

Institut für Experimentelle Medizin –
AG Systematische Proteomics und Bioanalytik
Prof. Dr. Andreas Tholey
an der Christian-Albrechts-Universität zu Kiel

**Methodische Ansätze zur Analyse posttranslationaler
Modifikationen mittels *Top-Down* und *Bottom-Up Proteomics***

Dissertation
zur
Erlangung der Doktorwürde
der Mathematisch-naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Alexander Becker
aus Kiel

Kiel 2020

Erster Gutachter:	Prof. Dr. Andreas Tholey
Zweite Gutachterin:	Prof. Dr. Thisbe Lindhorst
Tag der mündlichen Prüfung:	06.10.2020
Zum Druck genehmigt am:	06.10.2020

Danksagung

Der höchste Dank gebührt meinen Eltern, deren kompromissloser Unterstützung ich mir stets sicher sein konnte. Ihr habt es überhaupt erst ermöglicht, dass ich den akademischen Weg einzuschlagen und bis zum Ende zu beschreiten vermochte. Danke für all die Schubser in die richtige Richtung und wegweisenden Ratschläge!

Die Liste der Dinge, für die ich mich ebenso bei meinem Bruder bedanken möchte, ist lang. Die kitschige Zusammenfassung ist: Ohne dich wäre ich nicht der Mensch, der ich heute bin. Danke, Bruderherz!

Ich möchte weiterhin meinem Ex-Komilitonen und bestem Freund Max danken. Die überaus kräftezehrenden Lernphasen während des Studiums hätte ich ohne dich bestimmt nicht überstanden.

Als nächstes möchte ich meinem Doktorvater und Mentor Andreas Tholey für die Zeit in seinem Arbeitskreis danken. Dazu zählen unzählige, konstruktive, fachliche sowie nicht-fachliche Diskussionen, die mich stets auf das Neue motiviert haben.

Für die gleiche Unterstützung möchte ich mich bei meinem Arbeitskreis bedanken, vor allem aber Dennis Linke, der mir in den Anfängen immer mit Rat und Tat zur Seite stand. In diesem Sinne möchte ich außerdem Tomas Koudelka, Joanna Tucher, Prasath Sumasundaram, Liam Cassidy, Christian Treitz, Chien-Wen Hung, Mohamed Mokhtar, Britta Steer und Jana Schlenk danken. Ihr wart mir immer eine große Hilfe.

Ich möchte mich ebenso bei Dr. Ulrich Gerstel und Prof. Dr. Jens-Michael Schröder für die gemeinsamen Projekte bedanken. Die zugehörigen Diskussionen waren überaus lehrreich. Ich möchte euch ebenfalls für das Korrekturlesen meiner Arbeit danken!

Zu guter letzt möchte ich mich bei Prof. Dr. Christoph Becker-Pauly für die Zusammenarbeit im SFB877 und die Hilfe im gemeinsamen Projekt bedanken.

Inhalt

Danksagung	4
1. Zusammenfassung/Abstract	9
a. Zusammenfassung.....	9
b. Abstract	12
2. Abkürzungen	14
3. Ein- und Dreibuchstaben-Code der proteinogenen Aminosäuren	16
4. Einleitung.....	17
a. Proteomik.....	17
b. Massenspektrometrie	18
c. Von Fragmentspektren zur Primärstruktur von Proteinen und Peptiden	19
d. Enzymatische Hydrolyse als regulatorischer Mechanismus	22
e. <i>Bottom-Up</i> vs. <i>Top-Down</i> Proteomics.....	23
5. Grundlagen.....	30
a. Flüssigkeitschromatographie	30
b. Prinzipien der Ionisation für die MS.....	34
c. Massenspektrometrie	35
d. Datenbanksuchen.....	38
6. Entwicklung einer Methode zur <i>Top-Down</i> -Analyse von N-glykosylierten Proteinen und Peptiden	43
a. Glykoproteine und –peptide in der Massenspektrometrie	43
b. Versuchsführung und Zielsetzung	45
c. Material und Methoden.....	46
i. Chemikalien und Zubehör	46
ii. Expression und Aufreinigung von Meprin- α - und – β und Progranulin	46
iii. Progranulin-Verdau mit Meprin- α - und – β	46
iv. Zeitabhängiger Verdau von Progranulin mit Meprin- α - und – β	48

v.	Größenausschluss mittels Amicon Zentrifugalfilter	49
vi.	Lectin-Festphasenextraktion	49
vii.	LC-ESI MS (Ultimate3000, Dionex/Q Exactive Plus, Thermo Scientific).....	49
viii.	Identifizierung von Peptiden und Spaltstellen mittels Datenbanksuche	50
d.	Resultate und Diskussion	51
e.	Zusammenfassung.....	76
7.	Anwendung von <i>Bottom-Up</i> Methoden zur Sequenzierung von antimikrobiellen Peptiden des <i>stratum corneum</i>	79
a.	Antimikrobielle Peptide der Haut	79
b.	Durchführung und Zielsetzung.....	80
c.	Material und Methoden	81
i.	Chemikalien und Zubehör	81
ii.	Enzymatischer Verdau der gegen <i>S. aureus</i> aktiven Faktoren	82
iii.	Oxidation mit Dimethylsulfoxid	82
iv.	Festphasen-Extraktion mit OASIS Kartuschen	83
v.	Verdau von synthetischen Peptiden mit Cathepsin D.....	83
vi.	Aktivitäts-Assays	84
vii.	Analyse der nativen Hornerin-Peptide	84
viii.	MALDI MS (TOF/TOF 5800, AB Sciex)	84
ix.	LC-ESI MS (Ultimate3000, Dionex/Q Exactive Plus, Thermo Scientific).....	84
x.	Identifizierung der Desmoplakin- und Hornerin-AMP mittels Datenbanksuchen	86
d.	Resultate und Diskussionen	86
i.	Desmoplakin	86
ii.	Hornerin.....	95
e.	Zusammenfassung.....	101
i.	Desmoplakin	101

ii. Hornerin.....	103
8. Weitere Projekte	105
a. Phosphorylierung von Peptiden durch eine Michael-Addition zur Generierung von Phosphocystein-Analoga	105
b. Disulfid-Mapping des <i>bone morphogenetic protein 1</i> (BMP1)	107
c. <i>Top-Down Proteomics</i> : Vorversuche	111
9. Referenzen	116
10. Anhang.....	129
a. Chromatogramme und Massenspektren der Progranulin-Versuche.....	129
b. Hornerin-Peptide.....	144
11. Erklärung zum Eigenanteil der Dissertation und Publikationen	152

1. Zusammenfassung/Abstract

a. Zusammenfassung

Die nachfolgende Arbeit lässt sich in zwei größere Abschnitte einteilen: die Entwicklung einer Methode zur Analyse glykosylierter Proteine und Hydrolyseprodukte ohne die Verwendung von Arbeitsproteasen und die Sequenzierung von antimikrobiellen Peptiden aus biologischen Proben.

Glykosylierte Proteine und Peptide stellen in der Analytik eine Herausforderung dar, da sie sich vor allem in der Chromatographie anders verhalten als ihre nicht-glykosylierten Analoga. Oligosaccharide binden sehr viel Wasser, was in der Umkehrphasen-Flüssigkeitschromatographie eine Bandenverbreiterung zur Folge hat. Weiterhin werden für glykosylierte Peptide und Proteine in der Tandem-Massenspektrometrie primär Glykanfragmente beobachtet. Dies erschwert die Bestimmung der Aminosäuresequenz. Die vorgestellte Methode für die Analyse von glykosylierten Proteinen wurde mit einem Meprin- α/β -Verdau des Glykoproteins Progranulin durchgeführt. Die Arbeitsvorschrift sieht die Verwendung von Festphasenextraktionskartuschen vor, die mit immobilisierten Lektinen modifiziert worden sind. Das gewählte Adsorbens ist spezifisch für das Monosaccharid *N*-Acetylglucosamin, welches in der Kernstruktur von *N*-Glykanen in den meisten Organismen vorkommt. Der Verdau wurde nach der enzymatisch katalysierten Deglykosylierung mit PNGase F über das Adsorbens geführt und das Eluat mit LC-ESI MS auf das Vorhandensein von Hydrolyse-Produkten und deren Termini untersucht. Um die Bedeutung der Methode einordnen zu können, wurden die gleichen Proben auch mit einem Standardworkflow der *Bottom-Up Proteomics* untersucht. Als *Readout* wurden die Anzahl der identifizierten Peptide und Spaltstellen sowie die Sequenzabdeckung herangezogen. Im Vergleich der vorgestellten *Top-Down* Methode mit einem klassischen *Bottom-Up* Ansatz wurden für beide Enzyme in etwa die gleiche Anzahl an Peptiden identifiziert. Hinsichtlich der Spaltstellen (bis zu 41 % mehr Identifizierungen als in der *Bottom-Up* Methode) und Sequenzabdeckung (eine bis zu 16 % höhere Abdeckung der Primärstruktur) war der Lektin-SPE-Ansatz der *Bottom-Up* Methode überlegen. Auch *in puncto* Zeitaufwand ist der *Top-Down* Ansatz die bessere Wahl. Wenn Hydrolyseprodukte und die zugehörigen Spaltstellen analysiert werden sollen, so erfordert die Verwendung von Arbeitsproteasen (*Bottom-Up Proteomics*) das vorherige Markieren der ursprünglichen Termini. Dieses Vorgehen erhöht die Anzahl der benötigten

Arbeitsschritte und die damit verbundene Arbeitszeit um mehr als sieben Stunden. Die verwendeten Lektin-basierten Festphasenextraktionskartuschen eignen sich nicht nur für die Abtrennung von glykosylierten Analyten, sondern ebenfalls für deren Anreicherung, z.B. in Experimenten, in denen Glykanstrukturen aufgeklärt werden sollen. Aufgrund der Spezifität der eingesetzten Lektine für *N*-Acetylglucosamin eignet sich der Workflow für die Bearbeitung von Proben vieler Organismen.

Die oberen Hautschichten des Menschen, primär das *stratum corneum*, sind von einer Vielzahl von kommensalen und pathogenen Bakterien besiedelt. Dennoch sind diese bei einem gesunden Hautbild sehr selten für Erkrankungen verantwortlich. Der Grund dafür sind u.a. auf der Haut vorkommende antimikrobielle Peptide, die die Hautflora kontrollieren und eine Überbevölkerung besagter Prokaryonten verhindern. Ein opportunistisches Pathogen, das primär in wässrigen Habitaten vorkommt und ebenfalls auf der Haut zu finden ist, ist *Pseudomonas aeruginosa*. Auf der Suche nach antimikrobiellen Peptiden, die das Wachstum dieses Bakteriums beeinflussen können, wurden im Arbeitskreis um Prof. Dr. Jens-Michael Schröder (Dermatologie, Uniklinikum Kiel) Faktoren aus vereinigten Kallusextrakten hinsichtlich ihrer Potenz gegenüber *P. aeruginosa* untersucht. In mehreren Fraktionen konnte eine Wirkung gegen das Bakterium mit Sequenzabschnitten des Proteins Hornerin korreliert werden. Die entsprechenden Fraktionen wurden flüssigkeitschromatographisch aufgetrennt und massenspektrometrisch identifiziert und sequenziert. In einer einzelnen Fraktion konnten mehr als 100 Peptide des Hornerins nachgewiesen werden. Unter diesen war ebenfalls ein Abschnitt, der in Versuchen der AG Schröder bereits eine potente Wirkung gegen *P. aeruginosa* gezeigt hatte. Neben Hornerin konnten in den biologisch aktiven Proben ebenfalls Sequenzen der Proteine Dermcidin, Filaggrin-2 und Calprotectin nachgewiesen werden, für welche eine antimikrobielle Aktivität bekannt ist. Neben den Hornerin-haltigen Fraktionen wurde zu späterer Retentionszeit desselben HPLC-Laufs außerdem ein Faktor identifiziert, der das Wachstum des Pathogens *Staphylococcus aureus* einschränken konnte. In Sequenzierungsexperimenten konnte diese Substanz als ein aus 19 Aminosäuren bestehendes Peptid identifiziert werden, welches eindeutig dem Protein Desmoplakin zugeordnet werden konnte. Die Spaltstellen des Peptids führten zu der Hypothese, dass die im Schweiß des Menschen vorkommende Protease Cathepsin D für dessen Degradierung verantwortlich sein könnte. In Hydrolyseexperimenten mit diesem Enzym konnte bestätigt

werden, dass Cathepsin D in der Lage ist, die beiden relevanten Peptidbindungen zu hydrolysieren.

b. Abstract

The thesis consists of two major chapters: the development of a method for the analysis of glycosylated proteins and hydrolysis products without the application of work-proteases and the sequencing of antimicrobial peptides derived from biological samples.

Glycosylated proteins and peptides represent a sub-group of analytes in the field of proteomics, whose behaviour differs from non-glycosylated molecules, e.g., in chromatographic separation. Oligosaccharides bind water, which leads to peak broadening in reversed-phase liquid chromatography. Furthermore, glycosidic bonds are more prone to fragmentation during activation cycles in the mass spectrometers than peptide bonds. Tandem mass spectra of glycosylated analytes are therefore dominated by glycan fragments, which hampers the identification of a peptide's or a protein's primary structure. The presented workflow was conducted with the meprin- α/β digest of the glycoprotein progranulin. It features the application of solid phase extraction cartridges that have been modified with immobilized lectins. The adsorbens is specific for the glycosid *N*-acetylglucosamine, which is part of the core structure of *N*-glycans in most organisms. The hydrolysis products were deglycosylated using PNGase F and run through the adsorbens. Finally, the flow through was analysed for the generated peptides and all meprin-generated cleavage sites using LC-ESI MS. In order to estimate the established workflow's capability, the same samples were investigated using a standard Bottom-Up workflow. The number of identified peptides and cleavage sites were used as the read-out. Furthermore, the sequence coverage was compared. Both approaches identified approximately the same number of peptides. However, concerning cleavage sites (up to 41 % more sites identified) and sequence coverage (up to 16 % better coverage), the Top-Down approach clearly outperformed the Bottom-Up methodology. Another important factor is the time needed to prepare samples for an LC-MS/MS experiment. Here, the Bottom-Up approach, which requires termini labeling due to the usage of work proteases, took an additional seven hours compared to the lectin-based approach. The applied solid phase extraction cartridges are not only suitable for the removal of glycosylated analytes but also for the enrichment thereof. This is especially interesting for studies, which aim for the identification of glycan structures, e.g., glycosidic bonds and type of glycosids. As the chosen lectin is specific for the glycosid *N*-acetylglucosamine, the work flow is suitable for samples from various organisms.

The upper layers of the human skin, primarily the *stratum corneum*, are populated by numerous commensales and pathogenic bacteria. However, as long as the skin stays healthy, infections of said procaryotes do rarely occur. The reason for this is the presence of host antimicrobial peptides, which control the skin microbiota and prevent the growth of such pathogens to a certain extent. One of these pathogens is *Pseudomonas aeruginosa*. It can primarily be found in aquatic environments, e.g., mud. Humans are repeatedly in contact with this opportunistic pathogen, especially at their feet, yet infections rarely occur. The working group of Prof. Dr. Jens-Michael Schröder (Dermatology department, Uniklinikum Kiel) investigated the extracts from combined kallus samples for antimicrobial peptides that specifically inhibit the growth of *P. aeruginosa*. Multiple fractions showed potency against *P. aeruginosa*. The biological activity could be assigned to peptides from hornerin, a component of the epidermal cornified envelope. More than 100 sequences were identified, of which one peptide has already been proven to be a potent antibiotic against *P. aeruginosa* in previous activity assays performed by Schröder *et al.* Besides hornerin, peptides from other antimicrobial proteins or peptides have been identified in the given samples, namely dermcidin, filaggrin-2 and calprotectin. Additionally, a yet unknown factor could be correlated with a strong activity against *Staphylococcus aureus*. Using liquid chromatography and mass spectrometry, the factor could be identified as a peptide consisting of 19 amino acids. It could be unambiguously assigned to desmoplakin, a component of the desmosomes. The termini of the peptide allowed the assumption that desmoplakin could have been degraded by cathepsin D, a protease found in human sweat. Hydrolysis experiments confirmed the capability of the enzyme to hydrolyse the relevant peptide bonds in the desmoplakin protein.

2. Abkürzungen

AA	<i>Amino acid</i>
AAS	Atomabsorptionsspektroskopie
ABC	Ammoniumbicarbonat (Ammoniumhydrogencarbonat)
ACN	Acetonitril
AGC	<i>Automatic gain control</i>
API	<i>Atmospheric pressure ion source</i>
APP	<i>Amyloid precursor protein</i>
ATP	Adenosintriphosphat
BisTris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan
BMP1	<i>Bone morphogenetic protein 1</i>
BUP	<i>Bottom-Up Proteomics</i>
CFU	<i>Colony forming unit</i>
CHCA	α -Cyano-4-hydroxymethylsäure
CID	<i>Collision-induced dissociation</i>
DDA	<i>Data dependend acquisition</i>
DIA	<i>Data independend acquisition</i>
DMSO	Dimethylsulfoxid
DTT	Dithiothreitol
ECD	<i>Electron capture dissociation</i>
ECM	<i>Extracellular matrix</i>
EDTA	Ethylendiamintetraessigsäure
ESI	<i>Electrospray ionization</i>
ETD	<i>Electron transfer dissociation</i>
FA	<i>Formic acid</i> (Ameisensäure)
FDR	<i>False discovery rate</i>
FLI	<i>Fixed laser intensity</i>
HCD	<i>Higher-energy collisional dissociation</i>
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure

HPLC	<i>High performance liquid chromatography</i>
HRNR	Hornerin
IAA	Iodacetamid
ICP	<i>Ion-coupled plasma</i>
i.D.	Innerer Durchmesser
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
LASER	<i>Light amplification by stimulated emission of radiation</i>
LB	<i>Lysogeny broth</i>
MALDI	<i>Matrix-assisted LASER desorption/ionization</i>
MMP2	Matrix-Metalloproteinase 2
MS	Massenspektrometrie
NCE	<i>Normalized collision energy</i>
NEM	N-Ethylmaleinimid
o.D.	Optische Dichte
PD	Proteome Discoverer
PGRN	Progranulin
PKA	Proteinkinase A
PSA	<i>Primary structure alteration</i>
PSM	<i>Peptide spectrum match</i>
PTM	Posttranslationale Modifikation
RP-LC	<i>Reversed-phase liquid chromatography</i>
SA	<i>Staphylococcus aureus</i>
TCEP	Tris(2-carboxyethyl)phosphin
TDP	<i>Top-Down-Proteomics</i>
TFA	<i>Trifluoroacetic acid</i> (Trifluoressigsäure)
TIC	<i>Total ion chromatogram</i>
TOF	<i>Time of flight</i>
TRIS	Tris(hydroxymethyl)-aminomethan
TSB	<i>Tryptic soy broth</i>
UPLC	<i>Ultra performance liquid chromatography</i>

WCX	<i>Weak cation exchange</i>
WGA	<i>Wheat germ agglutinin</i>
XIC	<i>Extracted ion chromatogram</i>

3. Ein- und Dreibuchstaben-Code der proteinogenen Aminosäuren

Aminosäure	Einbuchstaben-Code	Dreibuchstaben-Code
Alanin	A	Ala
Cystein	C	Cys
Aspartat	D	Asp
Glutamat	E	Glu
Phenylalanin	F	Phe
Glycin	G	Gly
Histidin	H	His
Isoleucin	I	Ile
Lysin	K	Lys
Leucin	L	Leu
Methionin	M	Met
Asparagin	N	Asn
Prolin	P	Pro
Glutamin	Q	Gln
Arginin	R	Arg
Serin	S	Ser
Threonin	T	Thr
Selenocystein	U	Sec
Valin	V	Val
Tryptophan	W	Trp
Tyrosin	Y	Tyr

4. Einleitung

a. Proteomik

Die Proteomik (aus dem Engl., *Proteomics*) beschreibt die Wissenschaft, die sich mit der Analyse des Proteoms befasst. Als Proteom wird die Gesamtheit aller Proteine einer einzigen Zelle zu einem bestimmten Zeitpunkt bezeichnet. Im Gegensatz zum Genom, welches für alle Zellen desselben Organismus identisch ist, stellt das Proteom ein dynamisches System dar [1]. Dessen Zusammensetzung wird von Proteinabbau und –synthese sowie durch posttranslationale Modifikationen bestimmt. Die Proteinsynthese wiederum wird durch die zeitabhängige Genexpression kontrolliert. Auf diese Weise reagiert die Zelle auf äußere Einflüsse, wie z.B. das Binden eines Peptidhormons an deren Rezeptor. Je nach Organismus kann ein Proteom mehrere Tausend Proteine umfassen. Diese lassen sich in verschiedene, durch reversible (posttranslationale) Modifikationen (PTM) hervorgerufene Isoformen infolge von Spleißvorgängen und verkürzte, durch enzymatische Hydrolyse entstandene Sequenzabschnitte weiter unterteilen. Im Laufe der Jahre wurde eine Vielzahl an Bezeichnungen vorgeschlagen, unter anderem findet man die Begriffe *protein species*, *isoform*, *variant* und andere. Keine Begrifflichkeit beschrieb die jeweiligen Formen, ohne dabei mehrdeutig zu sein oder den Regeln der IUPAC zu widersprechen. Im Jahre 2013 wurde daher vom Konsortium für *Top-Down* Proteomik (aus dem Engl., *The Consortium for Top-Down Proteomics*) der Begriff „Proteoform“ definiert, der nun fortlaufend verwendet werden soll [2]. Er beschreibt alle molekularen Formen, in der ein Protein eines einzigen Gens auftreten kann. Dies beinhaltet genetische Variationen, RNA-Transkripte, die infolge des Spleißens auftreten, sowie sämtliche posttranslationale Modifikationen. Die Proteomik beschäftigt sich mit der Identifizierung, Quantifizierung und Charakterisierung dieser Proteoformen und bedient sich dabei unter anderem verschiedener Separationsmethoden sowie der Massenspektrometrie [3]. Das Arbeitsfeld kann in die Bereiche *Bottom-Up Proteomics* (BUP) und *Top-Down Proteomics* (TDP) unterteilt werden und unterscheidet sich in der Verwendung von Proteasen zur Hydrolyse der Analyten [4]. Diese Enzyme besitzen meistens eine hohe Spezifität in der Erkennung ihrer Substrate und hydrolysieren Peptidbindungen systematisch vor oder nach einigen wenigen Aminosäuren. Während die BUP die in einer Probe enthaltenen Proteine und Peptide mit solchen Proteasen in kleinere Analyten (Peptide) spaltet und somit die Komplexität weiter erhöht, wird in der TDP auf

diesen Schritt verzichtet. Die Analyten werden hier intakt analysiert. Die Verwendung dieser Proteasen liegt darin begründet, dass Peptide in der Chromatographie ein anderes Verhalten zeigen als Proteine und sich i.d.R. einfacher auftrennen lassen. Auf die jeweiligen Vor- und Nachteile der beiden Ansätze soll in Kapitel 1.e eingegangen werden.

b. Massenspektrometrie

Die Massenspektrometrie hat sich als Methode der Wahl für die Identifizierung und Quantifizierung von Proteinen und Peptiden etabliert [3]. Ermöglicht wurde sie durch die Verfügbarkeit von Ionisierungstechniken wie der *Electrospray Ionization* (ESI) [5] und *Matrix-Assisted LASER Desorption/Ionization* (MALDI) [6], die sich auch für die Sequenzierung von großen Biopolymeren eignen. Die Vorteile der Massenspektrometrie sind vor allem die niedrigeren Nachweisgrenzen im Femtomol-Bereich sowie die Geschwindigkeit, mit der Fragmentspektren aufgenommen werden können.

Ein Massenspektrometer lässt sich grob in drei Elemente einteilen: eine Ionenquelle, einen Massenanalysator und einen Detektor (**Abbildung 1**).

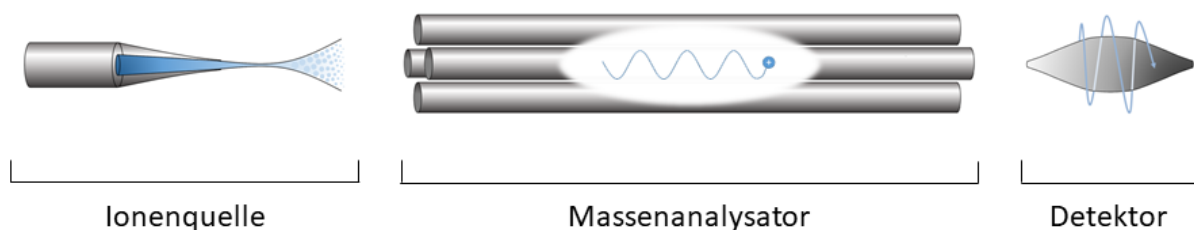


Abbildung 1 Aufbau eines Massenspektrometers (Schema) Ein Massenspektrometer kann in drei Elemente unterteilt werden: die Ionenquelle (hier am Beispiel einer ESI-Quelle), den Massenanalysator, der die Ionen anhand ihres m/z -Verhältnisses auftrennt (hier: Quadrupol [7]) und den Detektor (hier: Orbitrap [8]).

Zur Erzeugung der Ionen bedient man sich in der Proteomik meistens entweder dem Prinzip der ESI oder der MALDI (siehe Kapitel 5.b). Während der Ionisierung entstehen einfach (MALDI) oder mehrfach (ESI) geladene Ionen, die anschließend im Massenanalysator anhand ihres Masse-zu-Ladungs-Verhältnisses (m/z) aufgetrennt werden. Daher kann der Massenanalysator nach der Flüssigkeitschromatographie als eine weitere Dimension der Separation verstanden werden. Eine Form des Analysators stellen Flugzeitmassenspektrometer dar (kurz: TOF, aus dem Engl. von *time of flight* [9]), bei denen man sich zunutze macht, dass Analyten unterschiedlicher Masse, aber gleicher kinetischer Energie unterschiedlich lange Flugzeiten aufweisen und so zu verschiedenen Zeitpunkten

den Detektor erreichen. Der Zusammenhang zwischen Flugzeit t und Masse m , bzw. Masse-zu-Ladung m/z wird durch die Formel

$$\frac{m}{z} = 2 \cdot e \cdot U \cdot \left(\frac{t}{L}\right)^2$$

beschrieben. e steht für die Elementarladung und ist eine Naturkonstante. Die Beschleunigungsspannung U wird vom Nutzer eingestellt und die Länge des Flugrohrs L ist eine gerätespezifische Konstante. Die Ladung z kann anhand des Isotopenmusters bestimmt werden, da die Abstände von Isotopensignalen umgekehrt proportional zur Ladung sind. Für einfach geladene Spezies beträgt der Abstand 1 m/z , für zweifach geladene 0,5 m/z usw. Als weiteres Beispiel für Massenanalysatoren seien die sehr weit verbreiteten Multipole und im speziellen der Quadrupol genannt, welcher aus vier Stabelektroden besteht [7]. Gegenüberliegende Elektroden befinden sich dabei auf dem gleichen Potential und zwischen benachbarten Stäben ist eine Spannung angelegt, die eine zeitabhängige Summe aus Gleich- und Wechselspannung darstellt:

$$u(t) = U + V \cdot \cos(\omega t)$$

Das Verhältnis zwischen Gleichspannung U und Wechselspannung V kann so eingestellt werden, dass einige wenige Analyten stabile Trajektorien durch den Quadrupol beschreiben. Auf diese Weise können bestimmte m/z Verhältnisse selektiert werden. Als Detektoren kommen heutzutage hauptsächlich Elektronenvervielfacher [10] und Ionenfallen wie die Orbitrap [8] zum Einsatz. Letztere bieten eine Reihe von Vorteilen. Dazu gehören unter anderem die Größe (mittlerweile gibt es platzsparende *Benchtop* Geräte), sowie das hohe Auflösungsvermögen und die sehr kurzen Messzyklen für Fragmentspektren, was sie auch für Hochdurchsatzanalysen attraktiv machen.

c. Von Fragmentspektren zur Primärstruktur von Proteinen und Peptiden

In der Proteomik wird die massenspektrometrische Analyse oft in die Ebenen MS^1 und MS^2 unterteilt. Erstere befasst sich mit der Messung von Intaktmassen, bzw. deren Masse-zu-Ladungsverhältnis, während MS^2 (oft auch MS/MS oder tandem MS) die Generierung und Analyse von Fragmentensignalen beschreibt. Weitere Fragmentierungsvorgänge werden als MS^n (n ist die Stufe der Aktivierung) bezeichnet. Werden beispielsweise die Fragmente eines MS^2 -Vorganges weiter fragmentiert, spricht man von MS^3 . Das Masse-zu-Ladungs-Verhältnis

eines Analyten auf der MS^1 Ebene entspricht dem direkten Resultat des Ionisierungsvorgangs und hängt von der gewählten Ionisierungsmethode ab. Während beim MALDI-Prozess einfach geladene Spezies $[M+X]^+$ beobachtet werden, können bei der Elektrosprayionisierung auch Ionen höherer Ladung $[M+nX]^{n+}$ entstehen. X ist hier Platzhalter für den oder die Ladungsträger. In den meisten Fällen handelt es sich dabei um Protonen, jedoch werden mitunter auch Alkalimetall- oder Ammoniumionen als Ladungsträger registriert. Bei datenabhängigen Experimenten (*data-dependent acquisition*) wird zunächst ein sogenannter *full scan* aufgenommen, der über einen großen Bereich (meistens etwa 200-2000 m/z) alle Ionen misst. Verfügt das Massenspektrometer über einen Quadrupol, können mit dessen Hilfe im Anschluss besonders intensive Ionen isoliert und fragmentiert werden. Die Fragmentierung ist eine Folge der Anregung, die entweder durch angeregte Elektronen (ECD [11], ETD [12]) oder durch Molekülstöße (CID [13, 14], HCD [15]) erreicht wird. In Abhängigkeit von der Aktivierungsmethode entstehen für das Protein oder Peptid charakteristische Fragmente, die nach der Roepstorff-Fohlmann-Biemann Nomenklatur [16, 17] annotiert werden und in **Abbildung 2** dargestellt sind.

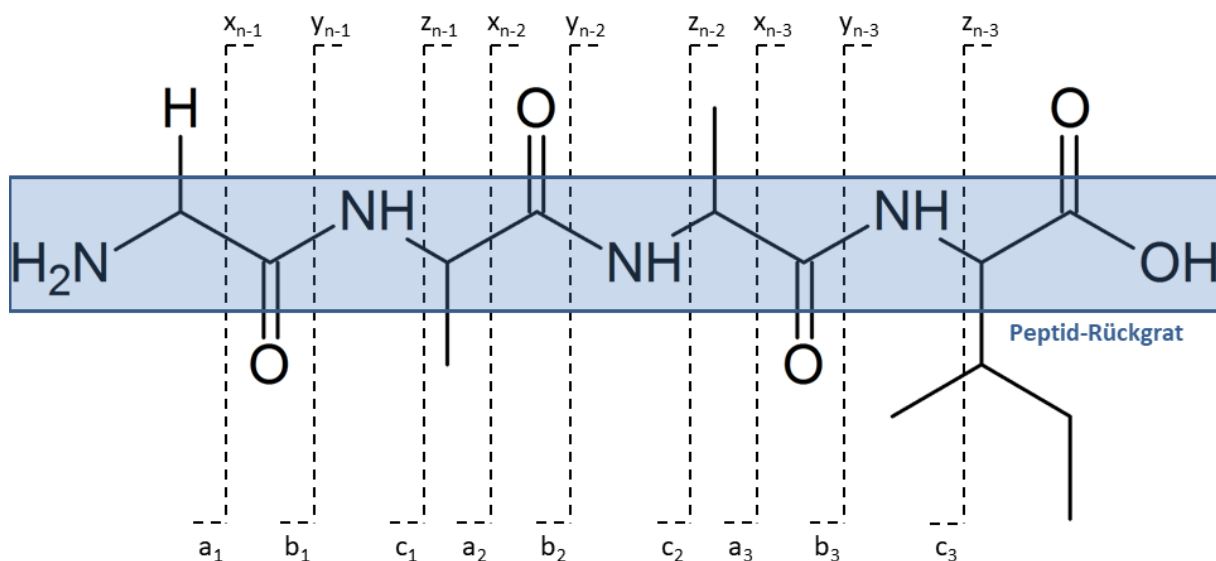


Abbildung 2 Roepstorff-Fohlmann-Biemann Nomenklatur am Beispiel des Tetrapeptids Gly-Ala-Ala-Ile Wird das Rückgrat eines Peptids fragmentiert, so entstehen für die Aktivierungsmethode charakteristische Ionen. Während bei stoßinduzierter Dissoziation hauptsächlich b-, y- und a-Ionen beobachtet werden, wird das Fragmentspektrum bei Anregung durch angeregte Elektronen von c- und z-Ionen dominiert. Die Nummerierung erfolgt für a-, b- und c-Ionen vom N-Terminus, für x-, y- und z-Ionen vom C-Terminus aus. n bezeichnet hier die Anzahl der Aminosäuren des Proteins oder Peptids.

Welche Fragmentsignale letztlich in einem MS^2 -Spektrum beobachtet werden, hängt von mehreren Faktoren ab. Zum einen wird die Intensität der Fragmentionen maßgeblich von der Intensität des Vorläuferions (*precursor*) beeinflusst. Ist dieses wenig intensiv, so sind

[illegible]

21

dass eine Ladung an dieser lokalisiert ist, dominieren y -Ionen das Spektrum. Dies ist ein Beispiel für die Sequenzabhängigkeit von detektierten Fragmentionen und deren Intensitäten (Vorläuferion: 390,228 m/z (2+), aufgenommen an der Q Exactive Plus, Strukturformel erstellt mit ACD/ChemSketch, version 2017.1.2, Advanced Chemistry Development, Inc., Toronto, On, Canada, www.acdslabs.com, 2019).

Die proteolytische Prozessierung gilt ebenfalls als PTM und soll im folgenden Kapitel gesondert betrachtet werden.

d. Enzymatische Hydrolyse als regulatorischer Mechanismus

Die enzymatisch katalysierte Kondensation von Aminosäuren zu Peptiden und Proteinen (Proteinbiosynthese) und deren Umkehrreaktion, die Hydrolyse von Peptidbindungen, sind sich ständig wiederholende Vorgänge, die in jedem Organismus ablaufen. Für letztere kann eine grobe Unterscheidung zwischen vollständiger Degradierung (z.B. im Verdauungstrakt) und gezielter Hydrolyse getroffen werden. Bei der gezielten Hydrolyse spielen oft spezifische Erkennungssequenzen eine Rolle und den Spaltprodukten (weiterhin als Peptide bezeichnet) kann zudem eine biologische Funktion innewohnen [20–22]. Um die Spezifität eines Enzyms zu bewerten, werden für eine identifizierte Spaltstelle meist die vier oder fünf nächsten Aminosäuren N- und C-terminal zur hydrolysierten Peptidbindung betrachtet (**Abbildung 4**).

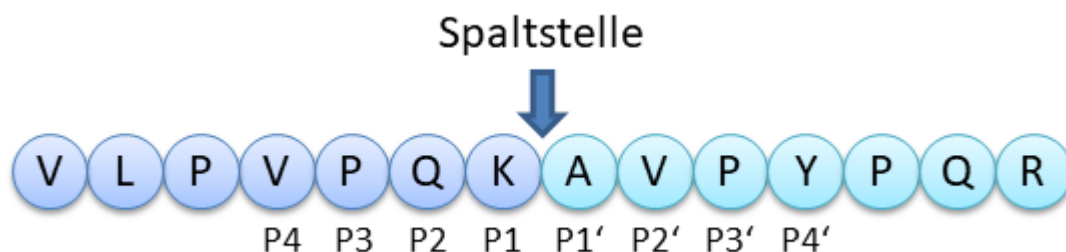


Abbildung 4 Bezeichnung der Aminosäuren einer Spaltstelle Die Aminosäuren in unmittelbarer Nähe einer identifizierten Spaltstelle werden oft herangezogen, um eine Aussage über die Substratspezifität einer Protease machen zu können. Dabei sind vor allem die inneren Aminosäuren der P4 bis P4' Positionen von Interesse [23].

Als Beispiel für biologisch relevante Spaltprodukte sei hier die Familie der antimikrobiellen Peptide (AMP) genannt, welche durch eine enzymatische Hydrolyse entstehen und als körpereigene Antibiotika verstanden werden können [20]. Diese wurden unter anderem in der äußersten Hautschicht, dem *stratum corneum*, nachgewiesen und verhindern das unkontrollierte Wachstum von Pathogenen auf der Haut, wie z.B. *Staphylococcus aureus* [24]. Peptide müssen jedoch nicht zwangsläufig eine positive Wirkung auf den Wirt haben. Ein relativ kurzer Abschnitt der extrazellulären Domäne des *amyloid precursor proteins* (APP), welcher durch oberflächliche Hydrolyse (sogenanntes *shedding*, bezeichnet die

Spaltung von membrangebundenen Proteinen nahe der Membranoberfläche) entsteht, gilt als mögliche Ursache für die Entstehung der Alzheimer-Krankheit [25].

Die Identifizierung der Spaltstelle(n) in Substraten könnte helfen, Hinweise auf die für die Hydrolyse verantwortliche Protease zu erhalten. Spaltprodukte werden heute in den meisten Studien mittels *Bottom-Up* Methoden identifiziert. Arbeits-Proteasen degradieren diese weiter und erleichtern so die (chromatographische) Auftrennung und die massenspektrometrische Analyse. Dieser Ansatz kann jedoch zu dem Problem führen, dass die Ionisierung der entstehenden Peptide aufgrund ihrer Aminosäuresequenz und/oder etwaiger PTM unzureichend ist. Infolgedessen werden die Termini, welche die ursprüngliche Spaltstelle anzeigen, oft nicht detektiert, was eine Identifizierung unmöglich macht (**Abbildung 5**). Dies gilt vor allem für C-terminale Spaltprodukte und solche, die über wenige basische und/oder viele saure funktionelle Gruppen verfügen. Vor allem die Seitenketten der Aminosäuren Aspartat und Glutamat erschweren eine Ionisierung im positiven Ionenmodus. Es wurde eine Reihe von chemischen Modifikationen entwickelt, die diese Schwierigkeit beheben sollen [26, 27]. *Top-Down* Ansätze verzichten hingegen vollständig auf weitere Hydrolyseschritte und erhalten die Spaltprodukte in ihrer ursprünglichen Aminosäuresequenz [28]. Dadurch wird die Komplexität einer Probe nicht erhöht und eine Identifizierung kann sowohl auf der Intaktebene als auch anhand des Fragmentspektrums erfolgen [29]. Eine Begründung ist in Kapitel 1.e gegeben.

e. *Bottom-Up vs. Top-Down Proteomics*

Als die Entwicklung der MALDI und ESI den Transfer von großen ionisierten Biomolekülen in die Gasphase ermöglichte, wurden zunächst einzelne kleine Proteine wie z.B. Ribonuklease A untersucht [30]. Die Analyse größerer Proteine folgte [31–34]. Etwa fünfzehn Jahre später liegen in der Massenspektrometrie die Limitationen in Bezug auf die Masse von intakten Proteinen bei etwa 200 kDa. Der Grund für die Beschränkung ist das abnehmende Signal-zu-Rausch-Verhältnis mit zunehmender Masse der Proteine [35]. Ein weiteres Problem stellen neben der Analytgröße auch komplexe Proben von Proteinen dar, deren Auftrennung unter anderem aufgrund von limitierter Löslichkeit eine Herausforderung darstellt. Beide Problematiken lassen sich durch die Verwendung von Enzymen zur Degradierung umgehen. Die bei dieser enzymatischen Hydrolyse entstehenden Peptide sind kleiner und zeichnen sich durch eine einfachere Handhabung in der Flüssigkeitschromatographie aus [36]. Im

Anschluss an die Chromatographie werden die generierten Peptide massenspektrometrisch analysiert und dem ursprünglichen Protein zugeordnet. Dieser letzte Schritt wird auch als „Proteininferenz“ (*protein inference*) bezeichnet [37]. Proteininferenz ist jedoch mit Problemen behaftet. Zunächst kann oft nicht zwischen Proteoformen unterschieden werden, da die Gründe für nicht identifizierte Sequenzen vielseitig sein können. Ob ein Peptid im Verlauf der Analytik verloren ging, z.B. durch Adsorption an Gefäßwänden oder der stationären Phase einer Säule, oder im ursprünglichen Zustand nicht vorhanden war, kann nicht mit Sicherheit festgestellt werden. Eine weitere Schwierigkeit stellen Sequenzabschnitte dar, die nicht einzigartig sind und daher mehreren Proteinen zugeordnet werden können (*protein inference problem*) [37]. Für eine zweifelsfreie Identifizierung sind daher Peptide nötig, deren Primärstruktur für das jeweilige Protein einzigartig ist [38]. Generell wird für *Bottom-Up* Ansätze in den seltensten Fällen eine vollständige Sequenzabdeckung erreicht. Dies kann bedeuten, dass Informationen über PTM oder Termini einer Spaltstelle fehlen. Die Identifizierung von Modifikationen oder Proteintermini mittels *Bottom-Up Proteomics* ist möglich, kann aber aus den im Folgenden genannten Gründen fehlschlagen (**Abbildung 5**). Gezeigt ist eine Proteoform, die nach der Translation an zwei Stellen proteolytisch prozessiert wurde und über eine Disulfidbrücke sowie eine Phosphorylierung und Glykosylierung verfügt. Der N-Terminus der kanonischen Sequenz, bei dem es sich z.B. um ein Signal- oder Propeptid handeln könnte, fehlt (**B**). Einem Standard *Bottom-Up* Workflow folgend, wird die Proteoform zunächst reduziert/alkyliert, mithilfe einer Arbeits-Protease verdaut (**C**) und anschließend im Massenspektrometer fragmentiert.

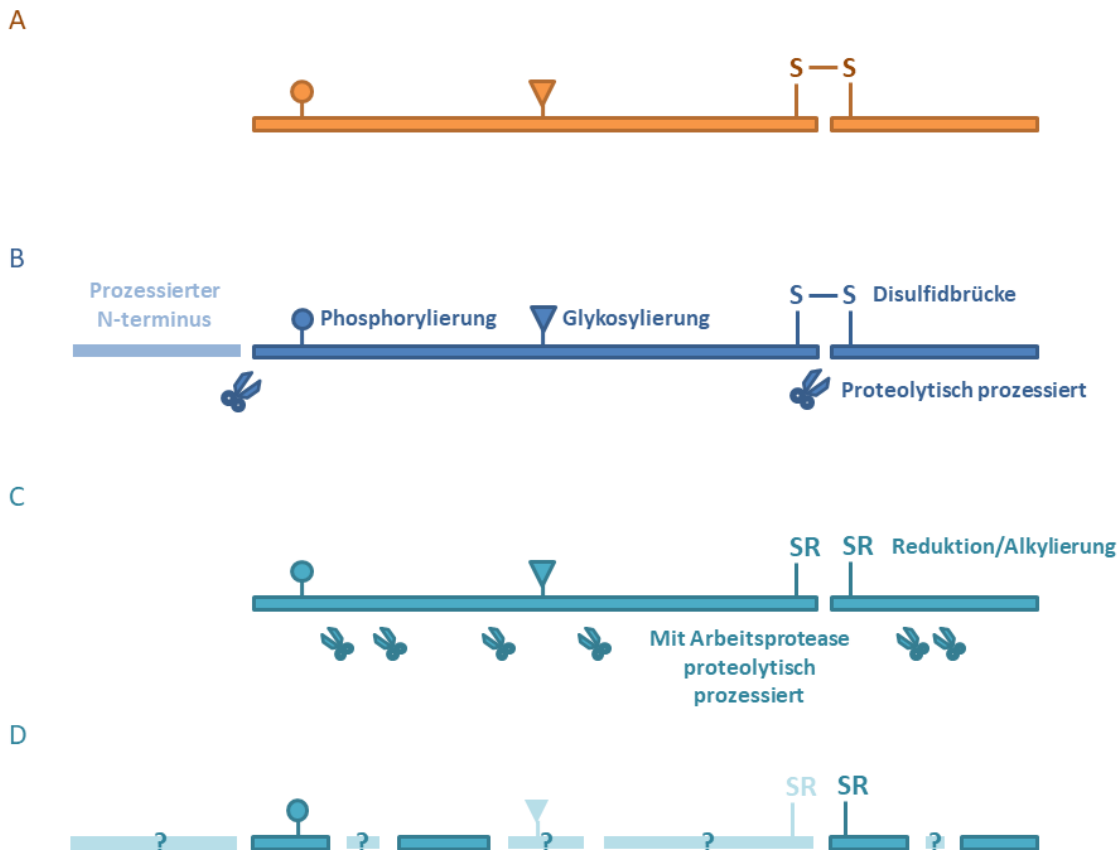


Abbildung 5 Schematische Darstellung der Bottom-Up und Top-Down Ansätze Viele Proteine werden nach ihrer Translation weiter modifiziert, beispielsweise durch enzymatische Hydrolyse, Glykosylierung oder Phosphorylierung. Das Resultat können mehrere Proteoformen (B) sein, die einem einzigen Gen entstammen. Bei Bottom-Up Methoden werden in der Regel zunächst alle Disulfidbrücken reduziert und alkylisiert. Im Anschluss erfolgt die weitere Hydrolyse mithilfe der Arbeitsprotease (C). Der Verdau wird dann massenspektrometrisch analysiert (D). Die Fragezeichen zeigen das Problem dieses Ansatzes auf: Eine Sequenzabdeckung von 100 % wird selten erreicht und Termini, der in (B) erzeugten Spaltstellen werden nicht identifiziert. Der Top-Down Ansatz verzichtet auf die Verwendung von Arbeitsproteasen und analysiert die Proteoform(en) in ihrer Gesamtheit (A).

