswinkel, A. N. Volkov, Y. G. J. Sterckx, A. Garcia-Pino, L. Buts, W. F. Vranken, J. Bouckaert, R. Roy, L. Wyns, N. A. J. van Nuland, *J. Med. Chem.* **2014**, *57*, 1416–1427.
[527] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.
[528] R. A. Bauer, *Drug Discov. Today* **2015**, *20*, 1061–1073.
[529] S. Leusmann, P. Ménová, E. Shanin, A. Titz, C. Rademacher, *Chem. Soc. Rev.* **2023**, *52*, 3663–3740.
[530] S. Wagner, D. Hauck, M. Hoffmann, R. Sommer, I. Joachim, R. Müller, A. Imberty, A. Varrot, A. Titz, *Angew. Chem. Int. Ed.* **2017**, *56*, 16559–16564.
[531] N. Gilboa-Garber, in *Methods Enzymol.*, Academic Press, **1982**, pp. 378–385.
[532] A. M. Spagnolo, M. Sartini, M. L. Cristina, *Rev. Res. Med. Microbiol.* **2021**, *32*, 169.

[533] D. Reynolds, M. Kollef, *Drugs* **2021**, *81*, 2117–2131.
[534] V. Aloush, S. Navon-Venezia, Y. Seigman-Igra, S. Cabili, Y. Carmeli, *Antimicrob. Agents Chemother.* **2006**, *50*, 43–48.
[535] A. J. Kunz Coyne, A. El Ghali, D. Holger, N. Rebold, M. J. Rybak, *Infect. Dis. Ther.* **2022**, *11*, 661–682.
[536] C. Chemani, A. Imberty, S. de Bentzmann, M. Pierre, M. Wimmerová, B. P. Guery, K. Faure, *Infect. Immun.* **2009**, *77*, 2065–2075.
[537] D. Tielker, S. Hacker, R. Loris, M. Strathmann, J. Wingender, S. Wilhelm, F. Rosenau, K.-E. Jaeger, *Microbiology* **2005**, *151*, 1313–1323.
[538] S. P. Diggle, R. E. Stacey, C. Dodd, M. Cámara, P. Williams, K. Winzer, *Environ. Microbiol.* **2006**, *8*, 1095–1104.
[539] T. Eierhoff, B. Bastian, R. Thuenauer, J. Madl, A. Audfray, S. Aigal, S. Juillot, G. E. Rydell, S. Müller, S. de Bentzmann, A. Imberty, C. Fleck, W. Römer, *Proc. Natl. Acad. Sci.* **2014**, *111*, 12895–12900.
[540] J.-L. Reymond, M. Bergmann, T. Darbre, *Chem. Soc. Rev.* **2013**, *42*, 4814–4822.
[541] S. Wagner, R. Sommer, S. Hinsberger, C. Lu, R. W. Hartmann, M. Empting, A. Titz, *J. Med. Chem.* **2016**, *59*, 5929–5969.
[542] I. Joachim, S. Rikker, D. Hauck, D. Ponader, S. Boden, R. Sommer, L. Hartmann, A. Titz, *Org. Biomol. Chem.* **2016**, *14*, 7933–7948.
[543] C. Grabosch, M. Hartmann, J. Schmidt-Lassen, T. K. Lindhorst, *ChemBioChem* **2011**, *12*, 1066–1074.
[544] L. F. Tietze, M. Arlt, M. Beller, K.-H. Gl üsenkamp, E. Jähde, M. F. Rajewsky, *Chem. Ber.* **1991**, *124*, 1215–1221.
[545] J. Kalia, R. T. Raines, *Curr. Org. Chem.* **2010**, *14*, 138–147.
[546] O. Sperling, A. Fuchs, T. K. Lindhorst, *Org. Biomol. Chem.* **2006**, *4*, 3913–3922.
[547] M. Wiegand, T. K. Lindhorst, *Eur. J. Org. Chem.* **2006**, *2006*, 4841–4851.
[548] T. K. Lindhorst, M. Märten, A. Fuchs, S. D. Knight, *Beilstein J. Org. Chem.* **2010**, *6*, 810–822.
[549] F. Beiroth, T. Koudelka, T. Overath, S. D. Knight, A. Tholey, T. K. Lindhorst, *Beilstein J. Org. Chem.* **2018**, *14*, 1890–1900.
[550] O. Jahn, K. Eckart, H. Tezval, J. Spiess, *Anal. Bioanal. Chem.* **2004**, *378*, 1031–1036.
[551] M. Hashimoto, Y. Hatanaka, *Eur. J. Org. Chem.* **2008**, *2008*, 2513–2523.
[552] D. P. Murale, S. C. Hong, M. M. Haque, J.-S. Lee, *Proteome Sci.* **2016**, *15*, 14.
[553] S. Sakamoto, I. Hamachi, *Anal. Sci.* **2019**, *35*, 5–27.
[554] J. N. deGruyter, L. R. Malins, P. S. Baran, *Biochemistry* **2017**, *56*, 3863–3873.
[555] D. Chen, M. M. Disotuar, X. Xiong, Y. Wang, D. H.-C. Chou, *Chem. Sci.* **2017**, *8*, 2717–2722.
[556] J.-R. Deng, N. C.-H. Lai, K. K.-Y. Kung, B. Yang, S.-F. Chung, A. S.-L. Leung, M.-C. Choi, Y.-C. Leung, M.-K. Wong, *Commun. Chem.* **2020**, *3*, 67.
[557] K. C. Tang, M. Raj, *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 1797–1805.
[558] J. I. MacDonald, H. K. Munch, T. Moore, M. B. Francis, *Nat. Chem. Biol.* **2015**, *11*, 326–331.
[559] A. Onoda, N. Inoue, E. Sumiyoshi, T. Hayashi, *Chembiochem Eur. J. Chem. Biol.* **2020**, *21*, 1274–1278.
[560] S. R. Maloy, V. J. Stewart, R. K. Taylor, *Genetic Analysis of Pathogenic Bacteria : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, NY, **1996**.
[561] A. Böck, K. Forchhammer, J. Heider, C. Baron, *Trends Biochem. Sci.* **1991**, *16*, 463–467.
[562] J. Normanly, L. G. Kleina, J.-M. Masson, J. Abelson, J. H. Miller, *J. Mol. Biol.* **1990**, *213*, 719–726.
[563] Q. Wang, A. R. Parrish, L. Wang, *Chem. Biol.* **2009**, *16*, 323–336.
[564] B. Israeli, D. S. Strugach, S. Gelkop, S. Weber, D. S. Gozlan, M. Amiram, *Adv. Funct. Mater.* **2021**, *31*, 2011276.
[565] C. Hoppmann, V. K. Lacey, G. V. Louie, J. Wei, J. P. Noel, L. Wang, *Angew. Chem. Int. Ed.* **2014**, *53*, 3932–3936.
[566] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027.
[567] *Proc. Chem. Soc.* **1961**, 357–396.

[568] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
[569] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
[570] Y. Nakamura, T. Gojobori, T. Ikemura, *Nucleic Acids Res.* **2000**, *28*, 292.
[571] H. R. Aerni, M. A. Shifman, S. Rogulina, P. O'Donoghue, J. Rinehart, *Nucleic Acids Res.* **2015**, *43*, e8.
[572] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, *342*, 357–360.
[573] D. H. Dube, C. R. Bertozzi, *Curr. Opin. Chem. Biol.* **2003**, *7*, 616–625.
[574] J. A. Prescher, D. H. Dube, C. R. Bertozzi, *Nature* **2004**, *430*, 873–877.
[575] N. G. Gurskaya, A. F. Fradkov, N. I. Pounkova, D. B. Staroverov, M. E. Bulina, Y. G. Yanushevich, Y. A. Labas, S. Lukyanov, K. A. Lukyanov, *Biochem. J.* **2003**, *373*, 403–408.
[576] C. Fessele, *Towards a Glycobiological Toolbox for the Investigation of Bacterial Adhesion*, Dissertation, Christiana Albertina University of Kiel, **2016**.
[577] Y. Mizunoe, S. N. Wai, *J. Infect. Chemother.* **1998**, *4*, 1–5.
[578] S. Hamada, A. Amano, S. Kimura, I. Nakagawa, S. Kawabata, I. Morisaki, *Oral Microbiol. Immunol.* **1998**, *13*, 129–138.
[579] K. A. Kline, S. Fälker, S. Dahlberg, S. Normark, B. Henriques-Normark, *Cell Host Microbe* **2009**, *5*, 580–592.
[580] M.-R. Yen, C. R. Peabody, S. M. Partovi, Y. Zhai, Y.-H. Tseng, M. H. Saier, *Biochim. Biophys. Acta BBA - Biomembr.* **2002**, *1562*, 6–31.
[581] D. J. Wurpel, M. Totsika, L. P. Allsopp, R. I. Webb, D. G. Moriel, M. A. Schembri, *J. Proteomics* **2016**, *131*, 177–189.
[582] L. Kuan, J. N. Schaffer, C. D. Zouzias, M. M. Pearson, *J. Med. Microbiol.* **2014**, *63*, 911–922.
[583] T. Golpasand, M. Keshvari, P. Behzadi, *BMC Microbiol.* **2024**, *24*, 344.
[584] R. R. Spurbeck, A. E. Stapleton, J. R. Johnson, S. T. Walk, T. M. Hooton, H. L. T. Mobley, *Infect. Immun.* **2011**, *79*, 4753–4763.
[585] T. K. Korhonen, V. Väisänen, H. Saxén, H. Hultberg, S. B. Svenson, *Infect. Immun.* **1982**, *37*, 286–291.
[586] G. Källenius, S. B. Svenson, H. Hultberg, R. Möllby, I. Helin, B. Cedergren, J. Winberg, *The Lancet* **1981**, *318*, 1369–1372.
[587] S. H. Jacobson, G. Källenius, L. E. Lins, S. B. Svenson, *J. Urol.* **1988**, *139*, 900–903.
[588] F. P. Lindberg, B. Lund, S. Normark, *EMBO J.* **1984**, *3*, 1167–1173.
[589] S. Normark, D. Lark, R. Hull, M. Norgren, M. Båga, P. O'Hanley, G. Schoolnik, S. Falkow, *Infect. Immun.* **1983**, *41*, 942–949.
[590] C. H. Jones, J. S. Dodson, S. J. Hultgren, *Structure, Function and Assembly of Adhesive P Pili*, ASM Press, Washington, D. C., **1996**.
[591] S. J. Hultgren, F. Lindberg, G. Magnusson, J. Kihlberg, J. M. Tennent, S. Normark, *Proc. Natl. Acad. Sci.* **1989**, *86*, 4357–4361.
[592] N. S. Omattage, Z. Deng, J. S. Pinkner, K. W. Dodson, F. Almqvist, P. Yuan, S. J. Hultgren, *Nat. Microbiol.* **2018**, *3*, 1362–1368.
[593] M. Du, Z. Yuan, G. T. Werneburg, N. S. Henderson, H. Chauhan, A. Kovach, G. Zhao, J. Johl, H. Li, D. G. Thanassi, *Nat. Commun.* **2021**, *12*, 5207.
[594] Y. Huang, B. S. Smith, L. X. Chen, R. H. G. Baxter, J. Deisenhofer, *Proc. Natl. Acad. Sci.* **2009**, *106*, 7403–7407.
[595] M. J. Kuehn, J. Heuser, S. Normark, S. J. Hultgren, *Nature* **1992**, *356*, 252–255.
[596] H. Hoschützky, F. Lottspeich, K. Jann, *Infect. Immun.* **1989**, *57*, 76–81.
[597] B. Lund, F. Lindberg, B. I. Marklund, S. Normark, *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 5898–5902.
[598] M. Båga, M. Norgren, S. Normark, *Cell* **1987**, *49*, 241–251.
[599] D. Verger, E. Miller, H. Remaut, G. Waksman, S. Hultgren, *EMBO Rep.* **2006**, *7*, 1228–1232.
[600] N. J. Holden, B. E. Uhlin, D. L. Gally, *Mol. Microbiol.* **2001**, *42*, 319–330.