Die erhaltenen Fragmentspektren werden mit theoretischen Spektren verglichen, die aus einem *in silico* (theoretischen) Verdau der Aminosäuresequenz der Proteoform mit der Arbeitsprotease resultieren (für Details siehe Kapitel 5.d). Je nach Güte des Spektrums gilt ein Peptid als identifiziert und dessen Sequenz als abgedeckt (**D**). Peptide, die nicht identifiziert wurden, sind in **Abbildung 5** mit einem Fragezeichen markiert. Der ursprünglich proteolytisch entfernte N-Terminus, sowie kleine Sequenzabschnitte, das glykosylierte Peptid und der C-Terminus der Spaltstelle aus (**B**) wurden in diesem Beispiel nicht identifiziert. Die Gründe für das Fehlen dieser Abschnitte sind vielfältig, können jedoch in einem einzigen *Bottom-Up* Experiment oft nicht ermittelt werden. Eine Möglichkeit, die Sequenzabdeckung zu erhöhen und damit die Identifizierung dieser fehlenden Peptide zu

ermöglichen, ist die Verwendung mehrerer Arbeitsproteasen unterschiedlicher Spezifität [39]. Die Vorteile des *Top-Down* Ansatzes sind ebenfalls in **Abbildung 5** dargestellt. Die Proteoform (**A**) wird hier intakt analysiert. Deren Masse kann mithilfe des Masse-zu-Ladungsverhältnisses und des Ladungszustandes des Signals im Massenspektrum berechnet werden. Die Masse des Präkursors beinhaltet alle Modifizierungen einer Proteoform, die unter der Bezeichnung *primary structure alterations* (PSA) zusammengefasst werden [40]: Substitutionen von Aminosäuren, posttranslationale Modifikationen und Verkürzungen der Aminosäuresequenz. Während erstere theoretisch auch in einem *Bottom-Up* Ansatz identifiziert werden können, so stellen fehlende Sequenzen ein großes Problem dieser Methode dar. Die Bestimmung der Intaktmasse mittels *Top-Down* Ansätzen hingegen gibt einen ersten Hinweis auf die Primärstruktur des Analyten. Wird durch ein Fragmentspektrum der intakten Proteoform eine ausreichend hohe Sequenzabdeckung erreicht, können so die Positionen der PSA bestimmt werden. Der technische Fortschritt in der Massenspektrometrie und Flüssigkeitschromatographie erlaubt mittlerweile die routinemäßige Sequenzierung von Proteinen aus komplexeren Proben bis etwa 30 kDa [41]. Dies ermöglicht den Verzicht auf die Verwendung von Proteasen und die direkte massenspektrometrische Analyse. Der *Top-Down* Ansatz bietet also im Vergleich zum etablierten *Bottom-Up* Verfahren einige Vorteile. Es gibt jedoch Gründe, warum nach wie vor ein Großteil der Proteomik-Fragestellungen unter Zuhilfenahme von Arbeitsproteasen gelöst wird. Zunächst sei hier die Löslichkeit von Proteinen genannt, die direkten Einfluss auf den Anteil einer Probe hat, welcher den Detektor überhaupt erreichen kann. Sie ist abhängig von Temperatur, Salzkonzentrationen, Gehalt an organischem Lösungsmittel und vor allem vom pH-Wert. Membran-gebundene Proteine, welche etwa 30 % des Proteoms ausmachen [42], sind ohne den Zusatz von Detergenzien in wässrigen Medien unlöslich. Zwar können auch Membranproteine durch Substanzen wie Natriumdodecylsulfat (SDS) oder Polyethylenglykolether in Lösung gebracht werden, allerdings werden Analysemethoden wie die Flüssigkeitschromatographie durch diese Substanzen erheblich gestört. SDS beeinträchtigt eine flüssigkeitschromatographische Trennung bereits ab Konzentrationen oberhalb von 0.02 % [43]. Der Einsatz von Detergenzien erfordert daher zusätzliche Arbeitsschritte, die diese Substanzen entfernen oder zumindest so weit verdünnen, dass die Analytik nicht beeinflusst wird. Leider geht das Entfernen, bzw. Verdünnen der Detergenzien mit Probenverlust und der Gefahr einher, dass (vor allem) Membranproteine ausfallen und

nicht wieder in Lösung gebracht werden können. Die Rekonstitution in kalter Ameisensäure gilt hierbei als vielversprechende Methode [43]. Ein weiterer Grund für die Verwendung von Proteasen zur Hydrolyse von Analyten ist das unterschiedliche Verhalten von Proteinen und Peptiden in der Flüssigkeitschromatographie. Diese ist aufgrund des Auflösungsvermögens, das mit HPLC und vor allem UPLC Geräten erreicht wird, einer der wichtigsten Separationsmethoden, um die Komplexität einer Probe zu verringern. Während elektrophoretische Methoden wie die Gelelektrophorese [44], die isoelektrische Fokussierung (IEF) [45] oder die kürzlich entwickelte *gel-eluted liquid fraction entrapment electrophoresis* (GELFrEE) [46] hauptsächlich für die Trennung von Proteinen verwendet werden, eignet sich die Flüssigkeitschromatographie für Proteine und Peptide gleichermaßen. Für eine hochaufgelöste Trennung von Peptiden ist u.a. die Flussrate entscheidend (Begründung in Kapitel 5.a). Die Separation von Proteinen wird vor allem durch die verwendeten Säulentypen und -materialien beeinflusst. Dabei haben sich monolithische Säulen und superporöse Partikel als bis dato beste Wahl empfohlen [43, 47, 48]. Trotz der mittlerweile realistischen *Peak*-Kapazitäten von 500 und mehr, ist eine eindimensionale Trennung von Proteomen, die mehrere Tausend Proteine umfassen können, nicht ausreichend. Selbst unter Verwendung von Säulen mit einer Länge von 1000 mm konnten aus komplexeren Proben keine Proteine oberhalb von 30.000 Da nachgewiesen werden [49]. Häufig werden daher flüssigkeitschromatographische mit gelelektrophoretischen Trennmethoden gekoppelt. Es konnte gezeigt werden, dass sich multidimensionale Ansätze für die Analyse von Proteomen eignen. Beispielweise wurden mit einer Kombination aus IEF, GELFrEE und LC-MS mehr als 5.000 Proteoformen aus humanen H1299 Zellen nachgewiesen; darunter solche mit Intaktmassen deutlich über 30.000 Da [50]. Dennoch liegt der Großteil der identifizierten Proteine in *Top-Down* Studien unterhalb dieser Grenze, weswegen die Entwicklung neuer Methoden für die Analyse größerer Proteoformen im Fokus aktueller Studien liegt und auch in dieser Arbeit diskutiert werden soll. Das Fehlen geeigneter Workflows ist jedoch nicht das einzige Problem, das die Identifizierung von Proteinen erschwert. Aufgenommene Spektren werden heutzutage meist computergestützt ausgewertet. Das hat sowohl mit der Komplexität als auch der Größe der Daten zu tun. Eine einzige Probe, die chromatographisch getrennt und massenspektrometrisch analysiert wird, führt nicht selten zu mehreren Zehntausend Fragmentspektren, die unmöglich alle manuell annotiert werden können. Man bedient sich daher leistungsfähigen Rechnern und

Algorithmen, die aufgenommene mit theoretischen Spektren abgleichen. Während es für die Analyse von Peptiden eine Vielzahl möglicher Algorithmen und Software-Pakete gibt, ist die Anzahl der für die Auswertung von *Top-Down* Datensätzen verfügbaren Programme begrenzt. Ähnlich wie die Chromatographie ist auch die Annotation von Proteinfragmentspektren deutlich komplizierter als die von Peptiden. Peptidfragmente sind selten größer als 1.500 Da und meistens ein- oder zweifach geladen. Die monoisotopischen *Peaks* eines Signals sind gleichzeitig auch die intensivsten. Dies erleichtert die Extraktion von monoisotopischen Massenlisten für den direkten Vergleich mit theoretischen Spektren. Für Proteine ergeben sich andere Größenordnungen. Die Intensitäten für Vorläuferionen verteilen sich oft über fünf und mehr Ladungszustände, was die Intensität einzelner Signale negativ beeinflusst [29]. Für Peptide werden meistens zwei oder drei Ladungen beobachtet (**Abbildung 6**). Ein weiteres Problem ist die abnehmende Sensitivität mit zunehmender Masse, weswegen für *Top-Down* Ansätze in der Regel mehr Ausgangsmaterial benötigt wird [51]. Außerdem sind selbst die Fragmente von Proteinen noch so groß, dass sie im Tandem-Massenspektrum sogenannte *isotopic envelopes* (Gruppe von Signalen, deren Intensität nach der Häufigkeit der Isotope verteilt sind) zeigen und monoisotopische Signale sehr schwach oder überhaupt nicht vorhanden, bzw. aufgelöst sind. Um eine Liste von monoisotopischen *m/z*-Werten zu erhalten, ist die Dekonvolution dieser Signalgruppen notwendig [40]. Da Signale einer Gruppe fehlen oder ganze *envelopes* überlappen können, ist die automatische Dekonvolution genauso wichtig wie fehleranfällig [52]. Weiterhin zeigen Proteine infolge ihrer Fragmentierung interne Spaltprodukte, deren Berücksichtigung die Sequenzabdeckung erheblich erhöhen kann [53]. Es gibt bis dato keine verfügbare Software, die in der Lage ist, diese internen Fragmente zu berücksichtigen.

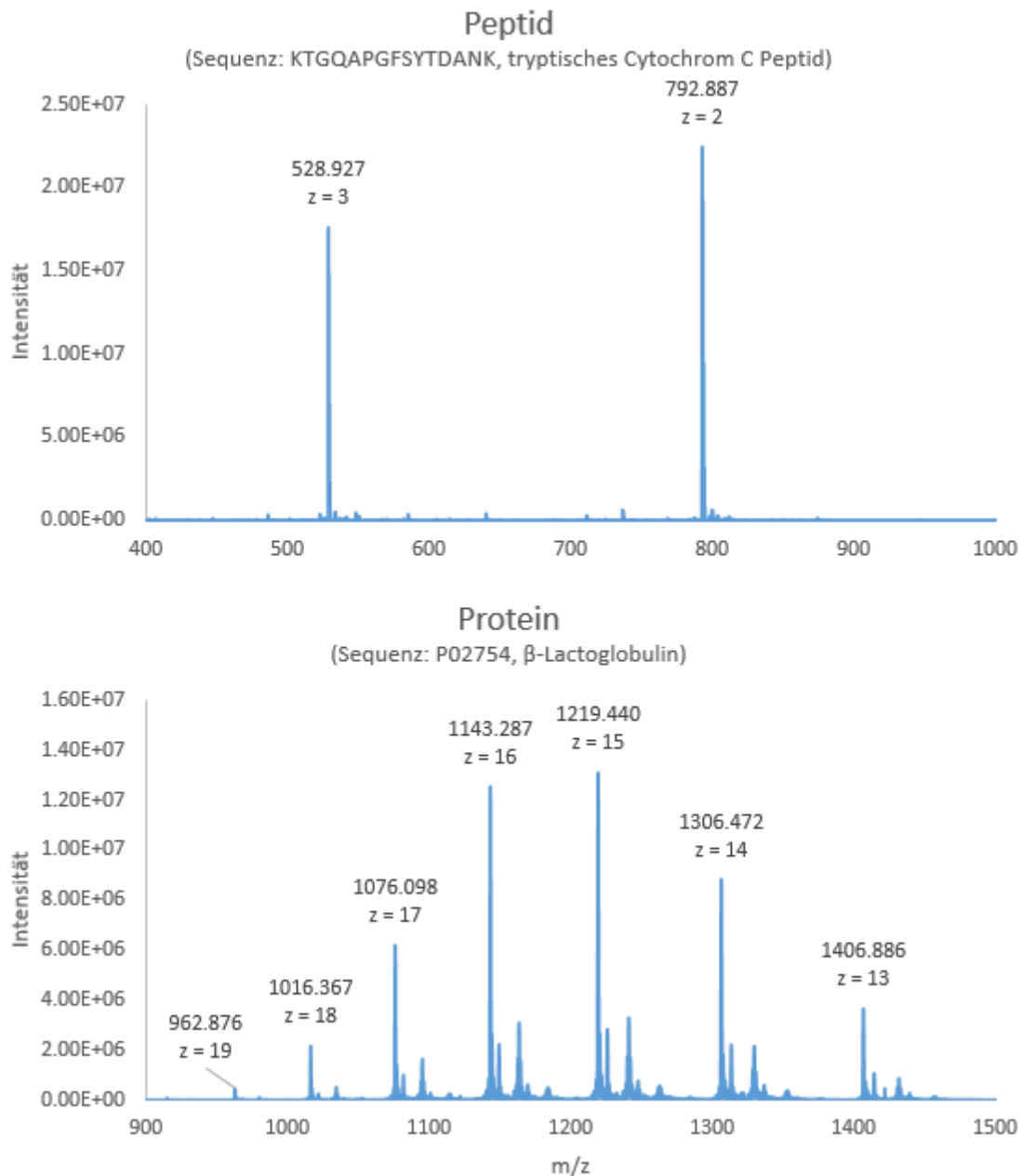


Abbildung 6 ESI-Massenspektren eines Peptids und eines intakten Proteins Gezeigt sind die Massenspektren der Sequenz KTGQAPGFSYTDANK (oben) und des β -Lactoglobulins (unten). Im Falle Cytochrom-C-Peptids verteilt sich die Gesamtintensität über zwei relevante Ladungszustände, während die des β -Lactoglobulins über mehr als fünf Zustände verteilt vorliegt. Würde man nun davon ausgehen, dass Peptid und Protein über die gleiche Gesamtintensität verfügen, so wären die absoluten Intensitäten der Proteinspezies ($z = +19, +18, +17$ usw.) niedriger als die der Peptidsignale ($z = +3$ und $+2$) im oberen Spektrum (Quelle: Eigene Messung).

Um die Proteomik weiterzuentwickeln, ist es also zunächst wichtig, eine Verbesserung der Analytik, vor allem aber der Separationsmethoden, sowie der Datenauswertung anzustreben. Trotz der Herausforderungen, die mit *Top-Down* Ansätzen verbunden sind, gibt es eine Vielzahl von Projekten, die die Vorteile und Wichtigkeit der Methodik unterstreichen [54].

5. Grundlagen

a. Flüssigkeitschromatographie

Die Flüssigkeitschromatographie trennt Analyten in Abhängigkeit von ihrem Adsorptionsvermögen an einer stationären Phase und dem Löslichkeitsverhalten in der mobilen Phase. Die Probe wird in gelöster Form auf die stationäre Phase der Säule gebracht und im Falle der *high performance liquid chromatography* (HPLC) bei hohen Flussraten über die Säule geführt. Die Dauer von der Injektion bis zur Elution ohne Wechselwirkung wird Totzeit t_0 genannt. Da Analyten i.d.R. mit der stationären Phase wechselwirken, verbleiben sie länger auf der Säule und besitzen eine Retentionszeit $t_R > t_0$. Je stärker die Wechselwirkung, desto größer ist auch t_R . Bei irreversibler Adsorption lässt sich keine Retentionszeit ermitteln. In diesem Falle ist entweder die Wechselwirkung zu stark oder das Elutionsvermögen des Laufmittels zu gering [55].

Die Trennleistung einer Säule wird durch die „Bodenhöhe“ H beschrieben, welche die Anzahl der theoretischen Böden N in folgender Weise mit der Länge L der Säule verknüpft [55]:

$$H = \frac{L}{N}$$

Die Größen H und N gehen auf die Beschreibung der fraktionierten Destillation zurück. Je mehr Böden eine Apparatur enthielt, desto besser war die Trennleistung und desto kleiner die Bodenhöhe. Da eine HPLC-Säule keine Böden besitzt, ist diese Gleichung für die Flüssigkeitschromatographie wenig anschaulich. J. J. VAN DEEMTER hat die Bodenhöhe daher durch die Summe der für das Verfahren relevanten Größen ausgedrückt [56]:

$$H = A + \frac{B}{u} + C \cdot u^2$$

A gibt die EDDY-Diffusion (Streudiffusion) wieder und ist eine für die Säule spezifische Konstante. Sie beschreibt den Einfluss von Partikelform/-größe und die Art und Weise, wie diese gepackt sind. B bezeichnet die longitudinale Diffusion und C die Wechselwirkung zwischen Analytmolekülen und der stationären Phase (Massentransfer). u ist die eingestellte Flussrate. Die Gleichung wird als VAN-DEEMTER-Gleichung bezeichnet. Sie begründet, warum Peptide einfacher zu handhaben sind als große Proteine: Die Bodenhöhe wird bei der Trennung von Peptiden hauptsächlich durch die longitudinale Diffusion B beeinflusst. Da der

Massentransferterm von der Anzahl und Intensität der Wechselwirkungen abhängt, die für Peptide eher klein sind, ist deren Massentransfer C nahezu vernachlässigbar. Dies bedeutet, dass eine hohe Trennleistung erreicht wird, indem Peptide bei höheren Flussraten aufgetrennt werden. Da die Flussrate im Term des Massentransfers quadratisch eingeht, existiert für Peptide eine theoretische optimale Flussrate, bei der die Bodenhöhe minimal ist. Das Auflösungsvermögen eines chromatographischen Systems wird bei der Trennung von Peptiden jedoch technisch durch den Gegendruck (*back pressure*) der Säule limitiert [49]. Für Proteine sind die Wechselwirkungen mit der stationären Phase wesentlich ausgeprägter und die longitudinale Diffusion weniger bestimmend. Für die Trennung von Proteinen existiert daher eine optimale und technisch meist realisierbare Flussrate u , für die die Bodenhöhe H minimal ist [55]. Der Massentransferterm ist jedoch nicht für alle Proteine gleich groß. Es gibt also ein Optimum für die Flussrate, das aber für jeden Analyten einen anderen Wert besitzen kann. Erhöht man also die Flussrate, so kann sich für verschiedene Analyten die chromatographische Auflösung verbessern, während sie sich für andere verschlechtert. Ein weiteres Problem ergibt sich aus der Tatsache, dass Proteine nicht nur mit der stationären Phase wechselwirken, sondern auch untereinander.

Peptide und Proteine werden hauptsächlich in Abhängigkeit von ihrer Hydrophobizität aufgetrennt. Als Säulenmaterialien kommen oft alkylierte Silicapartikel zum Einsatz, die über die Länge der Alkylkette für den Anwendungsbereich optimiert werden können. Gängige Kettenlänge reichen von Butylresten (C4) bis hin zu C18-modifizierten Partikeln [49]. Je länger die Alkylkette, desto stärker die Wechselwirkung mit hydrophoben Analyten. Für Peptide haben sich C18-Säulen als Standard durchgesetzt, während dieses Säulenmaterial für die Trennung von Proteinen oftmals ungeeignet ist. Besonders hydrophobe Proteoformen adsorbieren irreversibel an der stationären Phase, weswegen für die Trennung von Protein hauptsächlich C4- und C8-Säulen verwendet werden. Eine Besonderheit stellen monolithische Säulen dar. Im Gegensatz zu partikelbasierten Säulen verfügen sie über ein Netzwerk aus Molekülen und können aus kondensierten Tetraalkylsilanen [57], polymerisierten organischen Verbindungen wie z.B. Styrol [58] oder Estern der Methacrylsäure [59] bestehen. Monolithische Säulen weisen sowohl weite als auch enge Kanäle (Makro-, bzw. Mikroporen) auf. Erstere bedingen einen geringeren Gegendruck als Partikel-basierte Säulenmaterialien. Weiterhin wird durch die kürzeren Wegstrecken, die die Analyten zurücklegen, der Massentransferterm C eher durch

Konvektionsströme als durch die Diffusion in und aus Poren charakterisiert. Sie eignen sich daher besonders gut für die Trennung von Proteinen und führten zu einer höheren chromatographischen Auflösung im Vergleich zu Partikel-basierten Säulen [47].

Da die Wechselwirkung der Analyten mit der stationären Phase durch Salze und Verunreinigungen gestört werden können, kommen sogenannte Vorsäulen zum Einsatz, die einen größeren Durchmesser besitzen. Da sie weniger leicht verstopfen können und sich so zur Vorreinigung einer Probe eignen, werden sie oftmals zusammen mit der analytischen Säule in einem Ventilsystem installiert, welches in **Abbildung 7** dargestellt ist. Vorsäulen (auch *trap columns* genannt) erhöhen so die Lebensdauer der analytischen Säule.

Im Folgenden sollen der Aufbau und die verwendeten Laufmittel bei einer Phasenumkehr-Flüssigkeitschromatographie erläutert werden:

Die Proben gelangen zunächst auf die Vorsäule, auf der sie zurückgehalten werden (**Abbildung 7A**). Um wasserlösliche Kontaminationen und Salze zu entfernen, wird die Vorsäule dann für einige Zeit mit dem eingesetzten (wässrigen) Ladepuffer gespült und der Durchfluss verworfen. Der Fluss der Nanopumpen (meist kommen zwei Pumpen für die Laufmittel und eine Pumpe für den Ladepuffer zum Einsatz) läuft in dieser Position nur über die analytische Säule. Wird das Ventil umgeschaltet, so wird der Fluss des Autosamplers verworfen und der Nanofluss läuft nacheinander über die Vorsäule (im sogenannten *backflush mode* [60], also in umgekehrter Flussrichtung) und die analytische Säule (**Abbildung 7B**). Die Trennung der Analyten erfolgt, indem der Anteil des organischen Laufmittels schrittweise erhöht (Gradiententrennung) wird. Dadurch gelangen sie auf die analytische Säule, auf der die eigentliche Trennung stattfindet. Je hydrophober ein Peptid oder Protein ist, desto länger verbleibt es auf der Säule. Nach der analytischen Säule passiert der Fluss meistens eine UV-Zelle, in der die Absorption der Probe einer oder mehrerer Wellenlängen gemessen werden kann. In der Proteomik wird die Absorption meistens bei etwa 214 nm (Peptidbindung) und 280 nm (aromatische Aminosäuren) gemessen [55]. Nach der Chromatographie kann die Probe direkt mithilfe einer ESI-Quelle ionisiert werden. Man spricht dann von auch von *online mass spectrometry*. Werden die Analyten stattdessen fraktioniert oder durch direktes *spotten* auf ein MALDI *target* aufgebracht, so verwendet man die Bezeichnung *offline mass spectrometry*.

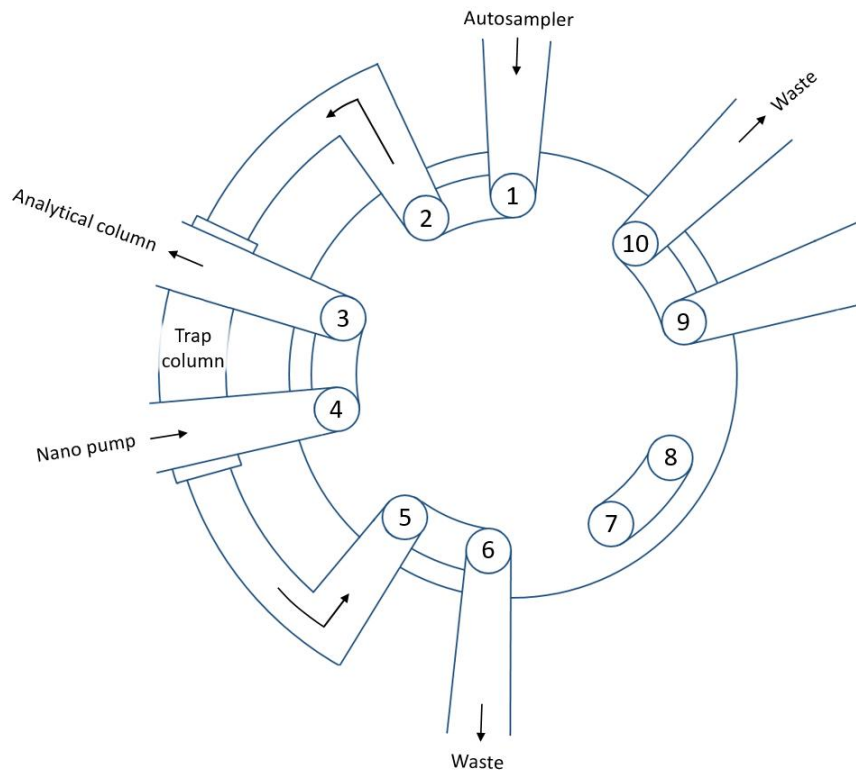
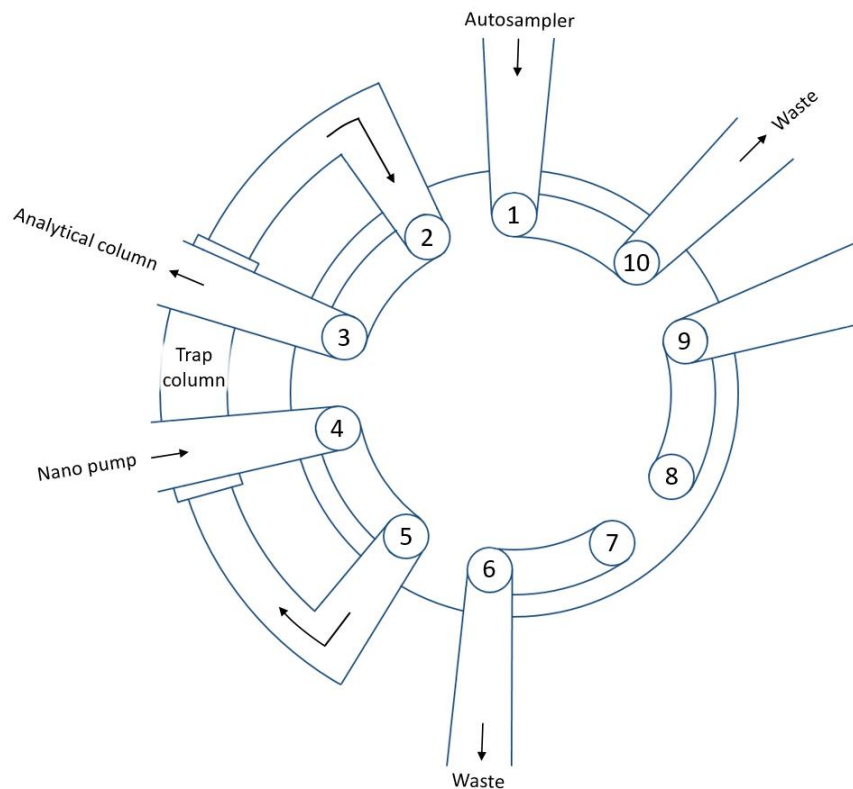
A**B**

Abbildung 7 Ventil-Konnektivität (10-Wege-Ventil) bei Nutzung einer Vorsäule Die injizierte Probe gelangt zunächst vom Autosampler auf die Vorsäule (Trap column) und wird von ihr zurückgehalten, während Salze und wasserlösliche Kontaminationen entfernt werden (A). Durch Umschalten des Ventils gelangt der Nanofluss in umgekehrter Richtung über die Vorsäule und im Anschluss auf die analytische Säule (B). Die Analyten werden in Abhängigkeit von ihrer Hydrophobizität aufgetrennt, im Anschluss direkt ionisiert und in das Massenspektrometer beschleunigt.

b. Prinzipien der Ionisation für die MS

In der Massenspektrometrie werden geladene Ionen in der Gasphase analysiert. Es gibt verschiedene Methoden der Ionisation, allerdings haben sich die MALDI [6] und ESI [5] wegen des schonenden Ionisierungsvorganges für die Analyse großer Biopolymere durchgesetzt. Da in dieser Dissertation primär eine Elektrosprayquelle verwendet wurde, soll die Funktionsweise der Ionisation in diesem Kapitel kurz erläutert werden. Wird die Flüssigkeitschromatographie direkt an das Massenspektrometer gekoppelt (*online*), so spricht man auch von LC-ESI Massenspektrometrie. Das hat zwar den Nachteil, dass die Probe nach der Injektion für weitere Messungen nicht mehr zur Verfügung steht, verringert aber gleichzeitig die Anzahl der Arbeitsschritte, wie sie bei der Fraktionierung hinzukommen und verkürzt somit die Analysenzeit. Des Weiteren können aufgenommene Spektren automatisch einer bestimmten Retentionszeit zugeordnet werden, was vor allem für die Identifizierung und Quantifizierung gleicher Analyten aus mehreren Injektionen/Proben genutzt wird. Nach der UV-Zelle gelangt die Probe direkt zur einer beschichteten Silicanadel, welche sich direkt vor der Eintritts-Kapillare des Massenspektrometers befindet (**Abbildung 8**) und einen Teil der Quelle darstellt.

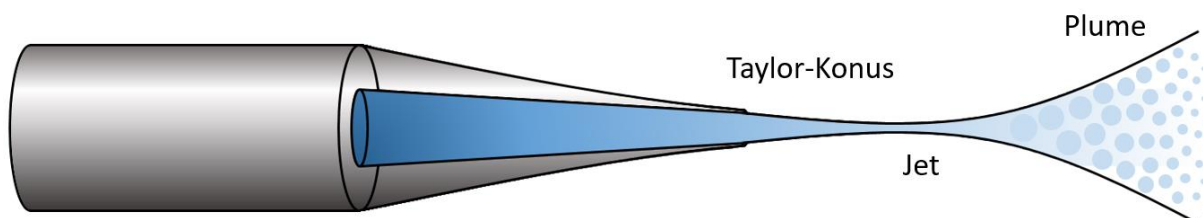


Abbildung 8 Schematische Darstellung des Prinzips der Elektrospray-Ionisation Der Nanofluss der Flüssigkeitschromatographie durchläuft die UV-Zelle und gelangt schließlich zur Quelle, einer ummantelten Silicaspitze. Zwischen dieser und der Eintrittskapillare des Massenspektrometers liegt eine Spannung an, die im Kilovoltbereich liegt. Aufgrund dieser werden die positiv geladenen Analyten zur Flüssigkeitsoberfläche forciert und es entsteht ein feines Spray, das in den Taylor-Konus [61], den Jetbereich und das sogenannte Plume eingeteilt werden kann.

Zwischen dieser und der Nadelspitze ist eine Spannung angelegt. Bei positiver Polarität werden positiv geladene Teilchen zur Flüssigkeitsoberfläche forciert und reichern sich dort an. In der Folge entsteht der sogenannte TAYLOR-Konus. Bei ausreichend hoher Spannung zwischen der Spitze und der Gegenelektrode (Eintrittskapillare des Massenspektrometers) werden kleine geladene Tröpfchen emittiert, deren Größe durch Verdampfungsprozesse weiter abnimmt. Während die Ladung Q konstant ist, nimmt der Radius dieser Tröpfchen r ab. Ist die Gleichung

$$Q^2 = 64\pi^2 \varepsilon_0 \gamma r^3$$

erfüllt, sind die Tröpfchen stabil. Sie ist auch als RAYLEIGH-Gleichung bekannt und gibt an, unter welchen Bedingungen die Oberflächenspannung γ die Ladung ausgleicht. Wird das sogenannte „RAYLEIGH-Limit“ unterschritten, d.h. ist die abstoßende Kraft von Teilchen gleicher Ladung größer als die Oberflächenspannung des Lösungsmittels, führt die Abstoßung zu einer „COULOMB-Explosion“. Dieser Vorgang wiederholt sich bis zur Desolvatisierung der Analyten [55].

Es gibt zwei Theorien, die den Übergang des gelösten Teilchens in die Gasphase beschreiben. Das erste ist als *charge residue model* (CRM) bekannt und bereits 1968 von MALCOLM DOLE beschrieben [62, 63] worden. Er nahm an, dass aufeinanderfolgende COULOMB-Explosion zu sehr kleinen Tröpfchen führen, die nur ein Analytmolekül enthalten. Die freien Ionen würden schließlich durch Kollision mit dem *curtain gas* entstehen. Das *curtain gas* ist ein Gasstrom an der Eintrittskapillare des Massenspektrometers, das gegenläufig zum Ionenstrom gerichtet ist und verhindern soll, dass Luft oder Lösungsmittel in das Massenspektrometer gelangen. Als Alternative zum *curtain gas* können beheizte Transferkapillaren eingesetzt werden, welche die Desolvatation unterstützen. Das zweite Modell wurde von J. R. IRIBARNE und B. A. THOMSON entwickelt und als *ion evaporation model* (IEM) bezeichnet [64]. Bei diesem wurde angenommen, dass Analytmoleküle direkt aus den Tröpfchen emittiert werden. Eine kontinuierliche Emission ist durch das Abnehmen des Tröpfchenradius infolge der Verdampfung gewährleistet. Bis heute konnte keines der Modelle widerlegt werden und verschiedene Beobachtungen lassen sich entweder mit dem einen noch dem anderen erklären. Im Vergleich zur MALDI ist die ESI das deutlich schonendere Ionisationsverfahren [55].

c. Massenspektrometrie

Nach der Ionisation gelangen die Moleküle in das Massenspektrometer. Der generelle Aufbau eines Massenspektrometers wurde bereits in Kapitel 1.b beschrieben. Da in den experimentellen Arbeiten hauptsächlich ein Q-Exactive Plus Massenspektrometer (Thermo Scientific, Bremen, Deutschland) verwendet wurde, soll anhand dessen der Verlauf der Ionen von der Quelle zum Detektor erklärt werden (**Abbildung 9**). Hinter der Eintrittskapillare befindet sich die *S-Lens*, welche aus ringförmigen Elektroden besteht, an denen eine Wechselspannung angelegt ist. Diese sorgt für eine Fokussierung der Ionen. Im Anschluss

werden die Analytmoleküle in einen gekrümmten Multipol, dem sogenannten *bent flatapole*, weitergeleitet.

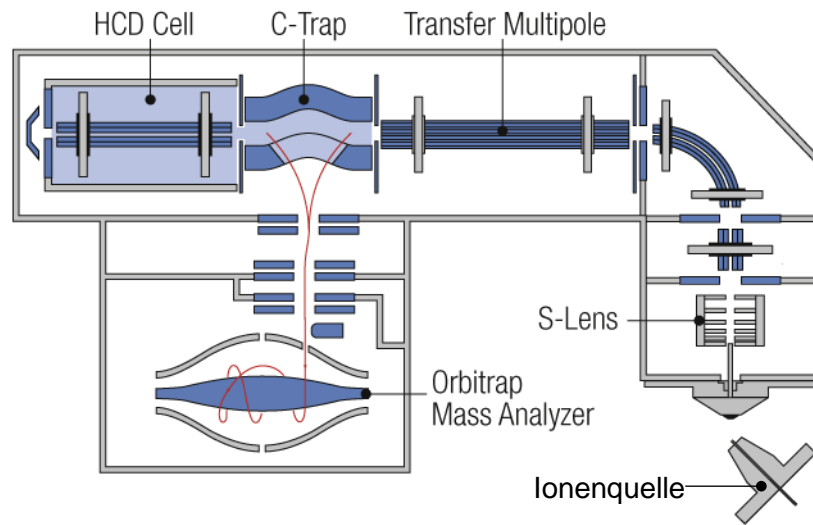


Abbildung 9 Q Exactive Plus (schematisch) Das Q Exactive Plus Massenspektrometer besteht grob aus sechs Abschnitten: Ionenquelle (API), der S-Lens und Ionenoptiken für Fokussierung und den Transfer der Analytmoleküle, einem Quadrupol für eine mögliche Isolation auf dem MS/MS Niveau, der C-Trap (Ionenfalle) zur Bündelung zu Ionenpaketen, der HCD-Zelle, in welcher die Ionen fragmentiert werden und der Orbitrap, welche als Massenanalysator und Detektor gleichermaßen dient (Quelle: Thermo Scientific Q-Exactive Plus, Thermo Fisher Scientific, Bremen, Deutschland).

Dessen Elektroden sind so angeordnet, dass Tröpfchen und ungeladene Teilchen den Multipol in Flugrichtung verlassen können. Nach dem gekrümmten Multipol folgt ein Quadrupol mit hyperbolischen Elektroden, in welchem Ionen anhand ihres Masse-zu-Ladungs-Verhältnisses selektiert werden können. Der Quadrupol fungiert als Massenanalysator des Systems. Die *C-Trap* bündelt schließlich die eintretenden Analytmoleküle zu Ionenpaketen und injiziert diese in die Orbitrap. Hinter der *C-Trap* befindet sich die HCD-Zelle. Die Fragmentierung von isolierten Ionen findet dort bei einem Druck von etwa 0,01 mbar statt, während das Hochvakuum im restlichen Massenspektrometer Werte von unter 10^{-10} mbar erreichen kann. In die HCD-Zelle wird Stickstoffgas eingeleitet, welche als Stoßmedium Verwendung findet. Über eine Änderung der Wechselspannung an den in der HCD-Zelle befindlichen Elektroden werden die Ionen zur Oszillation gezwungen und kollidieren mit den Stickstoffmolekülen. Die übertragene Energie hängt von der Stärke des Stoßes ab und wird über die *normalized collision energy* (NCE) eingestellt. Die Fragmente gelangen dann zurück in die *C-Trap* und werden zur Orbitrap weitergeleitet. Diese besteht aus einer mittig angeordneten ovalen Elektrode, um die die

Ionen in Abhängigkeit ihres Masse-zu-Ladungsverhältnisses unterschiedlich schnell oszillieren. Die Oszillationsfrequenz ist über die Gleichung

$$\omega = \sqrt{\frac{k}{m/z}}$$

mit dem m/z -Wert verknüpft [8]. k bezeichnet hier eine für das Gerät spezifische Konstante. Die sich überlagernden Frequenzen der einzelnen Ionen werden über eine Fourier-Transformation in das eigentliche Massenspektrum umgewandelt.

Für die massenspektrometrische Analyse gibt es die Möglichkeiten der *data-dependent acquisition* (DDA, Ionen eines bestimmten m/z -Wertes werden isoliert und fragmentiert) und der *data-independent acquisition* (DIA, alle Ionen des gewählten m/z -Bereichs werden simultan fragmentiert). An dieser Stelle soll detaillierter auf den in den Arbeiten verwendeten *data-dependent mode* eingegangen werden. In kurzen Intervallen wird während der chromatographischen Trennung ein *full scan* über einen großen Masse-zu-Ladungs-Bereich aufgenommen. Das bedeutet, dass alle Ionen den Quadrupol passieren und in der *C-Trap* gebündelt werden. Zwei Parameter, nämlich die *automatic gain control* (AGC) und die maximale Injektionszeit, begrenzen die Anzahl der Ionen, die zeitgleich von der *C-Trap* in die Orbitrap injiziert werden. Die maximale Injektionszeit beträgt etwa 100 ms für den *full scan*. Die Besonderheit einer Orbitrap ist, dass das Auflösungsvermögen

$$R = \frac{m}{\Delta m}$$

von der Messzeit abhängt. Je länger die Messzeit, desto höher ist das Auflösungsvermögen. Moderne Massenspektrometer erreichten bei 200 m/z erstmals R -Werte von 1.000.000 [65]. Im *full scan*, bzw. MS^1 -Spektrum werden die m/z -Werte der N intensivsten Ionen ermittelt (Top N), die den vom Nutzer festgelegten Kriterien entsprechen. Typische Einstellungen sind Ladungszustände von zwei und höher sowie ein minimales Signal-zu-Rausch-Verhältnis. Die Spannungen im Quadrupol werden dann so angepasst, dass er nur von solchen Ionen passiert werden kann, die dem ermittelten m/z -Wert plus/minus einer vom Nutzer eingestellten Toleranz entsprechen. Diese passieren die *C-Trap* und werden in der HCD-Zelle mit der ebenfalls manuell eingestellten Aktivierungsenergie fragmentiert. Die entstandenen Fragmentationen gelangen zurück in die *C-Trap* und von dort zur Orbitrap, wo die

Oszillationsfrequenzen und über eine Fourier-Transformation die Masse-zu-Ladungs-Verhältnisse der Fragmentionen ermittelt werden. Die Spannungen im Quadrupol werden im Anschluss so angepasst, dass das zweitintensivste Ion isoliert und fragmentiert werden kann. Dieser Vorgang wiederholt sich so lange, bis alle N Vorläuferionen fragmentiert und die zugehörigen Fragmentspektren berechnet wurden. Im Anschluss erfolgt der nächste *full scan*. Je nach eingestelltem Auflösungsvermögen (ergo Messdauer pro Spektrum) und der Anzahl N der zu fragmentierenden Ionen ergeben sich beliebig lange, sogenannte *duty cycles*. Als *duty cycle* bezeichnet man die Dauer, die für die Messung eines MS^1 -Spektrums und der anschließenden Fragmentspektren (MS/MS) erforderlich ist. Für ein noch besseres Signal-zu-Rausch-Verhältnis können mehrere Subspektren (auch als *micro scans* bezeichnet) akkumuliert werden, was gleichzeitig den *duty cycle* erhöht. Zusammenfassend kann man die Dauer eines Messzyklus' manuell beeinflussen, indem man das Auflösungsvermögen, die Anzahl N der zu fragmentierenden Vorläuferionen und die Anzahl der *micro scans* anpasst. Zwar sind hohe Werte für das Signal-zu-Rausch-Verhältnis, die Auflösung und die Fragmentspektren pro *full scan* wünschenswert, es muss jedoch stets die Peakbreite der Signale im Chromatogramm bedacht werden. Eine hohe chromatographische Auflösung entspricht sehr schmalen Peaks. Nimmt man die o.a. Einstellungen so vor, dass lange Messzyklen erreicht werden, so verringert das die Zahl der möglichen Fragmentspektren pro Zeiteinheit. Es muss also stets ein Kompromiss zwischen Auflösungsvermögen/Signal-zu-Rausch-Verhältnis und Informationsgehalt gefunden werden. Moderne Orbitrapgeräte schaffen bei niedrigem Auflösungsvermögen etwa bis zu 40 Spektren pro Sekunde [66].

d. Datenbanksuchen

Die Geschwindigkeit, mit der heutzutage Massenspektren und vor allem Fragmentspektren aufgenommen werden können, führt für eine einzige Analyse je nach Dauer der flüssigkeitschromatographischen Trennung zu Rohdaten, die im Gigabytebereich liegen und mehrere Zehntausend Spektren umfassen können. Da eine manuelle Auswertung aller Spektren nicht möglich ist, erfolgt die Annotation mithilfe von leistungsfähigen Algorithmen. Der typische Verlauf einer solchen Suche soll in diesem Kapitel kurz vorgestellt werden.

Jedes Fragmentspektrum ist mit dem m/z -Verhältnis des Vorläuferions verknüpft, aus dem es nach der Anregung entstanden ist. Eine Rohdatei wird zunächst nach allen Vorläufern gefiltert, welche ein minimales Signal-zu-Rausch-Verhältnis besitzen. Das ist wichtig, um zum

einen eine ausreichende Intensität der Signale im Fragmentspektrum zu gewährleisten und um zum anderen die Suchzeit zu minimieren. Weiterhin kann hier auch eine Begrenzung des m/z -Bereichs stattfinden. Nachdem die Rohdatei vorgefiltert wurde, erfolgt die eigentliche Datenbanksuche. Als Datenbank werden meistens FASTA-Dateien verwendet, die neben der Primärstruktur im 1-Buchstabencode auch die Proteinbezeichnung und den Organismus enthalten (**Abbildung 10**).

```
>sp|P02666|CASB_BOVIN Beta-casein OS=Bos taurus GN=CSN2 PE=1 SV=2
MKVLILACLVALALARELEELNVPGEIVESLSSEESITRINKKIEKFQSEEQQQTEDEL
QDKIHFAQTQSLVYPFPGPIPNLSPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPK
HKEMPFPPKYPVEPFTEQSLSLTLDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLSL
SQSKVLPVPQKAVPYPQRDMPPIQAFLLYQEPVLGPVRGPFPIIV
```

Abbildung 10 Fasta-Datei des β -Caseins (*Bos taurus*) Eine Datenbank enthält kanonische Aminosäuresequenzen, welche im 1-Buchstaben-Code ohne Leerzeichen aneinander gereiht sind. In der Kopfzeile sind außerdem wichtige Informationen wie der Proteinname inkl. Abkürzung und der Organismus enthalten.

Anhand der Enzym-Spezifität, welche sich aus der Experimentführung ergibt, generiert die Software aus diesen Listen potentiell in der Probe enthaltene Peptide. In *Bottom-Up* Ansätzen wird in der Regel das Enzym Trypsin verwendet, welches Proteine und Peptide C-terminal zu den basischen Aminosäuren Lysin und Arginin hydrolysiert. Die durch Trypsin generierten Peptide enthalten – mit Ausnahme des Protein-C-Terminus, welcher eine davon abweichende Aminosäure vorweisen kann – mindestens eine protonenaffine Aminosäure und lassen sich leichter ionisieren. Der tryptische *in silico* Verdau des β -Caseins ist in **Tabelle 1** gelistet.

Tabelle 1 Tryptischer *in silico* Verdau des β -Caseins (*Bos taurus*) Die Tabelle zeigt die Peptide über 500 Da, die generiert werden, wenn β -Casein (*Bos taurus*) mit Trypsin verdaut wird. Alle Sequenzen (bis auf den Protein-C-Terminus) enden auf eine basische Aminosäure. Die theoretische Sequenzabdeckung beträgt 94.3 %.