[601] B. A. Braaten, L. B. Blyn, B. S. Skinner, D. A. Low, *J. Bacteriol.* **1991**, *173*, 1789–1800.
[602] Y. Xia, D. Gally, K. Forsman-Semb, B. E. Uhlin, *EMBO J.* **2000**, *19*, 1450–1457.
[603] B. I. Marklund, J. M. Tennent, E. Garcia, A. Hamers, M. Båga, F. Lindberg, W. Gaastra, S. Normark, *Mol. Microbiol.* **1992**, *6*, 2225–2242.
[604] H. Schmidt, M. Hensel, *Clin. Microbiol. Rev.* **2004**, *17*, 14–56.
[605] N. Strömberg, B. I. Marklund, B. Lund, D. Ilver, A. Hamers, W. Gaastra, K. A. Karlsson, S. Normark, *EMBO J.* **1990**, *9*, 2001–2010.
[606] J. R. Johnson, *Clin. Microbiol. Rev.* **1991**, *4*, 80–128.
[607] A. B. Celi, J. Goldstein, M. V. Rosato-Siri, A. Pinto, *Front. Mol. Biosci.* **2022**, *9*, 813637.
[608] J. R. Johnson, T. A. Russo, J. J. Brown, A. Stapleton, *J. Infect. Dis.* **1998**, *177*, 97–101.
[609] J. R. Johnson, J. J. Brown, J. N. Maslow, *J. Infect. Dis.* **1998**, *177*, 651–661.
[610] J. R. Johnson, *Infect. Immun.* **1998**, *66*, 4568–4571.
[611] R. Lindstedt, N. Baker, P. Falk, R. Hull, S. Hull, J. Karr, H. Leffler, C. Svanborg Edén, G. Larson, *Infect. Immun.* **1989**, *57*, 3389–3394.
[612] J. F. Karr, B. J. Nowicki, L. D. Truong, R. A. Hull, J. J. Moulds, S. I. Hull, *Infect. Immun.* **1990**, *58*, 4055–4062.
[613] N. Strömberg, P. G. Nyholm, I. Pascher, S. Normark, *Proc. Natl. Acad. Sci.* **1991**, *88*, 9340–9344.
[614] A. E. Stapleton, M. R. Stroud, S. I. Hakomori, W. E. Stamm, *Infect. Immun.* **1998**, *66*, 3856–3861.
[615] N. Legros, S. Ptascheck, G. Pohlentz, H. Karch, U. Dobrindt, J. Müthing, *Glycobiology* **2019**, *29*, 789–802.
[616] J. Kihlberg, S. J. Hultgren, S. Normark, G. Magnusson, *J. Am. Chem. Soc.* **1989**, *111*, 6364–6368.
[617] R. Striker, U. Nilsson, A. Stonecipher, G. Magnusson, S. J. Hultgren, *Mol. Microbiol.* **1995**, *16*, 1021–1029.
[618] U. Nilsson, R. T. Striker, S. J. Hultgren, G. Magnusson, *Bioorg. Med. Chem.* **1996**, *4*, 1809–1817.
[619] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson, *ChemBioChem* **2002**, *3*, 772–779.
[620] A. Salminen, V. Loimaranta, J. A. F. Joosten, A. S. Khan, J. Hacker, R. J. Pieters, J. Finne, *J. Antimicrob. Chemother.* **2007**, *60*, 495–501.
[621] L. Hagberg, U. Jodal, T. K. Korhonen, G. Lidin-Janson, U. Lindberg, C. Svanborg Edén, *Infect. Immun.* **1981**, *31*, 564–570.
[622] R. Möllby, G. Källenius, T. K. Korhonen, J. Winberg, S. B. Svenson, *Infection* **1983**, *11*, 68–72.
[623] N. Agata, M. Ohta, H. Miyazawa, M. Mori, N. Kido, N. Kato, *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* **1989**, *8*, 156–159.
[624] S. Bosley, C. G. Krueger, A. Birmingham, A. B. Howell, J. D. Reed, *J. Diet. Suppl.* **2024**, *21*, 327–343.
[625] S. D. Manning, L. Zhang, B. Foxman, A. Spindler, P. Tallman, C. F. Marrs, *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 637–640.
[626] J. R. Johnson, A. L. Stell, N. Kaster, C. Fasching, T. T. O'Bryan, *Infect. Immun.* **2001**, *69*, 2318–2327.
[627] J. Jass, S. Schedin, E. Fällman, J. Ohlsson, U. J. Nilsson, B. E. Uhlin, O. Axner, *Biophys. J.* **2004**, *87*, 4271–4283.
[628] M. Andersson, E. Fällman, B. E. Uhlin, O. Axner, *Biophys. J.* **2006**, *90*, 1521–1534.
[629] N. Mortezaei, B. Singh, E. Bullitt, B. E. Uhlin, M. Andersson, *Sci. Rep.* **2013**, *3*, 3393.
[630] J. E. Klinth, M. Castelain, B. E. Uhlin, O. Axner, *PLOS ONE* **2012**, *7*, e38548.
[631] S. Balser, M. Röhrl, C. Spormann, T. K. Lindhorst, A. Terfort, *ACS Appl. Mater. Interfaces* **2024**, *16*, 14243–14251.
[632] L. Yang, R. Bashir, *Biotechnol. Adv.* **2008**, *26*, 135–150.
[633] S. Balser, J. Bernd, L. Fritsche, A. Terfort, *ACS Appl. Polym. Mater.* **2022**, *4*, 8344–8356.
[634] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, R. Y. Tsien, *Nat. Biotechnol.* **2004**, *22*, 1567–1572.
[635] E. M. Lederberg, J. Lederberg, *Genetics* **1953**, *38*, 51–64.
[636] J. L. Hartley, G. F. Temple, M. A. Brasch, *Genome Res.* **2000**, *10*, 1788–1795.
[637] L. Thomason, D. L. Court, M. Bubunenko, N. Costantino, H. Wilson, S. Datta, A. Oppenheim, *Curr. Protoc. Mol. Biol.* **2007**, *78*, 1.16.1-1.16.24.

[638] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court, *Nat. Protoc.* **2009**, *4*, 206–223.
[639] A. Haldimann, B. L. Wanner, *J. Bacteriol.* **2001**, *183*, 6384–6393.
[640] L. W. Enquist, A. Kikuchi, R. A. Weisberg, *Cold Spring Harb. Symp. Quant. Biol.* **1979**, *43 Pt 2*, 1115–1120.
[641] W. Ross, A. Landy, *Proc. Natl. Acad. Sci.* **1982**, *79*, 7724–7728.
[642] M. P. Molloy, B. R. Herbert, M. B. Slade, T. Rabilloud, A. S. Nouwens, K. L. Williams, A. A. Gooley, *Eur. J. Biochem.* **2000**, *267*, 2871–2881.
[643] L. Serre, K. Pereira de Jesus, C. Zelwer, N. Bureaud, F. Schoentgen, H. Bénédetti, *J. Mol. Biol.* **2001**, *310*, 617–634.
[644] A. Landy, W. Ross, *Science* **1977**, *197*, 1147–1160.
[645] Y. Kikuchi, H. A. Nash, *Proc. Natl. Acad. Sci.* **1979**, *76*, 3760–3764.
[646] V. C. Bode, A. D. Kaiser, *J. Mol. Biol.* **1965**, *14*, 399–417.
[647] M. V. Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov, S. A. Lukyanov, *Nat. Biotechnol.* **1999**, *17*, 969–973.
[648] B. J. Bevis, B. S. Glick, *Nat. Biotechnol.* **2002**, *20*, 83–87.
[649] F. St-Pierre, L. Cui, D. G. Priest, D. Endy, I. B. Dodd, K. E. Shearwin, *ACS Synth. Biol.* **2013**, *2*, 537–541.
[650] T. Miki, K. Yoshioka, T. Horiuchi, *J. Mol. Biol.* **1984**, *174*, 605–625.
[651] R. P. Shetty, D. Endy, T. F. Knight, *J. Biol. Eng.* **2008**, *2*, 5.
[652] L. Cui, A. Vigouroux, F. Rousset, H. Varet, V. Khanna, D. Bikard, *Nat. Commun.* **2018**, *9*, 1912.
[653] N. C. Shaner, *Clin. Chem.* **2013**, *59*, 440–441.
[654] X. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien, S. J. Remington, *Biochemistry* **2006**, *45*, 9639–9647.
[655] B. P. Cormack, R. H. Valdivia, S. Falkow, *Gene* **1996**, *173*, 33–38.
[656] H. Jeong, Y. M. Sim, H. J. Kim, S. J. Lee, *Front. Microbiol.* **2017**, *8*, 585.
[657] H. A. Nash, *Nature* **1974**, *247*, 543–545.
[658] H. A. Nash, C. A. Robertson, *J. Biol. Chem.* **1981**, *256*, 9246–9253.
[659] N. L. Craig, H. A. Nash, *Cell* **1984**, *39*, 707–716.
[660] M. Ptashne, K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer, R. T. Sauer, *Science* **1976**, *194*, 156–161.
[661] R. J. Penfold, J. M. Pemberton, *Gene* **1992**, *118*, 145–146.
[662] L. Cui, K. E. Shearwin, in *Synth. DNA Methods Protoc.* (Ed.: R.A. Hughes), Springer, New York, NY, **2017**, pp. 139–155.
[663] G. Guarneros, H. Echols, *J. Mol. Biol.* **1970**, *47*, 565–574.
[664] K. Abremski, S. Gottesman, *J. Biol. Chem.* **1982**, *257*, 9658–9662.
[665] L. Meyer-Leon, J. F. Senecoff, R. C. Bruckner, M. M. Cox, *Cold Spring Harb. Symp. Quant. Biol.* **1984**, *49*, 797–804.
[666] J. R. Broach, V. R. Guarascio, M. Jayaram, *Cell* **1982**, *29*, 227–234.
[667] B. J. Andrews, G. A. Proteau, L. G. Beatty, P. D. Sadowski, *Cell* **1985**, *40*, 795–803.
[668] X. D. Zhu, P. D. Sadowski, *J. Biol. Chem.* **1995**, *270*, 23044–23054.
[669] Y.-N. Park, D. Masison, E. Eisenberg, L. E. Greene, *Yeast* **2011**, *28*, 673–681.
[670] J. Dan, H. Deng, Y. Xia, Y. Zhan, N. Tang, Y. Wang, M. Cao, *Open Life Sci.* **2022**, *17*, 172–179.
[671] T. Hashimoto-Gotoh, F. C. H. Franklin, A. Nordheim, K. N. Timmis, *Gene* **1981**, *16*, 227–235.
[672] R. Giraldo, M. E. Fernández-Tresguerres, *Plasmid* **2004**, *52*, 69–83.
[673] M. Hartmann, H. Papavlassopoulos, V. Chandrasekaran, C. Grabosch, F. Beiroth, T. K. Lindhorst, C. Röhl, *FEBS Lett.* **2012**, *586*, 1459–1465.
[674] F. Bolivar, R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, S. Falkow, *Gene* **1977**, *2*, 95–113.
[675] I. Boros, G. Pósfai, P. Venetianer, *Gene* **1984**, *30*, 257–260.
[676] P. B. Olsen, P. Klemm, *FEMS Microbiol. Lett.* **1994**, *116*, 95–100.
[677] I. C. M. Dea, A. Morrison, in *Adv. Carbohydr. Chem. Biochem.* (Eds.: R.S. Tipson, D. Horton), Academic Press, **1975**, pp. 241–312.
[678] L. R. S. Moreira, E. X. F. Filho, *Appl. Microbiol. Biotechnol.* **2008**, *79*, 165–178.

[679] L. Melton, B. G. Smith, R. Ibrahim, R. Schröder, L. Melton, P. Harris, U. Schmitt, *N. Z. J. For. Sci.* **2009**.
[680] M. Mori, Y. Ito, T. Ogawa, *Carbohydr. Res.* **1990**, *195*, 199–224.
[681] T. Nagase, S. Shinkai, I. Hamachi, *Chem. Commun.* **2001**, 229–230.
[682] D. A. Guk, R. O. Burlutskiy, D. A. Lemenovskiy, E. K. Beloglazkina, *Mendeleev Commun.* **2023**, *33*, 14–16.
[683] A. Müller, H. Kobarg, V. Chandrasekaran, J. Gronow, F. D. Sönnichsen, T. K. Lindhorst, *Chem. – Eur. J.* **2015**, *21*, 13723–13731.
[684] P. Cattaneo, M. Persico, *Phys. Chem. Chem. Phys.* **1999**, *1*, 4739–4743.
[685] M. P. Jacobson, R. A. Friesner, Z. Xiang, B. Honig, *J. Mol. Biol.* **2002**, *320*, 597–608.
[686] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, R. A. Friesner, *Proteins* **2004**, *55*, 351–367.
[687] Schrödinger Release 2024-2: Glide, **2024**.
[688] Schrödinger Release 2024-2: MacroModel, **2024**.
[689] J. Brekalo, G. Despras, T. K. Lindhorst, *Org. Biomol. Chem.* **2019**, *17*, 5929–5942.
[690] C. T. Chung, S. L. Niemela, R. H. Miller, *Proc. Natl. Acad. Sci.* **1989**, *86*, 2172–2175.
[691] Addgene: Vector Database - pQE30. https://www.addgene.org/vector-database/3883/ [08 July 2024].
[692] lifescience-market.com - pQE-30. https://lifescience-market.com/plasmid-c-94/pqe30-p-126561.html [08 July 2024].
[693] K. A. Datsenko, B. L. Wanner, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640–6645.
[694] B. P. Callen, K. E. Shearwin, J. B. Egan, *Mol. Cell* **2004**, *14*, 647–656.
[695] J. R. Johnson, J. J. Brown, *J. Infect. Dis.* **1996**, *173*, 920–926.
[696] K. Okonechnikov, O. Golosova, M. Fursov, the UGENE team, *Bioinformatics* **2012**, *28*, 1166–1167.
[697] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 5463–5467.
[698] C. C. Richardson, in *The Enzymes* (Ed.: P.D. Boyer), Academic Press, **1981**, pp. 299–314.
[699] M. J. Engler, C. C. Richardson, in *The Enzymes* (Ed.: P.D. Boyer), Academic Press, **1982**, pp. 3–29.
[700] M. Mandel, A. Higa, *J. Mol. Biol.* **1970**, *53*, 159–162.
[701] J. Liu, W. Chang, L. Pan, X. Liu, L. Su, W. Zhang, Q. Li, Y. Zheng, *Indian J. Microbiol.* **2018**, *58*, 448–456.
[702] U. K. Laemmli, *Nature* **1970**, *227*, 680–685.
[703] T. S. Meyer, B. L. Lamberts, *Biochim. Biophys. Acta BBA - Gen. Subj.* **1965**, *107*, 144–145.
[704] A. M. Altschul, W. J. Evans, in *Methods Enzymol.*, Academic Press, **1967**, pp. 179–186.
[705] P. K. Smith, R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, D. C. Klenk, *Anal. Biochem.* **1985**, *150*, 76–85.
[706] M. Vaara, *Drugs Exp. Clin. Res.* **1991**, *17*, 437–443.
[707] Y. Kimura, H. Matsunaga, M. Vaara, *J. Antibiot. (Tokyo)* **1992**, *45*, 742–749.
[708] M. Schindler, M. J. Osborn, *Biochemistry* **1979**, *18*, 4425–4430.
[709] T. Katsu, S. Yoshimura, Y. Fujita, *FEBS Lett.* **1984**, *166*, 175–178.
[710] D. Wessel, U. I. Flügge, *Anal. Biochem.* **1984**, *138*, 141–143.
[711] K. Jeong, J. Kim, M. Gaikwad, S. N. Hidayah, L. Heikaus, H. Schlüter, O. Kohlbacher, *Cell Syst.* **2020**, *10*, 213-218.e6.
[712] H. L. Röst, U. Schmitt, R. Aebersold, L. Malmström, *Proteomics* **2014**, *14*, 74–77.
[713] S. Turan, J. Bode, *FASEB J.* **2011**, *25*, 4088–4107.
[714] P. P. Cherepanov, W. Wackernagel, *Gene* **1995**, *158*, 9–14.

Danksagung

An erster Stelle möchte ich selbstverständlich Frau Prof. Thisbe K. Lindhorst für die freundliche Aufnahme in den Arbeitskreis, sowie ihre gute Betreuung und Expertise in Ihrer Rolle als Doktormutter und Referentin danken. Ebenso danke ich ihr für die Bereitstellung des für mich überaus spannenden Themas dieser Dissertation, sowie weiterer Arbeiten, die es aus Platzgründen nicht mehr in diese Arbeit geschafft haben. Die Bearbeitung meines Herzensprojektes, das im Wesentlichen den gesamten Inhalt dieser umfangreichen Dissertation ausmacht, erwies sich zwar als deutlich herausfordernder und komplexer als zunächst angenommen aber hat mir im Großen und Ganzen viel Freude bereitet. Insbesondere, dass ich so tiefgreifende Einblicke und Erfahrungen in den verschiedensten Bereichen, angefangen von der Organischen Chemie, Photochemie und Proteinbiochemie, bis hin zur Gentechnik, Glycobiologie und in biologischen Assaysystemen gewinnen konnte, ist einfach großartig und wird sich auf meinem weiteren Weg sicher als nützlich erweisen. Herrn Prof. Axel J. Scheidig danke ich vielmals für die Übernahme der Betreuung eines Teils dieser Arbeit sowie für die vielfältige, jahrelange, freundliche Unterstützung des Arbeitskreises. Die Möglichkeit, einen nicht unwesentlichen Teil der gentechnischen Arbeiten, sowie der Proteinexpression und -Reinigung im Labor des AK Scheidig durchführen zu können, haben sich als überaus wertvoll erwiesen. Darüber hinaus haben die vielen inhaltlichen Tipps und Hilfestellungen von den Mitarbeitern des Arbeitskreises weitgehend dazu beigetragen, das Projekt sinnvoll und bis zum Ende durchführen zu können. All das ist alles andere als selbstverständlich und ich möchte dafür an dieser Stelle nochmals meinen Dank aussprechen.

Den aktuellen und früheren Mitarbeitern in beiden Arbeitskreisen danke ich für die gegenseitige Hilfe, eine tolle Zeit und viele schöne, einprägsame Erinnerungen an die vergangenen Jahre. In beiden Arbeitskreisen habe ich wertvolle Freundschaften geschlossen, die hoffentlich noch weit über die Promotionszeit hinaus bestehen werden. Prof. Andreas Tholey und Dr. Philipp Kaulich danke ich ganz besonders für die massenspektrometrischen Untersuchungen, die die Qualität dieser Dissertation erheblich gesteigert haben. Ebenso möchte ich allen weiteren Kooperationspartnern in Deutschland, Österreich und Frankreich danken, ohne die die insgesamt fünf veröffentlichten Publikationen nicht zustande gekommen wären. Meinen F3-Praktikanten Sönke Sdunnus und Jule Bergfeld, sowie meiner fleißigen Bachelorandin Sofie Todt danke ich für ihre Hilfe und ihren Beitrag zu dieser Arbeit. Natürlich gilt mein Dank auch allen weiteren ehemaligen und aktuellen Mitgliedern des Otto Diels-Instituts, ganz besonders Christine Haug und den Mitarbeitern aus der spektroskopischen Abteilung. Dr. Claus Bier danke ich außerdem für viele tiefgründige Diskussionen und unzählige aufbauende Worte. Herrn Prof. Rudi Glockshuber und seiner Mitarbeiterin Verena Eggli danke ich für die freundliche Bereitstellung des *E. coli*-Stammes HM125 und des mitgelieferten Plasmids zur Expression von $FimH_{trc}$. Wie immer gilt mein größter Dank meiner Familie, insbesondere natürlich meinen Eltern, die mich in den letzten Jahren sowohl mental als auch finanziell erheblich unterstützt haben. Mein Freund Cornelius hat mir in den letzten Jahren den Rücken freigehalten und war mir dadurch eine unverzichtbare Stütze in allen Lebenslagen und mein sicherer Fels in der Brandung. Meiner während der Promotionszeit leider verstorbenen Oma Brunhild bin ich von ganzen Herzen dankbar, weil sie immer an mich geglaubt und mir meine positiven Eigenschaften, wie meinen herausragenden Fleiß, meine Leidenschaft und mein Durchhaltevermögen vor Augen geführt hat. Ihr möchte ich diese Dissertation in liebevoller Erinnerung widmen.

Appendix

DNA sequences

Sequence of pPKL241

> [pQE-30-FimH-truncate - 3988 bp]
; FimH-truncate [CDS]: 108 - 668 - Selection From [1] To [3988]: partial T5/lacO Deletion: 30 - 61

ctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattatatagattcaattgtgagcggataacaatttcacacagaattcaggaggagattgtaatgaaacgtgttattaccctgtttgc
tgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaaccgccaatggtaccgctatccctattggcggtggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaa
aacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaata
tagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgtttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaa
agctggctcattaattgccgtgcttattttgcgacagaccaacaactataacagcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcaccatc
actaaggatccgcatgcgagctcggtaccccgggtcgacctgcagccaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggt
tgccgccgggcgttttttattggtgagaatccaagctagcttggcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaag
aacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgccc
gcctgatgaatgctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaata
ccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcac
cagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggct
tccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacgcctggggtaatgactctctagcttgaggcatcaaa
taaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccctctagagctgcctcgcgcgtttcggtgatgacggtgaaaacctc
tgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggcgcagccatgacccagtcacgtagc
gatagcggagtgtatactggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgctcttccgcttcc
tcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccag
caaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggataccgtccgccttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgg
tgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca
ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctg
acgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc
tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctg
caatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgg
gaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgt
aagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgct
catcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgag
caaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacata
tttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttt
cgtcttcac

Sequence of pPKL241 N136C

> [pQE-30-FimH-truncate-N136C - 3988 bp]
; FimH-truncate-N136C [CDS]: 108 - 668 - Selection From [1] To [3988]: partial T5/lacO Deletion: 30 - 61

ctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattatatagattcaattgtgagcggataacaatttcacacagaattcaggaggagattgtaatgaaacgtgttattaccctgtttgc
tgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaaccgccaatggtaccgctatccctattggcggtggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaa
aacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaata
tagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgtttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaa
agctggctcattaattgccgtgcttattttgcgacagaccaactgctataacagcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcaccatc
actaaggatccgcatgcgagctcggtaccccgggtcgacctgcagccaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggt
tgccgccgggcgttttttattggtgagaatccaagctagcttggcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaag
aacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgccc
gcctgatgaatgctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaata
ccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcac
cagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggct
tccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacgcctggggtaatgactctctagcttgaggcatcaaa
taaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccctctagagctgcctcgcgcgtttcggtgatgacggtgaaaacctc
tgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggcgcagccatgacccagtcacgtagc
gatagcggagtgtatactggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgctcttccgcttcc
tcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccag
caaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggataccgtccgccttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgg
tgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca
ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctg
acgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc
tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctg
caatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgg
gaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac

atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgt
aagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgct
catcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgag
caaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacata
tttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttt
cgtcttcac