[M+H] ⁺	Position	Peptid-Sequenz
6359,256	129-184	YPVEPFTEQSLSLTLDVENLHLPLLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSK
5316,853	64-112	IHPFAQTQSLVYPFPGPIPNLSPQNIPPLTQTPVVVPPFLQPEVMGVSK
2646,299	17-40	ELEELNVPGEIVESLSSEESITR
2186,168	199-217	DMPPIQAFLLYQEPVLGPVR
1981,862	48-63	FQSEEQQQTEDELQDK
830,452	192-198	AVPYPQR
780,498	185-191	VLPVPQK
748,370	123-128	EMPPFK
742,450	218-224	GPFIIV
646,323	115-120	EAMAPK

Mit Ausnahme des Protein-C-Terminus enden die Sequenzen, wie es für tryptische Peptide charakteristisch ist, entweder auf Arginin und Lysin. Sollten alle diese Peptide in einem Versuch identifiziert werden, so würde die Sequenzabdeckung 94.3 % betragen und einen der Nachteile der BUP aufzeigen: eine vollständige Sequenzabdeckung wird nur in Ausnahmefällen erreicht. Da ein Großteil aller Proteine posttranslational modifiziert wird, werden vom Anwender vor der Suche dynamische und feste Modifikationen spezifiziert. Feste Modifikationen schließen das Vorkommen einer Aminosäure in ihrer nativen Form aus. Diese Einstellung wird meistens dann verwendet, wenn durch das Experiment bedingte Veränderungen, wie z.B. die Alkylierung von Cystein-Seitenketten, auftreten. Dynamische Modifikationen sind solche, die vorkommen können, aber nicht müssen. Funktionelle Gruppen, die entweder nicht an allen Aminosäuren eines Typs auftreten oder in biologischen Systemen reversibel angefügt werden, zählen ebenfalls zu den Modifikationen, die in einer Datenbanksuche als dynamisch klassifiziert werden. Im Falle des β -Caseins sind einige Serin-Seitenketten phosphoryliert [67]. Gibt man die Veresterung eines Phosphatrests an Serin als dynamische Modifikation an, so erhöht sich die Anzahl der theoretischen tryptischen Peptide von zehn (siehe **Tabelle 1**) auf 176. Würde man in diesem Beispiel gleichzeitig die Oxidation der Methionin-Seitenketten als dynamische Modifikation zulassen, erhöht sich die Anzahl der möglichen tryptischen Peptide auf 571. Die Gesamtheit aller möglichen Sequenzen wird auch als *search space* bezeichnet [68]. Nun kann es vorkommen, dass die ursprüngliche Probe bereits einer anderen unbekannten Protease ausgesetzt war und Peptide enthält, deren Termini nicht der tryptischen Spezifität entsprechen. Für diese Fälle gibt es die Option, nach semi-tryptischen Peptiden zu suchen, welche der Spezifität von Trypsin nur an einem der beiden Termini entsprechen. Das würde die o.a. Zahl der möglichen Peptide weiter erhöhen. Stelle man sich nun vor, dass für die Datenbanksuche eines Vollproteoms mehrere hunderte bis tausende Proteinsequenzen als Datenbank verwendet werden, wird auch ersichtlich, warum die Angabe von dynamischen Modifikationen und der Enzym-Spezifität mit Bedacht erfolgen sollte. Selbst leistungsfähigste Rechner brauchen für eine komplexe Datenbanksuche mitunter mehrere Tage. Nachdem die Liste aller möglichen Peptide und deren m/z -Werte generiert wurden, werden diese mit den vorher gefilterten Rohdaten verglichen. Stimmt ein ermitteltes Vorläufer-Ion mit einem der generierten, theoretischen Werte überein, werden die Fragmentspektren abgeglichen. Dazu wird aus dem entsprechenden Peptid der Datenbank mithilfe der in Kapitel c beschriebenen Nomenklatur

(Abbildung 2) ein theoretisches Tandem-MS-Spektrum generiert und mit dem experimentell gemessenen verglichen. Die Massentoleranz der Vorläufer- und Fragmentionen werden vom Nutzer festgelegt und können aufgrund der hohen Massengenauigkeit moderner Geräte unter 10 ppm eingestellt werden. Aus der Massenabweichung aller Werte und der Anzahl und Intensität der übereinstimmenden Fragmentionensignale berechnen Algorithmen die Güte (*score*) einer Identifikation. Stimmt ein gemessenes Spektrum mit einem theoretischen überein, spricht man auch von einem *peptide spectrum match* (PSM). Ist das gemessene Fragmentspektrum von geringer Qualität (schwache Intensitäten, wenige identifizierte Fragmente) oder das entsprechende Peptid gar nicht in der Datenbank enthalten, so kann es sich bei einem PSM auch um eine Falschzuweisung handeln (*false positive*). Um die Zahl der Falschzuweisungen zu minimieren, kann folgender statistischer Ansatz verfolgt werden: Aus allen theoretisch möglichen Peptiden werden weitere Sequenzen generiert, die beispielsweise der inversen Primärstruktur entsprechen und als *decoys* (Lockvögel) bezeichnet werden (Abbildung 11). Die aufgenommenen Spektren werden dann sowohl mit der richtigen Liste als auch mit den *decoys* abgeglichen. Stimmt ein Spektrum mit einem der *decoys* überein, so kann dieser PSM mit Sicherheit als *false positive* eingestuft werden.

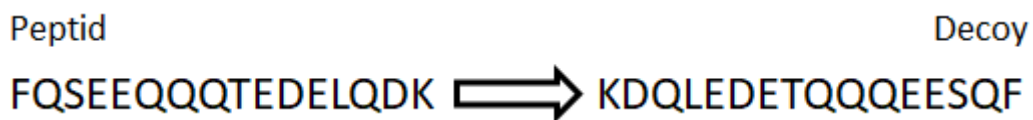


Abbildung 11 Generierung eines Lockvogels Für die Evaluierung der Ergebnisse der Datenbanksuche wird für jedes in der Datenbank gelistete Peptid ein Lockvogel generiert, dessen Primärstruktur der inversen Aminosäuresequenz entspricht.

Der Nutzer legt sogenannte *false discovery rates* (FDRs) fest, welche den Prozentsatz der Zuordnungen angeben, bei denen es sich statistisch gesehen um eine Falschzuweisung handelt. Üblicherweise werden eine strikte (0,01) und eine gelockerte (0,05) FDR angegeben. Mithilfe dieser Werte können die PSMs nun klassifiziert werden:

- (1) *high confidence*: Zuordnungen genügen den strikten Kriterien; statistisch gesehen handelt es sich bei 1 % der PSMs um Falschzuweisungen.
- (2) *medium confidence*: Zuordnungen genügen den gelockerten Kriterien; statistisch gesehen handelt es sich bei 5 % der PSMs um Falschzuweisungen.
- (3) *low confidence*: alle PSMs

Eine anschauliche Skizze zu dieser statistischen Betrachtung ist in **Abbildung 12** gezeigt.

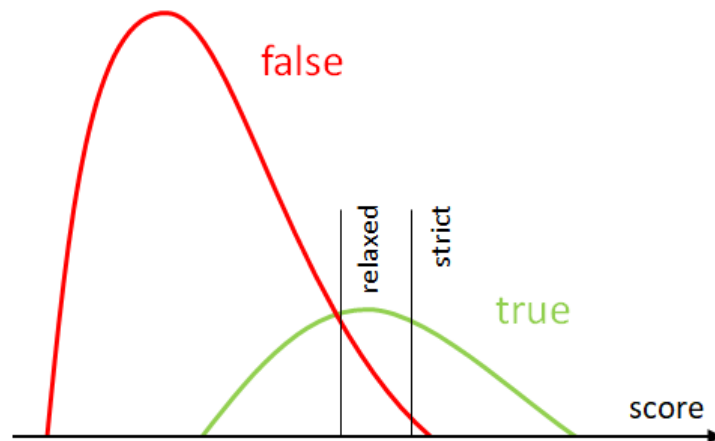


Abbildung 12 Graphische Darstellung der FDR Die false discovery rate (FDR) bestimmt den prozentualen Anteil aller PSM, die falsch sind. Eine niedrige FDR erhöht die Verlässlichkeit der Ergebnisse, schließt aber gleichzeitig auch richtige Zuweisungen aus [69].

Meistens werden die Ergebnisse einer Suche nach den *high confident peptides* gefiltert. Alle anderen PSM sollten im Zweifelsfall manuell annotiert und auf deren Richtigkeit geprüft werden. Vor allem bei sehr kleinen Datenbanken kann es passieren, dass der PSM mit der höchsten Güte eine wenig verlässliche Identifizierung darstellt [70].

Es ist wichtig zu erwähnen, dass der Nutzer bei einer konventionellen Datenbanksuche den *search space* definiert. Demzufolge kann nur identifiziert werden, was in diesem enthalten ist. Das bedeutet auch, dass nur solche Proteine identifiziert werden können, von denen man glaubt, dass sie in der Probe enthalten sind. Gleiches gilt für PTM. Eine Modifikation kann nur dann nachgewiesen werden, wenn sie vorher entweder als dynamische oder statische angegeben wurde. Es gibt jedoch Algorithmen, die nach einem *mass shift* suchen und so unerwartete Modifikationen oder fehlende Aminosäure-Sequenzen nachweisen können [71]. Eine solche Suche erhöht aber die Suchzeit weiter und wird daher meistens zielgerichtet für einzelne oder wenige Protein-Sequenzen durchgeführt.

6. Entwicklung einer Methode zur *Top-Down*-Analyse von N-glykosylierten Proteinen und Peptiden

a. Glykoproteine und –peptide in der Massenspektrometrie

Die Glykosylierung stellt eine häufig auftretende posttranslationale Modifizierung dar, die die strukturellen und physikochemischen Eigenschaften eines Proteins oder Peptides maßgeblich beeinflusst [72]. Die häufigsten Modifizierungen dieses Typs stellen O- und N-Glykosylierungen dar – eine Bezeichnung, die sich auf die Atome bezieht, über die die Bindung ausgebildet wird: das Sauerstoffatom der Hydroxylgruppen von Serin- und Threonin-Seitenketten oder das Stickstoffatom des primären Amids der Asparagin-Seitenkette. Seltener werden Glykane auch über ein Kohlenstoffatom des Tryptophans gebunden (C-Glykosylierung) [73, 74]. O-Glykane können verschiedene Kernstrukturen aufweisen [73], während N-Glykane eine einzige Kernstruktur besitzen, die vom Organismus abhängig ist [75]. Beim Menschen besteht diese aus zwei Einheiten N-Acetylglucosamin, drei Einheiten Mannose und einem Fucose-Molekül (**Abbildung 13**). Die Abbildung zeigt ebenfalls eine Beispielstruktur für eine solche Modifikation. Kohlenhydrate binden oft das zwei- bis vierfache der eigenen Masse an Wasser [76]. Dass sich dadurch die Hydrophobizität eines glykosylierten Analyten verringert, könnte die verringerte Wechselwirkung dergleichen mit den Umkehrphasen in der *reversed-phase liquid chromatography* (RP-LC) erklären. Aus diesem Grund werden Glykoproteine vor der flüssigkeiteschromatographischen Trennung meistens derivatisiert [77]. Auch in der Massenspektrometrie sind N-glykosylierte Analyten problematisch. Zum einen wird die Ionisierung durch die erschwerte Desolvatisierung während des ESI-Prozesses negativ beeinflusst [78] und zum anderen sind Glykoproteine oft mit einer Mischung verschiedener N-Glykanstrukturen modifiziert. Diese „Mikroheterogenität“ führt zu einer Anhäufung von Signalen, die, vor allem für intakte Proteine beobachtet, oft nicht aufgelöst werden kann [79]. Ein weiteres Problem dieser Modifikation ist, dass Protonen, welche im positiven Ionenmodus in erster Linie als Ladungsträger fungieren, die glykosidische Bindungen destabilisieren können. Durch Wechselwirkung des Protons mit den freien Elektronenpaaren des Sauerstoffatoms der glykosidischen Bindung wird somit die Aktivierungsenergie für die Dissoziation herabgesetzt [80].

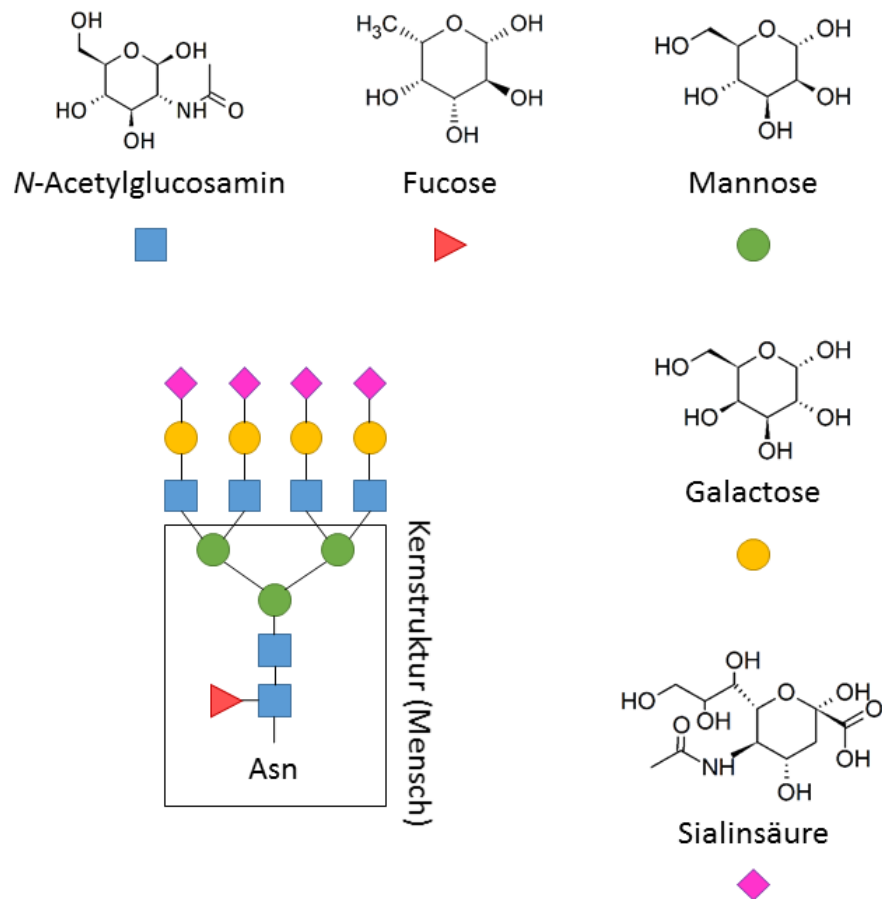


Abbildung 13 Struktureller Aufbau von N-Glykanen beim Menschen N-Glykane besitzen eine Kernstruktur, die sich für die meisten Organismen unterscheidet. Beim Menschen verbindet ein N-Acetylglucosamin das Protein oder Peptid mit dem Oligosaccharid. Dieses ist mit einem weiteren N-Acetylglucosamin und einem Molekül Fucose verknüpft. Das zweite N-Acetylglucosamin ist weiterhin mit einem Molekül Mannose verbunden, welches mit zwei weiteren Mannose-Einheiten verknüpft ist.

Dies hat zur Folge, dass kaum Bindungsbrüche im Peptidrückgrat erfolgen und Peptidfragmentsignale wenig intensiv sind. Stattdessen wird das Tandem-Massenspektrum eines glykosylierten Peptids oder Proteins von Signalen dominiert, die aus Bindungsbrüchen des Oligosaccharids resultieren [81]. Der Typ der Aktivierung (Stoßaktivierung oder Anregung mit energiereichen Elektronen) sowie die Aktivierungsdauer haben hierbei einen entscheidenden Einfluss auf die entstehenden Fragmente. Anregungsverfahren wie die ETD, welche mit angeregten Elektronen erfolgen, haben sich als für Glykanstrukturen geeignet erwiesen [80, 82].

b. Versuchsführung und Zielsetzung

Humanes Progranulin (PGRN) wurde mit Meprin- α oder Meprin- β verdaut. PGRN verfügt über mindestens vier glykosylierte Asparagin-Seitenketten [83] und eine weitere potentielle Glykosylierungsstelle (Konsensussequenz N-X-S/T). Progranulin ist eventuell an Prozessen in der Alzheimer-Krankheit beteiligt [84–86]. Es enthält außerdem sieben Sequenzen, die, durch Hydrolyse einzeln freigesetzt, als Granuline bezeichnet werden [87] (**Abbildung 14**).

10	20	30	40	50	60	70
MWTLVSWVAL	TAGLVAGTRC	PDGQFCPVAC	CLDPGGASYS	CCRPLLDKWP	TTLSRHI	GGP CQVDAHCSAG
80	90	100	110	120	130	140
HSCIFTVSGT	SSCCPFPEAV	ACGDGHCCP	RGFHCSADGR	SCFQ <u>RS</u> GNNS	VGAIQCPDSQ	FECPDFSTCC
150	160	170	180	190	200	210
VMVDGSWGCC	PMPQASCCED	RVHCCPHGAF	CDLVHTRCIT	PTGTHPLAKK	LPAQRTNRAV	ALSSSVMCPD
220	230	240	250	260	270	280
ARSRCPDGST	CCELPSGKYG	CCPMPNATCC	SDHLHCCPQD	TVCDLIQSKC	LSKENAT <u>TD</u> L	LTKLPAHTVG
290	300	310	320	330	340	350
DVKCDMEVSC	PDGYTCCRLQ	SGANGCCPFT	QAVCCEDHIH	CCPAGFTCDT	QKGTCE <u>Q</u> GPH	QVPWMEKAPA
360	370	380	390	400	410	420
HLSLPDPQAL	KRDV <u>PC</u> DNVS	SCPSSDTCCQ	LTSGEWGCCP	IPEAVCCSDH	QHCCPQGYTC	VAEGQCQ <u>R</u> GS
430	440	450	460	470	480	490
EIVAGLEKMP	ARRASLSHPR	DIGCDQHTSC	PVGQTCCPSL	GGSWACCQLP	HAVCCEDRQH	CCPAGYTCNV
500	510	520	530	540	550	560
KARSCEKEVV	SAQPATFLAR	SPHVGVDVE	CGEGHFCHDN	QTCCRDNRQG	WACCPYRQGV	CCADRRHCCP
570	580	590				
AGFRCAARGT	KCLR <u>RE</u> APRW	DAPLRDPALR	QLL			

Abbildung 14 Kanonische Sequenz des humanen Progranulin Im reifen PGRN fehlt das Signalpeptid (AA 1-17). Besonderes Augenmerk soll in dieser Abbildung auf die fett hervorgehobenen Abschnitte gelegt werden, welche die Begrenzungen der Granuline aufzeigt. Die Granuline werden anhand ihrer Position in der Sequenz nummeriert: Granulin-1 zwischen Gly-58 und Phe-113 (genauer Abschnitt unbekannt), Granulin-2 zwischen Ala-123 und Ile-179 usw. Ebenfalls unbekannt sind der C-terminale Abschnitt des Granulin-6 sowie die exakte Länge des Granulin-7. Die Abbildung zeigt die Granuline wie sie in der Uniprot-Datenbank (letzter Zugriff: 05.12.2017) hinterlegt sind.

Für vier Granuline konnte eine biologische Funktion nachgewiesen werden [88–90], weswegen das Studium der Prozessierung des Proteins und die dafür verantwortliche(n) Protease(n) von Bedeutung sind. Da bisher weder die exakte Sequenz aller Granuline, noch alle an der Degradierung beteiligte(n) Protease(n) identifiziert werden konnten, sollten in

diesem Versuch neben der Etablierung der Methode ebenfalls die Spaltprodukte des Progranulin/Meprin-Paares ermittelt werden.

c. Material und Methoden

i. Chemikalien und Zubehör

PNGase F wurde von New England BioLabs (Massachusetts, USA) erworben, Trypsin von Promega (Madison, USA), cOmplete Proteaseinhibitor von Hoffmann-La Roche (Basel, Schweiz), Salzsäure von Fluka (St. Gallen, Schweiz) und EDTA von Roth (Karlsruhe, Deutschland). Sep-Pak C18 Kartuschen stammten von Waters (Milford, USA), Lektin-SPE-Kartuschen und empfohlene Adsorptions- und Elutionspuffer von GALAB Technologies (Hamburg, Deutschland) und Amicon Zentrifugalfilter sowie TCEP von Merck (Darmstadt, Deutschland). Alle anderen Chemikalien wurden von Sigma-Aldrich (Steinheim, Deutschland) erworben. Entionisiertes Wasser wurde mittels eines arium611 VF Systems von Sartorius (Göttingen, Deutschland) generiert.

ii. Expression und Aufreinigung von Meprin- α - und - β und Progranulin

Rekombinantes humanes Meprin- α und humanes Progranulin wurden von der AG Becker-Pauly (CAU) aus Insektenzellen unter Verwendung des Baculovirus-Expressions-Systems nach dem Protokoll in [91] gewonnen. Meprin- α wurde mittels N-terminalem Strep-Tag und darauffolgender Affinitätschromatographie aufgereinigt und wie in [92] beschrieben mit Trypsin aktiviert. Progranulin wurde C-terminal mit einem His-Tag markiert und durch Ni-NTA Chromatographie aufgereinigt. Die Proteine wurden mittels Amicon Zentrifugalfilter (Merck, Darmstadt, Deutschland) konzentriert und in 25 mM HEPES Puffer (pH 7,5) bei -80 °C gelagert.

iii. Progranulin-Verdau mit Meprin- α - und - β

Das Progranulin wurde zunächst entweder mit Meprin- α oder Meprin- β verdaut. Anschließend wurden etwaige Disulfidbrücken reduziert, freie Cystein-Seitenketten alkyliert und Glykanstrukturen mithilfe von PNGase F entfernt. Die erhaltenen Proben wurden dann geteilt und entweder einem *Bottom-Up* oder *Top-Down* Ansatz unterzogen. Eine schematische Übersicht des *Workflows* ist in **Abbildung 15** gezeigt.

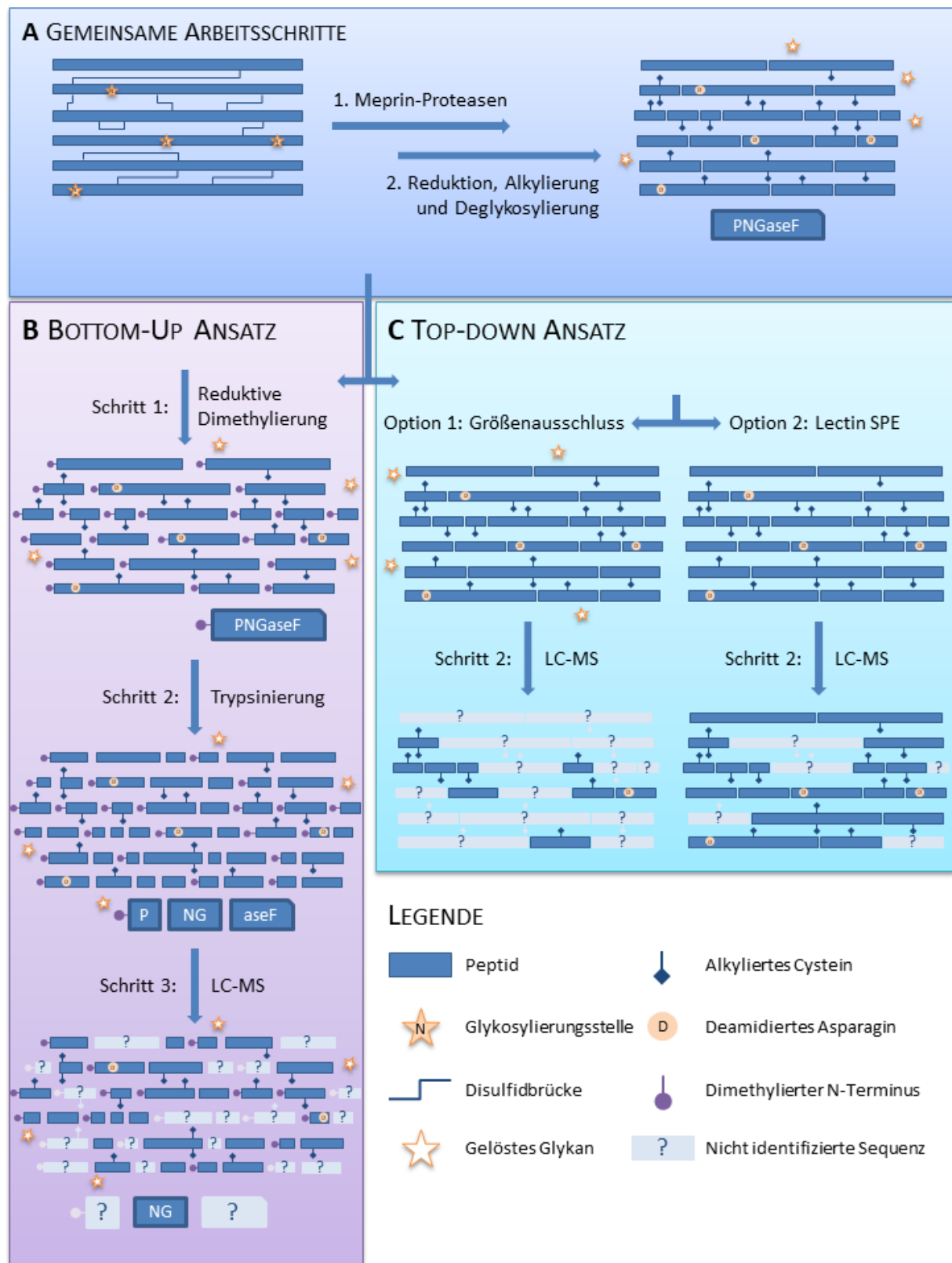


Abbildung 15 Workflow Die Arbeitsschritte für den Bottom-Up- und Top-Down-Ansatz waren zunächst identisch. Dies betrifft den Verdau mit Meprin- α/β , die Reduktion und Alkylierung sowie die Deglykosylierung mittels PNGase F (A). Im Anschluss wurden die Proben geteilt und entweder markiert und mit Trypsin verdaut (B) oder mittels Größenausschluss, bzw. Lectin-Festphasenextraktion von der nach wie vor intakten PNGase F und löslichen Glykanen sowie verbleibenden glykosylierten Peptiden abgetrennt (C).

Drei Mal 3,75 µg des intakten Progranulin wurden in 30 µL HEPES Puffer (50 mM, pH 7.4) gelöst. Zu diesen Proben wurden jeweils 300 ng (entspricht einem Enzym-zu-Substrat-Verhältnis von etwa 1:12) Meprin-α, Meprin-β oder das äquivalente Volumen HEPES Puffer (Kontrolle) hinzugegeben und für 30 min (Probenentnahme nach einer halben Stunde, nur Meprin-α) und 4 h (beide Proteasen) inkubiert. Nach dem Verdau wurden 0,1 µmol Tris(2-carboxyethyl)phosphin (TCEP) hinzugefügt und für 20 min auf 70 °C erhitzt. Den abgekühlten Proben wurden dann 0,1 µmol Iodoacetamid hinzugefügt und unter Ausschluss von Licht bei Raumtemperatur für 30 min inkubiert. Im Anschluss wurden die Ansätze mit 250 u PNGase F versetzt und bei 37 °C für 60 min auf dem Thermomixer geschüttelt. Nach diesem Schritt wurden die Proben aufgeteilt:

Zehn Mikroliter wurden für den *Bottom-Up* Ansatz entnommen. Zu diesen wurden Formaldehyd und Natriumcyanoborhydrid hinzugegeben, sodass deren finale Konzentration 20 mol/L betrug. Die Ansätze wurden bei 37 °C für drei Stunden inkubiert. Der Überschuss an Formaldehyd und Cyanoborhydrid wurde mittels C18 Sep-Paks nach den Angaben des Herstellers entfernt. Die von den Kartuschen eluierten Peptide wurden in einem Concentrator Plus von Eppendorf (Hamburg, Deutschland) eingeeengt und die Analyten in 50 µL HEPES Puffer wiederaufgenommen. Der pH-Wert wurde kontrolliert und mit 0,1 M Natronlauge auf etwa 7,0 eingestellt. Im Anschluss wurden jedem Ansatz 50 ng Trypsin hinzugefügt und bei 37 °C für vier Stunden verdaut. Das Enzym wurde durch die Zugabe von 1 µL Trifluoressigsäure (TFA) inaktiviert.

Das verbleibende Volumen der PNGase-F-behandelten Proben wurde halbiert und entweder durch Größenausschluss oder Lektin-SPE aufgereinigt (Kapitel 6.c.v und vi).

iv. Zeitabhängiger Verdau von Progranulin mit Meprin-α- und -β

Da im vorherigen Experiment (Kapitel 6.c.iii) tryptische Spaltstellen identifiziert werden konnten, die nicht mit der Spezifität von Meprin übereinstimmen, sollten im zweiten Ansatz etwaige Fremdproteasen inhibiert werden. Dazu wurde dem verwendeten HEPES Puffer (50 mM, pH 7.4) eine Tablette des EDTA-freien cOmplete Proteaseinhibitorcocktails zugesetzt. Laut Hersteller werden Metalloproteasen, zu denen Meprin-α und Meprin-β zählen, durch den Verzicht auf Ethylendiamintetraessigsäure (EDTA) nicht inhibiert. Zwei Mal 8.4 µg des intakten Progranulins wurden in diesem Puffer aufgenommen. Jeweils 600 ng (Enzym-zu-Substrat-Verhältnis etwa 1:14) Meprin-α, Meprin-β oder ein äquivalentes

Volumen HEPES Puffer (Kontrolle) wurden hinzugefügt und die Proben bei 37 °C inkubiert. Nach 2, 5, 10, 20, 30, 45 und 60 min wurden Teilproben von 12 µL entnommen, direkt mit 12 µL einer erhitzten (70 °C) Lösung von 160 mM EDTA und 10 mM TCEP vermischt und für 20 min auf 70 °C erhitzt. Der Denaturierung und Reduktion folgten die Alkylierung und Deglykosylierung wie in Kapitel 6.c.iii beschrieben.

v. Größenausschluss mittels Amicon Zentrifugalfilter

Um Rückstände von Glycerin zu entfernen, wurde die Filtereinheiten (Amicon Ultra-0.5 mL 30k, Merck, Darmstadt, Deutschland) zunächst mit 250 µL entionisiertem Wasser beladen und bei 14.000 × g für 20 min zentrifugiert. Anschließend wurden sie mit 250 µL BisTris-Puffer (pH 6,0) equilibriert (20 min bei 14.000 × g). Die kombinierten Filtrate wurden verworfen, die Einheiten mit der Probe beladen und bei 14.000 × g für 20 min zentrifugiert. Das Filtrat wurde bei -80 °C eingefroren und über Nacht gefriergetrocknet.

vi. Lektin-Festphasenextraktion

Die Kartuschen (AffiSep-WGA SPE, GALAB Technologies, Hamburg, Deutschland) wurden zwei Mal mit jeweils 500 µL Adsorptionspuffer (Bis-Tris, pH 6,0, ohne Konzentrationsangabe des Herstellers, GALAB Technologies, Hamburg, Deutschland) gespült. Im Anschluss wurde die Probe auf das Adsorbens aufgetragen und mittels Unterdruck langsam durch das Material geführt. Das Adsorbens wurde erneut zwei Mal mit jeweils 500 µL Adsorptionspuffer gespült und der Durchfluss sowie diese beiden Waschfraktionen miteinander kombiniert. Die vereinigten Lösungen wurden bei -80 °C eingefroren und über Nacht gefriergetrocknet. Vom Adsorbens zurückgehaltene Oligosaccharide sowie glykosylierte Peptide und Proteine wurden mit jeweils 500 µL Elutionspuffer (N-Acetylglucosamin in Bis-Tris-Puffer, pH 6,0, keine Konzentrationsangabe vom Hersteller) desorbiert und das Adsorbens mit zwei Säulenvolumina (zwei Mal 500 µL) Adsorptionspuffer regeneriert.

vii. LC-ESI MS (Ultimate3000, Dionex/Q Exactive Plus, Thermo Scientific)

Als Flüssigkeitschromatograph kam eine Ultimate3000 UPLC (Dionex, Idstein, Deutschland) zum Einsatz, die *online* an ein Q Exactive Plus Massenspektrometer (Thermo Scientific, Bremen, Deutschland) gekoppelt war. Die UPLC wurde mithilfe von DCMS Link 2.14 Build 3818 (Dionex) kontrolliert, das Massenspektrometer mit XCalibur 3.1.66.10 (Thermo Scientific). Als Laufmittel wurden 0,1 % TFA in 3 % ACN (Ladepuffer), 0,05 % FA (Eluent A)

0,04 % FA in 80 % ACN (Eluent B) verwendet. Wenn nicht anders angegeben, wurden die Flussraten auf $30 \mu\text{L min}^{-1}$ für den Autosampler und 250 nL min^{-1} für die Trennung der Analyten eingestellt. Für die *Bottom-Up* Proben wurde der Anteil des Laufmittels B innerhalb von 28 min von 4 auf 50 % erhöht. Für alle anderen Proben wurde der Anteil des Laufmittels B innerhalb von 58 min von 4 auf 60 % erhöht. Als Vorsäule wurde eine C4 Pepmap300 (Thermo Scientific, Bremen, Germany) verwendet (Partikelgröße: $5 \mu\text{m}$, Porengröße: 300 \AA , Säulenabmessungen: $300 \mu\text{m i.D.} \times 5 \text{ mm}$). Eine Accucore 150-C4 (Thermo Scientific, Bremen, Germany) kam als analytische Säule (Partikelgröße: $2,6 \mu\text{m}$, Porengröße: 150 \AA , Säulenabmessungen: $75 \mu\text{m i.D.} \times 500 \text{ mm}$) zum Einsatz. Aufgrund technischer Schwierigkeiten (Totvolumina) wurde auf die Absorptionsmessung im ultravioletten Bereich verzichtet. Die analytische Säule wurde online an das Massenspektrometer gekoppelt und die Probe mithilfe einer beschichteten Siliciumdioxidnadel (PicoEmitter SilicaTip, New Objective, USA) ionisiert. Die Ionisationsspannung betrug $1,55 \text{ kV}$. Der *full scan* wurde im Bereich $200\text{-}2.000 \text{ m/z}$ mit einem Auflösungsvermögen von 140.000 aufgenommen. Die AGC (*automatic gain control*), welche die Zahl der Ladungen begrenzt, wurde auf 3×10^6 eingestellt. Die maximale Injektionszeit betrug 100 ms . Für jede MS^1 -Messung wurden die drei intensivsten Ionen zur Fragmentierung ausgewählt und in einem Fenster von 3 m/z isoliert. Einfach geladene Ionen wurden ausgeschlossen. Die Vorläufer-Ionen wurden mittels HCD aktiviert. Wenn möglich, wurden sie bei normalisierten Kollisionsenergien (NCE) von 23 und 28 % fragmentiert. Bezüglich der Fragmentspektren wurden jeweils zwei *micro scans* akkumuliert. Das Auflösungsvermögen betrug auch hier 140.000 . Die AGC der Fragmentspektren wurde auf 1×10^5 eingestellt. Wurde ein Vorläuferion zur Fragmentierung ausgewählt, wurde dessen m/z -Wert für die nächsten 30 Sekunden nicht wieder ausgewählt. Diese Einstellung ermöglichte, dass im Chromatogramm weniger intensive Ionen fragmentiert wurden.

viii. Identifizierung von Peptiden und Spaltstellen mittels Datenbanksuche

Für die Datenbanksuche kam die Proteome Discoverer Software (version 2.2.0.388, Thermo Scientific, Bremen, Deutschland) zum Einsatz. Die Rohdaten wurden nach Vorläufern mit einer Masse zwischen 350 und 10.000 Da sowie einem minimalen Signal-zu-Rausch-Verhältnis von $1,5$ vorgefiltert. Als Datenbank wurde eine FASTA-Datei (Version: 01.01.2012) verwendet, die die Sequenzen des humanen Progranulins (Uniprot Zugriffsnummer: P28799), der humanen Meprin- α und Meprin- β (Uniprot Zugriffsnummern: Q16819 und

Q16820), der PNGase F des Organismus *Elizabethkingia miricola* (Uniprot Zugriffsnummer: P21163) sowie des *common Repository of Adventitious Proteins* (cRAP) enthielt. Es wurde kein Enzym spezifiziert. Die Peptidlänge wurde auf 6-144 Aminosäuren beschränkt. Die Toleranzen wurden zu 5 ppm (MS^1) und 0,02 Da (MS^2) festgelegt. Als feste Modifikation wurde die Carbamidomethylierung an Cystein-Seitenketten berücksichtigt. Die Oxidation von Methionin und die Deamidierung von Asparagin und Glutamin wurden als dynamische Modifikation zugelassen. Für die *Bottom-Up* Ansätze wurden mögliche PSM auf semi-tryptische Peptide eingegrenzt und die Dimethylierung des Peptid-N-Terminus vorausgesetzt (feste Modifikation). Als Nebenreaktion wurde die Dimethylierung auch an Lysin-Seitenketten zugelassen (dynamische Modifikation). Die PSM wurden mit FDR von 0,05 (relaxiert) und 0,01 (strikt) evaluiert. Ausschließlich Peptide mit einer hohen Zuverlässigkeit (*high confidence*) wurden berücksichtigt. Für eine genäherte Quantifizierung wurden die Peakflächen der Peptide mithilfe der Minora-Erweiterung für PD extrahiert, wenn diese im Chromatogramm über mindestens fünf Punkte verfügten, diese mindestens zwei isotope Signale aufwiesen und die maximale Retentionszeit Verzögerung dieser Isotope 12 Sekunden betrug. Alle bereits in der Kontrollprobe identifizierten Progranulin-Peptide wurden von der Auswertung ausgeschlossen.

d. Resultate und Diskussion

Für die Experimente stand zunächst nur Meprin- α zur Verfügung. Die Versuche mit Meprin- β wurden daher nicht zeitgleich durchgeführt, sondern zu einem späteren Zeitpunkt nachgeholt. Im ersten Versuch (Mep- α) wurde nach 30 min ein Teil der Probe entnommen und dem vorgestellten Workflow (**Abbildung 15**) unterzogen. Der 30-minütige Verdau mit Meprin- α stellte einen Vorversuch dar und wurde durchgeführt, um die Aktivität des Enzyms einschätzen und ein geeignetes Enzym-zu-Substrat-Verhältnis ermitteln zu können. Der vierstündige Verdau identifizierte im Vgl. zur 30-minütigen Inkubationszeit deutlich mehr Peptide und erzielte ebenfalls eine höhere Sequenzabdeckung (vgl. **Tabelle 3** und **Tabelle 4**). Da aufgrund der geringeren Spezifität des Meprin- α im Vergleich zu Meprin- β [93, 94] davon auszugehen war, dass Meprin- β nach 30-minütiger Inkubationszeit noch weniger Peptide generieren würde, wurde im nachfolgenden Versuch mit Meprin- β darauf verzichtet, eine Probe nach 30 Minuten zu entnehmen. Diese Vermutung wurde – wie **Tabelle 4** zeigt – bestätigt.

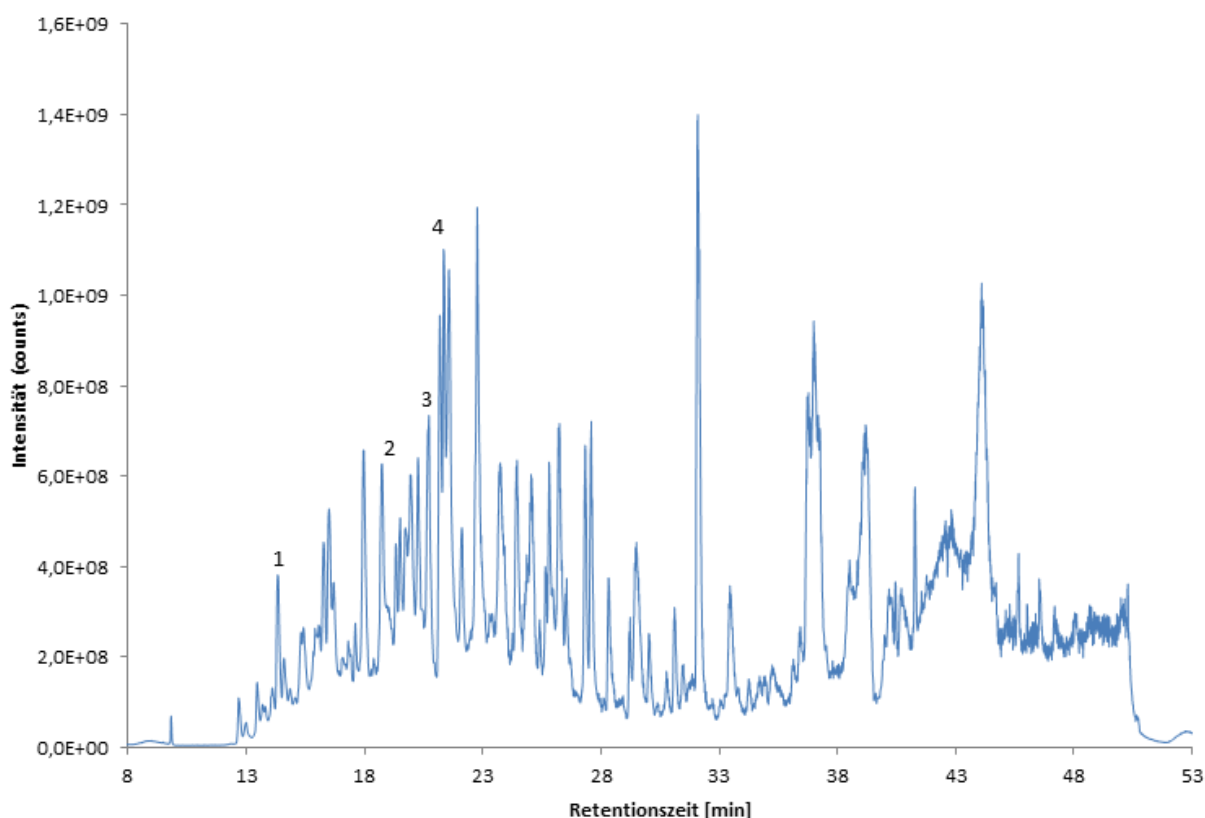


Abbildung 16 TIC des 30-minütigen Meprin- α -Verdaus (*Bottom-Up*) Einige von Meprin- α generierte PGRN-Peptide wurden beschriftet. Neben diesen traten auch Signale auf, die zu nicht dimethylierten Peptiden gehörten. Selbige wurden bei der Auswertung nicht berücksichtigt.

Zunächst sollen die Ergebnisse des 30-minütigen Meprin- α -Verdaus (Kapitel 6.c.iii) detaillierter betrachtet und die *Bottom-Up*- mit der *Top-Down*-Methode verglichen werden. In **Abbildung 16** ist das Chromatogramm des *Bottom-Up* Ansatzes gezeigt. Einige der intensiveren PGRN-Peptide, welche eine Meprin- α -generierte Spaltstelle besaßen (d.h. dimethyliert vorlagen), sind in **Tabelle 2** aufgelistet.

Tabelle 2 PGRN-Peptide nach 30-minütigem Meprin- α Verdau (*Bottom-Up*) Alle in dieser Tabelle gelisteten Peptide sind am N-Terminus dimethyliert. Die Reaktionsführung (**Abbildung 15**) bedingt, dass nur die Peptide als von Meprin- α verdaut gewertet werden können, welche über eine Dimethylierung verfügen.

#	Sequenz	m/z	z	Abweichung [ppm]	RT [min]
1	AKKLPAQR	332,56	3	-0,16	14,13
2	LPDPQALKR	365,23	3	6,24	18,75
3	QPATFLAR	466,27	2	1,68	20,62
4	AQPATFLAR	501,79	2	-0,46	21,65

Peptide, die nicht dimethyliert waren, wurden in dieser Auswertung nicht berücksichtigt und sind daher in **Abbildung 16** nicht annotiert. Einige der intensiveren Signale verfügten entweder über einen Ladungszustand von +1 und wurden daher nicht für eine Fragmentierung isoliert oder zeigten nach der Fragmentierung keine b- oder y-Ionen, wie sie für stoßaktivierte Peptide üblich sind.

Die Chromatogramme der *Top-Down* Ansätze (Lektin-Anreicherung und Größenausschluss) unterschieden sich von dem der *Bottom-Up* Methode (**Abbildung 17**).

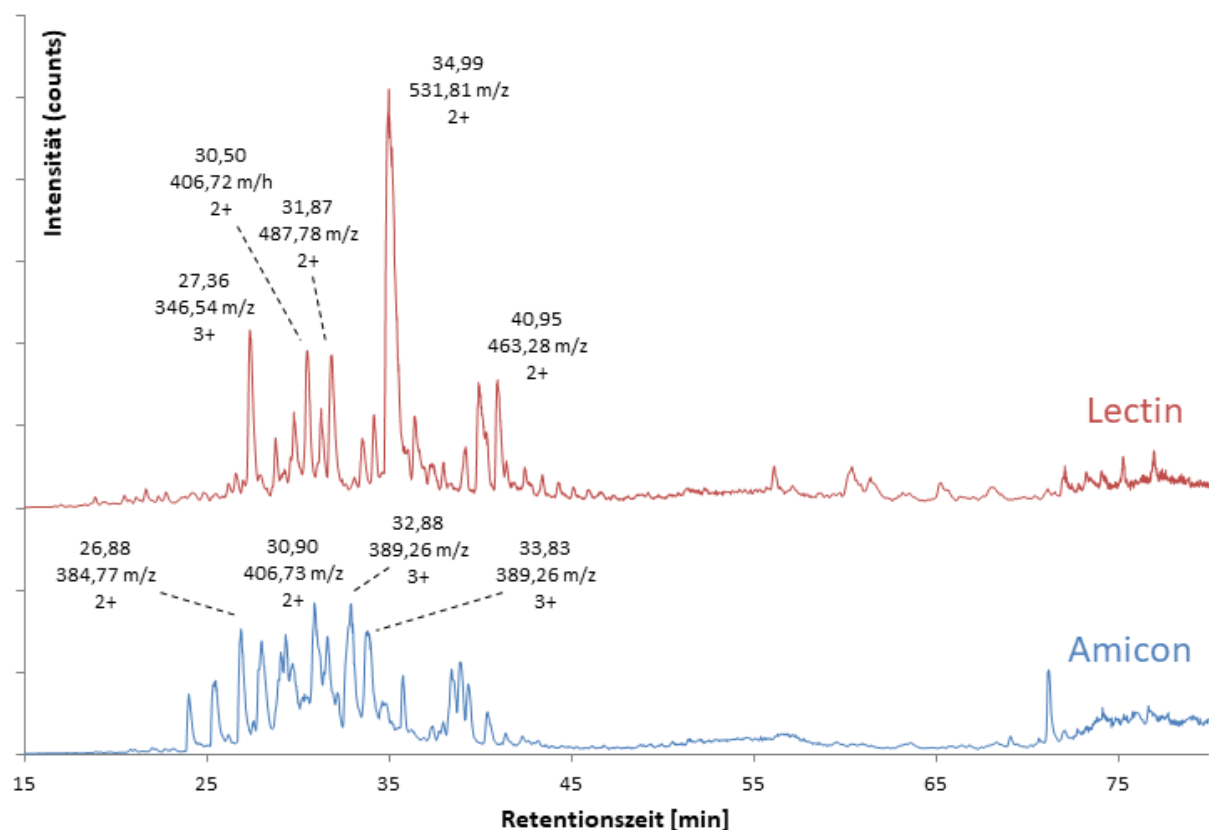


Abbildung 17 TIC des 30-minütigen Meprin- α -Verdaus (*Top-Down*) Trotz des etwas längeren Gradienten eluierten alle Analyten zwischen etwa 20 und 45 min. Im direkten Vergleich zum Bottom-Up Ansatz (**Abbildung 16**) fällt auf, dass die Chromatogramme insgesamt weniger Peaks zeigen. Diese Beobachtung könnte mit dem Verzicht auf den zusätzlichen tryptischen Verdau erklärt werden.