Sequence of pPKL241 S139C

> [pQE-30-FimH-truncate-S139C - 3988 bp]
; FimH-truncate-S139C [CDS]: 108 - 668 - Selection From [1] To [3988]: partial T5/lacO Deletion: 30 - 61
ctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattatagattcaattgtgagcggataacaatttcacacagaattcaggaggagattgtaatgaaacgtgttattaccctgtttgc
tgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaaccgccaatggtaccgctatccctattggcggtggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaa
aacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaata
tagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgtttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaa
agctggctcattaattgccgtgcttattttgcgacagaccaacaactataactgcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcaccatc
actaaggatccgcatgcgagctcggtaccccgggtcgacctgcagccaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggt
tgccgccgggcgtttttttattggtgagaatccaagctagcttggcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaag
aacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgccc
gcctgatgaatgctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcaccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaata
ccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcac
cagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggct
tccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacgcctggggtaatgactctctagcttgaggcatcaaa
taaaacgaaaggctcagtcgaaagactgggccttttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccctctagagctgcctcgcgcgtttcggtgatgacggtgaaaacctc
tgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggcgcagccatgacccagtcacgtagc
gatagcggagtgtatactggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgctcttccgcttcc
tcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccag
caaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgg
tgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcca
ctggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctg
acgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtc
tgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctg
caatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaacttatccgcctccatccagtctattaattgttgccgg
gaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgt
aagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgct
catcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgag
caaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacata
tttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttt
cgtcttcac

Sequence of pPKL241 ΔN-sigpep N136C

> [pQE-30-FimH-truncate-N136C - Delta N-Signalpeptid - 3927 bp]
FimH-truncate ΔN-sigpep - N136C [CDS]: 107 - 610 - Selection From [1] To [3927]
ctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattatagattcaattgtgagcggataacaatttcacacagaattcaggaggagattgtatgttcgcctgtaaaaccgccaatgg
taccgctatccctattggcggtggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaaaccatta
cagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgtttataattcg
agaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaactgctataacagcgatgat
ttccagtttgtgtggaatatttacgccaataatgatgtggtggtcctactggcggccatcaccatcactaaggatccgcatgcgagctcggtaccccgggtcgacctgcagccaagcttaattagc
tgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcttggcgagattttcaggagct
aaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctgga
tattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatat
gggatagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaa
aacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggca
aatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcgggg
cgtaatttttttaaggcagttattggtgcccttaaacgcctggggtaatgactctctagcttgaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaac
gctctcctgagtaggacaaatccgccctctagagctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcaga
caagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggcgcagccatgacccagtcacgtagcgatagcggagtgtatactggcttaactatgcggcatcagagcagattgtactgagagtgca
ccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaa
ggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccc
ctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccgga
tacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgc
gccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtg
gcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtt
tgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggat
cttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatc
catagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccag
ccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacagg
catcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaa

gtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggc
gaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatcca
gttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttg
aatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgcc
acctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtcttcac

Sequence of pFimHs-NT-trc

> [pFimHs-NT-trc - 5264 bp
; FimH [CDS]: 322 - 864 - Selection From [1] To [5264]
gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttct
ggataatgttttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagacc
atggaattcgagctcggtacccggccggggatcctctagaggagggatgattgtaatgaaacgagttattaccctgtttgctgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaacc
gccaatggtaccgcaattccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatctttgccataacgattatccgg
aaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgt
ttataattcgagaacggataagccgtggccggtggcgcttttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaacaactataa
cagcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggctaagcttggctgttttggcggatgagagaagaagattttcagcctgatacagattaaatcagaacgcagaag
cggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaac
tgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcggatttgaacgttgcgaagcaacg
gcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggcctttttgcgtttctacaaactcttttgtttatttttctaaatacattcaaatat
gtatccgctcatggatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttct
cgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccct
gatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggatttgccgatttcggcctat
tggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattca
aatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctc
acccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatg
agcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttа
cggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggа
tcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactc
tagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggta
tcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattgg
taactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactg
agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaact
ctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttacc
agtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacc
tacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag
ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctg
gccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagc
gaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatc
gctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcaga
ggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaacctttcgcggtatggcatgatagcgcccg
gaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcg
ggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaa
attgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggct
gatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaagacggta
cgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctcgctcgtctggctggctggcataaatatctcactcgcaatcaaattca
gccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatg
cgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaacaggattttcgcctgctggggca
aaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcg
ttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctg

Sequence of pFimH-His$_6$-NT-trc

> [pFimH-His6-NT-trc - 5285 bp]
; FimH-His6-truncate [CDS]: 322 - 882 - Selection From [1] To [5285]
gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttct
ggataatgttttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagacc
atggaattcgagctcggtacccggccggggatcctctagaggagggatgattgtaatgaaacgagttattaccctgtttgctgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaacc
gccaatggtaccgcaattccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatctttgccataacgattatccgg
aaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgt
ttataattcgagaacggataagccgtggccggtggcgcttttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaacaactataa
cagcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcaccatcactaagcttggctgttttggcggatgagagaagattttcagcctgatаca
gattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtct
ccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcgg
atttgaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggccttttgcgtttctacaaactcttttgttt
atttttctaaatacattcaaatatgtatccgctcatggatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgct
cctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatg
gttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttata
agggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccct
atttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttg
cggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcg

ccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtac
tcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagcta
accgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgca
aactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgag
ataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaa
atcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggt
ttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct
acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgc
acacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaa
caggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaac
gccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccg
aacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtatttttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgc
atagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtg
accgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaac
ctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaacca
ggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctc
cagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcac
aatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaa
cagtattatttttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctgg
cataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgc
caacgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaacccacca
tcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcc
caatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctg

Sequence of pFimH-His$_6$-NT-trc

> [pFimH-His6-NT-trc-R60P - 5285 bp]
; FimH-R60P-His6-truncate [CDS]: 322 - 882 - Selection From [1] To [5285]

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttct
ggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagacc
atggaattcgagctcggtacccggccggggatcctctagaggagggatgattgtaatgaaacgagttattaccctgtttgctgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaacc
gccaatggtaccgcaattccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatctttgccataacgattatccgg
aaaccattacagactatgtcacactgcaaccaggctcggcttatggcggcgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgt
ttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaacaactataa
cagcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcaccatcactaagcttggctgttttggcggatgagagaagattttcagcctgataca
gattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtct
ccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcgg
atttgaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggcctttttgcgtttctacaaactctttttgttt
atttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttg
cggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcg
ccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtac
tcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagcta
accgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccatacccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgca
aactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctgg
agccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgag
ataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaa
atcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggt
ttgtttgccggatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct
acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgc
acacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaa
caggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaac
gccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccg
aacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtatttttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgc
atagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtg
accgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaac
ctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaacca
ggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctc
cagtctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcac
aatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaa
cagtattatttttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctgg
cataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgc
caacgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaacccacca
tcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcc
caatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctg

Sequence of pFimH-His$_6$-NT-trc

> [pFimH-His6-NT-trc-N136C - 5285 bp]
; FimH-N136C-His6-truncate [CDS]: 322 - 882 - Selection From [1] To [5285]

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttct
ggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagacc
atggaattcgagctcggtacccggccggggatcctctagaggagggatgattgtaatgaaacgagttattaccctgtttgctgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaacc
gccaatggtaccgcaattccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatctttgccataacgattatccgg
aaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgt
ttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaactgctataac
agcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcactaagcttggctgttttggcggatgagagaagattttcagcctgatacag
attaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctc
cccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcgga
tttgaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggcctttttgcgtttctacaaactcttttgtttta
ttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcgg
cattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccc
cgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactca
ccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaacc
gcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaac
tattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagc
cggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagata
ggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcc
cttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgt
ttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacat
acctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacac
agcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacagg
agagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaacgccag
caacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacga
ccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagtt
aagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtc
tccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaacctttcgc
ggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccag
ccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctg
gccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttct
cgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtatt
attttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctggcataaat
atctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgat
cagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaac
aggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatac
gcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctg

Sequence of pFimH-His$_6$-NT-trc S139C

> [pFimH-His6-NT-trc-S139C - 5285 bp]
; FimH-S139C-His6-truncate [CDS]: 322 - 882 - Selection From [1] To [5285]

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttct
ggataatgtttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacaggaaacagacc
atggaattcgagctcggtacccggccggggatcctctagaggagggatgattgtaatgaaacgagttattaccctgtttgctgtactgctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaacc
gccaatggtaccgcaattccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtggggcaaaacctggtcgtggatctttcgacgcaaatctttgccataacgattatccgg
aaaccattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaatttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacgccgcgcgttgt
ttataattcgagaacggataagccgtggccggtggcgctttatttgacgcctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttattttgcgacagaccaacaactataa
ctgcgatgatttccagtttgtgtggaatatttacgccaataatgatgtggtggtgcctactggcggccatcaccatcactaagcttggctgttttggcggatgagagaagattttcagcctgatacag
attaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctc
cccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcgga
tttgaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatggcctttttgcgtttctacaaactcttttgtttta
ttttctaaatacattcaaatatgtatccgctcatggatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctc
ctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatcccttttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggt
tcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataag
ggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaaccccctatt
tgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcgg
cattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccc
cgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactca
ccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaacc
gcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaac
tattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagc
cggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagata
ggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcc
cttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgt

ttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacat
acctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacac
agcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacagg
agagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccag
caacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacga
ccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagtt
aagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtc
tccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggtgcaaaacctttcgc
ggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccag
ccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctg
gccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttct
cgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtatt
attttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctggctggctggcataaat
atctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgat
cagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaac
aggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatac
gcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattgatctg

Sequence of the *E. coli* K12 chromosomal *attB* site with inserted pOSIP-KL-*mcherry*

> [attb site (+ 100 nt downstream+upstream) + POSIP-Kl-*mCherry*- inserted - 8800 bp] - Selection From [1] To [8800]

agtctggtctggtagcaatgattttataggcatcacggcaatatacattttataacttaacggctgacatggccgtatgaaggcacccagaccgggaattaattcttgaagacgctactatttcctacttccta
tctagctaaaatctactaacactaggtcattaattcagaggatgaggattttattagtagaaccaagagtgggcggaacggaagctaacctatgtcttgaagaaaaatgataatggctgggataagtatag
tagattaaggaaataaagttaataataaacaatggcgtgataagtcggtatgtaaggtaagtagttgaagaacctatgctcttattaattagaaaaatttttctattattaatttaagaggggagaaaataa
agttgattaagtatggagcaggtgccgatgcttattaattagaattcagtagcatttgtcgtagccggtgcaaacacatacgcttcttgctgagtacgtgagttcaccacgcggaattcggtgttatcgaaca
ccactgcgccgcgaccagaaacgatatccacatccccttcaatgtagctgttggtcaccagcgtacgcggctgacgattcgtttccagacggttctgcacaccgctgttggtgacaaagaaggtgttctgac
gaccgagaatgttaacgttgttaatctgtacctggtcaccatcagtacgcagtgccaccgccggatggttacctgcatctacgctatcgcccagcgtgttttcgatggtcagattttgcagttgcaggccattg
ttttgtgaccagaagaccgcagagcagagaacaccgatactgtcgctgcgtttgctctggcagctatcgtacatataccacgctggtttacctggcatatatttgccgcgcgggttgacgtcgtgacgccagt
cggcagggctcatgccaccatcaagggaaagcccaatcttcacatcaatcggtttttcacctgtaccgtacagagtaattccacccggagcggcagggacatataccgttccctgatactcaccaggcatc
acggcaatatactggcgcttgttggtacgcttgataattgccgcatctaccgccgcctgaatcgtggtatgcgttacaccttgagtgcccgccgggccgacaacaaagtcaggttgcgcaggcagggtaatc
ggggaaggattccacgctgcagcacctggtgtcagggatgcaaaatagtgttgagcatcgaaattctgcgcttcttttgccgacagaatcgggcgagaagaggtaccaggcgcggtttgatcagaaggac
gttgatcgggcggggttgagctacaggcggtcagcgtcacgccaaaagccaatgccagcgccagacgggaaactgaaaatgtgttcacaggttgctccgggctatgaaatagaaaaatgaatccgttga
agcctgctttttatactaacttggcattataaaaaagcattgcttatcaatttgttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgatttcaattttgtcccactccctgcctctgtcatcac
gatactgtgatgccatggcatatggaagttcctatactttctagagaataggaacttcggaataggaacttcctaattcccatgtcagccgttaagtgttcctgtgtcactcaaaattgctttgagaggctcta
agggcttctcagtgcgttacatccctggcttgttgtccacaaccgttaaaccttaaaagctttaaaagccttatatattctttttttcttatataaacttaaaaccttagaggctatttaagttgctgatttatatt
aattttattgttcaaacatgagagcttagtacgtgaaacatgagagcttagtacgttagccatgagagcttagtacgttagccatgagggtttagttcgttaaacatgagagcttagtacgttaaacatgaga
gcttagtacgtgaaacatgagagcttagtacgtactatcaacaggttgaactgctgatcttcagatcctctacgccggacgcatcgtggccggatcttgcggccgcactaggtcgacgctcactcaaaggcg
gtaatcagataaaaaaaatccttagctttcgctaaggatgatttctgctagtattattagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaa
gccgtttctgtaatgaaggagaaaactcaccgaggcagttccaaagaatggcaaggtcctggtaacggtctgcgattccgacccgtccaacatcaatacaacctattaatttcccctcgtcaaaaataagg
ttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaagagcttgtgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaa
ccgttattcatgcgttgattgcgcctgagcaagacgaaatacacgatcgctgttaaaaggacaattacaaacaggaatcgaatgtaaccggcgcaggaacacggccagcgcatcaacaatattttcacctg
aatcaggatattcttctaatacctggaaggctgttttcccaggaatcgcggtggtgagtaaccacgcatcatcaggagtacggataaaatgcttgatggtcgggagaggcataaactccgtcagccagttg
agacggaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaagcgatagattgtcgcacctgattgcccgacattatcgcgagccc
atttatacccatataaatcagcgtccatgttggagtttaagcgcggacgggagcaagacgtttcccgttgaatatggctcataacaccccttgtattactgtttatgtaagcagacagttttattgttcatgatg
atatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtgcgttacatatcaaagggaaaactgtccatacccatgggctagctgatcagccagtgccaagcttggcattgcttatcaatttt
gttgcaacgaacaggtcactatcagtcaaaataaaatcattatttgatttcaattttgtcccactccctgcctctgtcatcacgatactgtgatgccatggtgtccgacttatgcccgagaagatgttgagcaa
acttatcgcttatctgcttctcatagagtcttgcagacaaactgcgcaactcgtgaaaggtaggcggatccccttcgaaggaaagacctgatgcttttcgtgcgcgcataaaatacctttgatactgtgccgga
tgaaagcggttcgcgacgagtagatgcaattatggtttctccgccaagaatctctttgcatttatcaagtgtttccttcattgatattccgagagcatcaatatgcaatgctgttgggatggcaatttttacgcc
tgttttgctttgctcgacataaagatatccatctacgatatcagaccacttcatttcgcataaatcaccaactcgttgcccggtaacaacagccagttccattgcaagtctgagccaacatggtgatgattctg
ctgcttgataaattttcaggtattcgtcagccgtaagtcttgatctccttacctctgattttgctgcgcgagtggcagcgacatggtttgttgttatatggccttcagctattgcctctcggaatgcatcgctcagt
gttgatctgattaacttggctgacgccgccttgccctcgtctatgtatccattgagcattgccgcaatttctttttgtggtgatgtcttcaagtggagcatcaggcagacccctccttattgctttaattttgctcat
gtaatttatgagtgtcttctgcttgattcctctgctggccaggatttttttcgtagcgatcaagccatgaatgtaacgtaacggaattatcactgttgattctcgctgtcagaggcttgtgtttgtgtcctgaaaat
aactcaatgttggcctgtatagcttcagtgattgcgattcgcctgtctctgcctaatccaaactctttacccgtccttgggtccctgtagcagtaatatccattgtttcttatataaaggttaggggggtaaatccc
ggcgctcatgacttcgccttcttcccatttctgatcctcttcaaaaggccacctgttactggtcgatttaagtcaacctttaccgctgattcgtggaacagatactctcttccatccttaaccggaggtgggaata
tcctgcattcccgaacccatcgacgacgcgtatacaacctccttagtacatgcaaccattatcaccgccagaggtaaaatagtcaacacgcacggtgttagatatttatccctgcggtgatagatttaacgt
atgagcacaaaaaagaaaccattaacacaagagcagcttgaggacgcacgtcgccttaaagcaatttatgaaaaaaagaaaaatgaacttggcttatcccaggaatctgtcgcagacaagatggggat
ggggcagtcaggcgttggtgctttatttaatggcatcaatgcattaaatgcttataacgccgcattgcttacaaaaattctcaaagttagcgttgaagaatttagcccttcaatcgccagagaaatctacgag
atgtatgaagcggttagtatgcagccgtcacttagaagtgagtatgagtaccctgttttttctcatgttcaggcagggatgttctcacctaagcttagaaccttttaccaaaggtgatgcggagagatgggtaa
gcacaaccaaaaaagccagtgattctgcattctggcttgaggttgaaggttaattccatgaccgcaccaacaggctccaagccaaagctttcctgacggaatgttaattctcgttgaccctgagcaggctgttg
agccaggtgatttctgcatagccagacttgggggtgatgagtttaccttcaagaaactgatcagggatagcggtcaggtgtttttacaaccactaaacccacagtacccaatgatcccatgcaatgagagtt
gttccgttgtggggaaagttatcgctagtcagtggcctgaagagacgtttggctgactagggaagttcctatactttctagagaataggaacttcggacgtctaagaaaccattattatcatgagggcccac
cggttaattaacggcaccaccgaaattcagtaagcagaaagtcaaaagcctccgaccggaggcttttgactattactcaacaggtaaggcgcgaggttttttcaattgttcgaaacgcaaaaaggccatcc
gtcaggatggccttctgcttaatttgatgcctgtcagtttaggttaggcgccatgcatctcgaggcatgcctgcattatttgtacagctcatccatgccaccggtagaatgacgaccctccgcgcgctcatatt
gctctacgatcgtataatcttcattatgagaggtgatgtccagtttaatattcacattgtacgcgccaggcagctgcacaggtttcttggctttgtacgtggttttcacttcagcgtcataatggccgccatcttt
cagtttcaggcgctgtttaatttcgcctttcagcgcaccatcttccggatacatgcgttcgctagacgcctcccaacccatcgtcttcttctgcatcaccgggccatcagatggaaaattagtaccacgcagttt
aactttatagatgaactcgccatcctgcagggaggagtcctgagtgacggtcacgacaccaccatcttcaaaattcattacgcgttcccacttgaaaccttccggaaaagacagcttcagatagtccgggat
atccgctgggtgtttaacataagctttagaaccgtacataaattgcgggctcaggatgtcccacgcaaaaggcagcgggccgcctttagtcactttcagtttggcggtctgggtgccttcatacggacggccc
tcgccttcgccttcgatctcgaactcgtgaccgttaacagaaccctccatgtgaactttgaagcgcatgaactctttaatgatagccatgttatcctcctcgcccttggaaaccatctagatttcccctctttaa
aggaattccggaaccattataggtatccatgaggtacctgtcaaggcctcgagtcgacagttcataggccgatccccaattcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttc
gcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggggaattaattcc
tggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcc
cagtctttcgactgagcctttcgttttatttgatgcctggggaattaattcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcc

tactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggggaattgggaattaattcttgaagacgaaagggcctcgtgatacg
cctatttttataggttaatgtcatgtgcagcatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgct
tacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagccggatcaattctcatgtttgacagcttatcactgatcagtgaattaatggcga
tgacgcatcctcacgataatatccgggtaggcgcaatcactttcgtctctactccgttacaaagcgaggctgggtatttcccggccttttctgttatccgaaatccactgaaagcacagcggctggctgaggag
ataaataataaacgaggggctgtatgcacaaagcatcttctgttgagttaagaacgagtatcgagatggcacatagccttgctcaaattggaatcaggtttgtgccaataccagtagaaacagacgaaga
agctagctttgcactggattgcgaggctttgtgcttctctggagtgcgacaggtttgatgacaaaaaattagcgcaagaagacaaaaatcaccttgcgctaatgctctgttacaggtcactaataccatcta
agtagttgattcatagtgactgcatatatgttgtgttttacagtattatgtagtctgttttttatgcaaaatctaatttaatatattgatatttatatcattttacgtttctcgttcagctttttatactaagttgagc
gaaacgggaaggtaaaaagacaaaaagttgtttttaatacctttaagtgataccagatggcattgcgccatctggcagagtgattaactaaacatcgcagtaatcgaggcgcttgccagagagtggaaat
gaacgttaaacccgaccatcgcgccgctggcaccttcatcgacatcaatacgttctatatccagcgcgtgaacggtaaaaatgtagcgatgagtttcgcctttcggcggtgctgcgccatcgtaccggtttt
accaaagtcggtacgcgtctgcaaaacgccgtctggcattgctaccagaccagagccaaaccccttgcggtaatacgcgggtatcagcgggtaagttaacaactacccagtgccaccagccggagccggtt
ggcgcatccgggtcgtagcaggtgacaacaaaactttcgttcccgcaggaacatcatcccacgccagatgcggtgaaatattatcgccatcgtaacccatgccgttaaagacatgacgatgcggcaattt
atcgccatcgcgcagatcgttactgatgagtttcatgaaccctcctttcttgtttgcagaaagtgtagccagaaaccctcacgcggacttctcgttattggcaaaaaaatgtttcatcctgtaccgcgcggtta
accgctgcggtcagacgctgcaactgttgcgggagaataatatagggcggcatcaggtaaatcagtttgccaaaaggccggatccagacaccctgttcgacaaagaatttttgcagcgccgccatattcac
cggatgagtggtttcgaccacgccaatggccccagtacgcgcacatcggcaaccatttcggcatcacgggcgggggcaagttgctcgcgcagctgtacttcaatatccgccacctgttgctgccagtcgcc
agattcgagaatcgccaggctggcgtttgctgccgcgcaggccagcggattgcccataaaagttggcccatgaaaataaaaaaacgagattatttttccaagcccccgagaacatacacaaatttagccc
gaacggattccctatattttata

Sequence of the *E. coli* K12 chromosomal *attB* site with inserted *mcherry* gene

> [attb site (+ 100 nt downstream+upstream) + POSIP-Kl-*mcherry*- inserted and flipped - 4790 bp] - Selection From [1] To [4790]