Es ist besonders auffällig, dass die *Top-Down* Chromatogramme im Vergleich über weniger Signale verfügen. Dies liegt vermutlich im Verzicht auf die Verwendung von Trypsin begründet, was zu einer insgesamt kleineren Anzahl an in der Probe enthaltenen Peptiden geführt hat. Die intensiveren, annotierten Signale konnten keinem Progranulin-Peptid zugeordnet werden. Es wird vermutet, dass diese Peptide von Proteinen stammen, die als Verunreinigung in der Progranulin-Probe enthalten waren. Es wäre zum Beispiel denkbar, dass es sich um Sequenzen des Organismus *Spodoptera frugiperda* handelt, in welchem

Progranulin überexprimiert worden ist. Da für diesen Organismus nur sehr wenige evaluierte Sequenzen zur Verfügung standen (< 100 Sequenzen in der *uniprot*-Datenbank) – von denen ribosomale Proteine den Großteil darstellten – wurde darauf verzichtet, Sequenzen des Organismus bei der Datenbanksuche einzubeziehen. Infolgedessen wurden den *Peaks* in der Abbildung keine Sequenzen zugeordnet. Wie schon im *Bottom-Up*-Ansatz wurden auch hier einfach geladene Verunreinigungen beobachtet, bei denen es sich unter anderem um unbekannte Polymere handeln könnte. Diese erkennt man sehr gut am periodischen Auftreten, z.B. im Lektin-Chromatogramm bei etwa 45 min. Ein Abgleich der Masse-zu-Ladungs-Verhältnisse der möglichen Polymere mit bekannten Verunreinigungen in der LC-MS Analytik führte zu keinen Ergebnissen.

Die Ergebnisse beider Ansätze sind in **Tabelle 3** zusammengefasst.

Tabelle 3 Ergebnisse der Datenbanksuchen (30-minütiger Meprin- α Verdau) Vergleicht man zunächst die beiden Zweige des Top-Down Ansatzes, so fällt auf, dass, obwohl die gleiche Menge an Material verwendet wurde, die Methode des Größenausschlusses zu signifikant weniger Identifizierungen (Peptide, Spaltstellen) und einer geringeren Sequenzabdeckung geführt hat. Vergleicht man den Bottom-Up Ansatz mit der Methode der Lektin-Festphasenextraktion, so wurde in etwa die gleiche Anzahl an Peptiden identifiziert, jedoch deutlich mehr Spaltstellen im Top-Down Ansatz. Gleichzeitig war auch die Sequenzabdeckung dieser Methode um 8,0 % höher. (*Die Sequenzabdeckung wurde für die Progranulin-Sequenz ohne Signalpeptid berechnet.)

Ansatz	Peptide	davon <i>unique</i>	Spaltstellen	Sequenz- abdeckung (%)*	Pot. Glykosylierungss tellen
BUP	53	46	35	33.0	1
TDP	16	14	24	26.2	1
Amicon					
TDP	54	51	55	41.0	1
Lectin					

Für die beiden *Top-Down* Varianten, Größenausschluss (Kapitel 6.c.v) und Lektin-Festphasenextraktion (Kapitel 6.c.vi), wurde die gleiche Menge an Ausgangsmaterial verwendet. Dennoch konnten mithilfe der Amicon-Kartuschen nur etwa ein Drittel der Peptide und knapp die Hälfte der Spaltstellen im Vergleich zur Lektin-Methode identifiziert werden. Die mittels Lektin-SPE beobachteten Spaltstellen und zugehörige Sequenzen sind in **Abbildung 18** gezeigt. Während die resultierende Sequenzabdeckung der

Größenausschlussmethode nur etwa 26 % betrug, wurden mittels Festphasenextraktion 41 % erreicht. Die Ergebnisse könnten in einem höheren Probenverlust während des Größenausschlusses begründet liegen. Die Filtereinheiten sind für Ansätze bis 500 µL ausgelegt (kleinere Einheiten werden vom Hersteller nicht vertrieben), während die Proben deutlich kleinere Volumina (< 100 µL) maßen. Um überhaupt einen Durchtritt der Probe zu gewährleisten, mussten die Volumina vor der Zentrifugation erhöht werden. Damit die Filter während des Vorgangs nicht austrocknen, befindet sich unterhalb der Membran ein Reservoir, das einen Teil der Probe zurückhält. Auf diesem Weg könnte möglicherweise ein Teil der Probe verloren gegangen sein. Ungeachtet der Dimension der Filtereinheiten stellt sich daher die Frage, ob die verwendeten Filtereinheiten, die einen kompletten Durchtritt der Probe pauschal verhindern, für eine Aufreinigung geringer Probenmengen überhaupt geeignet sind. Daher sollte die Wahl der Filtereinheiten infrage gestellt werden und nicht zwangsläufig die Methode des Größenausschlusses. Ein erneutes Experiment mit für kleine Volumina, bzw. Probenmengen geeigneten Filtern wäre notwendig, um die beiden Methoden (Affinitätschromatographie und Größenausschluss) miteinander vergleichen zu können. Da die Probenmengen des Progranulins begrenzt waren, wurden diese Versuche jedoch nicht durchgeführt und im weiteren Verlauf der Experimente (Kapitel 6.c.iv) auf die Verwendung der Zentrifugalfilter verzichtet. Die deglykosylierten Proben wurden fortan ausschließlich mittels Lektin-Festphasenextraktion aufgereinigt, welche die gewünschten Resultate erzielten.



Abbildung 18 Identifizierte Spaltstellen mittels *Top-Down-Ansatz* (Lektin-SPE) nach dem 30-minütigen Verdau (*Meprin-α*) Der Großteil aller Spaltstellen wurde im C-terminalen Teil des Proteins identifiziert. Nur sehr wenige, ausschließlich kurze Peptide vom N-terminalen Teil des Proteins wurden detektiert. Dies könnte ein Hinweis dafür sein, dass der N-Terminus im nativen Protein für Proteasen nicht oder erschwert zugänglich ist.

Vergleicht man die Daten der Lektin-*Top-Down* mit der klassischen *Bottom-Up* Methode (**Tabelle 3**), so wurden in beiden etwa 50 Peptide identifiziert. Die Anzahl der identifizierten Spaltstellen (*Top-Down*: 55, *Bottom-Up*: 35) und die Sequenzabdeckung (*Top-Down*: 41,0 %, *Bottom-Up*: 33,0 %) sprechen jedoch für die neu etablierte Methode. Die geringere Anzahl an identifizierten Spaltstellen lässt sich durch die Verwendung der Arbeitsprotease begründen. Während eines klassischen *Bottom-Up* Ansatzes werden die Spaltprodukte durch eine Arbeitsprotease (hier: Trypsin) weiter verdaut und somit artifizielle Spaltstellen generiert. Um diese sogenannten neo-N-Termini von denen zu unterscheiden, die aus dem Verdau mit dem zu untersuchenden Enzym (hier: Meprin- α/β) stammen, müssen die Termini der Meprin-Peptide vorher mittels Modifizierung markiert werden (hier: durch reduktive Dimethylierung). In der anschließenden Auswertung können nur solche Termini als von Meprin generierte Spaltstellen gezählt werden, die diese Markierung tragen. Aus diesem Grund wurde in der Datenbanksuche des *Bottom-Up* Ansatzes die Dimethylierung des N-

Terminus als feste Modifikation festgelegt. Die C-Termini dieser Peptide wurden nicht als Meprin-Spaltstelle gewertet, da nicht ausgeschlossen werden konnte, dass diese von einer durch Trypsin oder etwaige Fremdproteasen, die in der Probe enthalten sein könnten, katalysierten Hydrolyse stammen. Für die Identifizierung einer Spaltstelle ist es daher notwendig, das N-terminal modifizierte Peptid (enthält P1' bis P4') nachzuweisen. Eine anschauliche Erklärung ist in **Abbildung 19** gezeigt.

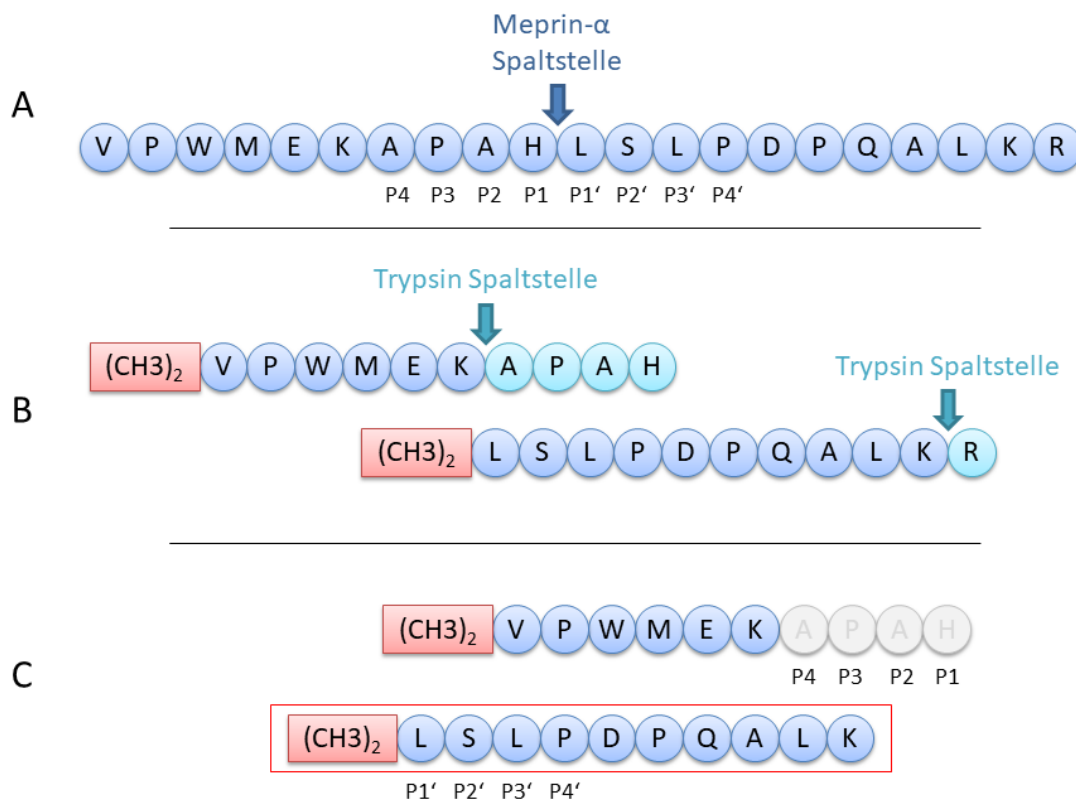


Abbildung 19 Prozessierung der Meprin-Peptide im Bottom-Up-Ansatz Nachdem Progranulin von Meprin-α verdaut worden ist (A), werden die N-Termini mittels Dimethylierung markiert. Im Anschluss erfolgt der Verdau mit Trypsin (B). Da dem ersten Peptid nach der „Trypsinierung“ der C-terminale Teil (ausgegraut) fehlt, der zur ursprünglichen Spaltstelle gehört, muss zwangsläufig das zweite, hier rot markierte Peptid nachgewiesen werden, um die Meprin-Spaltstelle zu identifizieren (C). Ansonsten wird die in (A) generierte Spaltstelle nicht nachgewiesen.

Dies könnte eine Erklärung für die deutlich kleinere Anzahl an identifizierten Spaltstellen für den *Bottom-Up* Ansatz sein. Die geringere Sequenzabdeckung lässt sich ebenfalls mit der Verwendung der Arbeitsprotease erklären. Wenn ein Protein oder Peptid sehr reich an basischen Aminosäuren ist, so enthält es sehr viele potentielle tryptische Spaltstellen. Bei einem Doppelverdau (Meprin/Trypsin) ist die Chance sehr groß, dass Analyten so weit hydrolysiert werden, dass Peptide entstehen, die zu klein sind (<6 AA), um in der LC-MS-

Analyse inklusive Datenbanksuche berücksichtigt zu werden. **Abbildung 19** zeigt ein solches semi-tryptisches Peptid (ausgegraute Sequenz APAH). Je kleiner ein Peptid ist, desto geringer ist zum einen die Chance, dass die Sequenz einzigartig ist und zum anderen die Güte des PSM [95].

Betrachtet man die Ergebnisse des 4-stündigen Verdau (siehe Chromatogramme im Anhang unter Kapitel 10.a), welcher sowohl mit Meprin- α als auch Meprin- β durchgeführt wurde, so ergibt sich ein ähnliches Bild (**Tabelle 4**). Meprin- α besitzt – wie bereits angesprochen – eine geringere Spezifität [93, 94], weswegen eine größere Anzahl an Spaltstellen zu erwarten war.

Tabelle 4 Ergebnisse der Datenbanksuchen (4-stündiger Verdau) Vergleicht man wieder die beiden Zweige des Top-Down Ansatzes (**Abbildung 15**) als erstes, so ergibt sich ein dem 30-minütigem Verdau ähnelndes Bild: die Methode des Größenausschlusses weist signifikant weniger Identifizierungen (Peptide, Spaltstellen) und eine geringere Sequenzabdeckung auf. Dies gilt für Meprin- α und Meprin- β gleichermaßen. Der Bottom-Up Ansatz identifizierte ähnlich viele Peptide, doch auch hier war die TDP in Bezug auf identifizierte Spaltstellen und Sequenzabdeckung überlegen. Weiterhin wurden für den Meprin- α -Verdau mittels Lektin-SPE vier potentielle Glykosylierungsstellen abgedeckt.

Ansatz	Peptide	davon <i>unique</i>	Spaltstellen	Sequenz- abdeckung (%)	Pot. Glykosylierungss tellen
Meprin- α					
BUP	189	182	116	74,8	2
TDP	72	71	75	54,2	1
Amicon					
TDP	178	170	135	90,8	4
Lectin					
Meprin- β					
BUP	97	91	66	54,0	1
TDP	11	11	18	17,5	0
Amicon					
TDP	99	93	93	66,3	2
Lectin					

Betrachtet man zunächst die Ergebnisse der beiden *Top-Down* Ansätze, so wurden mittels Größenausschluss auch hier deutlich weniger Peptide und Spaltstellen identifiziert, was für

den bereits erwähnten Probenverlust spricht. In der weiteren Diskussion sollen daher nur noch die Daten der Lektin-Festphasenextraktion beleuchtet werden. Für Meprin- α wurden nach vier Stunden 178 (TDP) und 189 (BUP) Peptide identifiziert, die aus 135 (TDP), bzw. 116 (BUP) Spaltstellen generiert wurden. Die Sequenzabdeckung des *Top-Down* Ansatzes war mit fast 91 % rund 16 % höher als die der klassischen *Bottom-Up* Methode. Ein weiterer nicht weniger wichtiger Aspekt waren die identifizierten PTM. Progranulin weist insgesamt fünf potentielle Glykosylierungsstellen auf. Für vier dieser Asparagine (Asn-118, Asn-265, Asn-368 und Asn-530) wurde eine Glykosylierung bereits nachgewiesen [83]. Eine fünfte (Asn-236) könnte ebenfalls modifiziert vorliegen. Im *Top-Down* Ansatz konnten vier der o.a. Asparagin-Seitenketten in der deamidierten Form in von Meprin- α generierten Peptiden nachgewiesen werden, während es für die *Bottom-Up* Methode nur zwei (Asn-265 und Asn-368) waren. Die Peptide wurden nicht manuell, sondern automatisch über die Datenbanksuche mit den in Kapitel 6.c.viii spezifizierten Suchparametern identifiziert. Die entsprechenden Massenspektren sind im Anhang zu finden (siehe **Abbildung 50**, **Abbildung 51**, **Abbildung 52** und **Abbildung 53** in Kapitel 10.a). Es sei darauf hingewiesen, dass die Asparagine in der deamidierten Form nachgewiesen wurden. Die Deamidierung ist das Resultat der enzymatischen Deglykosylierung und kein Nachweis für das Vorhandensein einer ursprünglichen Glykosylierung, da die Deamidierung auch für native Asparagine auftreten kann [96]. Ein möglicher Ansatz für zukünftige Arbeiten wäre es, die enzymatische Deglykosylierung in schwerem Wasser durchzuführen, um die aus der Reaktion resultierende Aspartat-Seitenkette von nativen Carbonsäuregruppen unterscheiden zu können.

Neben der Anzahl an identifizierten Peptiden und Spaltstellen sowie der erreichten Sequenzabdeckung sollten ebenfalls die Spezifitäten von Meprin- α und - β untersucht werden. Die von den Enzymen präferierten Aminosäuren der Positionen P4- bis P4'-Positionen sind für Peptide und intakte Substrate ermittelt worden [93]. Die Spaltstellen der mit der Lektin-basierten *Top-Down* Methode identifizierten Peptide wurden für jede Aminosäure an allen acht möglichen Positionen aufsummiert. Dieser Datensatz wurde verwendet, um ein IceLogo [97] zu generieren (**Abbildung 20**). Es sei an dieser Stelle darauf hingewiesen, dass die Häufigkeit der Aminosäuren an den einzelnen Positionen für ein einzelnes Substrat von der relativen Häufigkeit in der Primärstruktur abhängt. Wird ein Substrat untersucht, das beispielsweise sehr reich an Serin ist, so wird diese Aminosäure ebenfalls häufig in unmittelbarer Nähe einer Spaltstelle auftauchen. Die in **Abbildung 20**

dargestellten IceLogos sollten somit nicht unbedingt als repräsentativ für die generelle Spezifität von Meprin- α und Meprin- β gewertet werden. Damit ein IceLogo aussagekräftig sein kann, werden deutlich größere Datenmengen von einer Vielzahl von Substraten benötigt. Diese Grundlage war in diesem Experiment nicht gegeben, weshalb die IceLogos nur begrenzte Aussagekraft besitzen.

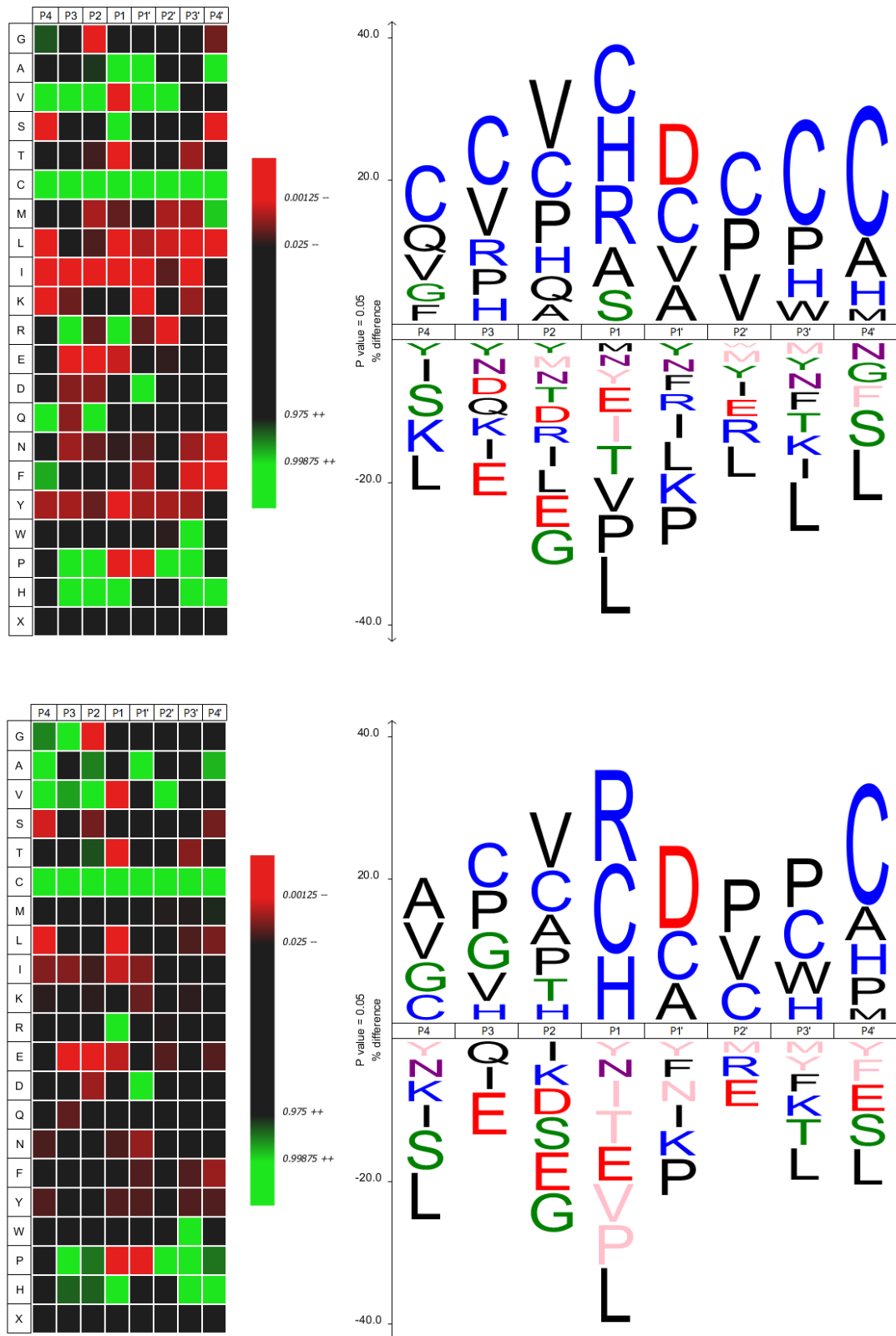


Abbildung 20 IceLogo der ermittelten Spaltstellen (Lektin-SPE, 4-stündiger Verdau) Meprin-α (oben) und -β (unten) weisen eine Präferenz für saure und kleine Aminosäuren in der P1'-Position auf. Die erhobenen Daten stimmen größtenteils mit diesen Ergebnissen überein. Einzig die basische Aminosäure Arginin in der P1-Position passt nicht zu den Ergebnissen von Becker-Pauly et al. [93].


Die Größe der Darstellung der Aminosäuren im Einbuchstabencode (Kapitel 3) in den IceLogos gibt die Häufigkeit wieder. Meprin- α und - β hydrolysieren Peptidbindungen von Peptiden und Proteinen präferiert vor den sauren Aminosäuren Aspartat und Glutamat sowie Alanin, Serin und Threonin. Dabei ist Meprin- β geringfügig spezifischer. Die Anzahl der in diesen Experimenten identifizierten Spaltstellen war für Meprin- β kleiner als die von Meprin- α , was mit diesem Befund übereinstimmt. Für beide Enzyme kam Cystein an allen Positionen sehr häufig vor. Dies hängt jedoch mit der relativen Häufigkeit der Aminosäure in der Primärstruktur des Progranulins zusammen (88 von 576 Aminosäuren sind Cysteine), nicht mit der Spezifität der Proteasen. Oft wird die Spezifität von Proteasen über die an den Positionen P1 und P1' auftretenden Aminosäuren definiert. An letzterer wurde Aspartat für beide Enzyme am häufigsten beobachtet, was mit den Arbeiten von Becker-Pauly *et al.* übereinstimmt [93]. Die zweithäufigste Aminosäure an der P1'-Position war Cystein, was aus dem o.a. Grund nicht weiter beachtet werden soll. Für Meprin- α waren die nächsthäufigsten Aminosäuren an dieser Stelle Valin und Alanin, was der ermittelten Spezifität der Protease für native Substrate entspricht. Für Meprin- β war Alanin die dritthäufigste Aminosäure. Auch dies stimmt mit den Befunden von Becker-Pauly *et al.* überein. Eine weitere Gemeinsamkeit war, dass eine Prolin-Seitenkette sowohl an der P1 als auch der P1'-Position ungünstig war, während diese Aminosäure sehr häufig an der P2'-Stelle beobachtet wurde. Eine ebenfalls an P2' häufig vorkommende Aminosäure war Valin, was mit vorherigen Ergebnissen konsistent ist. Der vermutlich größte Unterschied der im Experiment bestimmten Spezifität zu den von Becker-Pauly *et al.* erhobenen Daten ist das Vorkommen der basischen Aminosäure Arginin an der P1-Position. Während hauptsächlich Alanin, Serin und Glutamat an P1 beobachtet wurden, waren in den hier identifizierten Spaltprodukten des Progranulins neben Cystein die Aminosäuren Arginin und Histidin auffällig häufig vertreten. Die Bindung des Substrats im aktiven Zentrum von Meprin- β erfolgt höchstwahrscheinlich über Coulomb-Wechselwirkungen mit einer Lysin- (Lys-213) und zwei Arginin-Seitenketten (Arg-146 und Arg-184) [94]. Es wird davon ausgegangen, dass Arg-184 die Seitenketten der Aminosäuren an P1 und P1' im Substrat bindet und für die hohe Spezifität für saure Aminosäuren nahe der Spaltstelle verantwortlich ist. In der Meprin- α Protease, deren Sequenz zu etwa 55 % mit der von Meprin- β übereinstimmt, sind die basischen Aminosäuren an den äquivalenten Positionen durch Tyrosin und Prolin (R146Y, R184Y und K213P) substituiert. Dies würde erklären, warum Meprin- α weniger spezifisch ist als Meprin- β . Da

Arg-146, Arg-184 und Lys-213 im Meprin- β für die Wechselwirkung mit dem Substrat verantwortlich zu sein scheinen [94], ist eine Arginin-Seitenkette an der P1-Position zumindest für Meprin- β unwahrscheinlich. Es wurde daher angenommen, dass die Spaltstellen eventuell durch die Aktivität von Fremdproteasen entstanden sind. Eine Möglichkeit wären Verunreinigungen von Trypsin, welches zur Aktivierung von Meprin- α verwendet wurde und zusammen mit Meprin- α in die Probe(n) gelangt sein könnte. Da jedoch auch in der Kontrollprobe PGRN-Peptide nachgewiesen werden konnten, kann nicht ausgeschlossen werden, dass auch die Progranulin-Probe Verunreinigungen unbekannter Enzyme enthielt oder PGRN bereits während der Gewinnung teilweise verdaut worden ist. Um die Aktivität unbekannter Proteasen während des Versuchs zu verhindern, wurde in weiteren Versuchen (zeitabhängiger Verdau, Kap. 6.c.iv) auf Proteaseinhibitoren zurückgegriffen. Diese inhibieren laut Hersteller ein breites Spektrum an Serin- und Cystein-Proteasen verschiedener Organismen. Der Inhibitor-Mix war frei von EDTA und sollte somit Metalloproteasen, wie Meprin- α und - β , laut Hersteller nicht beeinflussen.

Im zweiten Experiment sollte die enzymatisch katalysierte Hydrolyse beider Proteasen, Meprin- α und Meprin- β , zeitabhängig und vor allem nach sehr kurzen Inkubationszeiten untersucht werden. Hierzu wurden den Ansätzen (wie in Kapitel 6.c.iv beschrieben) zu verschiedenen Zeitpunkten Aliquote entnommen. Die zu untersuchenden Zeitpunkte wurden zu 2, 5, 10, 20, 30, 45 und 60 min festgelegt und lieferten somit sieben Datenpunkte. Da nach sehr kurzen Inkubationszeiten bereits Peptide infolge der Hydrolyse entstehen können, waren die sehr kurzen Zeitpunkte von zwei und fünf Minuten von besonderem Interesse. In Proteolyseexperimenten werden oftmals feste Zeitpunkte gewählt, obwohl zeitnah auftretende Intermediate weiter degradiert werden können. Auf diese Weise können wichtige Informationen verloren gehen. Beispielsweise ist der Nachweis der als erstes auftretenden Spaltprodukte ein Hinweis dafür, welche Peptidbindungen im Substrat unmittelbar hydrolysiert werden und somit, welche Bereiche im nativen Zustand für die Protease zunächst verfügbar sind. Für die im zeitaufgelösten Verdau (Kapitel 6.c.iv) identifizierten Peptide wurden die chromatographischen Peaks integriert. Die erhaltenen Werte wurden, sofern ein Peptid zu mehreren Zeitpunkten identifiziert wurde, miteinander verglichen. So konnte eine Aussage darüber getroffen werden, zu welchem Zeitpunkt das Peptid auftritt und ob es eventuell weiter degradiert wurde. Die nach zwei Minuten Meprin- α -Verdau intensivsten Peptide sind in **Tabelle 5** aufgelistet.

Tabelle 5 Top-Down Methode: Peptide mit intensivstem Signal nach 2-minütigem Verdau (Meprin-α) Die Sequenzen wurden nach ihrer Häufigkeit nach zwei Minuten Verdau sortiert. Neben sehr kurzen Peptiden wurden fünf Sequenzen identifiziert, die über 50+ Aminosäuren verfügen. Diese wurden im Verlauf der Reaktion weiter hydrolysiert. Von den 21 gelisteten Peptiden waren nur die Sequenzen 2 und 5 über den gesamten Zeitraum nahezu konstant. Die Intensität der anderen 19 Peptide ändert sich mit der Zeit. In den meisten Fällen nahmen sie ab. Aber auch die Zunahme von Analyten wurde beobachtet (z.B. Peptid 3, 7 und 13).

Peptidsequenz		Inkubationszeit						
		2 min	5 min	10 min	20 min	30 min	45 min	60 min
1	P28799 [342-350] [Q].VPWMEKAPA.[H]	6.00E+08	8.57E+08	1.08E+09	1.37E+09	1.59E+09	1.43E+09	1.15E+09
2	P28799 [342-351] [Q].VPWMEKAPAH.[L]	1.96E+08	2.43E+08	2.27E+08	2.85E+08	3.00E+08	2.51E+08	1.59E+08
3	P28799 [352-362] [H].LSLPDPQALKR.[D]	1.03E+08	1.62E+08	3.36E+08	4.21E+08	5.63E+08	6.96E+08	7.67E+08
4	P28799 [580-585] [R].WDAPLR.[D]	9.94E+07	3.49E+07	2.15E+07	1.57E+07	5.12E+07	7.18E+07	2.27E+07
5	P28799 [351-362] [A].HLSLPDPQALKR.[D]	7.41E+07	1.19E+08	1.91E+08	1.54E+08	1.49E+08	1.19E+08	7.48E+07
6	P28799 [342-352] [Q].VPWMEKAPAH.L[S]	5.23E+07	6.85E+07	6.36E+07	5.96E+07	4.62E+07	3.66E+07	2.71E+07
7	P28799 [353-362] [L].SLPDPQALKR.[D]	5.06E+07	5.70E+07	9.51E+07	1.17E+08	1.96E+08	2.42E+08	2.22E+08
8	P28799 [438-501] [S].HPRDIGCDQHTSCPVGQTCCPSLGGSWACCQL PHAVCCEDRQHCCPAGYTCNVKARSCEKEVVS.[A]	4.47E+07	1.22E+07	2.22E+07	7.48E+06	1.15E+07	1.15E+06	8.39E+06
9	P28799 [581-589] [W].DAPLRDPAL.[R]	4.42E+07	1.52E+07	5.06E+06	6.27E+05	1.35E+05	1.41E+05	
10	P28799 [342-350] [Q].VPWMEKAPA.[H]	4.04E+07	1.55E+07	2.72E+07	1.87E+07	2.05E+07	1.54E+07	1.20E+07
11	P28799 [441-497] [R].DIGCDQHTSCPVGQTCCPSLGGSWACCQLPHA VCCEDRQHCCPAGYTCNVKARSCEK.[E]	3.64E+07	5.49E+06	4.49E+06				
12	P28799 [441-493] [R].DIGCDQHTSCPVGQTCCPSLGGSWACCQLPHA VCCEDRQHCCPAGYTCNVKAR.[S]	3.38E+07	4.37E+06	5.06E+06				
13	P28799 [341-350] [H].QVPWMEKAPA.[H]	2.92E+07	4.43E+07	6.71E+07	1.04E+08	1.40E+08	1.64E+08	1.66E+08
14	P28799 [438-501] [S].HPRDIGCDQHTSCPVGQTCCPSLGGSWACCQL PHAVCCEDRQHCCPAGYTCNVKARSCEKEVVS.[A]	2.88E+07	5.82E+05	9.27E+05	3.98E+05	5.23E+05		2.18E+06
15	P28799 [269-277] [T].DLLTKLPAH.[T]	1.76E+07	3.04E+07	6.90E+07	8.52E+07	1.27E+08	1.72E+08	2.04E+08
16	P28799 [351-359] [A].HLSLPDPQA.[L]	1.39E+07	1.99E+07	3.78E+07	3.62E+07	3.93E+07	3.04E+07	2.03E+07
17	P28799 [342-351] [Q].VPWMEKAPAH.[L]	1.18E+07	1.01E+07	1.50E+07	1.28E+07	1.42E+07	9.07E+06	6.68E+06
18	P28799 [341-351] [H].QVPWMEKAPAH.[L]	1.16E+07	1.75E+07	1.95E+07	2.92E+07	3.84E+07	4.87E+07	4.24E+07
19	P28799 [441-501] [R].DIGCDQHTSCPVGQTCCPSLGGSWACCQLPHA VCCEDRQHCCPAGYTCNVKARSCEKEVVS.[A]	1.12E+07	2.55E+05	4.52E+05				
20	P28799 [580-589] [R].WDAPLRDPAL.[R]	1.04E+07	3.84E+06	8.66E+05	6.62E+04	5.49E+04	3.38E+04	
21	P28799 [278-298] [H].TVGDVKCDMEVSCPDGYTCCR.[L]	1.01E+07	1.96E+06	6.51E+05	2.01E+05			



Niedrigster Wert
Höchster Wert

In dieser Liste sind die Peptide 8, 11, 12, 14 und 19 besonders auffällig, da sie mit 50 und mehr Aminosäuren deutlich länger sind als die übrigen Einträge. Da die Intensität dieser Sequenzen mit der Zeit abnahm, kann vermutet werden, dass sie entweder weiter hydrolysiert oder aufgrund der erhöhten Temperatur (37 °C) während des Verdaus zumindest teilweise oxidiert wurden. Entsprechend modifizierte Peptide konnten jedoch nicht identifiziert werden. Auch eine Deamidierung der Asparagin- und Glutamin-Seitenketten ist möglich, aufgrund der sehr langen Halbwertszeit mehrerer Tage *in vitro* jedoch unwahrscheinlich [96]. Betrachtet man den Ursprung aller Peptide (zweite Spalte in

der Tabelle), so stammen mehr als die Hälfte (11) der abundantesten Spaltprodukte aus der Region (AA 337-364 im Progranulin) zwischen den Bereichen von Granulin-4 und Granulin-5 (**Abbildung 21**). Drei weitere konnten dem Protein-C-Terminus zugeordnet werden. Die langen Abschnitte rührten mit einer Ausnahme (Peptid 12) von einer Hydrolyse zwischen den Bereichen von Granulin-5 und Granulin-6 (AA 418-441 im Progranulin) und zwischen Granulin-6 und Granulin-7 (AA 496-517 im Progranulin) her. Auffällig war, dass Peptide des Protein-N-Terminus nach zwei Minuten nur mit niedriger Intensität auftraten, was dafür sprechen könnte, dass der N-Terminus im nativen Substrat für Proteasen nicht zugänglich ist. Die Daten deuten weiterhin darauf hin, dass der Protein-C-Terminus sowie der Abschnitt zwischen Granulin-4 und Granulin-5 in der nativen Struktur ausgesetzt, d.h. im gefalteten Protein oberflächlich angeordnet sind und somit bevorzugt hydrolysiert werden. Die den Peptiden in **Tabelle 5** zugehörigen Spaltstellen zeigten elf Arginine in der P1-Position an, welche für die bereits angesprochene Verunreinigung an Trypsin sprechen könnte. Die C-terminalen Sequenzen 9 und 20 wiesen außerdem Arginin-Seitenketten an der P1'-Position auf. Progranulin wurde mittels C-terminalem His-Tag und anschließender Nickel-basierter Affinitätschromatographie aufgereinigt. Diese Termini entstanden vermutlich während der Entfernung des Tags und sollen nicht weiter berücksichtigt werden. Die übrigen Spaltstellen stimmten mit der Spezifität von Meprin- α überein.

10	20	30	40	50	60	70
MWTLVSWVAL	TAGLVAGTRC	PDGQFCPVAC	CLDPGGASYS	CCRPLLDKWP	TTLRHLGGP	CQVDAHCSAG
80	90	100	110	120	130	140
HSCIFTVSGT	SSCCPFPEAV	ACGDGHHCCP	RGFHCSADGR	SCFQ^QRS^NGNN^S	VGAIQCPDSQ	FECPDFSTCC
150	160	170	180	190	200	210
VMVDGSWGCC	PMPQASCCED	RVHCCPHGAF	CDLVHTRCIT	PTGTHPLAKK	LPAQRTNRAV	ALSSSV^MMCPD
220	230	240	250	260	270	280
ARSRCPDGST	CCELP^SSGKYG	CCPM^NNATCC	SDHLHCCPQD	TVCDLIQSKC	LSKENATTDL	LT^KLP^AHTVG
290	300	310	320	330	340	350
DVKCDMEVSC	PDGYTCCRLQ	SGAWGCCPFT	QAVCCEDHIH	CCPAGFTCDT	QKGTCEQGPH	QVPWMEKAPA
360	370	380	390	400	410	420
HLSLPDPQAL	KRDVPCD^NNVS	SCPSSDTCCQ	LTSGEWGCCP	IPEAVCCSDH	QHCCPQGYTC	VAEGQCQ^RRS
430	440	450	460	470	480	490
EIVAGLEKMP	ARRASLSHPR	DIGCDQHTSC	PVGQTCCPSL	GGSWACCQLP	HAVCCEDRQH	CCPAGYTCNV
500	510	520	530	540	550	560
KARSCEKEVV	SAQPATFLAR	SPHVGK^DDVE	CGEGHFCHDN	QTCCRDNRQG	WACCPYRQGV	CCADRRHCCP
570	580	590				
AGFRCAARGT	KCLRREAPRW	DAPLRDPALR	QLL			

Abbildung 21 In der nativen Struktur zugängliche Regionen des PGRN (Meprin- α -Verdau nach 2 min) Die Spaltstellen, aus denen die Peptide in **Tabelle 5** entstanden sind, wurden mit dunkelblauen Pfeilen gekennzeichnet. Die zugehörigen Sequenzen (z.T. überlappend) sind hellblau hervorgehoben. Die Granulin-Abschnitte sind mit Fettschrift markiert.

Die häufigsten von Meprin- α generierten Spaltprodukte nach 60-minütigem Verdau wurden ebenfalls nach dem Wert ihrer Intensität im Chromatogramm (XIC) in absteigender Reihenfolge sortiert und in **Tabelle 6** dargestellt. Die meisten dieser Sequenzen wurden mit fortschreitender Inkubationszeit intensiver, was vermuten lässt, dass sie von Meprin- α nicht weiter degradiert wurden. Interessant ist, dass die beiden Tabellen insgesamt zwölf Peptide gemeinsam haben, von denen zehn aus der Region (AA 337-364 im Progranulin) stammen. Dies lässt zwei Folgerungen zu: zum einen scheint dieser Abschnitt im nativen PGRN für Proteasen zugänglich zu sein und zum anderen werden die dort entstehenden Peptide nicht weiter hydrolysiert. Die Sequenzen des Übergangs zwischen Granulin-4 und Granulin-5 enthalten mit einem Aspartat (Asp-356) zwar eine potentielle Schnittstelle, jedoch weist diese ebenfalls eine Prolin-Seitenkette an der P1-Position auf, welche in vorherigen Studien nur selten beobachtet wurde [93]. Dieser Umstand könnte deren Stabilität und damit verbunden hohe Intensität (XIC) im Chromatogramm erklären. Eine weitere stabile Sequenz

(AA 580-585 im Progranulin) stammt vom Protein-C-Terminus. Peptide, die in **Tabelle 5** noch nicht gelistet waren und somit nach zwei Minuten Verdau noch nicht oder nur sehr wenig intensiv beobachtet wurden, sind mit einem Stern markiert.

Tabelle 6 Top-Down Methode: Peptide mit intensivstem Signal nach 60-minütigem Verdau (Meprin-α) Die Sequenzen wurden nach ihrer Häufigkeit nach 60 Minuten Verdau sortiert. Wie zu erwarten war, traten keine längeren Peptide mehr auf, was auf eine sequentielle Hydrolyse der Sequenzen 8, 11, 12, 14 und 19 aus **Tabelle 5** hinweist. Ein Großteil der gezeigten Peptide war bereits nach zwei Minuten Verdau besonders dominant. Alle anderen wurden mit einem * markiert und heben die Abschnitte hervor, die erst im späteren Verlauf des Verdaus an Intensität gewannen.

Peptidsequenz		Inkubationszeit						
		2 min	5 min	10 min	20 min	30 min	45 min	60 min
1	P28799 [342-350] [Q].VPWMEKAPA.[H]	6.00E+08	8.57E+08	1.08E+09	1.37E+09	1.59E+09	1.43E+09	1.15E+09
2	P28799 [352-362] [H].LSLPDPQALKR.[D]	1.03E+08	1.62E+08	3.36E+08	4.21E+08	5.63E+08	6.96E+08	7.67E+08
3	P28799 [353-362] [L].SLPDPQALKR.[D]	5.06E+07	5.70E+07	9.51E+07	1.17E+08	1.96E+08	2.42E+08	2.22E+08
4	P28799 [269-277] [T].DLLTKLPAH.[T]	1.76E+07	3.04E+07	6.90E+07	8.52E+07	1.27E+08	1.72E+08	2.04E+08
5	P28799 [341-350] [H].QVPWMEKAPA.[H]	2.92E+07	4.43E+07	6.71E+07	1.04E+08	1.40E+08	1.64E+08	1.66E+08
6	P28799 [342-351] [Q].VPWMEKAPAH.[L]	1.96E+08	2.43E+08	2.27E+08	2.85E+08	3.00E+08	2.51E+08	1.59E+08
7	P28799 [47-56] [L].DKWPTTSLRH.[L]*	6.81E+06	9.17E+06	1.73E+07	3.21E+07	6.33E+07	1.18E+08	1.09E+08
8	P28799 [46-55] [L].LDKWPTTSLR.[H]*	1.90E+06	5.35E+06	1.72E+07	2.77E+07	4.44E+07	6.76E+07	7.50E+07
9	P28799 [351-362] [A].HLSLPDPQALKR.[D]	7.41E+07	1.19E+08	1.91E+08	1.54E+08	1.49E+08	1.19E+08	7.48E+07
10	P28799 [266-277] [N].ATTDLLTKLPAH.[T]*	9.72E+06	1.46E+07	2.76E+07	3.19E+07	4.13E+07	4.68E+07	4.75E+07
11	P28799 [341-351] [H].QVPWMEKAPAH.[L]	1.16E+07	1.75E+07	1.95E+07	2.92E+07	3.84E+07	4.87E+07	4.24E+07
12	P28799 [267-277] [A].TTDLLTKLPAH.[T]*	5.43E+06	8.19E+06	1.63E+07	1.87E+07	2.61E+07	3.38E+07	3.83E+07
13	P28799 [46-56] [L].LDKWPTTSLRH.[L]*	8.85E+05	2.28E+06	4.48E+06	7.47E+06	1.53E+07	2.82E+07	3.03E+07
14	P28799 [342-352] [Q].VPWMEKAPAH.[S]	5.23E+07	6.85E+07	6.36E+07	5.96E+07	4.62E+07	3.66E+07	2.71E+07
15	P28799 [355-362] [L].PDPQALKR.[D]*	3.38E+06	6.27E+06	1.11E+07	1.43E+07	2.50E+07	3.25E+07	2.62E+07
16	P28799 [580-585] [R].WDAPLR.[D]	9.94E+07	3.49E+07	2.15E+07	1.57E+07	5.12E+07	7.18E+07	2.27E+07
17	P28799 [351-359] [A].HLSLPDPQA.[L]	1.39E+07	1.99E+07	3.78E+07	3.62E+07	3.93E+07	3.04E+07	2.03E+07
18	P28799 [353-363] [L].SLPDPQALKRD.[V]*	4.82E+06	5.21E+06	7.95E+06	1.14E+07	1.76E+07	2.03E+07	1.99E+07
19	P28799 [421-435] [S].EIVAGLEKMPARRAS.[L]*	2.21E+06	3.72E+06	5.51E+06	8.72E+06	1.18E+07	1.65E+07	1.73E+07
20	P28799 [420-435] [G].SEIVAGLEKMPARRAS.[L]*	2.59E+06	4.00E+06	6.89E+06	9.53E+06	1.36E+07	1.63E+07	1.71E+07
21	P28799 [419-428] [R].GSEIVAGLEK.[M]*	5.35E+06	3.74E+06	5.28E+06	6.55E+06	1.40E+07	2.11E+07	1.37E+07
22	P28799 [342-350] [Q].VPWMEKAPA.[H]	4.04E+07	1.55E+07	2.72E+07	1.87E+07	2.05E+07	1.54E+07	1.20E+07
23	P28799 [341-347] [H].QVPWMEK.[A]*	2.08E+06	2.83E+06	9.90E+06	5.28E+06	9.09E+06	1.34E+07	1.04E+07



Drei dieser Abschnitte stammen vom Protein-N-Terminus. Deren Intensität war nach zwei Minuten relativ gering, was die Vermutung unterstützt, dass Teile des Proteins erst im Verlauf des Verdaus zugänglich werden. Ebenfalls mit einem Stern markiert sind Peptide der Region, welche zwischen Granulin-5 und Granulin-6 (AA 419-435 im Progranulin) zu finden

sind. Dies könnte ein Hinweis dafür sein, dass die potentiellen Spaltstellen C-terminal zu Granulin-5 erst zugänglich werden, nachdem der Abschnitt C-terminal zur Spaltstelle Asp-441 (siehe **Abbildung 21**) proteolytisch entfernt worden ist. Oder mit anderen Worten: sobald die langen Sequenzen aus **Tabelle 5** hydrolytisch entfernt worden sind, wird der Bereich AA 419-435 für Proteasen zugänglich. Infolgedessen wären die Peptide 19-21 (**Tabelle 6**) erst entstanden, nachdem der C-terminale Abschnitt entfernt wurde. Dieser Befund weist ebenfalls darauf hin, dass der C-Terminus im nativen Progranulin für Proteasen zugänglich ist und andere Abschnitte des Proteins schützt. Um diese These zu stützen, könnte der PGRN-Verdau mit anderen Endoproteasen wiederholt werden. Ungeachtet der Spezifität sollten nach kurzen Inkubationszeiten ebenfalls primär Peptide des C-Terminus beobachtet werden. Die partielle Hydrolyse als Möglichkeit der Strukturanalyse von Proteinen wurde in der Literatur bereits beschrieben [98–100]. Die enzymatische Hydrolyse von gefalteten Proteinen ist zumindest teilweise sterisch gehindert. Dennoch besitzen diese meist Elemente, die für Proteasen verfügbar sind. Dazu zählen z.B. ausgesetzte *Linker*-Regionen und flexible Schleifen [98]. Die Identifizierung der Hydrolyseprodukte nach kurzen Inkubationszeiten kann also helfen, diese Elemente eines Proteins zu identifizieren.