tctggtctggtagcaatgattttataggcatcacggcaatatacattttatacgtatgaaggcacccagaccgggaattaattcttgaagacgctactatttctacttcagtagatttgtcgtagccggtgcaa
acacatacgcttcttgctgagtacgtgagttcaccacgcggaattcggtgttatcgaacaccactgcgccgcgaccagaaacgatatccacatccccttcaatgtagctgttggtcaccagcgtacgcggct
gacgattcgtttccagacggttctgcacaccgctgttggtgacaaagaaggtgttctgacgaccgagaatgttaacgttgttaatctgtacctggtcaccatcagtacgcagtgccaccgccggatggttacc
tgcatctacgctatcgcccagcgtgttttcgatggtcagattttgcagttgcaggccattgttttgtgaccagaagaccgcagagcagagaacaccgatactgtcgctgcgtttgctctggcagctatcgtaca
tataccacgctggtttacctggcatatatttgccgcgcgggttgacgtcgtgacgccagtcggcagggctcatgccaccatcaagggaaagcccaatcttcacatcaatcggtttttcacctgtaccgtacag
agtaattccacccggagcggcagggacatataccgttccctgatactcaccaggcatcacggcaatatactggcgcttgttggtacgcttgataattgccgcatctaccgccgcctgaatcgtggtatgcgtt
acaccttgagtgcccgccgggccgacaacaaagtcaggttgcgcaggcagggtaatcggggaaggattccacgctgcagcacctggtgtcagggatgcaaaatagtgttgagcatcgaaattctgcgctt
cttttgccgacagaatcgggcgagaagaggtaccaggcgcggtttgatcagaaggacgttgatcgggcggggttgagctacaggcggtcagcgtcacgccaaaagccaatgccagcgccagacgggaa
actgaaaatgtgttcacaggttgctccgggctatgaaatagaaaaatgaatccgttgaagcctgcttttttatactaacttggcattatataaaaaagcattgcttatcaatttgttgcaacgaacaggtcactat
cagtcaaaataaaatcattatttgatttcaattttgtcccactccctgccctctgtcatcacgatactgtgatgccatggccatatggaagttcctatactttctagagaataggaacttcggacgtctaagaaacc
attattatcatgagggcccaccggttaattaacggcaccaccgaaattcagtaagcagaaagtcaaaagcctccgaccggaggcttttgactattactcaacaggtaaggcgcgaggttttttcaattgttc
gaaacgcaaaaaggccatccgtcaggatggccttctgcttaatttgatgcctgtcagtttaggttaggcgccatgcatctcgaggcatgcctgcattatttgtacagctcatccatgccaccggtagaatgac
gaccctccgcgcgctcatattgctctacgatcgtataatcttcattatgagaggtgatgtccagtttaatattcacattgtacgcgccaggcagctgcacaggtttcttggctttgtacgtggttttcacttcagc
gtcataatggccgccatctttcagtttcaggcgctgtttaatttcgcctttcagcgcaccatcttccggatacatgcgttcgctagacgcctcccaacccatcgtcttcttctgcatcaccgggccatcagatgg
aaaattagtaccacgcagtttaactttatagatgaactcgccatcctgcagggaggagtcctgagtgacggtcacgacaccaccatcttcaaaattcattacgcgttcccacttgaaaccttccggaaaaga
cagcttcagatagtccgggatatccgctgggtgtttaacataagctttagaaccgtacataaattgcgggctcaggatgtcccacgcaaaaggcagcgggccgccttttagtcactttcagtttggcggtctg
ggtgccttcatacggacggccctcgccttcgccttcgatctcgaactcgtgaccgttaacagaaccctccatgtgaacttgaagcgcatgaactctttaatgatagccatgttatcctcctcgccccttggaaa
ccatctagatttcccctctttaaaggaattccggaaccattataggtatccatgaggtacctgtcaaggcctcgagtcgacagttcataggccgatccccaattcctggcagtttatggcgggcgtcctgccc
gccaccctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttat
ttgatgcctggggaattaattcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaa
caacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggggaattaattcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttcgcaacgttcaa
atccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggggaattgggaattaattcttgaa
gacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgtgcagcatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgg
gcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagccggatcaattctcatgtttgacagcttatc
actgatcagtgaattaatggcgatgacgcatcctcacgataatatccgggtaggcgcaatcactttcgtctctactccgttacaaagcgaggctgggtatttcccggccttttctgttatccgaaatccactgaa
agcacagcggctggctgaggagataaataataaacgaggggctgtatgcacaaagcatcttctgttgagttaagaacgagtatcgagatggcacatagccttgctcaaattggaatcaggtttgtgccaa
taccagtagaaacagacgaagaagctagctttgcactggattgcgaggctttgtgcttctctggagtgcgacaggtttgatgacaaaaaattagcgcaagaagacaaaaatcaccttgcgctaatgctctg
ttacaggtcactaataccatctaagtagttgattcatagtgactgcatatatgttgtgttttacagtattatgtagtctgttttttatgcaaaatctaatttaatatattgatatttatatcatttacgtttctcgttc
agctttttatactaagttgagcgaaacgggaaggtaaaaagacaaaaagttgtttttaatacctttaagtgataccagatggcattgcgccatctggcagagtgattaactaaacatcgcagtaatcgaggg
cgcttgccagagagtggaaatgaacgttaaacccgaccatcgcgccgctggcaccttcatcgacatcaatacgttctatatccagcgcgtgaacggtaaaaatgtagcgatgagtttcgcctttcggcggtg
ctgcgccatcgtacccggttttaccaaagtcggtacgcgtctgcaaaacgccgtctggcattgctaccagaccagagccaaaccccttgcggtaatacgcgggtatcagcgggtaagttaacaactacccag
tgccaccagccggagccggttggcgcatccgggtcgtagcaggtgacaacaaaactttcgttcccgcaggaacatcatcccacgccagatgcggtgaaatattatcgccatcgtaacccatgccgttaaa
gacatgacgatgcggcaatttatcgccatcgcgcagatcgttactgatgagtttcatgaaccctccttttcttgtttgcagaaagtgtagccagaaaccctcacgcggacttctcgttattggcaaaaaaatgtt
tcatcctgtaccgcgcggttaaccgctgcggtcagacgctgcaactgttgcgggagaataatatagggcggcatcaggtaaatcagtttgccaaaaggccggatccagacaccctgttcgacaaagaatttt
ttgcagcgccgccatatttcaccggatgagtggtttcgaccacgccaatggcccccagtacgcgcacatcggcaaccatttcggcatcacgggcgggggcaagttgctcgcgcagctgtacttcaatatccg
ccacctgttgctgccagtcgccagattcgagaatcgccaggctggcgtttgctgccgcgcaggccagcggattgcccataaaagttggcccatgaaaaacgaaagtaaaatcctattcgcaaaattcctac
tatattttata

Sequence alignment of the *fimH* inserts on pPKL241 and pFimH-His$_6$-NT-trc

```
pFimH-His6-NT-trc      G  M  I  V  fM  K  R  V  I  T  L  F  A  V  L
               310     gggatgattgtaatgaaacgagttattaccctgtttgctgtactg   354
                       ||###|||||||||||||||#|||||||||||||||||||||||
                98     gga--gattgtaatgaaacgtgttattaccctgtttgctgtactg   140
pPKL241                G  -  I  V  fM  K  R  V  I  T  L  F  A  V  L

          L  M  G  W  S  V  N  A  W  S  F  A  C  K  T  A  N  G  T  A
 355      ctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaaccgccaatggtaccgca   414
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
 141      ctgatgggctggtcggtaaatgcctggtcattcgcctgtaaaaccgccaatggtaccgct   200
          L  M  G  W  S  V  N  A  W  S  F  A  C  K  T  A  N  G  T  A

          I  P  I  G  G  G  S  A  N  V  Y  V  N  L  A  P  V  V  N  V
 415      attccgatcggtggcggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtg   474
          ||#||#||#||#||#|||||||||||||||||||||||||||||||||||||||||||
 201      atccctattggcggtggcagcgccaatgtttatgtaaaccttgcgcccgtcgtgaatgtg   260
          I  P  I  G  G  G  S  A  N  V  Y  V  N  L  A  P  V  V  N  V

          G  Q  N  L  V  V  D  L  S  T  Q  I  F  C  H  N  D  Y  P  E
 475      gggcaaaacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaa   534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 261      gggcaaaacctggtcgtggatctttcgacgcaaatcttttgccataacgattatccggaa   320
          G  Q  N  L  V  V  D  L  S  T  Q  I  F  C  H  N  D  Y  P  E

          T  I  T  D  Y  V  T  L  Q  R  G  S  A  Y  G  G  V  L  S  N
 535      accattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaat   594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 321      accattacagactatgtcacactgcaacgaggctcggcttatggcggcgtgttatctaat   380
          T  I  T  D  Y  V  T  L  Q  R  G  S  A  Y  G  G  V  L  S  N

          F  S  G  T  V  K  Y  S  G  S  S  Y  P  F  P  T  T  S  E  T
 595      ttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacg   654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 381      ttttccgggaccgtaaaatatagtggcagtagctatccatttcctaccaccagcgaaacg   440
          F  S  G  T  V  K  Y  S  G  S  S  Y  P  F  P  T  T  S  E  T

          P  R  V  V  Y  N  S  R  T  D  K  P  W  P  V  A  L  Y  L  T
 655      ccgcgcgttgtttataattcgagaacggataagccgtggccggtggcgctttatttgacg   714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 441      ccgcgcgttgtttataattcgagaacggataagccgtggccggtggcgctttatttgacg   500
          P  R  V  V  Y  N  S  R  T  D  K  P  W  P  V  A  L  Y  L  T

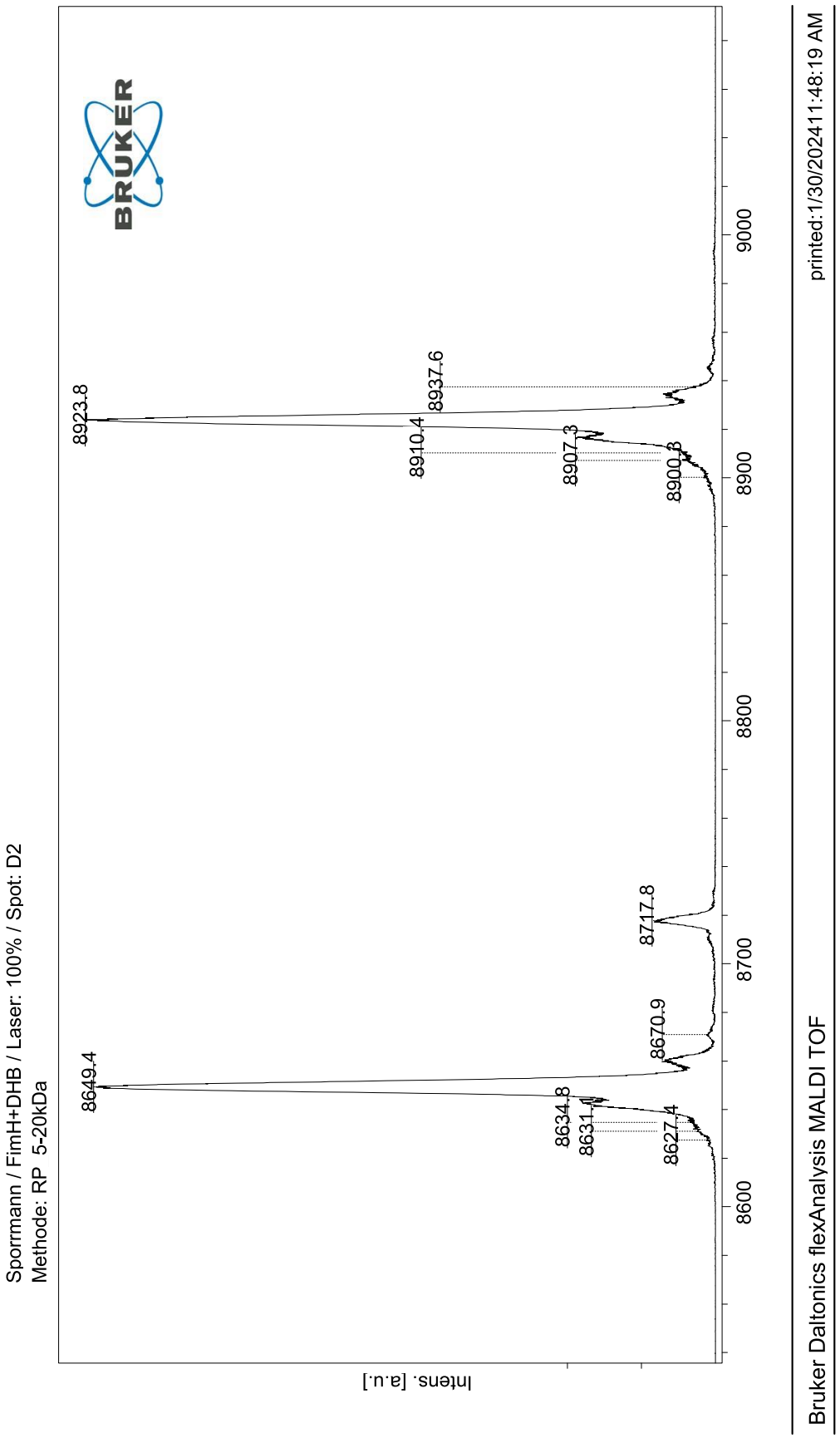
          P  V  S  S  A  G  G  V  A  I  K  A  G  S  L  I  A  V  L  I
 715      cctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttatt   774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 501      cctgtgagcagtgcgggcggggtggcgattaaagctggctcattaattgccgtgcttatt   560
          P  V  S  S  A  G  G  V  A  I  K  A  G  S  L  I  A  V  L  I

          L  R  Q  T  N  N  Y  N  S  D  D  F  Q  F  V  W  N  I  Y  A
 775      ttgcgacagaccaacaactataacagcgatgatttccagtttgtgtggaatatttacgcc   834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 561      ttgcgacagaccaacaactataacagcgatgatttccagtttgtgtggaatatttacgcc   620
          L  R  Q  T  N  N  Y  N  S  D  D  F  Q  F  V  W  N  I  Y  A

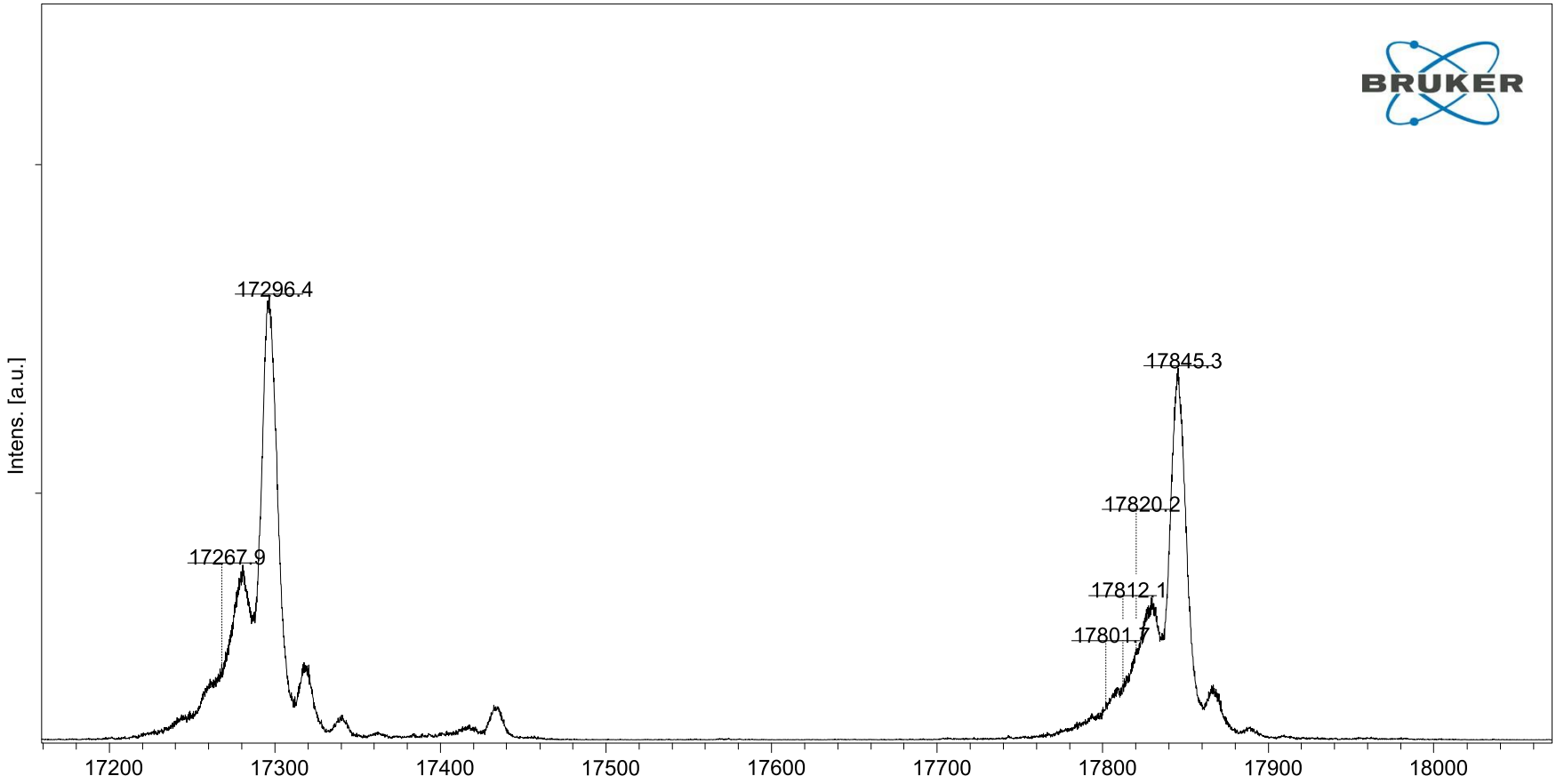
          N  N  D  V  V  V  P  T  G  G  H  H  H  H  H  H  *
 835      aataatgatgtggtggtgcctactggcggccatcaccatcaccatcactaa            886
          ||||||||||||||||||||||||||||||||||||||||||||||||||
 621      aataatgatgtggtggtgcctactggcggccatcaccatcaccatcactaa            672
          N  N  D  V  V  V  P  T  G  G  H  H  H  H  H  H  *
```

MALDI mass spectra of FimH variants

MALDI mass spectra from the MALDI-TOF analysis of $FimH_{trc}$



Sporrmann / FimH+DHB / Laser: 100% / Spot: D2
Methode: RP_5-20kDa

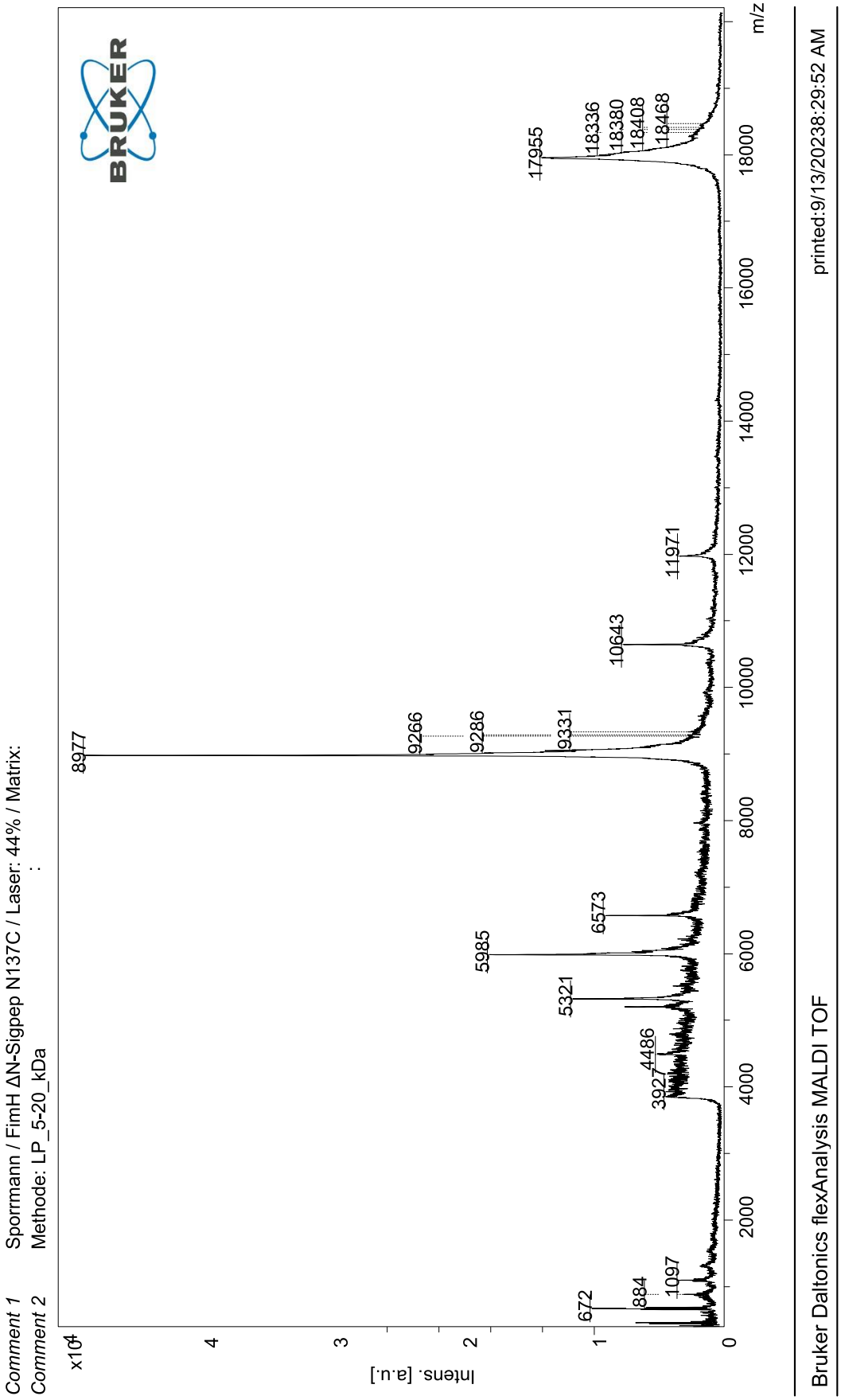


Bruker Daltonics flexAnalysis MALDI TOF printed:1/30/202411:49:18 AM

m/z list from the MALDI-TOF analysis of $FimH_{trc}$

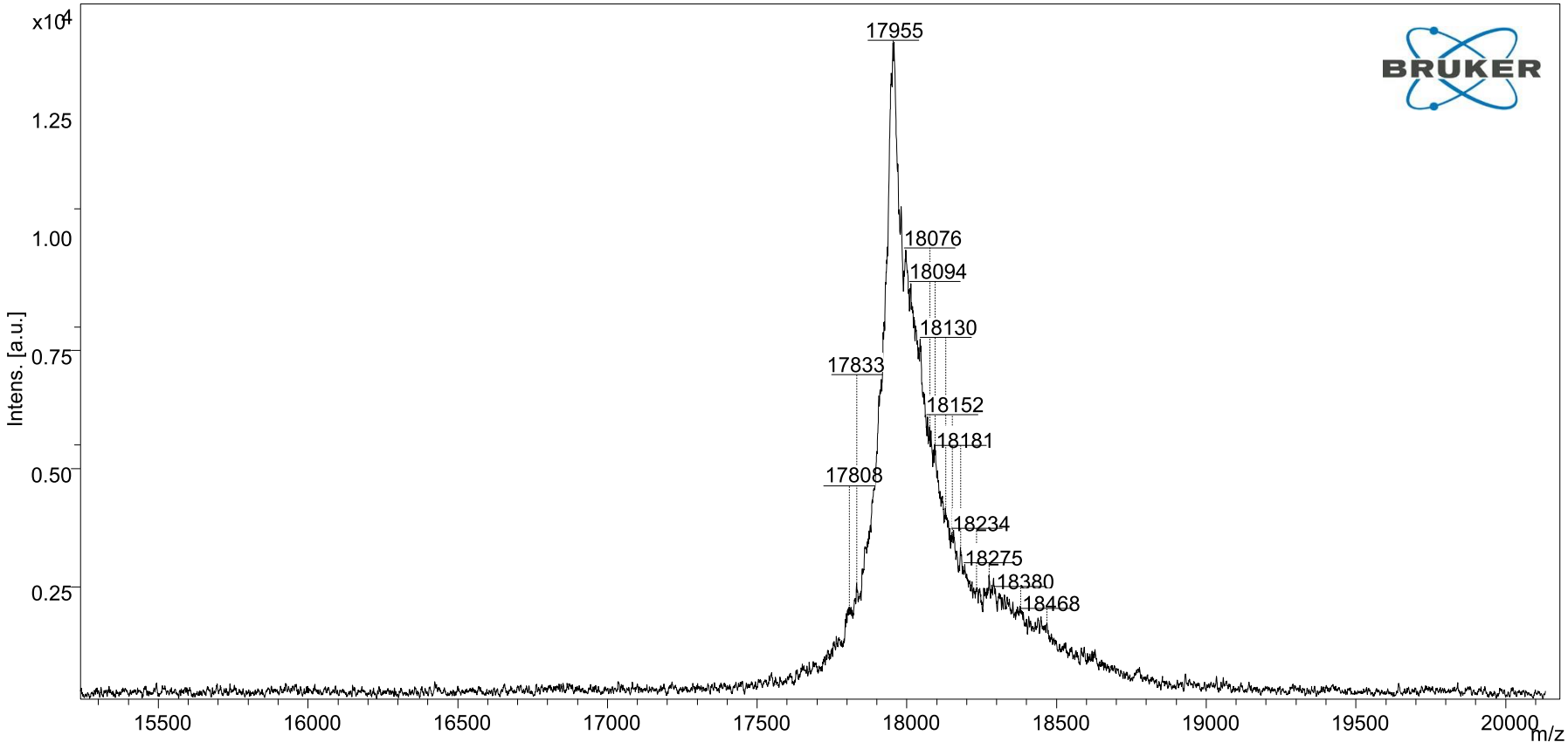
m/z	S/N	Res.	Intens.
5766.4	5	2870	41
5948.8	7	1042	58
5949.7	7	7390	55
8625.2	4	20201	41
8627.4	5	46647	49
8628.2	5	15050	47
8631.1	11	10857	101
8632.5	12	16918	111
8634.8	15	9190	140
8642.6	76	2623	717
8649.4	356	1791	3360
8659.0	25	6758	236
8660.4	29	3540	275
8670.9	5	6020	49
8713.4	5	10007	47
8717.8	33	2188	316
8717.8	35	2209	326
8900.3	5	15984	65
8901.2	5	16524	58
8906.2	11	13835	136
8907.3	15	9971	178
8908.3	13	9525	159
8910.4	15	6458	182
8916.7	61	3186	748
8923.8	278	1850	3405
8933.2	22	19003	268
8934.6	23	9838	284
8937.6	11	6889	129
17267.9	8	78830	367
17280.6	19	9784	851
17296.4	48	3041	2158
17317.6	8	6041	377
17801.7	4	23080	181
17809.0	6	13273	255
17812.1	6	24543	266
17820.2	10	14415	442
17829.7	16	12304	689
17832.0	15	12251	646
17835.8	12	19816	519
17845.3	43	1844	1809
17863.2	5	9877	199
17866.8	6	13136	267