Die Spaltstellen der in **Tabelle 6** gelisteten Abschnitte sind in **Abbildung 22** dargestellt. Interessant ist, dass alle 23 Peptide außerhalb der vermuteten Granulinsequenzen liegen. Das aus der Hydrolyse eines Granulins entstandene Peptid mit dem höchsten XIC umfasste die Aminosäuren Leu-299 bis Gln-341. Es war nach einer Inkubationszeit von 60 Minuten mit einem Wert von $6,71E+5$ mehr als drei Größenordnungen weniger intensiv als Peptid 1 in **Tabelle 6**. Eine mögliche Erklärung für diese Beobachtung ist das Fehlen von Disulfidbrücken in den Intersektionen der Granuline und eine damit verbunden hohe Zugänglichkeit für Proteasen. Obwohl Cystein mit einer relativen Abundanz von über 15 % die häufigste Aminosäure des Progranulins ist, kommt sie in den Bereichen Gln-114 bis Gly-122, Thr-180 bis Ser-205, Ser-262 bis Gly-280, Gln-337 bis Asp-363, Arg-418 bis Asp-441, Lys-497 bis Lys-517 sowie im C-Terminus (Arg-574 bis Leu-593) nicht vor. Die Granuline sind demzufolge überproportional reich an Cystein-Seitenketten. Von den im reifen Progranulin vorhandenen 88 Cysteinen kommen nur sechs im N-terminalen Teil des Proteins vor. Somit verteilen sich 82 Cystein-Reste auf die sieben Granuline (390 Aminosäuren insgesamt), was einer effektiven Häufigkeit von über 21 % entspricht. Für zwölf der 88 Cysteine wurde eine Verknüpfung in Disulfid-Brücken nachgewiesen [101]. Um aufzuklären, wie viele der Cysteine

in Disulfidbrücken verknüpft sind, wurde die Kontroll-Probe ebenfalls analysiert (**Abbildung 23**).

10	20	30	40	50	60	70
MWTLVSWVAL	TAGLVAGTRC	PDGQFCPVAC	CLDPGGASYS	CCRPLLDKWP	TTLSRHLGGP	CQVDAHCSAG
				↑↑	↑↑	
80	90	100	110	120	130	140
HSCIFTVSGT	SSCCPFPEAV	ACGDGHCCP	RGFHCSADGR	SCFQ R SGNNS	VGAIQCPDSQ	FECPDFSTCC
150	160	170	180	190	200	210
VMVDGSWGCC	PMPQASCCED	RVHCCPHGAF	CDLVHTRCIT	PTGTHPLAKK	LPAQRTNRAV	ALSSSV M C P D
220	230	240	250	260	270	280
ARSRCPDGST	CCELPSGKYG	CCPMPNATCC	SDHLHCCPQD	TVCDLIQSKC	LSKEN A T T D L	LTKLPAH T VG
					↑↑	↑
290	300	310	320	330	340	350
DVKCDMEVSC	PDGYTC C R L Q	SGANGCCPFT	QAVCCEDHIH	CCPAGFTCDT	QKGTCEQ G PH	QVPWMEK A P A
					↑↑	↑↑
360	370	380	390	400	410	420
HLSLPDPQAL	K R DVPCDNVS	SCPSSDTCCQ	LTSGEWGCCP	IPEAVCCSDH	QHCCPQGYTC	VAEGQCQ R G S
↑↑	↑	↑↑				↑↑
430	440	450	460	470	480	490
EIVAGLEKMP	ARRASLSHPR	DIGCDQHTSC	PVGQTCCPSL	GGSWACCQLP	HAVCCEDRQH	CCPAGYTCNV
↑	↑					
500	510	520	530	540	550	560
KARSCEKEVV	SAQPATFLAR	SPHVGKDV E	CGEGHFCHDN _	QTCCRDNRQG	WACCPYRQGV	CCADRRHCCP
570	580	590				
AGFRCAARGT	KCLRREAPRW	DAPLRDPALR	QLL			
	↑	↑				

Abbildung 22 Top-Down Methode: Intensivste Peptide nach 60-minütigem Verdau (Meprin-α) und zugehörige Spaltstellen Die nach zwei Minuten intensiven langen Abschnitte (siehe **Tabelle 5**) wurden nach einer Stunde Inkubation nicht mehr oder nur noch sehr wenig intensiv beobachtet. Stattdessen wurden Peptide detektiert, die vom Protein-N-Terminus stammen. Dieser Befund deutet darauf hin, dass der N-Terminus im nativen Protein nicht zugänglich ist und vermutlich vom C-terminalen Abschnitt geschützt wird.

Die Kontrolle wurde nicht mit Meprin-Proteasen behandelt, jedoch ebenfalls reduziert, alkyliert und deglykosyliert. Da das Massenspektrum zwischen etwa 19 und 24 min Retentionszeit nicht ausreichend hoch aufgelöst war, konnte anhand dieser Ergebnisse leider keine Aussage über die Anzahl der Disulfid-Brücken des intakten Proteins gemacht werden.

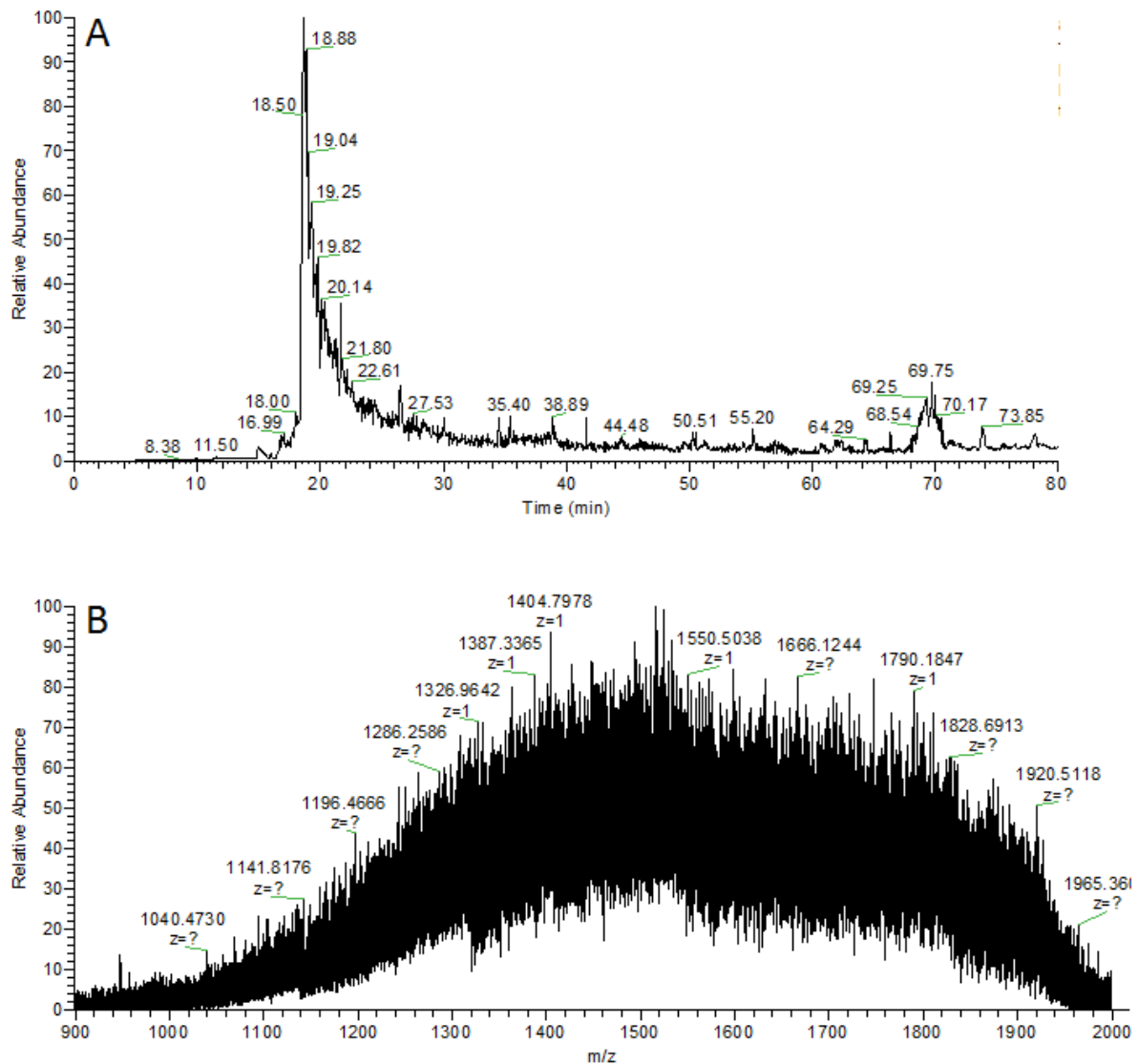


Abbildung 23 LC-ESI Massenspektrum des intakten, deglykosylierten Progranulins Die ungefähre Intaktmasse des Progranulins beträgt etwa 61.787 Da (berechnet ohne Signalpeptid 1-17 oder Modifikationen). Im Chromatogramm ist ein intensives Signal zu sehen, welches jedoch Tailing zeigte. Im Massenspektrum können die Signale zu dieser Retentionszeit nicht aufgelöst werden.

Das Fehlen von intensiven Peptiden, deren Ursprung innerhalb der Granulinabschnitte liegt, könnte auf die angesprochene überdurchschnittliche Häufigkeit von Cystein zurückzuführen sein. Erfahrungsgemäß erschwert diese Aminosäure die Ionisation von Peptiden, weswegen in der Literatur bereits Methoden vorgestellt wurden, die die Ionisationseffizienz erhöhen sollen [102, 103]. Es kann außerdem nicht ausgeschlossen werden, dass es neben den bereits annotierten [101] weitere Disulfidbrücken im Progranulin gibt, die die Hydrolyse in diesen Abschnitten erschwert haben könnten.

Ein weiterer Aspekt, der untersucht werden sollte, war die Entwicklung der längeren Sequenzen aus **Tabelle 5**. Da diese nach 60 Minuten deutlich an Intensität verloren hatten,

kann davon ausgegangen werden, dass sie als Intermediate auftraten und entweder weiter degradiert oder anderweitig modifiziert worden sind. Zwei mögliche detektierte Folgeprodukte des Peptids 11 aus **Tabelle 5** sind in **Abbildung 24** dargestellt.

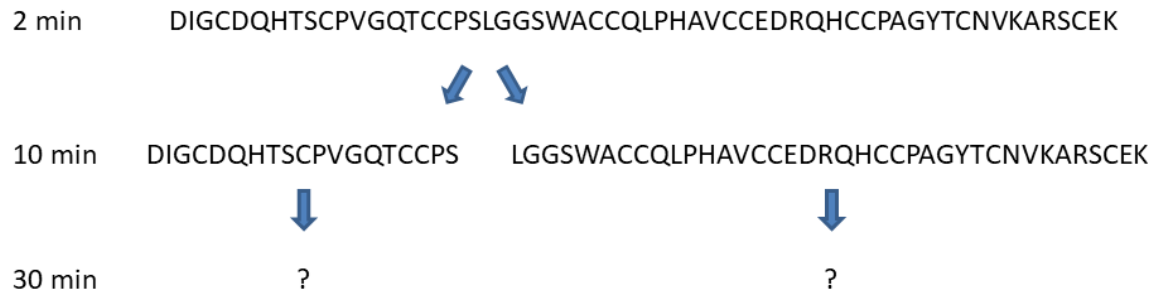


Abbildung 24 Sequentielle Degradierung Nach zwei Minuten wurde Peptid 11 (**Tabelle 5**) als Analyt mit hoher Intensität im XIC detektiert. Dessen Intensität nahm im Verlauf des Verdaus ab, weswegen davon auszugehen ist, dass es weiter degradiert wurde. Zwei mögliche Spaltprodukte aus diesem wurden nach 10 Minuten identifiziert. Doch auch diese traten im späteren Verlauf nicht mehr auf, weswegen nicht auszuschließen ist, dass eine weitere Degradierung stattfand.

Für diese konnten jedoch nur beim Zeitpunkt von 10 Minuten XIC-Werte ermittelt werden, was vermuten lässt, dass sie weiter degradiert wurden. Da in der Region Asp-441 bis Lys-497 insgesamt 14 Spaltstellen identifiziert wurden (**Abbildung 18**), kann nicht mit Sicherheit gesagt werden, welche Spaltprodukte aus welchen Intermediaten entstanden sind. Vermutlich ist es eine Kombination aus den verschiedensten Peptiden, was auch erklären würde, warum nach einer Inkubationszeit von 60 Minuten keine kleineren Peptide identifiziert wurden als das in **Abbildung 24** gezeigte (Asp-441 bis Ser-459). Wenn aus einem einzigen Vorläuferpeptid parallel verschiedene Spaltprodukte entstehen können, so ist deren Abundanz zwangsläufig geringer. Dies gilt auch für alle weiteren Hydrolysevorgänge und deren Produkte.

Die Ergebnisse des zeitabhängigen Meprin- β -Verdau unterschieden sich von denen der Meprin- α -Protease und deuteten auf die geringfügig höhere Spezifität des Enzyms hin [93, 94]. Die PGRN-Peptide mit den nach 2-minütigem Verdau höchsten XIC-Werten sind in **Tabelle 7** dargestellt. In der Probe, welche nach 45 Minuten entnommen wurde, konnten aus unbekannten Gründen nur sehr wenige Peptide (< 15) identifiziert werden. Diese traten außerdem mit Intensitäten auf, die z.T. um zwei Größenordnungen kleiner waren als in den Proben $t = 30 \text{ min}$ und $t = 60 \text{ min}$. Aus diesem Grund ist dieser Zeitpunkt nicht aufgeführt. Im Gegensatz zu Meprin- α , nahmen die meisten Peptide, welche nach zwei Minuten die höchste Intensität aufwiesen, in ihrer Abundanz weiter zu. Davon ausgenommen sind die

langen Sequenzen 1 und 17, welche vermutlich aufgrund ihrer Länge und der vielen potentiellen Spaltstellen weiter hydrolysiert wurden, sowie die Peptide 3, 12, 13, 15 und 16. Entweder waren diese über den gesamten Zeitraum nahezu konstant (3 und 12) oder traten unregelmäßig auf (13, 15 und 16).

Tabelle 7 Top-Down Methode: Peptide mit intensivstem Signal nach 2-minütigem Verdau (Meprin-β) Die Sequenzen wurden nach ihrer Häufigkeit nach zwei Minuten verdau sortiert. Neben hauptsächlich sehr kurzen Peptiden wurden zwei Sequenzen deutlich längere Sequenzen identifiziert, die über 50+ Aminosäuren verfügten. Diese wurden jedoch auch von Meprin-β weiter degradiert.

Peptidsequenz			Inkubationszeit					
			2 min	5 min	10 min	20 min	30 min	60 min
1	P28799 [438-501]	[S].HPRDIGCDQHTSCPVGQTCCPSLGGSWACCQL PHA VCCEDRQHCCPAGYTCNVKARSCEKEVVS.[A]	2,43E+06	4,10E+06	1,48E+06	3,37E+06	2,08E+05	
2	P28799 [352-362]	[H].LSLPDPQALKR.[D]	1,56E+06	2,53E+06	3,57E+06	6,05E+06	1,01E+07	1,91E+07
3	P28799 [581-589]	[W].DAPLRDPAL.[R]	1,04E+06	1,29E+06	9,34E+05	8,15E+05	1,11E+06	7,13E+05
4	P28799 [582-592]	[D].APLRDPALRQL.[L]	1,03E+06	9,31E+05	1,03E+06	3,26E+06	4,25E+06	7,05E+06
5	P28799 [345-361]	[W].MEKAPAHLSLPDPQALK.[R]	8,11E+05	1,15E+06	1,76E+06	1,74E+06	4,11E+06	8,17E+06
6	P28799 [341-353]	[H].QVPWMEKAPAHLS.[L]	6,29E+05	1,31E+06	1,70E+06	3,06E+06	5,52E+06	9,83E+06
7	P28799 [345-362]	[W].MEKAPAHLSLPDPQALKR.[D]	6,07E+05	2,12E+06	4,01E+06	5,06E+06	1,67E+07	3,75E+07
8	P28799 [351-362]	[A].HLSLPDPQALKR.[D]	5,12E+05	6,96E+05	1,16E+06	2,53E+06	5,20E+06	1,08E+07
9	P28799 [583-590]	[A].PLRDPALR.[Q]	4,04E+05	6,34E+05	4,56E+05	4,13E+05	1,65E+06	9,03E+05
10	P28799 [336-351]	[C].EQGPHQVPWMEKAPAH.[L]	2,88E+05	4,60E+05	6,14E+05	1,07E+06	1,70E+06	3,28E+06
11	P28799 [498-517]	[K].EVVSAQPATFLARSPHVGVK.[D]	2,60E+05	2,16E+05		6,48E+05	6,16E+05	7,80E+05
12	P28799 [421-432]	[S].EIVAGLEKMPAR.[R]	2,39E+05	7,89E+04	1,90E+05	2,69E+05	1,38E+05	1,72E+05
13	P28799 [47-56]	[L].DKWPTTLSRH.[L]	2,09E+05	6,73E+04		2,42E+04	1,41E+04	
14	P28799 [353-362]	[L].SLPDPQALKR.[D]	1,31E+05	4,02E+05	6,31E+05	1,26E+06	3,03E+06	7,03E+06
15	P28799 [416-435]	[Q].CQRGSEIVAGLEKMPARRAS.[L]	1,20E+05	8,70E+04		5,41E+04		
16	P28799 [496-510]	[C].EKEVVSAQPATFLAR.[S]	1,19E+05			1,15E+05	1,38E+05	6,18E+04
17	P28799 [442-501]	[D].IGCDQHTSCPVGQTCCPSLGGSWACCQLPHA V CCEDRQHCCPAGYTCNVKARSCEKEVVS.[A]	1,11E+05	1,17E+05	5,50E+04			
18	P28799 [355-362]	[L].PDPQALKR.[D]	1,07E+05	1,62E+05	1,26E+05	1,50E+05	4,15E+05	6,23E+05
19	P28799 [266-277]	[N].ATTDLLTKLPAH.[T]	1,02E+05	2,22E+05	1,91E+05	6,50E+05	1,17E+06	2,78E+06



Dass die übrigen Sequenzen in ihrer Intensität zunahmen, könnte daran liegen, dass das Substrat Progranulin von Meprin-β langsamer hydrolysiert wurde (Kinetik). Die nach zwei Minuten identifizierten Spaltstellen sind in **Abbildung 25** gezeigt. Trotz der verwendeten Proteaseinhibitoren wurden auch für Meprin-β untypische Aminosäuren an der P1-Position nachgewiesen. Beispielsweise zeigen die Peptide 2, 7, 8, 9, 11, 14, 16 und 18 zwar die charakteristischen Aminosäuren (Aspartat, Glutamin oder Serin) in der P1'-Position, jedoch standen diesen die basischen Seitenketten eines Arginins oder Lysins in der P1-Position

gegenüber, was den bisherigen Befunden für diese Protease [93] widerspricht. Basische Aminosäuren am C-Terminus eines Peptids könnten beispielsweise für eine Aktivität von Serin-Proteasen sprechen, die Trypsin-ähnliche Spezifitäten aufweisen. Weiterhin wurden Spaltstellen identifiziert, an denen ein Arginin an der P1'-Position auftrat. Dies gilt für die Peptide 3, 5 und 12. Aufgrund der bereits diskutierten Struktur des aktiven Zentrums von Meprin-β sind basische Aminosäuren an der P1'-Position äußerst unwahrscheinlich.

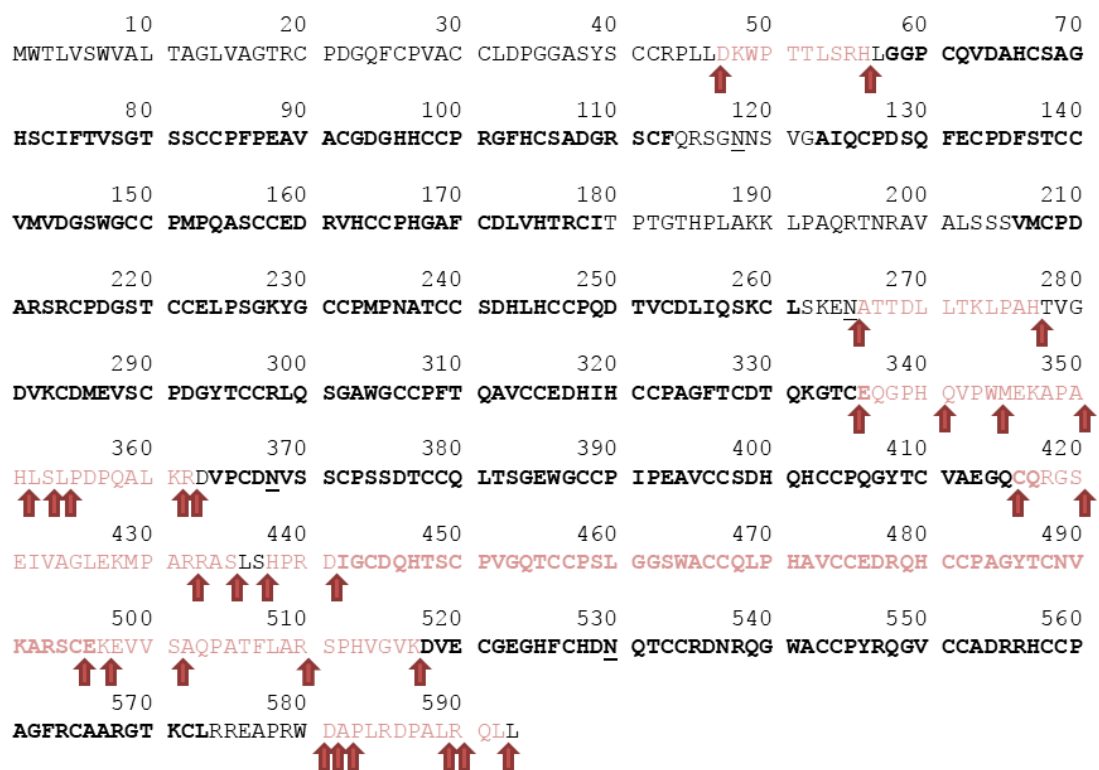


Abbildung 25 Top-Down Methode: In der nativen Struktur zugängliche Regionen des PGRN (Meprin-β-Verdau nach 2 min) Die Spaltstellen, aus denen die Peptide in Tabelle 7 entstanden sind, wurden mit dunkelroten Pfeilen gekennzeichnet. Die zugehörigen Sequenzen (z.T. überlappend) sind hellrot hervorgehoben. Die Granulinabschnitte sind mit Fettbuchstaben markiert.

Da während der *in vitro*-Hydrolyse Proteaseinhibitoren verwendet wurden, kann davon ausgegangen werden, dass die Peptide mit für Meprin-β ungewöhnlichen Termini bereits vor dem Meprin-Verdau in der Probe enthalten waren. Zwar wurden Kontrollproben gemessen und alle dort aufgedeckten Sequenzen aus der Analyse entfernt, jedoch kann somit nicht sichergestellt werden, dass alle in der ursprünglichen Probe enthaltenen Peptide auch identifiziert wurden. Es könnten größere Peptide in der Probe enthalten gewesen sein, die beispielsweise aufgrund der bereits angesprochenen Häufigkeit von Cystein in einigen

Abschnitten nicht effektiv ionisiert werden konnten. Würden diese Sequenzen im weiteren Verlauf des Experiments von Meprin- α/β hydrolysiert, könnten Peptide entstanden sein, die sich einfacher ionisieren und detektieren lassen. Diese Peptide würden nun aber einen Terminus aufweisen, der nicht mit der Spezifität der Meprine übereinstimmt. Aus diesem Grund müssen nicht alle aufgeführten Spaltstellen – trotz der Tatsache, dass Kontrollproben gemessen wurden – von einer Hydrolyse durch Meprin- α/β herrühren. Als Ursache für das Vorhandensein von PGRN-Peptiden in der Kontrolle kommt wiederum die Aktivität von Fremdproteasen während der Expression, bzw. während der Aufreinigung in Frage.

Die von Meprin- β generierten Peptide, welche nach 2 min Inkubationszeit die intensivsten Signale zeigten, ähneln den Sequenzen, die schon für Meprin- α beobachtet wurden. Sie stammen aus der Region (AA 337-441 im Progranulin) zwischen Granulin-4 und Granulin-5 oder vom Protein-C-Terminus. Eine Ausnahme bildet das N-terminale Peptid 13 (**Tabelle 7**). Intensive Signale von N-terminalen Peptiden wurden für Meprin- α erst nach längerer Inkubationszeit beobachtet. **Tabelle 5** und **Tabelle 7** wurden jedoch nach den relativen Signalintensitäten sortiert sind, weswegen die Position der Sequenzen in den Tabellen nicht direkt verglichen werden können. Vielmehr sollten zu diesem Zweck die absoluten Werte herangezogen werden. Die Integrale der Sequenz (AA 47-56 im Progranulin) betrugen $6,81\text{E}+6$ (Meprin- α , 2 min) und $2,09\text{E}+5$ (Meprin- β , 2 min). Somit war das Peptidsignal im Meprin- α -Verdau nach zwei Minuten bereits um mehr als eine Größenordnung intensiver. Dass dieselbe Sequenz im Meprin- β -Verdau relativ gesehen intensiver ist, unterstreicht die geringfügig höhere Spezifität des Enzyms. Dafür spricht auch, dass nach zwei Minuten für Meprin- β nur 40 Peptide identifiziert wurden, während es für Meprin- α insgesamt 180 Sequenzen waren. Nichtsdestotrotz widerspricht das Auftreten N-terminaler Peptide nach zwei Minuten Meprin- β -Verdau der zunächst angenommenen sterischen Hinderung des Protein-N-Terminus. Oder mit anderen Worten: der Protein-N-Terminus scheint im nativen Protein bereits für Proteasen zugänglich zu sein.

Tabelle 8 Top-Down Methode: Peptide mit intensivstem Signal nach 60-minütigem Verdau (Meprin-β) Die Sequenzen wurden nach der Signal-Intensität im Chromatogramm nach 60 Minuten Verdau sortiert. Wie zu erwarten war, traten keine längeren Peptide mehr auf, was auf eine sequentielle Degradierung der beiden längeren Sequenzen aus **Tabelle 7** hinweist. Im Gegensatz zum Meprin-α-Verdau wurden nach 60 Minuten Peptide identifiziert, die nach zwei Minuten noch nicht oder nicht ausreichend abundant vorhanden waren. Alle Sequenzen, die mit einem * markiert sind, kamen in **Tabelle 7** noch nicht vor und sind demnach erst im späteren Verlauf des Verdaus entstanden.

Peptidsequenz		Inkubationszeit					
		2 min	5 min	10 min	20 min	30 min	60 min
1	P28799 [345-362] [W].MEKAPAHLSLPDPQALKR.[D]	6,07E+05	2,12E+06	4,01E+06	5,06E+06	1,67E+07	3,75E+07
2	P28799 [352-362] [H].LSLPDPQALKR.[D]	1,56E+06	2,53E+06	3,57E+06	6,05E+06	1,01E+07	1,91E+07
3	P28799 [351-362] [A].HLSLPDPQALKR.[D]	5,12E+05	6,96E+05	1,16E+06	2,53E+06	5,20E+06	1,08E+07
4	P28799 [341-353] [H].QVPWMEKAPAHLS.[L]	6,29E+05	1,31E+06	1,70E+06	3,06E+06	5,52E+06	9,83E+06
5	P28799 [345-361] [W].MEKAPAHLSLPDPQALK.[R]	8,11E+05	1,15E+06	1,76E+06	1,74E+06	4,11E+06	8,17E+06
6	P28799 [582-592] [D].APLRDPALRQL.[L]	1,03E+06	9,31E+05	1,03E+06	3,26E+06	4,25E+06	7,05E+06
7	P28799 [353-362] [L].SLPDPQALKR.[D]	1,31E+05	4,02E+05	6,31E+05	1,26E+06	3,03E+06	7,03E+06
8	P28799 [346-362] [M].EKAPAHLSLPDPQALKR.[D]*	7,87E+04	2,95E+05	7,73E+05	1,45E+06	2,83E+06	5,84E+06
9	P28799 [336-351] [C].EQGPHQVPWMEKAPAH.[L]	2,88E+05	4,60E+05	6,14E+05	1,07E+06	1,70E+06	3,28E+06
10	P28799 [266-277] [N].ATDILLTKLPAH.[T]	1,02E+05	2,22E+05	1,91E+05	6,50E+05	1,17E+06	2,78E+06
11	P28799 [341-351] [H].QVPWMEKAPAH.[L]*				3,44E+05	7,70E+05	1,78E+06
12	P28799 [345-362] [W].MEKAPAHLSLPDPQALKR.[D]*		7,68E+04	2,52E+05	8,90E+05	8,26E+05	1,53E+06
13	P28799 [336-344] [C].EQGPHQVPW.[M]*		7,45E+04	9,57E+04	2,30E+05	5,51E+05	1,28E+06
14	P28799 [341-352] [H].QVPWMEKAPAH.[S]*			1,20E+05	2,69E+05	5,11E+05	1,10E+06
15	P28799 [342-352] [Q].VPWMEKAPAH.[S]*				1,14E+05	2,35E+05	9,08E+05
16	P28799 [583-590] [A].PLRDPALR.[Q]	4,04E+05	6,34E+05	4,56E+05	4,13E+05	1,65E+06	9,03E+05
17	P28799 [341-362] [H].QVPWMEKAPAHLSLPDPQALKR.[D]*			9,39E+04	1,26E+05	3,98E+05	8,23E+05
18	P28799 [498-517] [K].EVVSAQPATFLARSPHVGVK.[D]	2,60E+05	2,16E+05		6,48E+05	6,16E+05	7,80E+05
19	P28799 [581-589] [W].DAPLRDPAL.[R]	1,04E+06	1,29E+06	9,34E+05	8,15E+05	1,11E+06	7,13E+05
20	P28799 [355-362] [L].PDPQALKR.[D]	1,07E+05	1,62E+05	1,26E+05	1,50E+05	4,15E+05	6,23E+05
21	P28799 [172-188] [C].DLVHTRCITPTGTHPLA.[K]*			4,88E+04	6,17E+04	8,89E+04	3,47E+05
22	P28799 [341-353] [H].QVPWMEKAPAHLS.[L]*	2,38E+04	3,57E+04	5,41E+04	7,03E+04	1,19E+05	2,65E+05
23	P28799 [351-359] [A].HLSLPDPQA.[L]*	6,69E+04	1,16E+05	6,23E+04	4,56E+04	1,68E+05	2,35E+05
24	P28799 [336-353] [C].EQGPHQVPWMEKAPAHLS.[L]*	2,74E+04	4,95E+04	7,11E+04	9,33E+04	1,64E+05	2,24E+05

Niedrigster Wert
Höchster Wert

Die 24 Peptide mit den intensivsten Signalen nach 60-minütiger Inkubation mit Meprin-β sind in **Tabelle 8** gelistet. Im Gegensatz zum Verdau mit Meprin-α traten hier sieben Peptide auf, die zum Zeitpunkt 2 min nicht nachgewiesen wurden. Besonders interessant sind die Peptide 13 und 21, da sie aus Regionen stammen, die vom Meprin-α-Verdau nicht abgedeckt wurden. Allerdings weist eines eine für beide Meprin-Proteasen untypische Aminosäure (Lysin) in der P1'-Position auf. Es ist nicht unmöglich, jedoch sehr unwahrscheinlich, dass die

Sequenz von Meprin- β generiert wurde. Von den 24 gelisteten Peptiden stammen 15 aus der auch für Meprin- α zugänglichen Region zwischen Granulin-4 und Granulin-5 (AA 337-364). Sie besitzen ebenfalls die höchsten Intensitäten, was die These unterstützt, dass diese Abschnitte im nativen Protein oberflächlich angeordnet sind. Drei weitere Sequenzen überlappen mit diesem Bereich und weisen mit Glu-336 ebenfalls eine für Meprin- β spezifische Spaltstelle auf. Vom Protein-C-Terminus wurden drei weitere Peptide nachgewiesen. Die Peptide des Meprin- β -Verdau, die nach 60 Minuten für die intensivsten Signale verantwortlich waren, und deren zugehörigen Spaltstellen sind in **Abbildung 26** gezeigt.

10	20	30	40	50	60	70
MWTLVSWVAL	TAGLVAGTRC	PDGQFCPVAC	CLDPGGASYS	CCRPLLDKWP	TTLSRHLGGP	CQVDAHCSAG
80	90	100	110	120	130	140
HSCIFTVSGT	SSCCPFPEAV	ACGDGHHCCP	RGFHCSADGR	SCFQ R SGNNS	VGAIQCPDSQ	FECPDFSTCC
150	160	170	180	190	200	210
VMVDGSWGCC	PMPQASCCED	RVHCCPHGAF	C D LVHTRCIT	PTGTHPLAKK	LPAQRTNRAV	ALSSSV M CPD
			↑	↑		
220	230	240	250	260	270	280
ARSRCPDGST	CCELP S GKYG	CCPMFNATCC	SDHLHCCPQD	TVCDLIQSKC	LSKEN A TTDL	LTKLPAHTVG
					↑	↑
290	300	310	320	330	340	350
DVKCDMEVSC	PDGYTCCRLQ	SGAWGCCPFT	QAVCCEDHIH	CCPAGFTCDT	QKGTCEQGP H	QVPWMEKAPA
					↑	↑↑↑
360	370	380	390	400	410	420
HLSLPDPQAL	KRDVPCDNVS	SCPSSDTCCQ	LTSGEWGCCP	IPEAVCCSDH	QHCCPQGYTC	VAEGQCQ R GS
↑↑↑↑	↑↑					
430	440	450	460	470	480	490
EIVAGLEKMP	ARRASLSHPR	DIGCDQHTSC	PVGQTC C PSL	GGSWACCQLP	HAVCCEDRQH	CCPAGYTCNV
500	510	520	530	540	550	560
KARSCEKEVV	SAQPATFLAR	SPHVG V KDVE	CGEGHFCHDN	QTCCRDNRQG	WACCPYRQGV	CCADRRHCCP
↑		↑				
570	580	590				
AGFRCAARGT	KCLRREAPRW	DAPLRDPALR	QLL			
		↑↑↑	↑↑↑			

Abbildung 26 Top-Down Methode: Peptide mit intensivstem Signal nach 60-minütigem Verdau (Meprin- β)
Die nach zwei Minuten intensiven langen Abschnitte (siehe **Tabelle 7**) wurden nach einer Stunde Inkubation nicht mehr beobachtet. Der Großteil der stabilen Peptide stammt von der ausgesetzten Intersektion AA 337-364 sowie vom Protein-C-Terminus.

e. Zusammenfassung

In diesen Versuchen sollte primär gezeigt werden, dass der Verzicht auf Arbeitsproteasen für die Identifizierung von Spaltstellen und Hydrolyseprodukten einen Vorteil gegenüber dem klassischen *Bottom-Up* Ansatz besitzt. Die Verwendung von Arbeitsproteasen führt

artifizielle Spaltstellen ein. Ursprünglich generierte Sequenzen müssen daher vorher markiert und die dafür benötigten Chemikalien im Anschluss entfernt werden. Weitere Arbeitsschritte bedeuten in der Proteomics aber eine längere Bearbeitungszeit sowie mögliche Probenverluste. Die hier vorgestellte Methode umgeht diese. Darüber hinaus konnte gezeigt werden, dass sich die angewandte Lektin-basierte Festphasenextraktion für Analyten eignet, die glykosyliert sind. *N*-Glykosylierte Proteine und Peptide führen in der Chromatographie sowie Massenspektrometrie zu einer schlechten Auflösung, Ionisation und Fragmentierung. Die Reduktion/Alkylierung sowie Deglykosylierung und die anschließende Festphasenextraktion können innerhalb von zwei Stunden durchgeführt werden. Die zusätzlichen Arbeitsschritte im *Bottom-Up* Ansatz erforderten für die Dimethylierung (3 h) und den Verdau (4 h) zusammen bereits sieben Stunden. Die anschließende Aufarbeitung ist dabei noch nicht berücksichtigt.

Untersucht wurde der Verdau von humanem Progranulin mit den Metalloproteasen Meprin- α und Meprin- β . Es konnte gezeigt werden, dass die vorgestellte Methode, die den Ausschluss von glykosylierten Analyten und freigesetzten Glykanen zum Ziel hatte, sich eignet, um die Spaltprodukte und deren Termini (Spaltstellen) von Proteolyseexperimenten zu identifizieren. Beide Methoden, Lektin-basierte TDP und klassische BUP, identifizierten für die untersuchten Proben in etwa die gleiche Anzahl an Peptiden, zeigten jedoch Unterschiede in den identifizierten Spaltstellen und der Sequenzabdeckung. Der Lektin-basierte *Top-Down* Ansatz identifizierte bis zu 40 % mehr Spaltstellen und erreichte eine um 21 % höhere Sequenzabdeckung bei gleichzeitig kürzerer Probenvorbereitung. Der Größenausschlussansatz hingegen identifizierte im Vergleich zum *Bottom-Up* Ansatz nur 68 % der Spaltstellen und deckte 79 % der Sequenz ab. Die Gründe für die geringeren Prozentsätze liegen, wie bereits erwähnt, vermutlich in der Größe der Filtereinheiten begründet und nicht unbedingt an der Methode des Größenausschlusses. Nichtsdestotrotz wurde die Methode des Größenausschlusses im weiteren Verlauf der Experimente nicht mehr berücksichtigt.

Mithilfe des Lektin-basierten *Top-Down* Ansatzes identifizierten Spaltstellen wurden mit der für Meprin-Proteasen bekannten Spezifität verglichen. Zwar wurden für die für Meprin- α und Meprin- β charakteristischen sauren Aminosäuren in der P1'-Position nachgewiesen, gleichzeitig aber auch eine hohe Abundanz von Cystein im gesamten Abschnitt (P4 bis P4')

und einige basische Aminosäuren in der P1-Position. Progranulin ist überdurchschnittlich reich an Cystein-Seitenketten, was den ersten Befund erklärt. Das Auftreten von tryptischen Peptiden (vor allem Arginin an der P1-Position) stimmt jedoch nicht mit der Spezifität der Meprin-Proteasen überein. Meprin- α wurde nach der Affinitätschromatographie mittels Trypsin aktiviert, was die Vermutung nahe legt, dass diese Spaltstellen durch Verunreinigungen an Trypsin hervorgerufen wurden. Weiterhin wurden in den Kontrollproben, die ausschließlich intaktes Progranulin enthalten sollten, Hydrolyseprodukte von PGRN nachgewiesen. Zwar wurden diese Sequenzen in der Analyse der Meprin-generierten Spaltprodukte ausgeschlossen, es kann jedoch nicht sichergestellt werden, dass alle in der Kontrolle vorhandenen Peptide nachgewiesen wurden. Demzufolge verbleibt für alle Peptide, die im Versuch als von Meprin- α oder Meprin- β generiert gewertet wurden, die Möglichkeit, während der Progranulin-Expressierung generiert worden und somit eine Falschzuweisung zu sein. Bezüglich der ermittelten Spezifität der beiden Proteasen sollte berücksichtigt werden, dass ein einziges Substrat zu keinen aussagekräftigen Ergebnissen führen kann. Vor allem dann nicht, wenn dessen Aminosäurezusammensetzung von den üblichen Häufigkeiten abweicht.

Die vorgestellte Methode sieht die Verwendung von immobilisierten WGA-Lektinen des Organismus *Triticum vulgaris* vor, die die Glukoside *N*-Acetylglucosamin und Sialinsäure spezifisch binden. *N*-Acetylglucosamin ist Teil der Kernstruktur von *N*-Glykosylierungen in allen Organismen, sodass sich die Methode für Proteinproben weitestgehend unabhängig derer Herkunft eignen sollte. Die Kartuschen wurden hier für das Entfernen von glykosylierten Proteinen und Peptiden sowie freien Glykanen verwendet. Es sollte jedoch auch möglich sein, sie für die Anreicherung der o.a. Stoffklassen zu verwenden. Dies könnte hilfreich sein, wenn z.B. die Struktur von Glykanen aufgeklärt werden soll.

7. Anwendung von *Bottom-Up* Methoden zur Sequenzierung von antimikrobiellen Peptiden des *stratum corneum*

a. Antimikrobielle Peptide der Haut

Die oberste Schicht der menschlichen Haut (*stratum corneum*) ist von einer Vielzahl von Bakterien besiedelt. Dabei handelt es sich vor allem um Kommensale, aber auch Pathogene wie z.B. *Pseudomonas aeruginosa* und *Staphylococcus aureus* [104, 105]. Letztere zählen zu den häufigsten Ursachen von Infektionskrankheiten der Haut in Krankenhäusern [106]. Die Tatsache, dass diese Bakterien Resistenzen gegen von Menschen verwendeten Antibiotika wie z.B. Vancomycin [107] entwickelt haben, erfordert die Suche nach neuen antimikrobiellen Substanzen. Da bekannt ist, dass antimikrobielle Peptide, die entweder vom Wirt gebildet werden oder deren Expression durch den Kontakt zu Pathogenen induziert wird [108], das Wachstum von Bakterien auf der Haut regulieren, liegt in der Identifizierung der natürlichen AMP und deren Wirkungsweise das Potential, neue chemotherapeutische Substanzen zu entwickeln. Es gibt eine Reihe von antimikrobiellen Peptiden nicht-humanen Ursprungs, die als Vorlage für sogenannte „Designer“-AMP verwendet wurden [109]. Da in der Anwendung solcher Wirkstoffe jedoch das Risiko einer Immunreaktion besteht [110] und es bis dato vergleichsweise wenige humane AMP gibt, die als Vorlagen verwendet wurden [111], liegt die Identifizierung von wirksamen humanen Sequenzen im Fokus aktueller Forschung. Von besonderem Interesse sind dabei kationische AMP, deren Aktivität auf der Interaktion mit der Zellmembran anstatt einzelnen Aminosäuren von Proteinen beruht. Auf diese Weise ist die Entwicklung von Resistenzen zwar möglich, jedoch unwahrscheinlich [112].

Ein Beispiel für ein auf der Haut wirksames antimikrobielles Peptid, welches als Vorläufer von den Schweißdrüsen produziert und proteolytisch aktiviert wird, ist Dermcidin [113]. Dermcidin enthält eine sich über 47 Aminosäuren erstreckende anionische Sequenz (AA 63-109), die gegen viele hautrelevante Bakterien wirkt, nicht jedoch *P. aeruginosa*. Weitere auf der Haut vorkommende Peptide, für die eine antimikrobielle Wirkung nachgewiesen werden konnte, sind Abschnitte des Filaggrin-2 [114], Psoriasin [115], RNase 7 [116] und die Familie der Defensine [117–120].

P. aeruginosa, das in wässrigen Habitaten und Böden zu finden ist [104, 121], ist ein opportunistisches Pathogen, das insbesondere bei Fehlen des kutanen Schutzes wie z.B. Brandverletzungen, zu schweren Infektionen führen kann [122, 123]. Besonders für Menschen, welche an Mukoviszidose erkrankt sind, stellt das Bakterium eine erhebliche Gefahr dar [124].

b. Durchführung und Zielsetzung

Hinweis: Die nachfolgenden Absätze beschreiben die Extraktion von AMP aus Kallusproben sowie durchgeführte mikrobiologische Assays und wurden im Arbeitskreis um Prof. Dr. Jens-Michael Schröder durchgeführt. Die Extraktion und Fraktionierung sowie die Tests hinsichtlich einer mikrobiologischen Aktivität waren nicht Teil der eigenen experimentellen Arbeiten.

Die Arbeitsgruppe um Prof. Dr. Jens-Michael Schröder (Dermatologie, Universitätsklinikum Kiel) vermutete, dass antimikrobielle Peptide, welche gegen *P. aeruginosa* aktiv sind, in den Bereichen der Haut vorkommen müssten, die oft Kontakt zu dem Pathogen haben. Da gerade feuchte Böden zu den natürlichen Habitaten von *P. aeruginosa* zählen, wurde vermutet, dass wirksame AMP vor allem an den Füßen zu finden sein müssten. Um diese These zu untersuchen, wurden kombinierte Kallusproben von der Ferse auf das Vorhandensein antimikrobiell wirksamer Peptide untersucht. Vorherige Versuche hatten gezeigt, dass die Abwehr gegen *P. aeruginosa* nur bei pH-Werten von 6 oder niedriger und in Abwesenheit von löslichen Nährstoffen aktiv ist. Aufgrund der Tatsache, dass diese Voraussetzungen in bisher angewandten Protokollen nicht gewährleistet waren, wurde diese antimikrobielle Aktivität bis dato nicht entdeckt. Während der Aufreinigung der Kallusextrakte, bzw. in den darauf folgenden Aktivitätstests wurde für einige Heparin-affine Fraktionen (**Abbildung 27**) eine Aktivität gegen *P. aeruginosa* nachgewiesen. Das Ziel des Teil-Versuchs war die Sequenzierung der für diese Aktivität verantwortlichen Peptide.

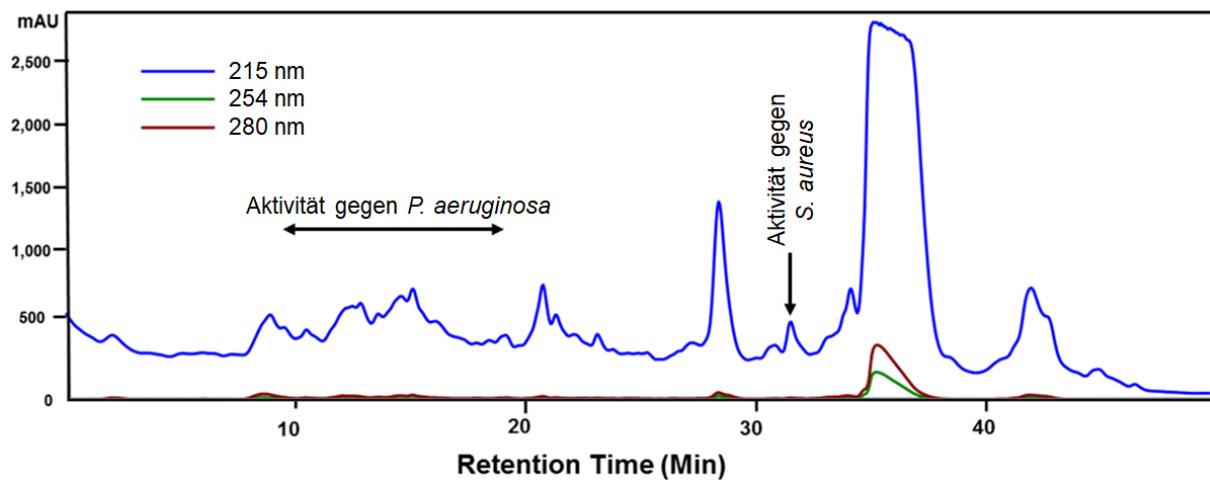


Abbildung 27 Umkehrphasen-Chromatographie der Heparin-affinen Fraktionen Die UV-Absorption wurde bei den in der Protein-Analytik üblichen Wellenlängen von 215, 254 und 280 nm gemessen. Mehrere Fraktionen zwischen etwa 10 und 20 min Retentionszeit waren antimikrobiell gegen *P. aeruginosa* wirksam. Für zwei weitere Faktoren, welche bei etwa 32 min eluierten, konnte eine hohe Aktivität gegen *S. aureus* festgestellt werden (mit freundlicher Genehmigung von Prof. Dr. Jens-Michael Schröder).

In weiteren Sichtungen derselben Fraktionen konnten zwei Faktoren (**Abbildung 27**) mit hoher gegen *Staphylococcus aureus* gerichteter Aktivität korreliert werden. Bei diesen handelte es sich vermutlich um Peptide bestehend aus etwa 20 Aminosäuren, die ein nahezu gleiches Elutionsverhalten in der Flüssigkeitschromatographie zeigten. In ersten massenspektrometrischen Untersuchungen der AG Schröder konnten die Peptide jedoch nicht sequenziert werden, da die entsprechenden Tandem-Massenspektren von einem einzigen intensiven Signal dominiert und für eine Sequenzierung nicht genügend Fragmente detektiert wurden. Mithilfe angepasster Methodik sollte im zweiten Teilversuch eine Identifizierung der beiden isolierten Faktoren erfolgen. Diese ist Gegenstand der vorliegenden Arbeit.

c. Material und Methoden

i. Chemikalien und Zubehör

Ameisensäure, Acetonitril, Ammoniumhydrogencarbonat, Trifluoressigsäure, Dithiothreitol, Iodacetamid, Chymotrypsin (*Bos taurus*), GluC (*Staphylococcus aureus*), LysC (*Lysobacter enzymogenes*), Pepsin (*Sus scrofa*), Calciumchlorid und Salzsäure wurden von Sigma-Aldrich (Steinheim, Deutschland) erworben, Trypsin (*Sus scrofa*) von Promega (Madison, USA), TRIS von Roth (Karlsruhe, Deutschland) und CHCA von LaserBio Labs (Antibes, Frankreich). Entionisiertes Wasser wurde mittels eines arium611 VF Systems von Sartorius (Göttingen, Deutschland) generiert. Die synthetischen Peptide (Desmoplakin AA 2414-2450, AA 2428-

2446 und AA 2425-2461, *H. sapiens*, siehe **Abbildung 34**) wurden von Biosynthan (Berlin, Deutschland) in 90-prozentiger Reinheit bezogen.

ii. Enzymatischer Verdau der gegen *S. aureus* aktiven Faktoren

Trypsin, Chymotrypsin, GluC und LysC. Zwei Mal 1 μL des gelösten Faktors (gegen *S. aureus* aktive Fraktion, siehe **Abbildung 27**) wurden mit jeweils 19 μL Ammoniumhydrogencarbonat-Puffer (50 mM, pH 7,0) verdünnt. Zu dieser Lösung wurden 4 μL Dithiothreitol-Stammlösung (200 mM in 50 mM ABC), bzw. 4 μL des verwendeten ABC-Puffers hinzugefügt und für 15 min auf 70 °C erhitzt. Nach Abkühlen wurden 8 μL Iodacetamid-Stammlösung (200 mM in 50 mM ABC), bzw. 8 μL des verwendeten ABC-Puffers hinzugegeben und unter Lichtausschluss für 30 min bei Raumtemperatur inkubiert. Zuletzt wurden 20 μL der DTT-Stammlösung, bzw. des ABC-Puffers hinzugefügt. Beide Ansätze wurden dann in vier Proben gleichen Volumens aufgeteilt. Zu diesen wurden 100 ng Trypsin, Chymotrypsin, GluC oder LysC hinzugegeben. Die beiden Chymotrypsin-Proben wurden weiterhin mit 1 μL Calciumchlorid (145 mM) versetzt. Die Ansätze wurden dann für 16 h bei 37 °C (Trypsin, GluC und LysC), bzw. 30 °C (Chymotrypsin) inkubiert. Der Verdau wurde durch die Zugabe von 1 μL Ameisensäure (10 %, v/v) gestoppt.

Pepsin. 1 μL des gelösten Faktors wurden mit 14 μL des ABC-Puffers verdünnt. Zu dieser Lösung wurde 1 μL DTT-Stammlösung gegeben und für 15 min auf 70 °C erhitzt. Nach Abkühlen wurde 1 μL IAA-Stammlösung hinzugefügt und der Ansatz unter Lichtausschluss für 30 min bei Raumtemperatur inkubiert. Anschließend wurden 3 μL DTT-Stammlösung hinzugegeben und der pH-Wert der Lösung mit 1 μL Salzsäure (500 mM) auf etwa 2 eingestellt. Zu diesem Ansatz wurden 100 ng Pepsin (in 10 mM HCl) hinzugegeben und für 24 h bei 37 °C inkubiert.

iii. Oxidation mit Dimethylsulfoxid

Die synthetisierten Peptide wurden mittels Dimethylsulfoxid nach [125] ohne den Zusatz von Essigsäure oxidiert. Dazu wurden wässrige Lösungen der Sequenzen mit DMSO versetzt, sodass die Konzentrationen des Analyten und des DMSO bei einem Gesamtvolumen von 100 μL in etwa 1 $\mu\text{g}/\mu\text{L}$, bzw. 10 % (v/v) betrugen. Nach vier, acht, 24 und 48 h wurden Aliquote entnommen, flüssigchromatographisch aufgetrennt und massenspektrometrisch analysiert.

iv. Festphasen-Extraktion mit OASIS Kartuschen

40 µL (entsprechen etwa 40 µg Peptid) des nach der Oxidation erhaltenen Ansatzes wurden 1:10 verdünnt und dem Protokoll des Herstellers folgend mittels OASIS HLB 1cc Kartuschen aufgereinigt. Die Kartuschen wurden zunächst mit 1 mL Acetonitril gespült und mit 1 mL entionisiertem Wasser equilibriert. Anschließend wurden die Proben unter leichtem Unterdruck über das Adsorbens geführt. Die Kartuschen wurden dann zwei Mal mit jeweils 500 µL entionisiertem Wasser gespült. Zuletzt wurden die Peptide mit 300 µL Acetonitril eluiert und bei Raumtemperatur im Vakuum (Concentrator Plus, Eppendorf, Hamburg, Deutschland) bis zur Trockene eingengt.

v. Verdau von synthetischen Peptiden mit Cathepsin D

Die oxidierten Peptide wurden in 60 µL Acetonitril wieder aufgenommen, in drei gleiche Volumina aufgetrennt (entsprechen jeweils 13 µg Peptid) und das Lösungsmittel im Vakuum entfernt. Von den nativen Proben wurden in ähnlicher Weise Aliquote angesetzt, sodass für den Versuch insgesamt $6 \times 13 \mu\text{g}$ Peptid (jeweils drei Mal nativ und drei Mal oxidiert) zur Verfügung standen (**Abbildung 28**).

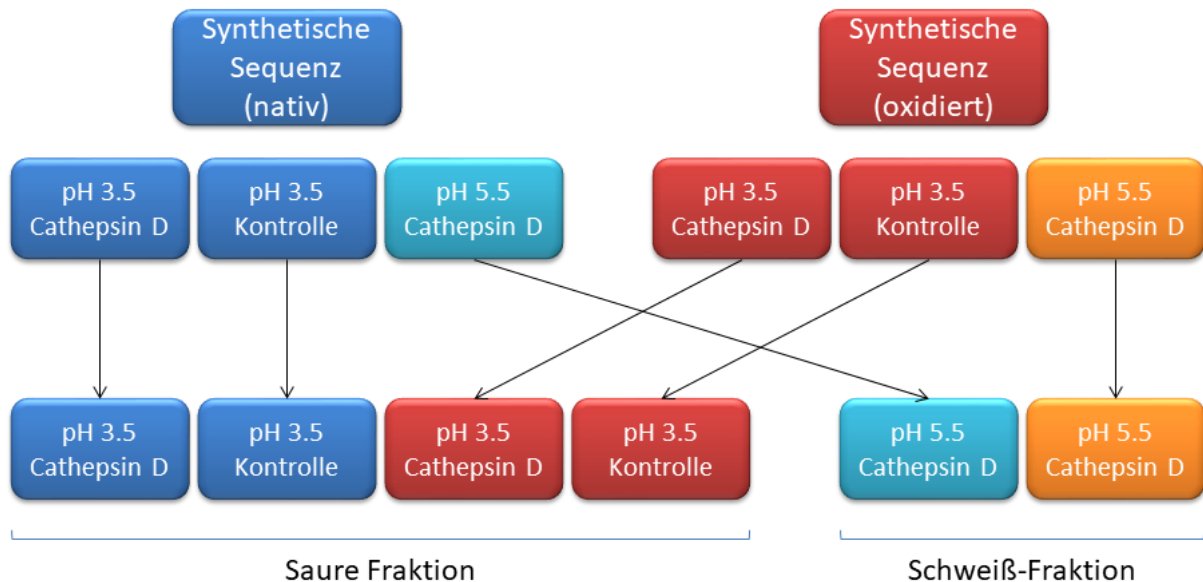


Abbildung 28 Proben-Übersicht des Cathepsin D Verdau Die Peptide wurden sowohl in reduzierter als auch oxidiert Form verwendet. Der Cathepsin D Verdau wurde bei den pH Werten 3,5 und 5,5 durchgeführt. Bei pH 3,5 wurden außerdem Kontrollproben untersucht.

Die Proben wurden in salzsaurem Glycin- (pH 3,5, 50 mM) oder Phosphat-Puffer (pH 5,5, 1 mM) aufgenommen und mit Cathepsin D (330 ng pro Aliquot, Enzym-zu-Substrat-Verhältnis etwa 1:40) versetzt. Die Ansätze wurden bei 37 °C inkubiert. Bei 5 und 30 min,

sowie 1, 2, 4, 8 und 24 h wurde 1 µL entnommen, mit 9 µL CHCA-Lösung (3,33 mg/mL in 70 % ACN) vermischt und direkt auf ein MALDI *target* gespottet. Die gespotteten Proben wurden dann an einem MALDI TOF/TOF 5800 (AB Sciex, Darmstadt, Deutschland) gemessen. Zusätzlich wurde 1 µL des 24 h Verdaus entnommen, mithilfe einer Ultimate3000 (Dionex, Idstein, Deutschland) aufgetrennt und an einer Q Exactive Plus (Thermo Scientific, Bremen, Deutschland) massenspektrometrisch analysiert.

vi. Aktivitäts-Assays

Die Aktivitäts-Assays gegen *S. aureus* wurden von Dr. Ulrich Gerstel im Arbeitskreis um Prof. Dr. Jens-Michael Schröder durchgeführt wurden. Sie waren nicht Teil der eigenen experimentellen Arbeiten.

vii. Analyse der nativen Hornerin-Peptide

Je 20 µL der HPLC-Fractionen, welche eine Aktivität gegen *P. aeruginosa* zeigten (siehe **Abbildung 27**), wurden im Arbeitskreis um Prof. Dr. Schröder vereinigt und lyophilisiert. Alle weiteren hier aufgeführten Schritte waren Teil der eigenen Experimente. Der Rückstand wurde in 50 µL einer 0,1-prozentigen TFA-Lösung in 3 % ACN aufgenommen und eine Verdünnungsreihe (1:10, 1:100 und 1:1000) angesetzt. Die Proben wurden dann mit steigender Konzentration flüssigkeitschromatographisch aufgetrennt und massenspektrometrisch sequenziert (Kapitel 7.c.ix).

viii. MALDI MS (TOF/TOF 5800, AB Sciex)

Der LASER emittierte elektromagnetische Strahlung einer Wellenlänge von 355 nm bei einer Wiederholrate von 1000 Hz. Für jedes Spektrum wurden vier Subspektren akkumuliert, die ihrerseits aus 250 Schüssen entstanden sind. Die *fixed laser intensity* (FLI) betrug 3.500. Die Spektren wurden im Bereich 1000-5000 m/z im positiven Ionenmodus aufgenommen. Als Matrix wurde α -Cyano-4-hydroxymizinsäure mit einer Konzentration von 3.33 mg/mL verwendet.

ix. LC-ESI MS (Ultimate3000, Dionex/Q Exactive Plus, Thermo Scientific)

Als Flüssigkeitschromatograph kam eine Ultimate3000 UPLC (Dionex, Idstein, Deutschland) zum Einsatz, die *online* an ein Q Exactive Plus Massenspektrometer (Thermo Scientific, Bremen, Deutschland) gekoppelt war. Die UPLC wurde mithilfe von DCMS Link 2.14 Build 3818 (Dionex) kontrolliert, das Massenspektrometer mit XCalibur 3.1.66.10 (Thermo

Scientific). Als Laufmittel wurden 0,1 % TFA in 3 % ACN (Ladepuffer), 0,05 % FA (Eluent A) 0,04 % FA in 80 % ACN (Eluent B) verwendet. Wenn nicht anders angegeben, wurden die Flussraten auf $30 \mu\text{L min}^{-1}$ für den Autosampler und 300 nL min^{-1} für die Trennung der Analyten eingestellt. Die Gradienten für die Trennung der Analyten sind in **Diagramm 1** (Desmoplakin) und **Diagramm 2** (Hornerin) gezeigt. Als Vorsäule wurde eine C18 Pepmap100 (Thermo Scientific, Bremen, Germany) verwendet (Partikelgröße: $5 \mu\text{m}$, Porengröße: 100 \AA , Säulenabmessungen: $300 \mu\text{m i.D.} \times 5 \text{ mm}$). Eine C18 Acclaim PepMap RSLC (Thermo Scientific, Bremen, Germany) kam als analytische Säule (Partikelgröße: $2,0 \mu\text{m}$, Porengröße: 100 \AA , Säulenabmessungen: $75 \mu\text{m i.D.} \times 150 \text{ mm}$) zum Einsatz.

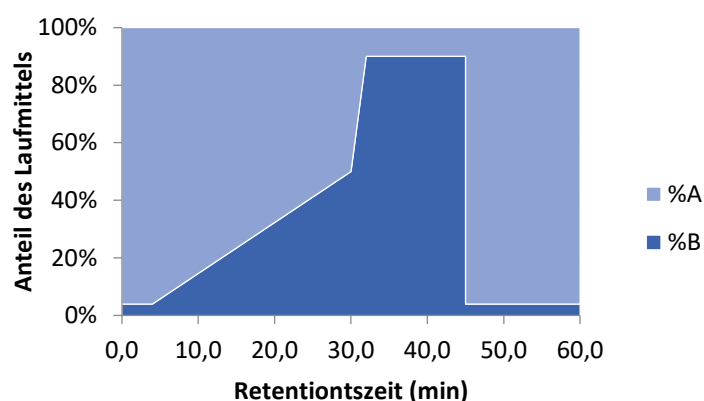


Diagramm 1 Gradient zur Trennung der Desmoplakin-Proben Die Vorsäule wurde zunächst für 4 min mit 4 % Eluent B gespült. Bei 4 min wurde dann das Ventil umgestellt (siehe **Abbildung 7**) und die Analyten getrennt, indem der Anteil an B in 26 min auf 50 % erhöht wurde. Hydrophobe Bestandteile wurden dann über 13 min mit 90 % B von der Säule eluiert und das System über 15 min mit 4 % B equilibriert.

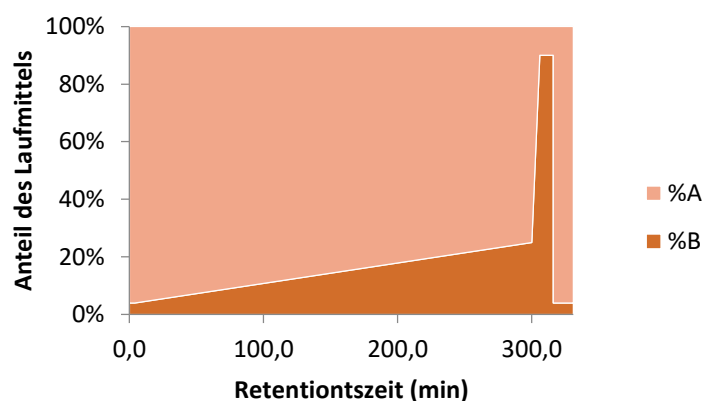


Diagramm 2 Gradient zur Trennung der Hornerin-Proben Die Vorsäule wurde zunächst für 4 min mit 4 % Eluent B gespült. Bei 4 min wurde dann das Ventil umgestellt (siehe **Abbildung 7**) und die Analyten getrennt, indem der Anteil an B in 296 min auf 25 % erhöht wurde. Anschließend wurde B in 6 min auf 90 % erhöht und hydrophobe Bestandteile über 10 min von der Säule eluiert. Zuletzt wurde das System über 15 min mit 4 % B equilibriert.

Aufgrund technischer Schwierigkeiten (Totvolumina) wurde auf die Absorptionsmessung im ultravioletten Bereich verzichtet. Die UPLC wurde online an das Massenspektrometer gekoppelt und die Probe mithilfe einer beschichteten Siliciumdioxid-Nadel (PicoEmitter SilicaTip, New Objective, USA) ionisiert. Die Ionisationsspannung betrug 2,41 kV. Der *full scan* wurde im Bereich 300-2000 m/z mit einem Auflösungsvermögen von 70.000 aufgenommen. Die AGC (*automatic gain control*), wurde auf 3e6 eingestellt. Die maximale Injektionszeit betrug 100 ms. Für jede MS¹-Messung wurden die zehn intensivsten Ionen zur Fragmentierung ausgewählt und in einem Fenster von 3 m/z isoliert. Einfach geladene Ionen und solche mit Ladungszuständen von acht oder höher wurden ausgeschlossen. Die Vorläuferionen wurden mittels HCD aktiviert und einer normalisierten Kollisionsenergie von 25 % fragmentiert. Das Auflösungsvermögen der Fragmentspektren betrug 17.500. Die AGC der Fragmentspektren wurde auf 1e5 eingestellt. Wurde ein Vorläuferion zur Fragmentierung ausgewählt, wurde dessen m/z-Wert für die nächsten 15 Sekunden nicht wieder ausgewählt.

x. Identifizierung der Desmoplakin- und Hornerin-AMP mittels Datenbanksuchen

Für die Datenbanksuche kam die Proteome Discoverer Software (version 1.4, Thermo Scientific, Bremen, Deutschland) zum Einsatz. Die Rohdaten wurden nach Vorläufern mit einer Masse zwischen 350 und 5.000 Da sowie einem minimalen Signal-zu-Rausch-Verhältnis von 1,5 vorgefiltert. Als Datenbank wurde das humane Proteom inklusive Isoformen (36.244 Sequenzen, Version: 04/2016) verwendet. Es wurde kein Enzym spezifiziert. Die Peptidlänge wurde auf 7-144 Aminosäuren beschränkt. Die Toleranzen wurden zu 10 ppm (MS¹) und 0,02 Da (MS²) festgelegt. Die Oxidation von Methionin wurde als dynamische Modifikation zugelassen. Die PSM wurden mit einer FDR von 0,01 (strikt) evaluiert.

d. Resultate und Diskussionen

i. Desmoplakin

Die Probe, welche die gegen *S. aureus* aktiven Peptide enthielt (**Abbildung 27**), wurde zunächst flüssigkeitschromatographisch aufgereinigt und online massenspektrometrisch untersucht. Das Chromatogramm zeigte zwei *Peaks*, die sich in ihrer Retentionszeit nur um etwa drei Minuten unterschieden (**Abbildung 29**). Die Intaktmassen konnten zu 2198,155

(Abbildung 29A) und 2181,132 Da (Abbildung 29B) bestimmt werden. Die Abweichungen der identifizierten Sequenzen betrugen 4,89 und 6,67 ppm.

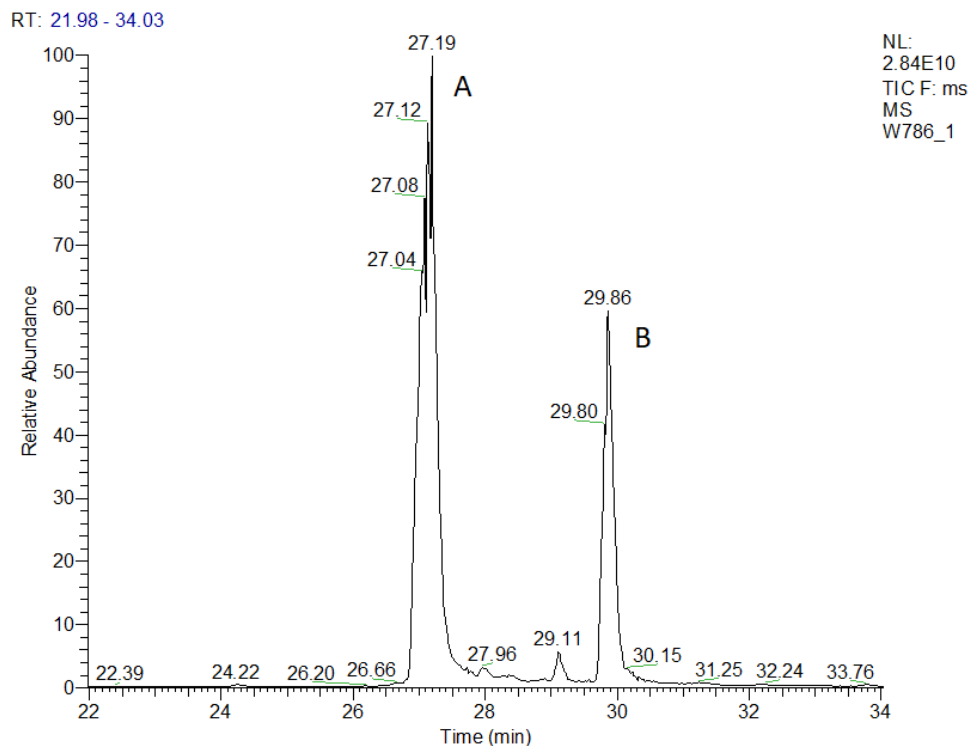


Abbildung 29 Total ion chromatogram (TIC) der aktiven Faktoren (LC-ESI MS, positiver Ionen-Modus) Die Probe enthielt zwei intensive Signale, die sich in der Retentionszeit um nur etwa drei Minuten und in der Masse um 17,023 Da unterschieden. Zwar sind noch weitere Analyten sichtbar, die biologische Aktivität konnte jedoch über mehrere Fraktionen hinweg mit diesen beiden Analyten korreliert werden.

Für beide Analyten, welche dreifach geladen auftraten, wurden ebenfalls Ionen detektiert, die der doppelten Masse entsprachen. Es wurde davon ausgegangen, dass es sich bei ihnen um Homodimere handelte. Nach Reduzierung und Alkylierung etwaiger Disulfidbrücken wurde die Probe erneut mittels LC-ESI MS untersucht. Die Massenverschiebung für Iodacetamid als Alkylierungsreagenz beträgt für einzelne Cysteine 57,022 Da und 116,059 Da für Disulfidbrücken. Für beide Peptide konnte eine Massenverschiebung von 116,059 Da und für die vermeintlichen Dimere von 232,117 Da nachgewiesen werden. Es konnte also davon ausgegangen werden, dass die Peptide über eine intramolekulare Disulfidbrücke verfügten. Das Tandem-Massenspektrum beider Analyten zeigte ein einziges intensives Signal bei etwa 229,155 Da (Abbildung 30). Der Vergleich des Masse-zu-Ladung-Verhältnisses mit den berechneten Werten aller proteinogenen Aminosäuren sprach für ein γ_2 -Ion bestehend aus Prolin und Leucin/Isoleucin.

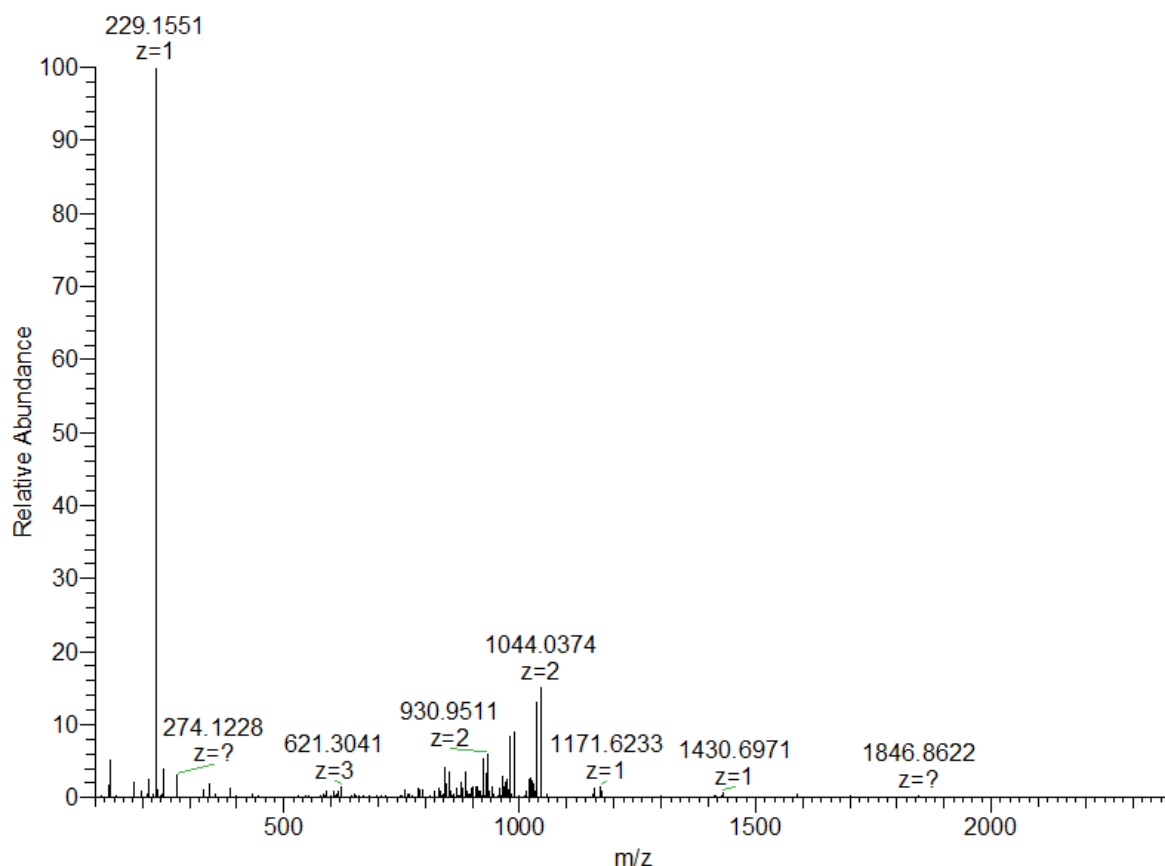


Abbildung 30 MS/MS-Spektrum des ersten Faktors (LC-ESI MS, positiver Ionenmodus, Vorläuferion: $[M+3H]^{3+} = 733,726 \text{ m/z}$) Neben einem sehr intensiven Signal bei 229 Da wurden kaum andere ausreichend intensive Signale detektiert. Eine Sequenzierung des Faktors ohne vorherige Modifizierung war daher nicht möglich.

Da ansonsten kaum intensive Signale detektiert wurden, war eine Sequenzierung nicht möglich. Aus diesem Grund wurden im Anschluss mehrere Proteasen unabhängig voneinander verwendet, um die beiden Sequenzen in kleinere Peptide zu hydrolysieren (Kaptiel 7.c.ii). Verwendet wurden Trypsin, Chymotrypsin, Pepsin, GluC und LysC. Trypsin spaltet Peptidbindungen C-terminal zu Arginin und Lysin, GluC nach Aspartat und Glutamat, LysC nach Lysin, Chymotrypsin nach den hydrophoben Aminosäuren Leucin, Phenylalanin, Tyrosin und Tryptophan. Pepsin hydrolysiert Peptidbindungen unspezifisch, nur nicht nach Cystein-Seitenketten. Es konnten für alle Enzyme kürzere Abschnitte nachgewiesen werden, welche durch Überlappung die Primärstruktur der beiden gesuchten Peptide aufklärten (**Tabelle 9**). Bei den Faktoren handelte es sich um einen kurzen Sequenzabschnitt von 19 Aminosäuren, der einzigartig dem strukturgebenden Protein Desmoplakin (AA 2428-2446) zugeordnet werden konnte. Desmoplakin agiert als Verbindung zwischen den Desmosomen und Filamentstrukturen, wie sie an den Kontaktflächen von Zellen

vorkommen. Mutationen im Desmoplakin-codierenden Gen sorgen für eine schwächere Bindung von Zellen und Zellverbänden [126].

Tabelle 9 Identifizierte Peptide nach Proteolyse Die beiden Faktoren wurden mit insgesamt fünf verschiedenen Enzymen verdaut. Mit einem * markierte Aminosäuren tragen eine Modifikation. Markierte Cysteine waren carbamidomethyliert, N-terminales Glutamat trat als zyklisches Pyro-Glutamat auf. Die Retentionszeiten beziehen sich auf den in **Diagramm 1** gezeigten Gradienten.

Protease	Generiertes Peptid	Retentionszeit (min)
Trypsin	QLKERC*IKDEETGLC*LLPL	42,06
	QLKERC*IKDEETGLC*LLPL	40,06
Chymotrypsin	QLKER C *IKDEETGLC*LLPL	37,05
GluC	QLKERC*IKDEETGLC*LLPL	41,78
	QLKERC*IKDEETGLC*LLPL	41,63
	QLKERC*IKDEETGLC*LLPL	21,35
	E* L KERC*IKDEETGLC*LLPL	25,23
LysC	QLKERC*IKDEETGLC*LLPL	39,56
Pepsin	QLKERC*IKDEETGL C *LLPL	26,64
	QLKERC*IKDEETGLC*LLPL	28,97
	QLKER C *IKDEETGLC*LLPL	14,83
	QLKERC*IKDEETGLC*LLPL	14,34
	E* L KERC*IKDEETGLC*LLPL	15,36

Besonders hilfreich waren die Ergebnisse des Verdaus der unspezifischen Protease Pepsin. Sie generierte zwei Sequenzen, die über fünf, bzw. sechs Aminosäuren verfügten und den C-Terminus der aktiven Faktoren abdeckte. Neben dem schon für die intakten Sequenzen sehr intensiven Signal bei 229,155 Da, bei dem es sich in der Tat um das γ_2 -Ion ^+PL-OH handelte, konnten weitere Fragmente detektiert werden, die halfen, die Aminosäure-Sequenz des C-Terminus zu lösen (**Abbildung 31**).

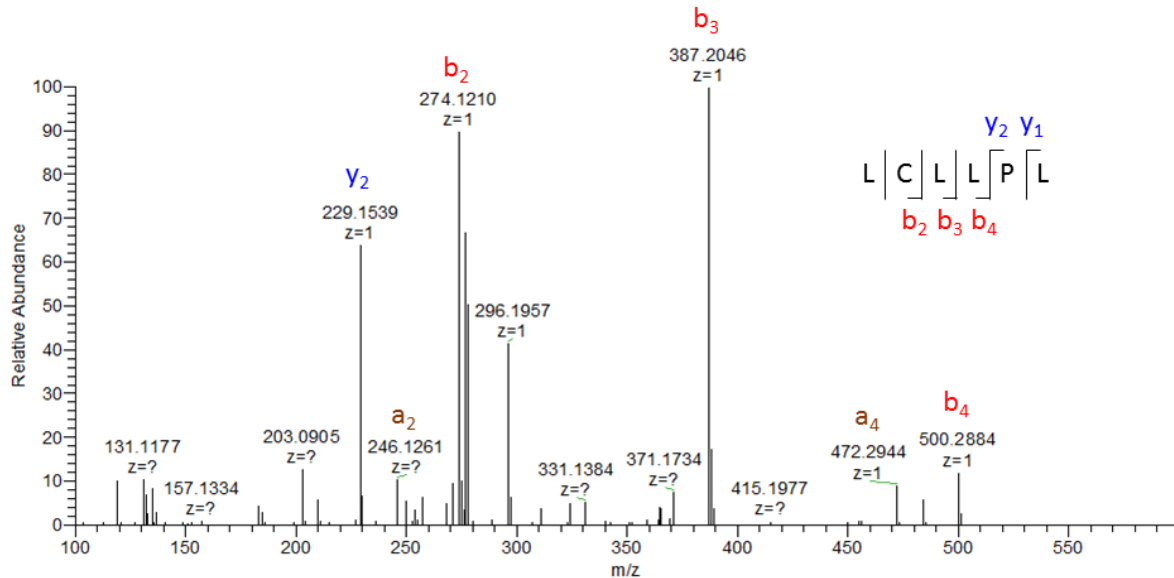


Abbildung 31 MS/MS-Spektrum der Sequenz LC_{alk}LLPL (LC-ESI MS, positiver Ionenmodus, Vorläuferion: 364.72 m/z) Für den verkürzten C-Terminus traten neben dem schon für die Faktoren dominanten y_2 -Ion weitere Fragmente auf, die es ermöglichten, die Primärstruktur aufzuklären. Neben dem y_2 -Ion traten ebenfalls die Fragmente b_2 , b_3 und b_4 mit hoher relativer Intensität auf.

Aufschlussreich waren weiterhin die Ergebnisse des Verdaus mit GluC. Sie lieferten die Primärstruktur des N-Terminus und konnten ebenfalls den geringen Massenunterschied der beiden Faktoren von etwa 17 Da erklären. Aufgrund der fast identischen Aminosäure-Sequenz konnte festgestellt werden, dass der Massenunterschied auf die N-terminale Aminosäure fiel. In der kanonischen Sequenz des Desmoplakin (AA 2428-2446) findet man an dieser Stelle ein Glutamin. N-terminale Glutamin-Seitenketten können durch Desaminierung in Pyro-Glutamat umgewandelt werden (**Abbildung 32**).

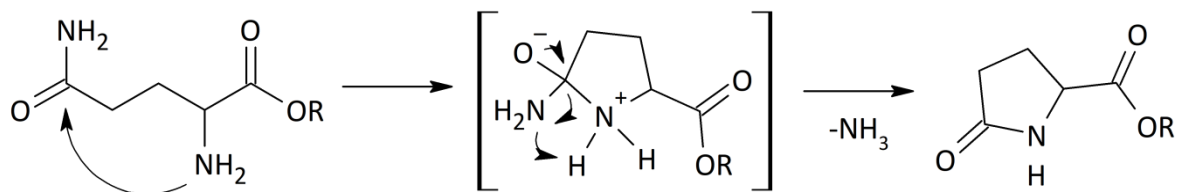


Abbildung 32 Desaminierung N-terminale Glutamin-Seitenketten können durch einen nukleophilen Angriff der freien Amino-Funktion unter Abspaltung eines Moleküls Ammoniak in Pyro-Glutamat umgewandelt werden. Diese Reaktion läuft spontan ab und ist eher temperatur- als pH-abhängig.

Es konnte gezeigt werden, dass diese Umwandlung hauptsächlich temperaturabhängig ist und spontan auftritt [127]. Dieser Umstand lässt vermuten, dass der Ringschluss am N-Terminus möglicherweise eine Folge der Proben-Lagerung ist. Diese Hypothese konnte jedoch nicht abschließend geklärt werden.

Nachdem die Sequenzen der Faktoren aufgeklärt worden waren, sollten die Schnittstellen im intakten Protein untersucht werden. An beiden befindet sich in der Sequenz eine Leucin-Seitenkette an der P1-Position. Um einen Hinweis über die dafür verantwortliche(n) Protease(n) zu erhalten, wurde die Merops-Datenbank (<https://www.ebi.ac.uk/merops/>) nach möglichen Einträgen durchsucht. Die Datenbank stellt eine Sammlung von Enzymen dar, für die (mögliche) Substrate und präferierte Erkennungssequenzen experimentell identifiziert worden sind. **Abbildung 33** zeigt die Einträge, die für die Eingaben X-X-X-L-|-Q-L-K-X (N-terminale Spaltstelle) sowie X-X-X-L-|-K-E-X-X (C-terminale Spaltstelle) ausgegeben wurden. X bezeichnet hier eine beliebige Aminosäure.

Database search with query XXX-XXX-XXX-LEU+GLN-LEU-LYS-XXX returned 2 hits

Substrate	Cleavage Site	Peptidase
Tyrosine-protein phosphatase non-receptor type 11	peptide-Leu210+Gln-peptide	A01.009 : cathepsin D
Stress-induced-phosphoprotein 1	peptide-Leu106+Gln-peptide	A01.010 : cathepsin E

Database search with query XXX-XXX-XXX-LEU+LYS-GLU-XXX-XXX returned 45 hits

Substrate	Cleavage Site	Peptidase
Arginase-1	peptide-Leu40+Lys-peptide	A01.009 : cathepsin D
Pyruvate kinase isozymes M1/M2	peptide-Leu61+Lys-peptide	A01.009 : cathepsin D
Rho GDP-dissociation inhibitor 2	peptide-Leu100+Lys-peptide	A01.009 : cathepsin D
Rho GDP-dissociation inhibitor 2	peptide-Leu100+Lys-peptide	A01.010 : cathepsin E
Pregnancy-associated glycoprotein	Peptide-Leu37+Lys-Peptide	A01.051 : pepsin F
ATP dependent DNA helicase 2 subunit 2	peptide-Leu306+Lys-peptide	A02.001 : HIV-1 retroviral

Abbildung 33 Ausgabe der Merops-Datenbanksuche Die Abbildung zeigt die Sucheingabe sowie die relevantesten Enzyme für die angegebenen Schnittstellen. Cathepsin D und E wurden als einzige Enzyme ausgegeben, die für die beiden Schnittstellen in Frage kommen könnten.

Für die zweite Eingabe wurden u.a. weitere Cathepsine (G, V und L), die Peptidyl-Lys-Metallopeptidase und Granzym M ausgegeben. Diese Listen wurden nach Enzymen durchsucht, für die eine Präsenz im *stratum corneum* realistisch ist. Cathepsin D, eine im Schweiß des Menschen nachgewiesene Endopeptidase, wurde als mögliches Enzym identifiziert, das für die Generierung der Faktoren verantwortlich sein könnte [128]. Es hydrolysiert Peptidbindungen primär C-terminal zu Leucin-Seitenketten und sekundär nach der ebenfalls hydrophoben Aminosäure Phenylalanin. Um die Fähigkeit der Protease zu untersuchen, die beiden Peptidbindungen zu spalten, wurden verlängerte Peptide (**Abbildung 34**) synthetisiert und mit humanem Cathepsin D inkubiert.

2420 2430 2440 2450 2460 2470
DTKGFFDPNT EENLTYLQLK **ERCIKDEETG** LCLLPLKEKK KQVQTSQKNT LRKRRVVIVD

2420 2430 2440 2450 2460 2470
DTKGFFDPNT EENLTYLQLK **ERCIKDEETG** LCLLPLKEKK KQVQTSQKNT LRKRRVVIVD

2420 2430 2440 2450 2460 2470
DTKGFFDPNT EENLTYLQLK **ERCIKDEETG** LCLLPLKEKK KQVQTSQKNT LRKRRVVIVD

Abbildung 34 Synthetische Desmoplakin-Peptide Um zu untersuchen, ob Cathepsin D in der Lage ist, die identifizierten Schnittstellen zwischen Leu-2428 und Gln-2429 sowie Leu-2446 und Lys-2447 zu hydrolysieren, wurden verlängerte Peptide (hier blau hinterlegt) synthetisiert und mit humanem Cathepsin D inkubiert. Die rot markierte Sequenz ist der nachgewiesene Faktor.

Die Sequenzen wurden in reduzierter Form synthetisiert und die im Faktor nachgewiesene Disulfidbrücke durch Inkubation in 10-prozentigem DMSO generiert (Kapitel 7.c.iii). DMSO gilt als schwaches Oxidationsmittel und die Kinetik der Reaktion erforderte für eine quantitative Umsetzung eine Inkubationszeit von 48 h (**Abbildung 35**). Die Reaktionsprodukte wiesen ausschließlich intramolekulare und keine intermolekularen Disulfidbrücken auf.

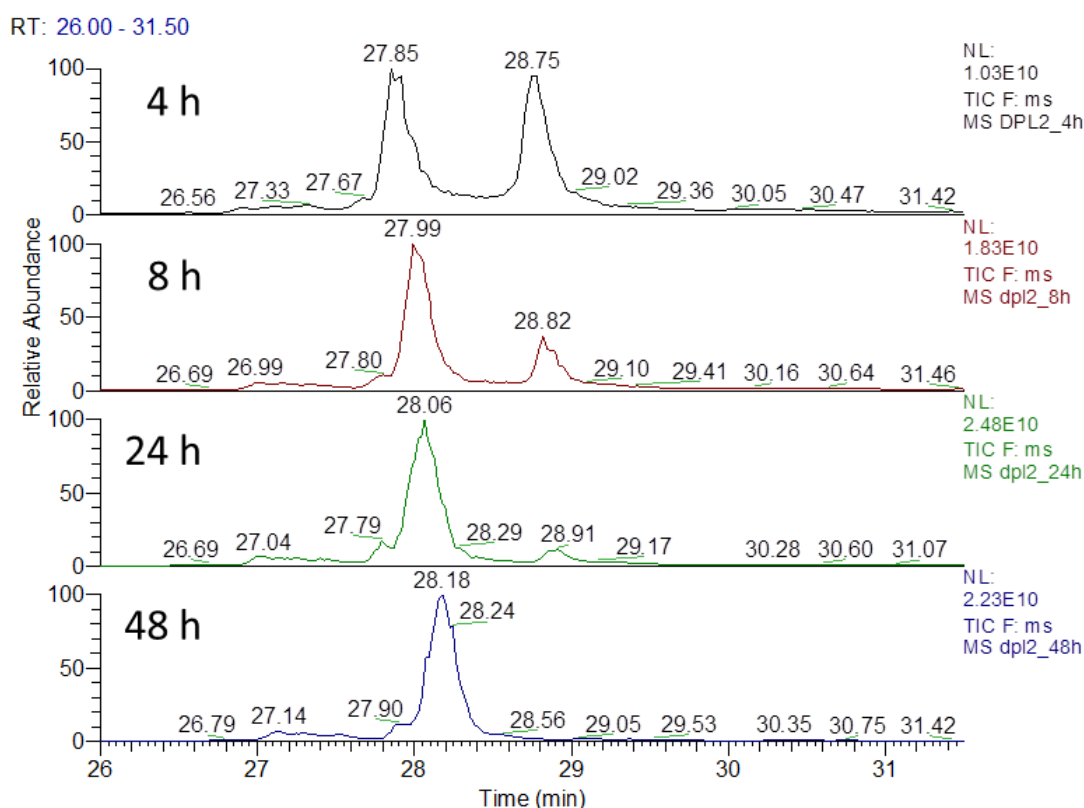


Abbildung 35 TIC der oxidierten Desmoplakin-Peptide (LC-ESI MS, positiver Ionenmodus) Die synthetisch generierten Desmoplakin-Peptide wurden bei Raumtemperatur in 10-prozentiger DMSO-Lösung inkubiert. Um den Fortschritt der Reaktion aufzuzeigen, wurden zu verschiedenen Zeitpunkten Proben entnommen und die Reaktionsprodukte aufgetrennt und analysiert. Für eine quantitative Umsetzung waren etwa zwei Tage nötig.