MALDI mass spectra from the MALDI-TOF analysis of $FimH_L$ N136C ΔN-sigpep



XLI

Comment 1 Sporrmann / FimH ΔN-Sigpep N137C / Laser: 44% / Matrix:
Comment 2 Methode: LP 5-20 kDa :

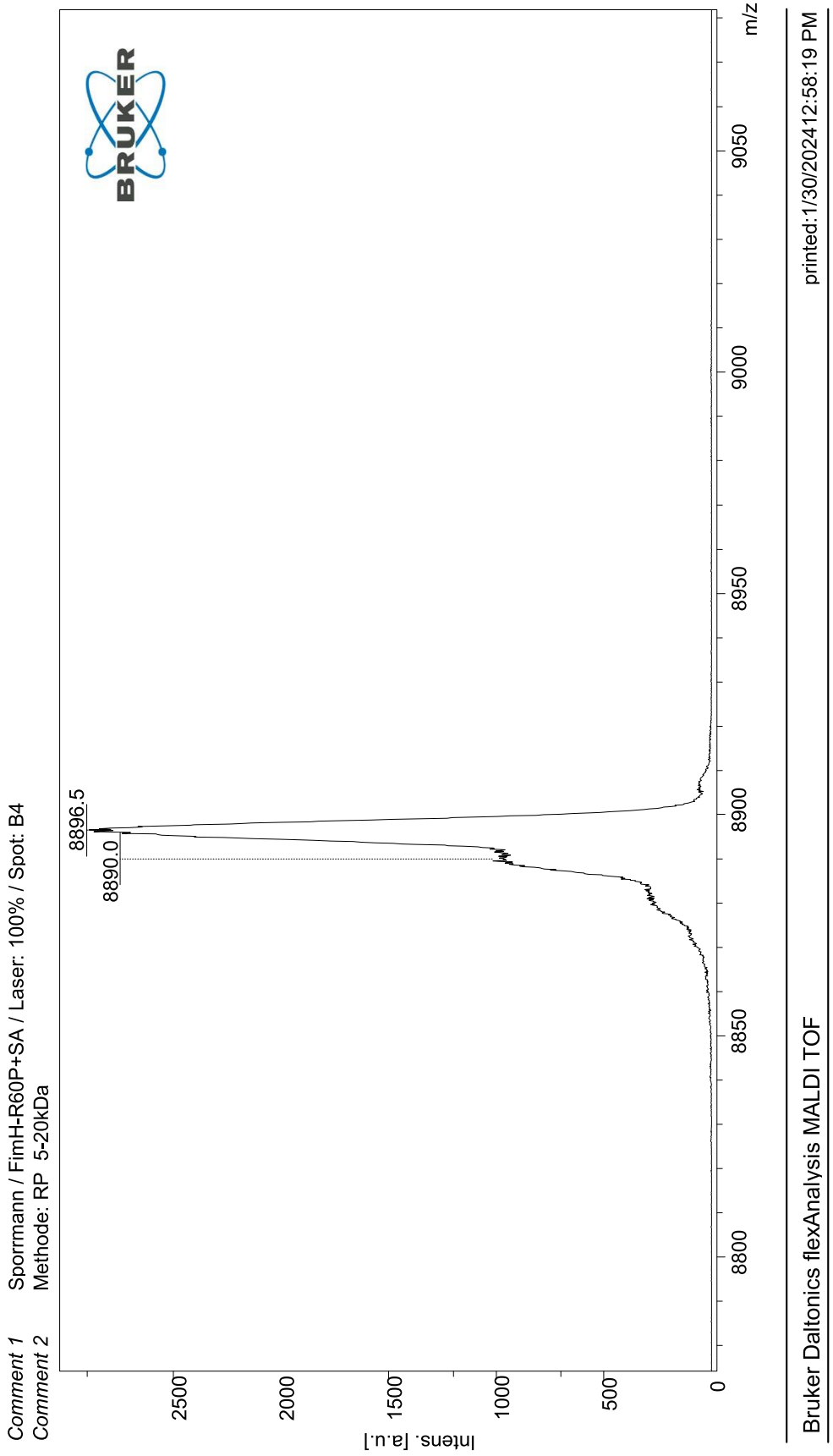


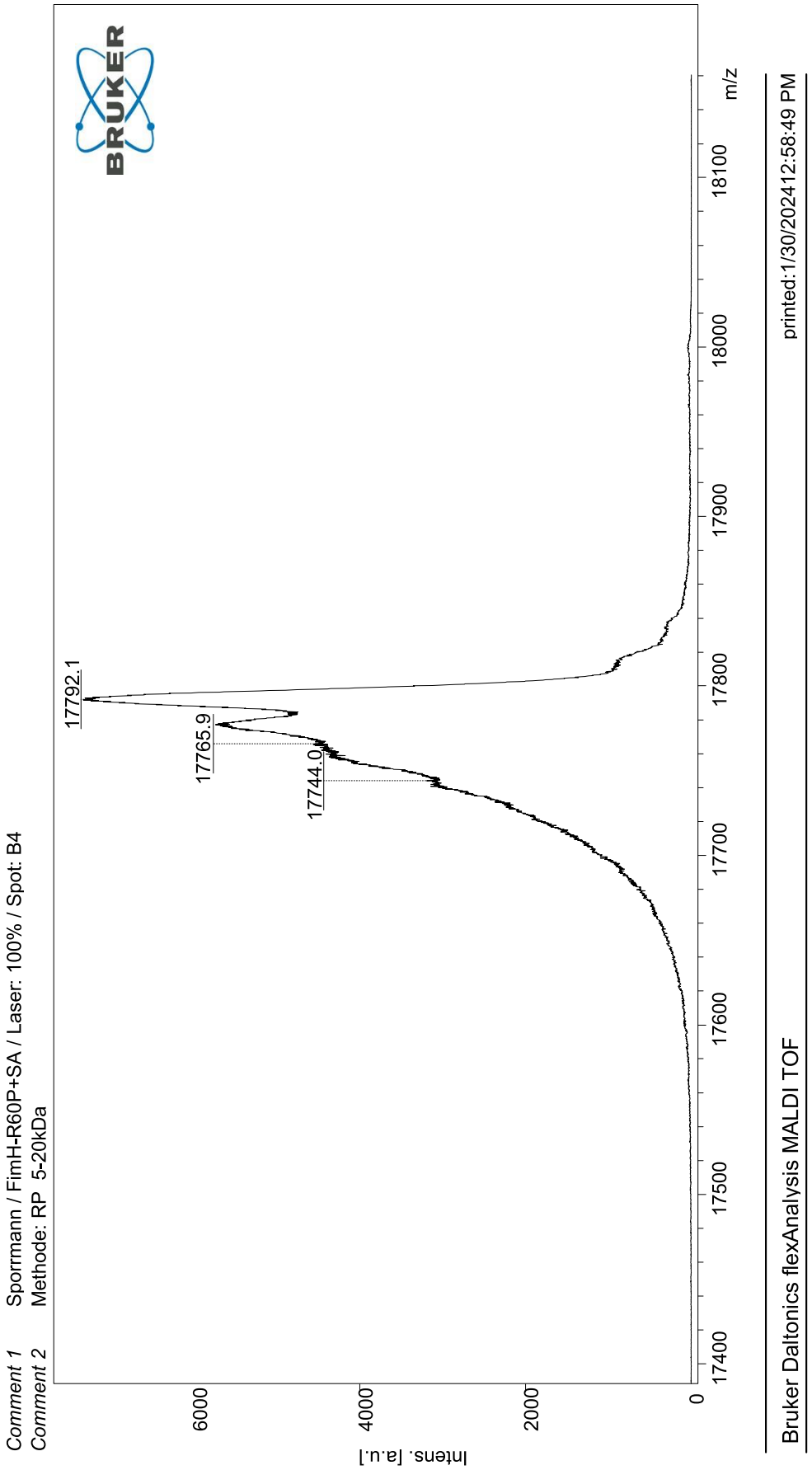
Bruker Daltonics flexAnalysis MALDI TOF printed:9/13/20238:30:19 AM

m/z list from the MALDI-TOF analysis of $FimH_L$ N136C ΔN-sigpep

m/z	S/N	Res.	Intens.
9266	3	632	2635
9286	4	2462	2686
9331	4	1458	2782
10643	22	1183	7761
10663	7	1079	3019
10685	5	1354	2453
10694	5	3253	2342
10705	5	967	2397
11971	10	693	3420
17808	10	727	2025
17833	14	1930	2570
17955	83	396	14010
17997	56	781	9624
18045	45	1584	7739
18076	34	2515	5885
18094	31	1149	5512
18130	23	1497	4165
18152	20	1322	3573
18181	18	1096	3244
18234	13	2518	2463
18244	13	2325	2438
18261	13	2239	2462
18275	15	5771	2737
18290	14	1430	2685
18309	12	1199	2361
18336	12	1662	2283
18380	10	538	2047
18408	9	2363	1875
18439	9	1487	1842
18449	9	4144	1865
18468	9	1749	1714

MALDI mass spectra from the MALDI-TOF analysis of $FimH_{trc}$ R60P




Comment 1 Sporrmann / FimH-R60P+SA / Laser: 100% / Spot: B4
Comment 2 Methode: RP 5-20kDa
17792.1
17765.9
17744.0
Intens. [a.u.]
6000
4000
2000
0
17400
17500
17600
17700
17800
17900
18000
18100
m/z
Bruker Daltonics flexAnalysis MALDI TOF
printed:1/30/202412:58:49 PM

m/z list from the MALDI-TOF analysis of $FimH_{trc}$ R60P

m/z	S/N	Res.	Intens.
8890.0	44	2541	1015
8896.5	125	1921	2892
17741.6	81	1165	3148
17744.0	82	22172	3197
17758.0	113	2397	4364
17765.9	118	4321	4549
17777.0	150	2368	5746
17792.1	191	1703	7297