Für den Verdau wurden zwei verschiedene Puffersysteme verwendet. Das pH-Optimum des Cathepsin D liegt bei etwa 3,5. Ein Ansatz wurde daher in salzsaurem Glycin-Puffer durchgeführt. Um den auf der Haut vorkommenden pH-Wert zu imitieren, wurde für den zweiten Ansatz ein Phosphat-Puffer verwendet, dessen pH-Wert auf pH 5,5 eingestellt wurde. Die entstandenen Peptide wurden anschließend flüssigkeitschromatographisch getrennt und massenspektrometrisch sequenziert. Es konnte gezeigt werden, dass Cathepsin D in der Lage ist, die Peptidbindungen zwischen Leu-2427 und Gln-2428 sowie zwischen Leu-2446 und Lys-2447 zu spalten (**Abbildung 36**).

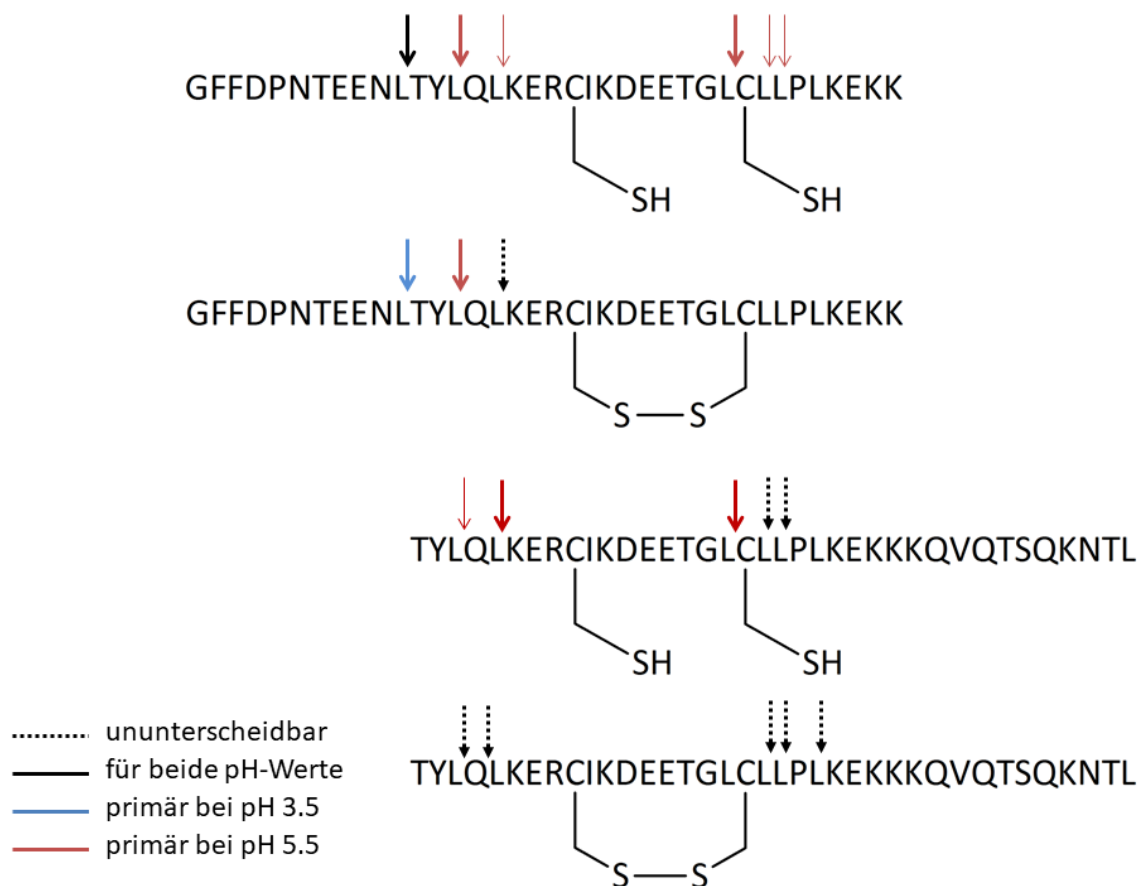


Abbildung 36 Identifizierte Spaltstellen des Desmoplakin/Cathepsin D Verdau Es konnte gezeigt werden, dass die Hydrolyse einiger Peptidbindungen bei verschiedenen pH-Werten präferiert wird. Die Spaltstelle zwischen Leu-2427 und Gln-2428 wurde bei beiden pH-Werten und für alle Spezies beobachtet, während die Hydrolyse zwischen Leu-2446 und Lys-2447 nur für die zweite, oxydierte Sequenz nachgewiesen werden konnte.

Während die Spaltprodukte der hydrolysierten Peptidbindung zwischen Leu-2427 und Gln-2428 sehr dominant auftraten, konnte das Peptid AA 2447-2461, welches ein Nachweis dafür ist, dass Cathepsin D auch die zweite Peptidbindung hydrolysieren kann, nur bei pH 3,5 und nur für die oxydierte Sequenz nachgewiesen werden (**Abbildung 37**).

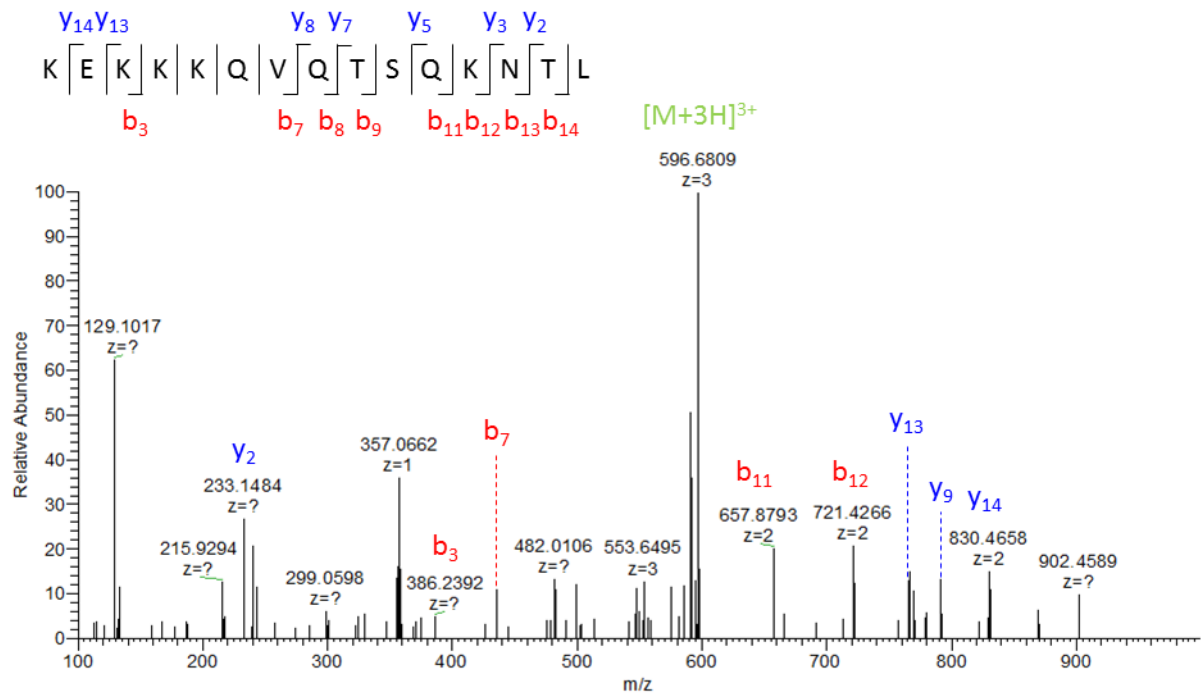


Abbildung 37 MS/MS-Spektrum der Sequenz KEKKKQVQTSQKNTL (LC-ESI MS, positiver Ionenmodus, Vorläuferion: $[M+3H]^{3+} = 596.681$ m/z) Das dominante Signal des Vorläufer-Ions weist darauf hin, dass die Aktivierungsenergie nicht hoch genug war, um das Peptid vollständig zu fragmentieren. Dementsprechend sind auch die Fragmentsignale wenig intensiv. Dennoch wurden, Neutralverluste einbezogen, 15 Fragmente beobachtet und das Peptid zweifelsfrei identifiziert.

Cathepsin D kam also für die Degradierung des Desmoplakins in Frage. Der Faktor wurde in Fraktionen nachgewiesen, die aus natürlichen Quellen stammten und somit eventuell auch andere für die biologische Aktivität wichtige Co-Faktoren enthalten haben könnte. Aus diesem Grund sollte der synthetisch generierte Faktor (Desmoplakin AA 2428-2446) in dessen reduzierter und oxidierter Form hinsichtlich seiner Fähigkeit untersucht werden, das Wachstum des Pathogens *S. aureus* zu inhibieren.

Hinweis: Der nachfolgende Absatz beschreibt die Aktivitätsassays der synthetischen Desmoplakin-Peptide. Diese Versuche wurden von Dr. Ulrich Gerstel im Arbeitskreis um Prof. Dr. Jens-Michael Schröder durchgeführt und waren nicht Teil der eigenen Arbeiten.

Das Peptid wurde den Medien hinzugefügt, die für das Wachstum von *S. aureus* in Mikrotiterplatten verwendet wurden. Da nicht ausgeschlossen werden konnte, dass Co-Faktoren für die Aktivität nötig sind und ähnliche Phänomene bereits für andere AMP nachgewiesen worden sind [129, 130], wurden die Aktivitätsassays mit und ohne den Zusatz von Metallionen durchgeführt. In Gegenwart von Spurenelementen konnten für die Faktoren (reduziert und oxidiert) in beiden Medien eine Inhibition beobachtet werden. Der

Versuch wurde mehrmals wiederholt, jedoch konnte das o.a. Ergebnis nicht verlässlich reproduziert werden. Aus diesem Grunde wurde das Projekt schließlich ohne ein endgültiges Ergebnis beendet.

ii. Hornerin

Vorversuche hatten gezeigt, dass die gegen *Pseudomonas aeruginosa* aktiven Kallusextrakte eine Vielzahl von Analyten enthielten, die sich chromatographisch kaum trennen ließen. Es konnte also davon ausgegangen werden, dass es sich um Peptide handelte, die sich in ihren physikochemischen Eigenschaften ähneln würden. Aus diesem Grund wurden die erhaltenen Extrakte über einen deutlich längeren Gradienten aufgetrennt als es für die Desmoplakin-Proben erforderlich war (vgl. **Diagramm 1** und **Diagramm 2** in Kapitel 7.c.ix). Da nicht vorhergesagt werden konnte, welche Enzyme an der Degradierung der ursprünglichen Proteine beteiligt waren, wurde in den Datenbanksuchen keine Enzymspezifität angegeben. Um neben den sehr abundanten Peptiden ebenfalls die weniger intensiven Analyten zu erfassen, wurde aus den Ergebnis-Dateien des ersten LC-ESI MS Laufs eine sogenannte *exclusion list* generiert, die die m/z -Verhältnisse aller identifizierten Peptide enthielt. Die gleiche Probe wurde unter Berücksichtigung dieser Liste erneut gemessen. Ein Vorläuferion, dessen m/z in der *exclusion list* enthalten war, wurde nicht erneut zur Fragmentierung ausgewählt, auch wenn es den Anforderungen (Ladungszustand größer gleich 2, minimales Signal-zu-Rausch-Verhältnis, den N intensivsten Ionen eines *full scans* zugehörig) entsprechen würde. Die im Folgenden diskutierten Ergebnisse liegen den kombinierten Resultaten beider Läufe zugrunde.

Mit insgesamt 122 identifizierten Peptiden (davon 117 *unique*) war Hornerin das Protein der Fraktion, für das die meisten Sequenzen identifiziert wurden. Die Primärstruktur des Hornerins (HRNR), welche zu 40 % abgedeckt wurde, umfasst 2850 Aminosäuren und besteht aus insgesamt 30 sich wiederholenden Sequenzen (**Abbildung 38**). Die Ähnlichkeit in der Aminosäuresequenz und die damit verbundenen ähnlichen physikochemischen Eigenschaften der Peptide könnten erklären, warum eine chromatographische Trennung der Analyten kaum möglich war. Interessanterweise enthielten alle nicht abgedeckten Abschnitte mindestens ein Cystein. Sollten einige oder alle dieser Cysteine modifiziert vorliegen, wie z.B. im Falle eine Palmitoylierung [131], so wären keine Peptide in diesen Abschnitten identifiziert worden. Das liegt daran, dass in der Datenbanksuche nur die

Oxidation von Methionin als mögliche dynamische Modifikation zugelassen wurde. Da die Datenbank weit über 30.000 Proteinsequenzen enthielt, musste die Zahl der dynamischen Modifikationen limitiert werden, um akzeptable Suchzeiten zu gewährleisten (Begründung siehe Kapitel 5.d). Auch würden palmitoylierte Cysteine unter Verwendung eines wässrigen Extraktionsmediums aufgrund ihres lipophilen Charakters vermutlich nicht extrahiert werden. Sollten sie wider Erwarten extrahiert worden sein, könnten diese Peptide während der flüssigkeitschromatographischen Trennung aufgrund der starken Wechselwirkung mit Umkehrphasen außerdem irreversibel adsorbiert worden sein. Die Gründe für das Fehlen dieser Sequenzabschnitte konnten nicht abschließend geklärt werden.

Viele der identifizierten Peptide überlappten sich in ihrer Sequenz, weswegen trotz der hohen Anzahl an Sequenzen (122) nur etwa 40 % der kanonischen Primärstruktur abgedeckt wurde. Ein Beispiel für diese Überschneidungen ist in **Abbildung 39** gezeigt. Die Auflistung aller identifizierten Peptide und deren Überlappung mit anderen Analyten sind im Anhang (Kapitel 10) zu finden.

	1	11	21	31	41	51	61	71	81
1	MPKLLQGVIT	VIDVIFYQYAT	QHGEYDTLTK	AELKELLENE	FHQILKNPND	PDTVDIILQS	LDRDHNKKVD	FTEYLLMIFK	LVQARNKIIG
91	KDYCQVSGSK	LADDTHQHQE	EQEETEKEEN	KRQESSFSHS	SWSAGENDSY	SRNVRGSLKP	GTESISRRLS	FQPDFSGQHN	SYSGQSSSYG
181	EQNSDSHQSS	GRGQCGSGSG	QSPNYGQHGS	GSGQSSSNDT	HGSGSGQSSG	FSQHKSSSGQ	SSGYSQHGSG	SGHSSSYGQH	GSRSGQSSRG
271	ERHRSSSGSS	SSYGQHGSGS	RQSLGHGRQG	SGSRQSPSHV	RHGSGSGHSS	SHGQHGSGSS	YSYSRGHYES	GSGQTSGFGQ	HESGSGQSSG
361	YSKHGSGSGH	SSSQGQHGST	SGQASSSGQH	GSSSRQSSSY	GQHEASRHS	SGRQGHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGRSS
451	SSGPYVSGSG	YSSGFGHHES	SSEHSSGYTQ	HGSGSGHSSG	HGQHGSRSGQ	SSRGSRQSSS	AGSSSSYGQH	GSGSRQSLGH	SRHGSGSGQS
541	PSPSRGRHES	GSRQSSSYGP	HGYGSGRSSH	RGPYESGSGH	SSGLGHQESR	SGQSSSYGQH	GSSSGHSSSTH	GQHGSTSGQS	SSCGQHGAIS
631	GQSSSHGQHG	SGSSQSSRYG	QQGSGSGQSP	SRGRHGSDFG	HSSSYQHGGS	GSGWSSSSNGP	HGVSQSQSSG	FGHKSGSQSS	SGYSQHGSGS
721	SHSSGYRKHG	SRSGQSSRSE	QHGSSSGLSS	SYGQHGSGSH	QSSGHGRQGS	GSGHSPSRVR	HGSSSGHSSS	HGQHGSGTSC	SSSCGHYESG
811	SGQASGFGQH	ESGSGQGYSQ	HGSASGHFSS	QGRHGSTSGQ	SSSSGQHDSS	SGQSSSYGQH	ESASHHASGR	GRHGSGSGQS	PGHGQRGSGS
901	GQSPSYGRHG	SGSGRSSSSG	RHGSGSGQSS	GFGHKSSSGQ	SSGYTQHGSG	SGHSSSYEQH	GSRSGQSSRS	EQHGSSSGSS	SSYGQHGSGS
991	RQSLGHGQHG	SGSGQSPSPS	RGRHGSGSQQ	SSSYGPHYRS	SGWSSSRGPY	ESGSGHSSGL	GHRESRSQSS	SGYQHGSSSS	GHSSTHGQHG
1081	STSGQSSSCG	QHGAASSGQS	SHGQHGSGSS	QSSGYGRQGS	GSGQSPGHGQ	RGSGSRQSPS	YGRHGSGSGR	SSSSQHGSGG	LGESSGFGHH
1171	ESSSQSSSYG	SGHGSGSGHS	SGYQHGSGRS	GQSSRGERHG	SSSGSSSHYG	QHGSGSRQSS	GHGRQSGSGG	HSPSRGRHGS	GLGHSSSHQQ
1261	HGSGSGRSSS	RGPYESRSGH	SSVFGQHESG	SGHSSAYSQH	GSGSGHFCSQ	GQHGSTSGQS	STFDQEGSST	GQSSSYGHRG	SGSSQSSGYG
1351	RHGAGSGQSP	SRGRHGSGSG	HSSSYQHGGS	GSGWSSSSGR	HGSGSGQSSG	FGHHESSSWQ	SSGCTQHGSG	SGHSSSYEQH	GSRSGQSSRG
1441	ERHGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS	RGRHGSGSQQ	SSSYSPYGS	SGWSSSRGPY	ESGSSHSSGL	GHRESRSQSS
1531	SGYGQHGSSS	GHSSTHGQHG	STSGQSSSCG	QHGAASSGQS	SHGQHGSGSS	QSSGYGRQGS	GSGQSPGHGQ	RGSGSRQSPS	YGRHGSGSGR
1621	SSSSQHGSGG	LGESSGFGHH	ESSSQSSSYG	SGHGSGSGHS	SGYQHGSGRS	GQSSRGERHG	SSSRSSSYRG	QHGSGSRQSS	GHGRQSGSGG
1711	QSPSRGRHGS	GLGHSSSHQQ	HGSGSGRSSS	RGPYESRSGH	SSVFGQHESG	SGHSSAYSQH	GSGSGHFCSQ	GQHGSTSGQS	STFDQEGSST
1801	GQSSSHGQHG	SGSSQSSSYG	QQGSGSGQSP	SRGRHGSGSG	HSSSYQHGGS	GSGWSSSSGR	HGSGSGQSSG	FGHHESSSWQ	SSGYTQHGSG
1891	SGHSSSYEQH	GSRSGQSSRG	EQHGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS	RGRHGSGSQQ	SSSYGPHYGS	SGWSSSRGPY
1981	ESGSGHSSGL	GHRESRSQSS	SGYQHGSGSS	GHSSTHGQHG	SASGQSSSCG	QHGAASSGQS	SHGQHGSGSS	QSSGYGRQGS	GSGQSPGHGQ
2071	RGSGSRQSPS	YGRHGSGSGR	SSSSQHGPGP	LGESSGFGHH	ESSSGQSSSY	SGHGSGSGHS	SGYGQHGSGS	GQSSRGERHG	SSSGSSSYRG
2161	QHGSGSRQSS	GHGRQSGSGG	HSPSRGRHGS	GSGHSSSHGQ	HGSGSGRSSS	RGPYESRSGH	SSVFGQHESG	SGHSSAYSQH	GSGSGHFCSQ
2251	GQHGSTSGQS	STFDQEGSST	GQSSSHGQHG	SGSSQSSSYG	QQGSGSGQSP	SRGRHGSGSG	HSSSYQHGGS	GSGWSSSSGR	HGSGSGQSSG
2341	FGHHESSSWQ	SSGYTQHGSG	SGHSSSYEQH	GSRSGQSSRG	ERHGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS	RGRHGSGSQQ
2431	SSSYSPYGS	SGWSSSRGPY	ESGSGHSSGL	GHRESRSQSS	SGYQHGSGSS	GHSSTHGQHG	STSGQSSSCG	QHGAASSGQS	SHGQHGSGSS
2521	QSSGYGRQGS	GSGQSPGHGQ	RGSGSRQSPS	YGRHGSGSGR	SSSSQHGSGG	LGESSGFGHH	ESSSQSSSYG	SGHGSGSGHS	SGYQHGSGRS
2611	GQSSRGERHG	SSSGSSSHYG	QHGSGSRQSS	GHGRQSGSGG	QSPSRGRHGS	GLGHSSSHQQ	HGSGSGRSSS	RGPYESRSLGH	SSVFGQHESG
2701	SGHSSAYSQH	GSGSGHFCSQ	GQHGSTSGQS	STFDQEGSST	GQSSSYGHRG	SGSSQSSGYG	RHGAGSGQSL	SHGRHGSGSG	QSSSYQHGGS
2791	GSGQSSGYSQ	HGSGSGQDGY	SYCKGGSNHD	GGSSGSYFLS	FPSSTSPYFY	VQEQRQCYFY			

Abbildung 38 Sequenzabdeckung des HRNR Obwohl mehr als 100 Peptide identifiziert worden sind, betrug die Sequenzabdeckung nur etwas über 40 %. Der Grund dafür sind überlappende Abschnitte und die Tatsache, dass sich das Protein aus insgesamt 30 sich wiederholenden Einheiten besteht.

Das Interessante an diesen Sequenzen ist, dass die Termini viele verschiedene Spezifitäten aufweisen, was dafür sprechen könnte, dass HRNR von mehreren Enzymen degradiert worden ist. Da sich viele der Peptide N- oder C-terminal nur um eine einzige Aminosäure unterschieden, wurde vermutet, dass neben Endo- auch Exopeptidasen HRNR hydrolysieren.

Trotz dieser Annahme wurden die identifizierten Peptide und die zugehörigen Termini hinsichtlich ihrer Häufigkeiten an den Positionen P4 bis P4' untersucht (**Abbildung 40**).

Abbildung 39 Überlappung von identifizierten Peptiden Viele der nachgewiesenen Sequenzen überschneiden sich z.T. vollständig mit anderen Peptiden. Die Struktur der N- und C-Termini dieser Sequenzen lässt die Vermutung zu, dass neben Endo- ebenfalls Exopeptidasen an der Degradierung des Hornerins beteiligt waren.

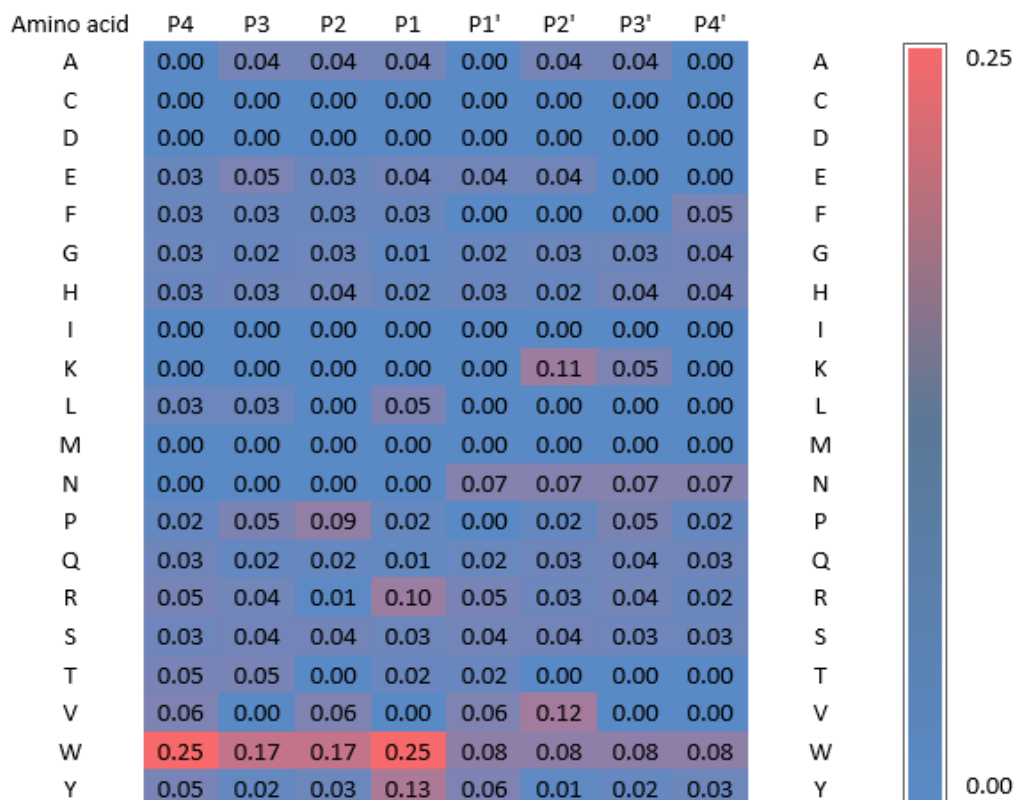
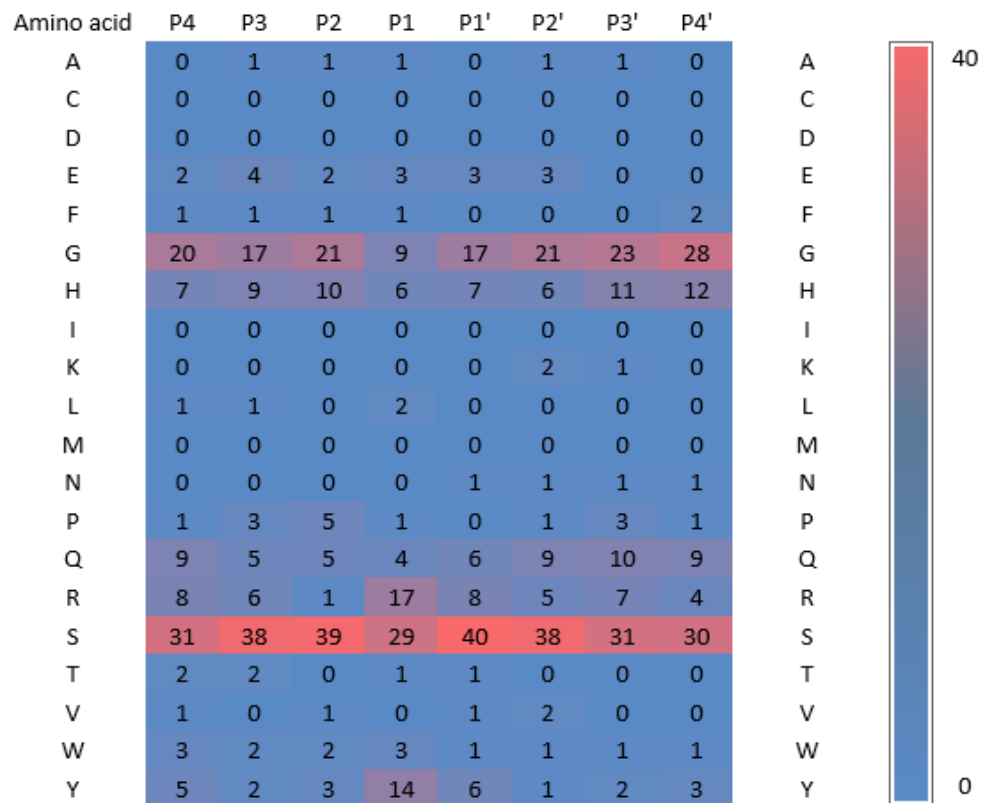


Abbildung 40 Absolute und relative Häufigkeit der Aminosäuren der Spaltstellen Die N- und C-Termini der 122 identifizierten Peptide wurden hinsichtlich der Aminosäure-Häufigkeit untersucht. Dabei wurde jede identifizierte Spaltstelle nur ein einziges Mal gezählt. Die Sequenzen stammen von ursprünglich 91 einzigartigen Spaltstellen im HRNR. Gezeigt sind die absoluten (oben) und relativen Häufigkeiten. Da das Protein sehr reich an Glycin und Serin ist, wurden diese entsprechend häufig beobachtet.

Obwohl Tryptophan im HRNR nur insgesamt zwölf Mal vorkommt, wurde diese Aminosäure drei Mal an der P1-Position identifiziert. Die Spaltstellen entsprachen entweder der Erkennungssequenz X-X-X-W-|-S-X-X-X oder X-X-X-W-|-Q-X-X-X. Die Suche in der Merops-Datenbank nach Enzymen, die in Experimenten gezeigt hatten, die Peptidbindung zwischen Tryptophan und Serin, bzw. Tryptophan und Glutamin hydrolysieren zu können, führte zu einer Liste von potentiellen Proteasen. Filtert man die Liste jedoch nach all den Enzymen, welche in oder auf der Haut vorkommen und nachgewiesenermaßen an der Degradierung von extrazellulären Matrixproteinen beteiligt sind, so kommen nur einige wenige in Frage. Dies wären zum einen die Matrixmetallopeptidase 2 (MMP2), welche für beide Sucheingaben ausgegeben wurde, und zum anderen Cathepsin D. Letzteres Enzym wurde bereits bei der Degradierung des Desmoplakins in Betracht gezogen. Die Matrix-Metallopeptidasen sind, wie der Name vermuten lässt, für die Degradierung von Proteinen verantwortlich, die sich in der extrazellulären Matrix (ECM) befinden [132]. Dazu zählen unter anderem Proteine, die an der Adhäsion zwischen Zellen beteiligt sind sowie alle strukturgebenden Moleküle der extrazellulären Matrix. Cathepsin D wurde, wie bereits in Kapitel 7.d.i erwähnt, im Schweiß nachgewiesen. Daher könnte auch dieses Enzym für die Degradierung von Proteinen im *cornified envelope* verantwortlich sein. Weiterhin wurde eine Tyrosin-Seitenkette in der P1-Position häufiger beobachtet, eine Spaltstelle, für die ebenfalls MMP2 in Frage kommt. C-terminale Tyrosin-Seitenketten könnten außerdem ein Hinweis für eine Aktivität der Profilaggrin Peptidase 1 sein, welche ihr Substrat Profilaggrin an benachbarten Tyrosin-Seitenketten hydrolysiert [133]. C-terminale Arginine wurden in den HRNR-Peptiden mit einer relativen Häufigkeit von etwa 10 % beobachtet. Basische Aminosäuren in dieser Position deuten möglicherweise auf die Aktivität von Serin-Proteasen hin, die eine dem Trypsin ähnliche Spezifität aufweisen. Dazu zählen viele Enzyme der Familie der Kallikreine. Zwei Proteasen, die dieser Gruppe zugehörig sind und im *stratum corneum* sowie *stratum granulosum* nachgewiesen wurden und dort u.a. eine Rolle in der antimikrobiellen Abwehr spielen, sind Kallikrein 5 und Kallikrein 7 [134, 135]. Sie sind im Gegensatz zu Trypsin spezifischer für Arginin an P1 und spalten nicht oder nur selten nach Lysin. Letztere kommt in der Sequenz des HRNR nur 19 Mal vor (< 1 %) und wurde nicht als C-terminale Aminosäure identifiziert. Relativ gesehen trat auch die hydrophobe Aminosäure Leucin häufig in der P1-Position auf, ein Befund, der ebenfalls für die vermutete Aktivität von

Cathepsin D spricht, jedoch wurde dieser Terminus nur in zwei von insgesamt 91 Peptiden beobachtet.

Unter den 122 identifizierten Sequenzen wurde auch der Abschnitt AA 2656-2684 nachgewiesen, welchen der Arbeitskreis um Prof. Dr. Schröder neben vielen anderen rekombinanten und synthetischen HRNR-Peptiden hinsichtlich einer Aktivität gegen *P. aeruginosa* untersucht hatte [136]. Konzentrationen der Peptide im niedrigen mikromolaren bis nanomolaren Bereich waren ausreichend, um das Pathogen zu eradizieren. Der Einfluss verschiedener HRNR-Peptide auf Kolonie-bildende Einheiten (CFU) von *P. aeruginosa* wurde von der AG Schröder in 10 mM Phosphat-Puffer (pH 5,5) in 1 %-iger TSB oder 0,25 %iger Glukose untersucht und waren nicht Teil der eigenen Arbeiten.

Neben HRNR wurden in den Kallus-Extrakten außerdem Peptide des Filaggrin-2, der S100-Proteine A8 und A9 sowie des Dermcidins nachgewiesen. Für C-terminale Abschnitte des Filaggrin-2 wurde eine antimikrobielle Wirkung gegen *P. aeruginosa* nachgewiesen [114]. Filaggrin-2, welches in intakter Form unlöslich ist, gilt als Substrat der Profilaggrin Peptidase 1. Dass Peptide beider Proteine, HRNR und Filaggrin-2, in der gleichen Fraktion nachgewiesen wurden, ist ein Hinweis dafür, dass HRNR auch von PEP1 degradiert werden könnte. Mit der Sequenz AA 2355-2375 wurde ein Abschnitt des FLG-2 nachgewiesen, der Teil des antimikrobiell äußerst potenten C-Terminus AA 2244-2391 ist [114]. Die beiden Proteine S100-A8 und S100-A9 sind insofern interessant, als dass sie als Komplex, auch bekannt als Calprotectin, antimikrobiell aktiv sind. Calprotectin konkurriert bei Infektionen mit dem Pathogen vor allem um das essentielle Spurenelement Mn^{2+} [129] und inhibiert so das bakterielle Wachstum. Dermcidin ist, wie bereits in Kapitel 7.a erwähnt, ein kleines Protein, dessen C-terminaler Abschnitt im Schweiß nachgewiesen wurde [113] und als Breitbandantibiotikum gegen eine Vielzahl von Bakterien aktiv ist, darunter *E. coli*, *E. faecalis*, *S. aureus* und *C. albicans*.

e. Zusammenfassung

i. Desmoplakin

Für verschiedene Fraktionen der Kallus-Extrakte konnte in vorherigen Versuchen die Wirksamkeit zweier Faktoren gegen *S. aureus* korreliert werden. Die Faktoren konnten unter Zuhilfenahme der Proteasen Trypsin, Chymotrypsin, Pepsin, GluC sowie LysC eindeutig

identifiziert werden. Sie bestanden jeweils aus 19 Aminosäuren und sind einzigartig für das strukturgebende humane Protein Desmoplakin (AA 2428-2446). Die entsprechenden Schnittstellen wiesen jeweils eine Leucin-Seitenkette in der P1-Position auf. In der Merops-Datenbank stimmten die dort angegebenen Spezifitäten der Proteasen Cathepsin D und E mit den identifizierten Spaltstellen (C-terminal zu Leucin) überein. Da Cathepsin D im Schweiß des Menschen nachgewiesen wurde, könnte dieses Enzym für die Degradierung des Desmoplakins verantwortlich sein. Um diese These zu untersuchen, wurden zwei verlängerte Desmoplakin-Sequenzen (AA 2414-2450 sowie zwischen Leu-2425-2461) mit humanem Cathepsin D bei dem für das Enzym optimalen pH-Wert von 3,5 und dem auf der Haut vorherrschenden Milieu (pH 5,5) inkubiert und die entstandenen Produkte analysiert. Es konnte gezeigt werden, dass Cathepsin D *in vitro* in der Lage ist, die beiden Spaltstellen zu generieren. Da nun sowohl die Sequenz als auch ein für die Generierung des potentiellen antimikrobiellen Faktors verantwortliches Enzym identifiziert worden waren, sollte das Peptid hinsichtlich dessen Potenz gegen das Pathogen *S. aureus* untersucht werden. Hierzu wurden die Bakterien von Dr. Ulrich Gerstel in Mikrotiter-Platten kultiviert und der synthetisierte Faktor (AA 2428-2446) nach Erreichen einer o.D. von 0,2 in reduzierter und oxidierter Form zugegeben. Da für den isolierten Faktor keine biologische Aktivität nachgewiesen werden konnte, wurden den Ansätzen im zweiten Versuch Spurenelemente hinzugefügt. Die Begründung für dieses Vorgehen war das Ergebnis von anderen Studien, die für andere AMP eine Abhängigkeit von Co-Faktoren beschrieben hatten. Nach Zugabe einer Spurenelementlösung konnte allerdings kein verlässlich reproduziertes Wachstum von SA dokumentiert werden. Die Aktivitätsassays haben gezeigt, dass die antimikrobielle Aktivität der synthetischen Peptide vermutlich von weiteren Faktoren abhängig ist, die in diesen Experimenten nicht eindeutig identifiziert werden konnten. Die Ergebnisse lassen jedoch die Vermutung zu, dass für die Wirksamkeit Metallionen nötig sind. Eine Möglichkeit, diese These zu untersuchen, wäre die Analyse der Fraktionen mit für Metallionen empfindlichen Methoden wie z.B. ICP MS oder AAS. Eine weitere wichtige Fragestellung ist der Verbleib anderer Desmoplakin-Peptide. Die identifizierten Sequenzen bestehen aus 19 Aminosäuren, während die kanonische Sequenz 2871 Aminosäuren zählt. Ähnlich wie das Hornerin besteht auch Desmoplakin aus mehreren *repeats*. Dieser Umstand könnte bedeuten, dass weitere DPL-Peptide, die über eine ähnliche Primärstruktur verfügen, ebenfalls antimikrobiell aktiv sein könnten. In einer Folgestudie könnte das angewandte Protokoll zur Extraktion

wiederholt und gezielt nach DPL-Peptiden gesucht werden. Die identifizierten Sequenzen verfügen, im Vergleich zu den Hornerin-Peptiden, neben den basischen Aminosäuren Lysin und Arginin mit drei Glutamat-Seitenketten und einem Aspartat auch über vier saure Aminosäuren. Es kann also spekuliert werden, dass die Aktivität gegen *S. aureus* auf einem anderen Mechanismus beruht. Um diese These zu untersuchen, sollten jedoch zunächst die für die antimikrobielle Aktivität notwendigen Abhängigkeiten identifiziert werden.

ii. Hornerin

Um die äußerste Hautschicht, das *stratum corneum*, auf antimikrobielle Peptide zu untersuchen, wurden von der AG Schröder vereinigte Kallusproben von der Ferse einer Extraktion unterzogen, die Extrakte mittels *weak cation exchange* (WCX) Flüssigkeitschromatographie gereinigt und an einer C8-Umkehrphase fraktioniert. Diese Fraktionen wurden auf eine Aktivität gegen das Pathogen *Pseudomonas aeruginosa* untersucht. Die aktiven Fraktionen wurden dann im Rahmen der vorliegenden Arbeit an einer C18-Umkehrphase chromatographisch weiter getrennt und massenspektrometrisch analysiert. In der Probe konnten insgesamt 122 Peptide des Hornerins nachgewiesen werden, wovon 117 Sequenzen für das Protein einzigartig waren. Die Sequenzabdeckung betrug etwa 40 %. Einige identifizierte Peptide waren nur um einige wenige Aminosäuren kürzer als andere, weswegen vermutet werden konnte, dass neben Endopeptidasen auch Exopeptidasen an der Degradierung des Hornerins beteiligt gewesen sind. Um einen Hinweis auf die für die Hydrolyse verantwortlichen Enzyme zu erhalten, wurden die Termini der 122 Sequenzen untersucht. Aufgrund der repetitiven Primärstruktur des Proteins wurden unter den theoretisch möglichen 244 Spaltstellen nur 91 einzigartige beobachtet. An den Peptidtermini waren die Aminosäuren Glycin und Serin besonders häufig. Da HRNR jedoch sehr reich an beiden ist (Glycin: 24,5 %, Serin: 33,6%), wurde dieser Befund nicht weiter erläutert. Um die absolute Häufigkeit der einzelnen Aminosäuren zu berücksichtigen, wurden die Ergebnisse relativiert. Eine sehr hohe relative Häufigkeit wiesen die Aminosäuren Tryptophan, Tyrosin, Arginin und Leucin an der P1-Position auf. Enzyme, die für die Degradierung in Frage kämen, waren zum einen Cathepsin D, eine im Schweiß des Menschen nachgewiesene Protease, die Matrix Metalloprotease 2 sowie die Profilaggrin Peptidase 1. Die basische Aminosäure Arginin, welche ebenfalls häufig am C-Terminus beobachtet wurde, spricht außerdem für Enzyme, deren Spezifität dem Trypsin ähneln. Zwei mögliche Proteasen gehören zur Familie der Kallikreine, nämlich Kallikrein-5 und Kallikrein-7.

Neben Peptiden des Hornerins, wurden außerdem Sequenzen der Proteine Filaggrin-2, S100-A8 und –A9 sowie Dermcidin identifiziert. Für diese Proteine oder Abschnitte dergleichen ist eine antimikrobielle Aktivität bekannt.

Dieser Versuch zeigte, dass das *stratum corneum* eine Vielzahl von aktiven Substanzen enthält, die ein unkontrolliertes Wachstum von pathogenen Bakterien und Kommensalen auf der Haut verhindern könnten. Die gegen *P. aeruginosa* aktiven Peptide konnten eindeutig dem Hornerin zugeordnet werden, welches vermutlich durch verschiedene Endo- und Exopeptidasen degradiert wird. Es ist durchaus wahrscheinlich, dass bei der Hydrolyse auch Proteasen mikrobiellen Ursprungs eine Rolle spielen. Es wurde bereits gezeigt, dass Kommensale wie z.B. *Fingoldia magna* proteolytische Aktivität freisetzen [137].

Die Identifizierung der Hornerin Peptide erfolgte in Proben, die von gesunden Patienten stammen. Da für einige von ihnen eine antimikrobielle Wirkung nachgewiesen werden konnte, stellt sich die Frage, ob ähnliche Sequenzen auch bei einer Erkrankung der Haut, wie z.B. atopischer Dermatitis, nachgewiesen werden können oder ob die durch die Erhöhung des pH-Wertes aktiven Serin-Proteasen den antimikrobiellen Schutz u.a. dadurch torpedieren, dass wirksame Peptide durch Hydrolyse inaktiviert werden. Eine solche Studie könnte durch die Applikation des verwendeten Extraktionsprotokolls für Kallus-Proben von verschiedenen, gesunden wie kranken Patienten realisiert werden und weitere Einblicke in die natürliche Schutzfunktion der Haut geben. Es könnte weiterhin zeigen, ob die Zusammensetzung und Primärstruktur der nachgewiesenen AMP auch zwischen Patienten der gleichen Gruppe (z.B. gesund vs. gesund) gleich sind oder sich individuell unterscheiden.

8. Weitere Projekte

In den folgenden beiden Kapiteln sollen experimentelle Arbeiten und deren Ergebnisse kurz vorgestellt werden, die entweder nicht umfangreich genug waren, um ein eigenes detailliertes Kapitel zu rechtfertigen, oder zu nicht erwarteten Ergebnissen geführt haben.

a. Phosphorylierung von Peptiden durch eine Michael-Addition zur Generierung von Phosphocystein-Analoga

Die Phosphorylierung von Proteinen ist eine posttranslationale Modifizierung, die unter anderem in der Signaltransduktion eine große Rolle spielt [138, 139]. Die Phosphorylierung wird von Enzymen katalysiert, die als Kinasen bezeichnet werden. Bei der Reaktion wird ein Molekül Phosphat von ATP oder GTP auf das Substrat übertragen. Ziel dieser Reaktion sind primär die Hydroxylgruppen der Aminosäuren Serin, Threonin und Tyrosin. Die Phosphorylierung anderer Aminosäuren fand aufgrund derer Säurelabilität lange wenig Beachtung. Die Laufmittel der in der Proteomik verwendeten flüssigkeitschromatographischen Systeme weisen heutzutage meist einen sauren pH-Wert auf, bei dem die Phosphor-Stickstoff-, bzw. Phosphor-Schwefel-Bindung anderer phosphorylierter Aminosäuren nicht stabil sind. Beispielweise konnte festgestellt werden, dass die Schwefel-Phosphor-Bindung in Thiophosphatestern bei pH-Werten zwischen 7 und 12 stabil ist, während die kürzesten Halbwertszeiten (um 15 min) bei pH 2-4 beobachtet wurden [140, 141]. Das Vorhandensein von säurelabilen Phosphorylierungen konnte durch die Adaption vorhandener Protokolle auch an den Aminosäuren Cystein und Histidin nachgewiesen werden [142, 143]. Da über das Verhalten dieser Modifizierungen in der Massenspektrometrie, insbesondere für Phosphocystein, wenig bekannt ist, sollten durch enzymatische und chemische Methoden kürzere Peptide generiert und durch verschiedene Fragmentierungsmethoden analysiert werden. Die Übertragung eines Phosphatrestes auf eine Cystein-Seitenkette sollte zunächst enzymatisch erreicht werden. Dazu wurden die Aminosäuresequenz LRRASLG, auch als „Kemptid“ und Substrat der Proteinkinase (PKA) bekannt [144], und deren Cystein-Analogon LRRACLG mit der PKA in einem geeigneten Reaktionspuffer (1 mM ATP und 10 mM MgCl_2 in 25 mM HEPES-Puffer, pH 8,0) umgesetzt. Da die Phosphorylierung des Cysteins auf diese Weise jedoch nicht erreicht werden konnte, wurde versucht, das Peptid chemisch zu phosphorylieren. Der Versuch sah die Übertragung eines Phosphatrests in zwei Schritten vor. Zunächst sollte die Serin-Seitenkette durch eine β -

Eliminierung und unter Katalyse von Ba^{2+} -Ionen in die Dehydroalanin-Spezies umgewandelt werden. Anschließend sollte eine Michael-Addition des Thiophosphats, wie in [145, 146] beschrieben, erfolgen (**Abbildung 41**). Als Substrat diente erneut das Kemptid, welches synthetisch erzeugt wurde (Biosynthan, Berlin, Deutschland).

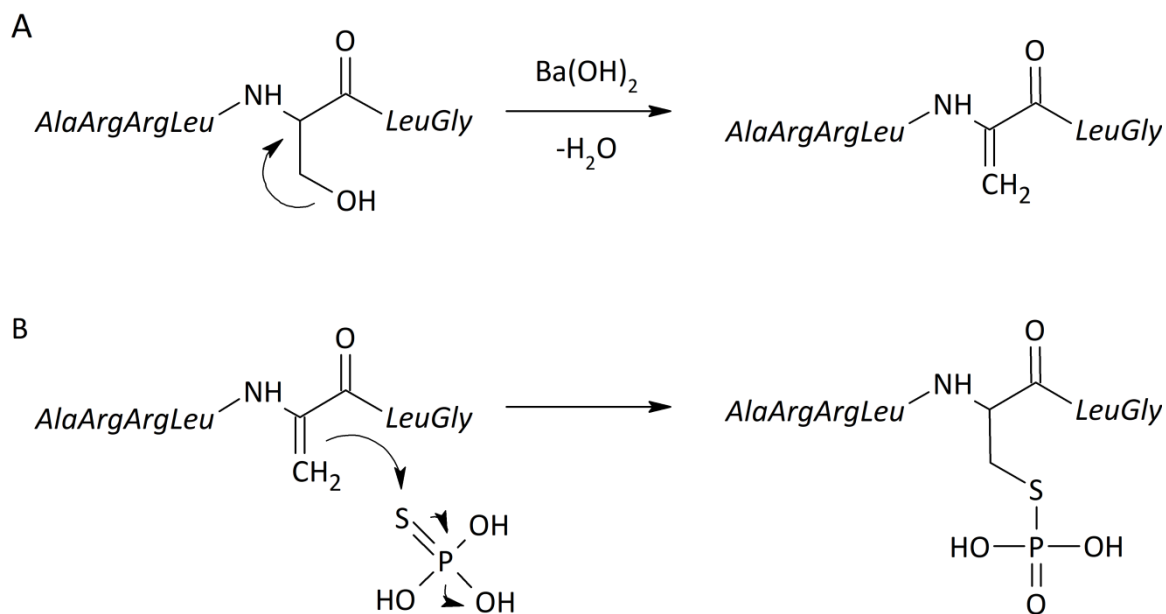


Abbildung 41 Umwandlung des Peptids in zwei Schritten Das Kemptid (LRRASLG) wurde zunächst in Erdalkalilauge dehydratisiert (A). Das Produkt dieser Reaktion, welches ein Dehydroalanin enthielt, sollte dann mit Thiophosphat umgesetzt werden, um die Phosphocystein-Spezies zu generieren (B).

Es sollte zunächst mit Bariumhydroxid dehydratisiert werden (**Abbildung 41A**). Die Umwandlung konnte durch MALDI MS bestätigt werden. Im Anschluss sollte das modifizierte Kemptid, welches nun ein Dehydroalanin enthielt, mit Natriumthiophosphat zur Phosphocystein-Spezies umgewandelt werden (**Abbildung 41B**). Thiophosphat hydrolysiert in wässrigen Lösungen leicht zu Phosphat und Schwefelwasserstoff (erkennbar am Geruch nach faulen Eiern, der nach einiger Zeit auftrat), weswegen das Salz im molaren Überschuss hinzugefügt wurde. Die chemische Phosphorylierung des Dehydroalanins konnte weder mit MALDI TOF-MS, noch mit LC-ESI Orbitrap MS nachgewiesen werden.

Das ursprüngliche Protokoll modifizierte mit Histon H3 ein intaktes Protein [146]. Leider wurden von der Arbeitsgruppe keine Fragmentspektren gezeigt, sodass eine genauere Untersuchung der Daten nicht möglich ist. Die eigenen Versuche lassen vermuten, dass sich die Methode ohne vorherige Anpassung nicht für die Modifizierung auf Peptidebene eignet.

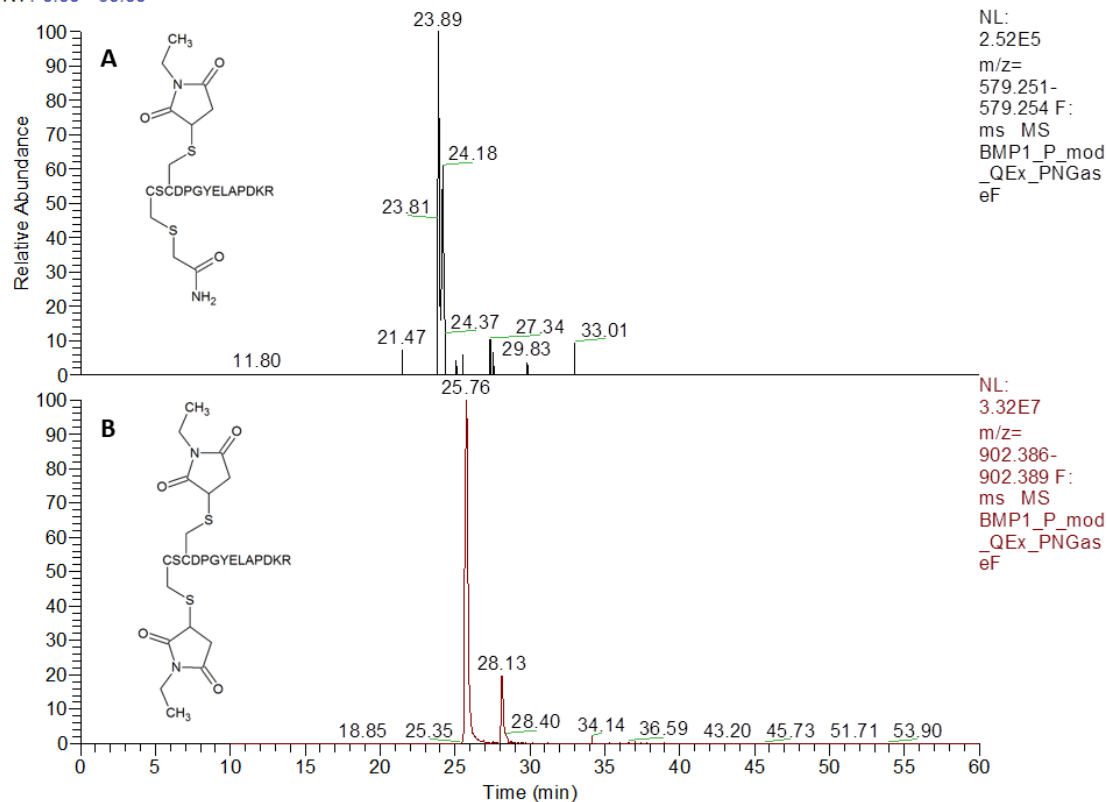
b. Disulfid-Mapping des *bone morphogenetic protein 1* (BMP1)

Das knochenmorphogenetische Protein 1 (aus dem Engl., *bone morphogenetic protein 1*) ist eine Protease, die für die Prozessierung der Prokollagene I-III verantwortlich ist und so eine elementare Rolle in der Formation der extrazellulären Matrix spielt [147]. Das prozessierte Protein enthält 24 Cysteine, von denen einige nachgewiesenermaßen in Disulfidbrücken verknüpft sind [148]. Da die Disulfidstruktur des BMP1 jedoch nicht endgültig aufgeklärt ist, sollte die Verknüpfung experimentell bestimmt werden. Die Versuche wurden primär von Chien-Wen Hung durchgeführt und ausgewertet. Die Ergebnisse wurden in *J Proteomics* veröffentlicht [149]. Aufgrund von Rückfragen während der Revision waren weitere Experimente nötig, die Teil der vorliegenden Arbeit sind. Im nativen Protein sollte untersucht werden, ob freie Cystein-Seitenketten vorhanden sind und wenn ja, um welche es sich handelt. Dazu wurde das Protein in HEPES-Puffer aufgenommen und freie Cystein-Seitenketten mit Iodacetamid alkyliert. Im Anschluss wurde überschüssiges IAA mittels SPE entfernt und vorhandene Disulfidbrücken mit TCEP reduziert. Die so generierten Thiol-Gruppen wurden mit NEM alkyliert. Auf diese Weise sollte ermöglicht werden, im intakten Protein zwischen freien und verknüpften Cysteinen zu unterscheiden. Alle ursprünglich freien Cysteine sollten nun carbamidomethyliert sein, während die Sulfhydrylgruppen der verbrückten Cystein-Seitenketten mit Maleinimid blockiert sein sollten. Im Anschluss folgte ein Doppelverdau mit Chymotrypsin und Trypsin über jeweils zwei Stunden. In den Arbeiten von Chien-Wen Hung konnten ebenfalls mehrere *N*-Glykosylierungen nachgewiesen werden, weswegen im letzten Schritt alle *N*-Glykane mit PNGase F entfernt und die Proben flüssigkeitschromatographisch aufgetrennt und massenspektrometrisch analysiert wurden. Eine Übersicht der identifizierten Cystein-Modifikationen ist in **Tabelle 10** dargestellt. Fünfzehn mit NEM modifizierte Cystein-Seitenketten wurden mit hoher Verlässlichkeit identifiziert. Für fünf weitere Cysteine, nämlich Cys-205, Cys-319, Cys-322, Cys-461 und Cys-559, wurde eine NEM-Modifizierung entweder mit geringer Güte oder gar nicht nachgewiesen. Eine manuelle Auswertung der Fragmentspektren durch Hung ergab, dass Cys-205 mit Cys-183, Cys-319 mit Cys-322, Cys-461 mit Cys-435 und Cys-559 entweder mit Cys-551 oder Cys-572 verknüpft ist. Für letztere wurde in den Alkylierungsexperimenten jedoch ebenfalls eine Carbamidomethylierung mit hoher Verlässlichkeit nachgewiesen, weswegen diese Aminosäure eventuell teilweise reduziert vorkommt.

Tabelle 10 Übersicht der Cystein-Modifizierungen der Doppel-Alkylierung Das native Protein wurde mit IAA alkylisiert, mittels SPE aufgereinigt, mit TCEP reduziert und die entstandenen freien Cysteine mit NEM alkylisiert. Das Protein wurde dann entweder einfach mit Trypsin oder doppelt mit Trypsin/Chymotrypsin (nacheinander) hydrolysiert und die Peptide mit LC-MS/MS analysiert. +: Peptide mit dieser Modifizierung wurden mit hoher Verlässlichkeit identifiziert (high confidence, FDR < 1 %); Low conf.: FDR > 5 % (wenige identifizierte Fragmentationen); -: nicht beobachtet; NEM: NEM und hydrolysiertes NEM.

	Signal (NEM-alkyliert)	Signal (IAA-alkyliert)	Verhältnis der XIC (NEM/IAA)	Anmerkung
Cys-163	+	Low conf.	//	
Cys-183	+	Low conf.	//	
Cys-185	+	Low conf.	//	IAA-Alkylierung nur an Cys 186 identifiziert (siehe Abbildung 42)
Cys-186	+	Low conf.	//	
Cys-205	Low conf.	Low conf.	//	
Cys-319	Low conf.	Low conf.	//	
Cys-322	Low conf.	Low conf.	//	
Cys-348	+	Low conf.	//	
Cys-375	+	-	//	
Cys-397	+	Low conf.	//	
Cys-435	+	Low conf.	//	
Cys-461	-	Low conf.	//	
Cys-488	+	Low conf.	//	
Cys-510	+	-	//	
Cys-551	+	Low conf.	//	
Cys-559	Low conf.	Low conf.	//	
Cys-563	+	Low conf.	//	
Cys-572	+	+	> 100	Beide Cysteine in einem Peptid identifiziert
Cys-574	+	+	> 100	(siehe Abbildung 42)
Cys-587	+	+	> 100	Beide Cysteine in einem Peptid
Cys-591	+	+	> 100	identifiziert
Cys-617	+	Low conf.	//	
Cys-644	+	Low conf.	//	
Cys-666	+	Low conf.	//	

RT: 0.00 - 60.00



RT: 0.00 - 60.00

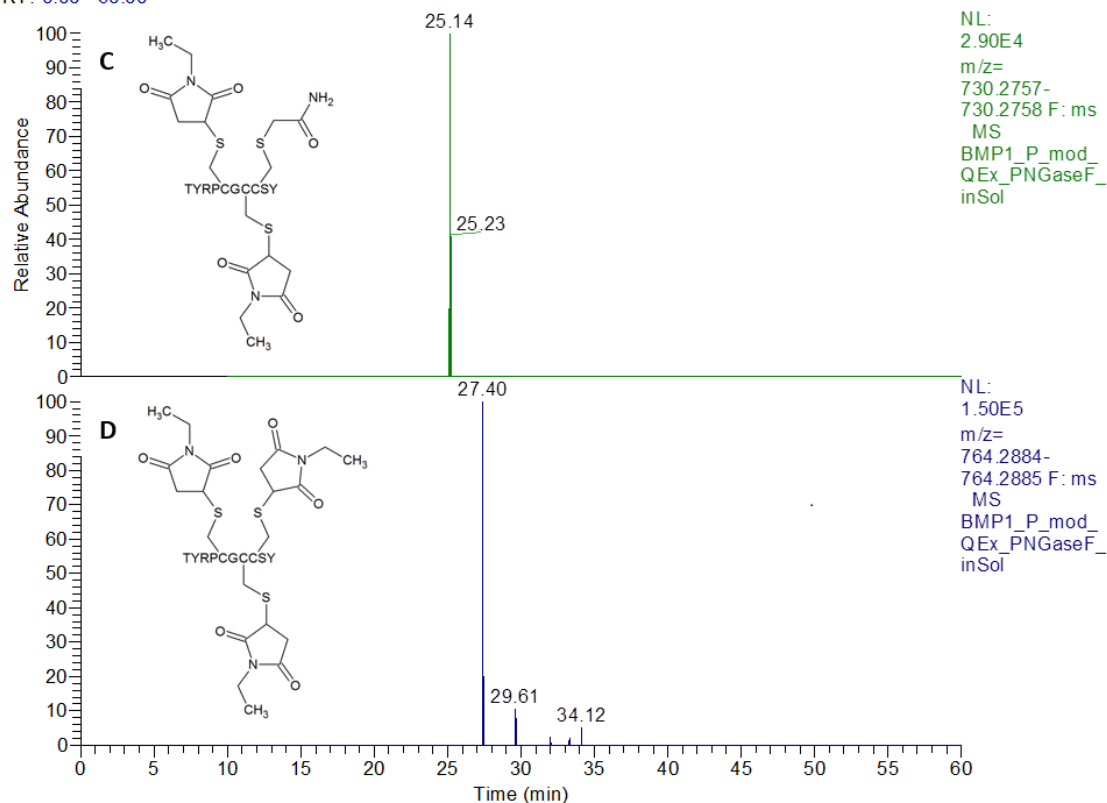


Abbildung 42 XIC der Schlüssel-Peptide aus Tabelle 10 Gezeigt sind die Peptide der Cysteine Cys-572 und Cys-574 (A/B) sowie Cys-183, Cys-185 und Cys-186 (C/D). In den oberen beiden XIC soll auf das Verhältnis der Intensitäten aufmerksam gemacht werden. Wie in **Tabelle 10** gezeigt, war das Verhältnis der mit NEM modifizierten Peptide um einen Faktor größer 100 höher als die IAA-Analoga. Dieses Verhältnis spricht für eine überwiegende Verknüpfung in Disulfidbrücken. Teilweise freie Cysteine können für diese Positionen jedoch nicht

ausgeschlossen werden. Ein mit IAA verknüpftes Cys-186 wurde mit geringer Verlässlichkeit identifiziert. Es wird jedoch vermutet, dass die Verknüpfung der Cysteine der vicinalen Disulfidbrücke dynamisch wechselt.

Neben Cys-572 sind weiterhin auch die Cysteine Cys-574, Cys-183 sowie die vicinalen Cys-185 und Cys-186 interessant. Diese wurden in gemeinsamen Sequenzen in verschiedenen Modifikationen nachgewiesen. Die entsprechenden Ionenchromatogramme (XIC) sind in **Abbildung 42** gezeigt. Die carbamidomethylierte Spezies der Sequenz in **Abbildung 42A** wurde mit hoher Verlässlichkeit identifiziert. Die Intensität des Vorläuferions im Chromatogramm war allerdings um einen Faktor 100 geringer als die zweifach maleinimidierte Spezies (**Abbildung 42B**). Cys-572 und Cys-574 liegen demnach primär in Disulfidbrücken verknüpft vor. Aufgrund der hohen Verlässlichkeit der Identifizierung von Sequenz **A** kann aber nicht ausgeschlossen werden, dass eine der beiden Aminosäuren zumindest teilweise reduziert vorliegt. Es sei an dieser Stelle angemerkt, dass wegen einer unzureichenden Ionenserie des Tandem-Massenspektrums die Position der Carbamidomethylierung weder Cys-572 noch Cys-574 zweifelsfrei zugeordnet werden konnte. Auch für Cys-587 und Cys-591 wurden Sequenzen mit einer carbamidomethylierten Cystein-Seitenkette verlässlich identifiziert. Auch hier sprach das Verhältnis der Intensitäten der XIC für eine primäre Verknüpfung in Disulfidbrücken. Die Sequenz AA 179-188 enthält insgesamt drei Cysteine, von denen Cys-185 und Cys-186 direkt benachbart sind. Ein Fragmentspektrum konnte mit hoher Verlässlichkeit der dreifach maleinimidierten Spezies zugeordnet werden (**Abbildung 42D**). Die Untersuchungen von Chien-Wen Hung zeigten, dass das Protein eine seltene Disulfidbrücke der o.a. benachbarten Cysteine enthält. Das zweifach NEM-modifizierte und einfach carbamidomethylierte Peptid (**Abbildung 42C**) wurde mit geringer Verlässlichkeit identifiziert. Es kann dennoch nicht ausgeschlossen werden, dass die Cysteine Cys-185 und Cys-186 teilweise reduziert vorliegen. Alle Ergebnisse können in [149] eingesehen werden.

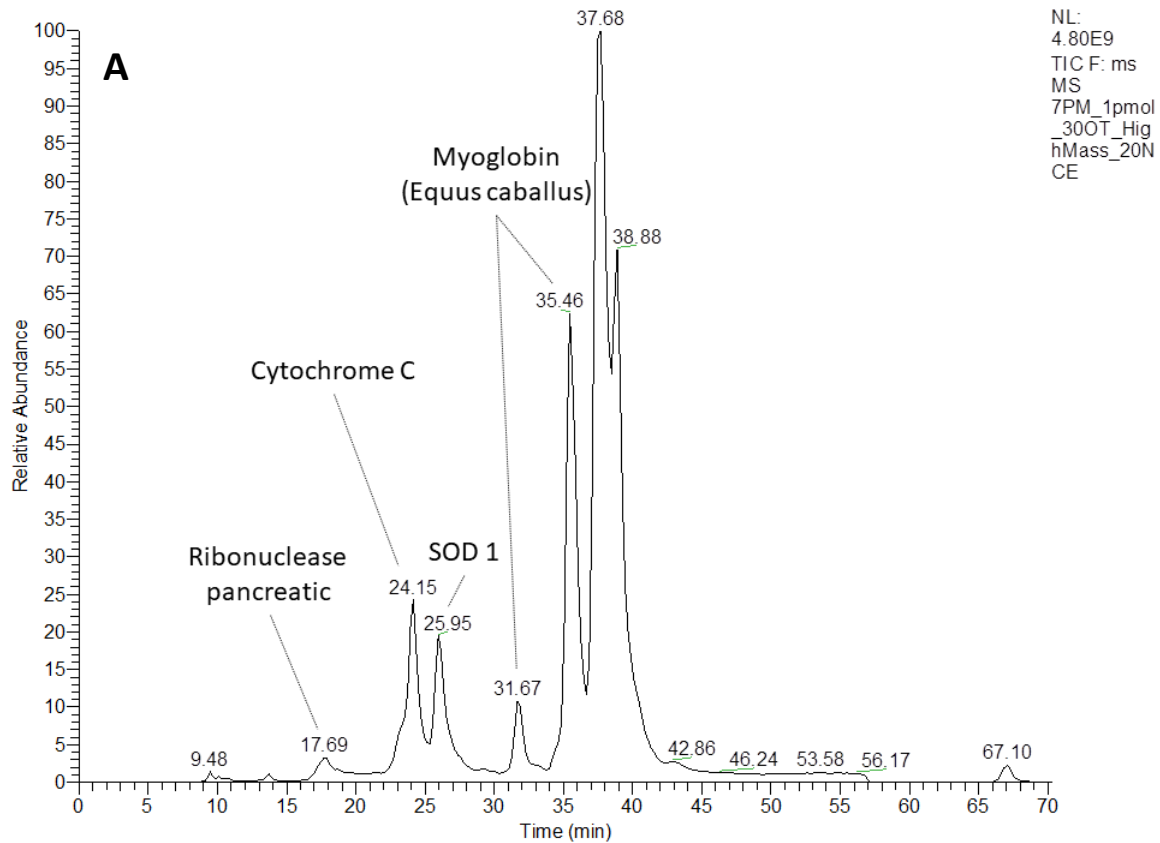
c. Top-Down Proteomics: Vorversuche

Im Vorfeld der in Kapitel 6 vorgestellten Versuche und Ergebnisse wurden verschiedene Feldversuche durchgeführt, um das Verhalten von intakten Proteinen in der Chromatographie (Interaktion mit Umkehrphasen) und Massenspektrometrie (Ionisationseffizienz, Fragmentierungsverhalten) zu untersuchen. Dazu wurde ein Proteinmix angesetzt, der die in **Tabelle 11** aufgeführten Analyten enthielt.

Tabelle 11 Übersicht der im Proteinmix enthaltenen Analyten Die Proteine deckten einen Massebereich von etwa 10.000 Da bis 80.000 Da ab. Durchgestrichene Proteine waren zunächst im Proteinmix enthalten, führten aber zu Problemen in der Chromatographie (Transferrin, Albumin) oder sind im gewählten sauren Milieu nicht löslich (Casein) und wurden daher im zweiten Ansatz nicht mehr verwendet.

Protein	Organismus	Molekule Masse $\left[\frac{g}{mol}\right]$
Cytochrom C	<i>Equus caballus</i> (Pferd)	11833
Albumin	<i>Bos taurus</i> (Rind)	69293
Carboanhydrase	<i>Bos taurus</i> (Rind)	28822
β-Casein	<i>Bos taurus</i> (Rind)	25107
Lysozym	<i>Gallus Gallus domesticus</i> (Huhn)	16239
Myoglobin	<i>Equus caballus</i> (Pferd)	17083
Transferrin	<i>Homo sapiens</i> (Mensch)	77064
β-Lactoglobulin	<i>Bos taurus</i> (Rind)	19883
α-Lactalbumin	<i>Bos taurus</i> (Rind)	16247
Ribonuclease A	<i>Bos taurus</i> (Rind)	16461

Aufgrund der schlechten Löslichkeit der Caseine im sauren Milieu wurde β-Casein nachträglich ausgeschlossen. Gleiches gilt für die sehr großen Proteine Albumin und Transferrin, die in den ersten Versuchen zu nicht auswertbaren Chromatogrammen führten. Die Probe wurde mit verschiedenen *setups* und Einstellungen untersucht. Zwei Beispiel-Chromatogramme sind in **Abbildung 43** gezeigt. Als Säulenmaterialien kamen C4-modifizierte Silicapartikel (**A**) und C4-modifizierte Silikapartikel mit *core shell* Technologie (**B**) zum Einsatz. Weiterhin wurden verschiedene Gradienten und Kollisionsenergien verwendet. Wie aus der Abbildung ersichtlich wird, konnten nicht alle aufgelisteten Proteine identifiziert werden. Als Gründe hierfür kommen niedrige Ionisationseffizienzen und/oder Ionisationssuppressionseffekte bei Co-Elution in Frage.



RT: 0.00 - 170.22

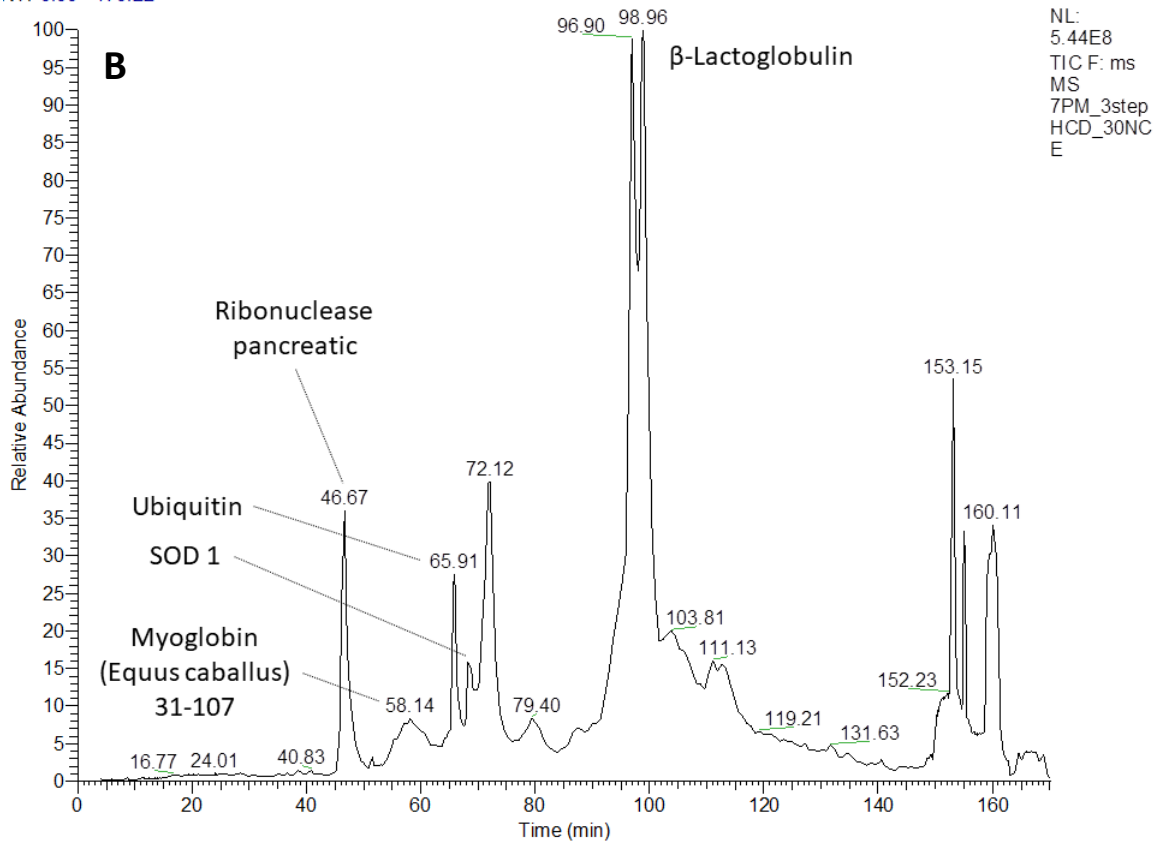


Abbildung 43 TICs des Proteinmixes Die Probe wurde mithilfe einer konventionellen porösen C4- (A) als auch mit einer C4-modifizierten Säule aufgetrennt, welche über sogenannte Core-Shell-Partikel verfügte (B). Letzte sollen sich aufgrund der kürzeren Diffusionswege vor allem für größere Peptide und Proteine eignen.

Wenn intensive Signale nicht zugeordnet wurden, lag dies daran, dass entweder keine aussagekräftigen MS¹-Spektren aufgenommen werden konnten (vgl. **Abbildung 23**) oder dass die stoßinduzierte Aktivierung zu Fragmentspektren führten, die nicht auswertbar waren (unzureichende Aktivierungsenergie, niedrige Intensität der Fragmentationen-Signale). Interessanterweise wurden Proteine identifiziert, welche zwar in **Tabelle 11** nicht aufgelistet, in den Standards jedoch als Verunreinigungen enthalten waren. Aus früheren Experimenten war beispielsweise bekannt, dass das verwendete Myoglobin sowohl die Superoxiddismutase 1 als auch Ubiquitin als Verunreinigung enthielt. Die Identifizierung erfolgte sowohl anhand der Intaktmasse (**Abbildung 44**), als auch mithilfe der Fragmentspektren.

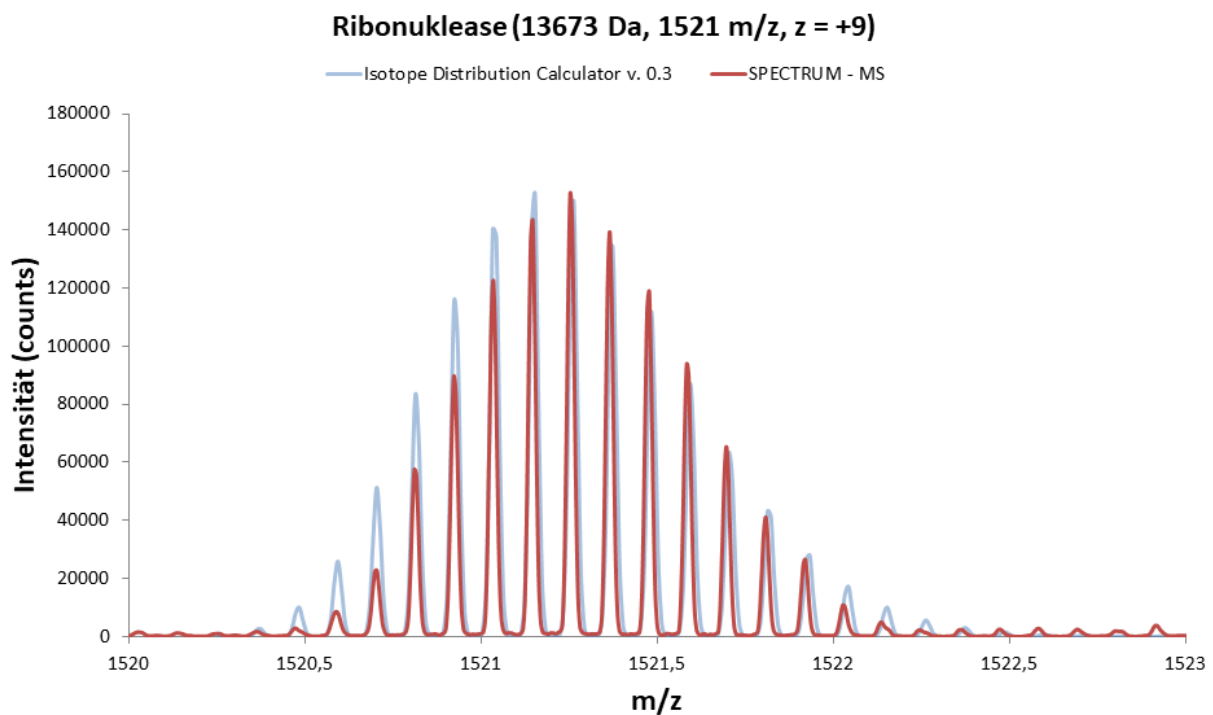


Abbildung 44 ESI MS-Spektrum der Ribonuklease (isoliert aus dem Pankreas, Vorläufer-Ion: 1521 m/z, Ladungszustand: +9) Das Spektrum zeigt die gemessene Isotopenverteilung (rot) als auch die berechnete Verteilung (blau). Für die Berechnung wurden die Aminosäuresequenz (ohne Signalpeptid, 27-150) und das Vorhandensein von vier Disulfidbrücken zugrunde gelegt [150].

Wie in Kapitel 4.e erklärt wurde, ist die Intaktmasse, bzw. die isotopische Verteilung eines Signals repräsentativ für alle PSA des Vorläuferions. Dies soll am Beispiel der Ribonuklease (Uniprot *accession number*: P61823) gezeigt werden. Dem reifen Protein fehlt beispielsweise das Signalpeptid (1-26) und acht der vorhandenen Cystein-Seitenketten sind in Disulfidbrücken involviert. Mithilfe eines geeigneten Algorithmus [151] kann nun die Isotopenverteilung der vorliegenden Sequenz berechnet und dargestellt werden. Vergleicht

man die berechnete Verteilung mit der gemessenen, so ergibt sich für die Ribonuklease eine Massenabweichung des intensivsten Signals von 0,009 Da (oder 5.6 ppm). Als nächstes sollen die identifizierten Fragmente diskutiert werden (**Abbildung 45**).



Abbildung 45 Identifizierte Fragmente der Ribonuklease Die Aminosäuren 1-26 (grau) stellen das Signalpeptid dar. Die laut [150] miteinander verknüpften Cysteine sind farbig hinterlegt. Die identifizierten Fragmente sind gekennzeichnet.

Insgesamt wurden 26 Fragmente identifiziert; davon 16 b- und 10 y-Ionen. Dass keine Fragmente detektiert wurden, die aus Bindungsbrüchen innerhalb des Proteins herrühren, hängt vermutlich mit dem Vorhandensein der Disulfidbrücken zusammen. Der große Vorteil der *Top-Down Proteomics* ist jedoch, dass terminale Fragmentationen in Verbindung mit der ermittelten Intaktmasse ausreichen können, um die Primärstruktur einer Proteoform erhalten zu können. Würde es Abweichung von der kanonischen Sequenz geben oder wären etwaige unbekannte PTM vorhanden, so würde die experimentell bestimmte Intaktmasse deutlich von der berechneten abweichen. Lediglich die exakte Lokalisation einer PTM ist bei unzureichender Detektion von Fragmentationen schwierig.

Wie im Folgenden gezeigt werden soll, kann die Anzahl der identifizierten Fragmentationen jedoch erhöht werden, indem die Aktivierung bei verschiedenen Energien erfolgt. Eines der

auch im Proteinmix enthaltenen Proteine, Cytochrom C, wurde offline, d.h. ohne vorherige flüssigkeitschromatographische Trennung, injiziert und durch momentane Änderung der Aktivierungsparameter fragmentiert. Aus **Abbildung 46** wird ersichtlich, dass Bindungen bereits bei niedrigen Aktivierungsenergien gebrochen werden, während andere deutlich höhere Energien erfordern.

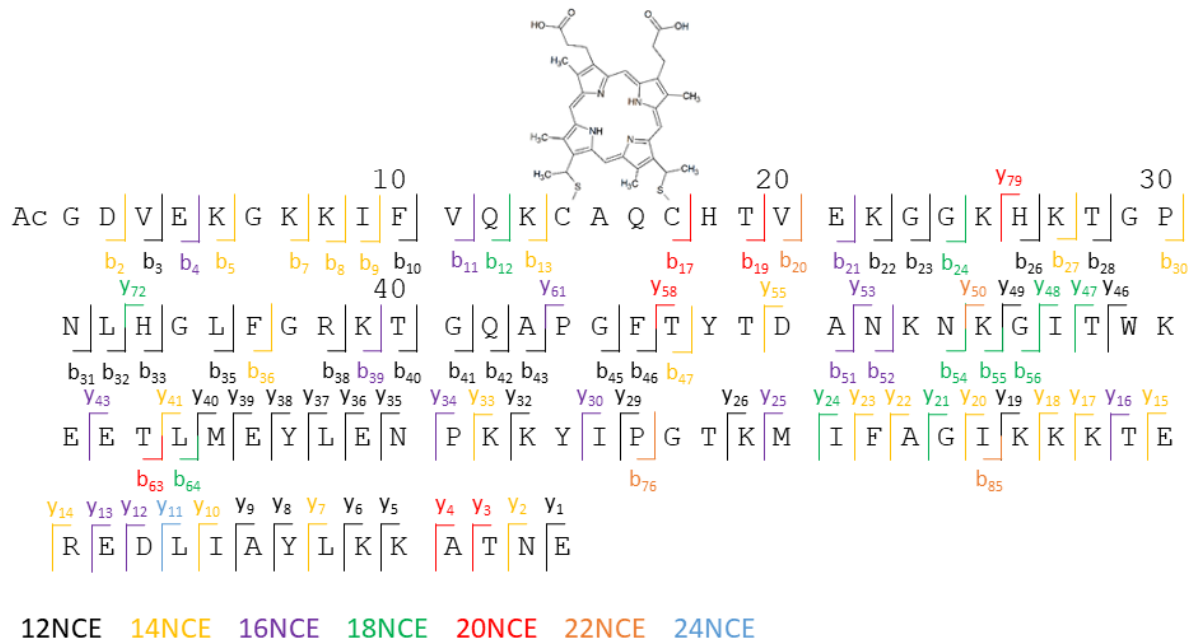


Abbildung 46 Identifizierte Fragmentationen bei verschiedenen normalisierten Kollisionsenergien (Auswertung eines ESI MS/MS Spektrums, Positiver Ionenmodus, Vorläufer-Ion: 589,500 m/z, z = +21) Cytochrom C wurde bei insgesamt sieben verschiedenen Aktivierungsenergien fragmentiert. Die bei den einzelnen Übergängen detektierten und identifizierten Fragmentationen sind entsprechend farblich kodiert.

Diese Daten zeigen, dass es für eine hohe Sequenzabdeckung hilfreich sein kann, Vorläuferionen bei mindestens zwei verschiedenen Aktivierungsenergien zu untersuchen. Im hier gewählten Beispiel konnte die Abdeckung von 32,7 % (bei 12 % NCE) auf 82,7 % (über alle NCE) erhöht werden. Es sei an dieser Stelle darauf hingewiesen, dass die Aktivierung bei mehreren NCE den in Kapitel 5.c angesprochenen *duty cycle* erhöht, sodass sich eine derartige Strategie für komplexe Proben und hohe chromatographische Auflösungen weniger eignet.

9. Referenzen

- [1] Beynon, R.J., The dynamics of the proteome: strategies for measuring protein turnover on a proteome-wide scale. *Brief. Funct. Genomic. Proteomic.* 2005, 3, 382–390.
- [2] Smith, L.M., Kelleher, N.L., Consortium for Top Down Proteomics, Proteoform: a single term describing protein complexity. *Nat. Methods* 2013, 10, 186–187.
- [3] Aebersold, R., Mann, M., Mass spectrometry-based proteomics. *Nature* 2003, 422, 198–207.
- [4] McLafferty, F.W., Breuker, K., Jin, M., Han, X., et al., Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J.* 2007, 274, 6256–6268.
- [5] Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., et al., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, 246, 64–71.
- [6] Hillenkamp, F., Karas, M., Beavis, R.C., Chait, B.T., Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 1991, 63, 1193A–1203A.
- [7] Yost, R.A., Enke, C.G., Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J. Am. Chem. Soc.* 1978, 100, 2274–2275.
- [8] Makarov, A., Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* 2000, 72, 1156–1162.
- [9] Stephens, W., A Pulsed Mass Spectrometer with Time Dispersion. *Proc. Am. Phys. Soc.* 1946, 69, 691.
- [10] Dietz, L.A., Basic Properties of Electron Multiplier Ion Detection and Pulse Counting Methods in Mass Spectrometry. *Rev. Sci. Instrum.* 1965, 36, 1763–1770.
- [11] Zubarev, R.A., Kelleher, N.L., McLafferty, F.W., Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* 1998, 120, 3265–3266.
- [12] Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., et al., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* 2004, 101, 9528–9533.
- [13] Scoble, H.A., Martin, S.A., Biemann, K., Peptide sequencing by magnetic deflection tandem mass spectrometry. *Biochem. J.* 1987, 245, 621–622.

- [14] Hunt, D.F., Yates, J.R., Shabanowitz, J., Winston, S., et al., Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 1986, *83*, 6233–6237.
- [15] Olsen, J.V., Macek, B., Lange, O., Makarov, A., et al., Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, *4*, 709–712.
- [16] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 1984, *11*, 601.
- [17] Biemann, K., Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* 1988, *16*, 99–111.
- [18] Humphrey, S.J., James, D.E., Mann, M., Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab. TEM* 2015, *26*, 676–687.
- [19] Berg, J.M., Tymoczko, J.L., Jr, G.J.G., Stryer, L., Stryer Biochemie, 8th ed., Springer Spektrum, 2018.
- [20] Pasupuleti, M., Schmidtchen, A., Malmsten, M., Antimicrobial peptides: key components of the innate immune system. *Crit. Rev. Biotechnol.* 2012, *32*, 143–171.
- [21] Debas, H.T., Gastrin. *Clin. Investig. Med. Med. Clin. Exp.* 1987, *10*, 222–225.
- [22] Weir, M.R., Dzau, V.J., The renin-angiotensin-aldosterone system: a specific target for hypertension management. *Am. J. Hypertens.* 1999, *12*, 205S–213S.
- [23] Schechter, I., Berger, A., On the size of the active site in proteases. I. Papain. 1967. *Biochem. Biophys. Res. Commun.* 2012, *425*, 497–502.
- [24] Harder, J., Schröder, J.-M., Psoriatic scales: a promising source for the isolation of human skin-derived antimicrobial proteins. *J. Leukoc. Biol.* 2005, *77*, 476–486.
- [25] Masters, C.L., Simms, G., Weinman, N.A., Multhaup, G., et al., Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 1985, *82*, 4245–4249.
- [26] Somasundaram, P., Koudelka, T., Linke, D., Tholey, A., C-Terminal Charge-Reversal Derivatization and Parallel Use of Multiple Proteases Facilitates Identification of Protein C-Termini by C-Terminomics. *J. Proteome Res.* 2016, *15*, 1369–1378.
- [27] Frey, B.L., Ladrör, D.T., Sondalle, S.B., Krusemark, C.J., et al., Chemical derivatization of peptide carboxyl groups for highly efficient electron transfer dissociation. *J. Am. Soc. Mass Spectrom.* 2013, *24*, 1710–1721.

- [28] Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaserud, D.J., et al., Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* 1999, *121*, 806–812.
- [29] Catherman, A.D., Skinner, O.S., Kelleher, N.L., Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.* 2014, *445*, 683–693.
- [30] Loo, J.A., Edmonds, C.G., Smith, R.D., Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* 1990, *248*, 201–204.
- [31] Loo, J.A., Edmonds, C.G., Smith, R.D., Tandem mass spectrometry of very large molecules: serum albumin sequence information from multiply charged ions formed by electrospray ionization. *Anal. Chem.* 1991, *63*, 2488–2499.
- [32] Forbes Andrew J., Mazur Matthew T., Patel Hiten M., Walsh Christopher T., et al., Toward efficient analysis of <70 kDa proteins with 100% sequence coverage. *PROTEOMICS* 2001, *1*, 927–933.
- [33] Ge, Y., Lawhorn, B.G., ElNaggar, M., Strauss, E., et al., Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J. Am. Chem. Soc.* 2002, *124*, 672–678.
- [34] Sze, S.K., Ge, Y., Oh, H., McLafferty, F.W., Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. *Proc. Natl. Acad. Sci. U. S. A.* 2002, *99*, 1774–1779.
- [35] Cai, W., Tucholski, T., Chen, B., Alpert, A.J., et al., Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal. Chem.* 2017, *89*, 5467–5475.
- [36] Gregorich, Z.R., Ge, Y., Top-down Proteomics in Health and Disease: Challenges and Opportunities. *Proteomics* 2014, *14*, 1195–1210.
- [37] Nesvizhskii, A.I., Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics MCP* 2005, *4*, 1419–1440.
- [38] Zhao, Y., Lin, Y.-H., Whole-cell protein identification using the concept of unique peptides. *Genomics Proteomics Bioinformatics* 2010, *8*, 33–41.
- [39] Linke, D., Koudelka, T., Becker, A., Tholey, A., Identification and relative quantification of phosphopeptides by a combination of multi-protease digestion and isobaric labeling. *Rapid Commun. Mass Spectrom. RCM* 2015, *29*, 919–926.

- [40] Kou, Q., Wu, S., Liu, X., Systematic Evaluation of Protein Sequence Filtering Algorithms for Proteoform Identification Using Top-Down Mass Spectrometry. *Proteomics* 2018, 18.
- [41] Fornelli, L., Toby, T.K., Schachner, L.F., Doubleday, P.F., et al., Top-down proteomics: Where we are, where we are going? *J. Proteomics* 2018, 175, 3–4.
- [42] Tan, S., Tan, H.T., Chung, M.C.M., Membrane proteins and membrane proteomics. *Proteomics* 2008, 8, 3924–3932.
- [43] Kachuk, C., Doucette, A.A., The benefits (and misfortunes) of SDS in top-down proteomics. *J. Proteomics* 2018, 175, 75–86.
- [44] Smithies, O., Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults. *Biochem. J.* 1955, 61, 629–641.
- [45] Bjellqvist, B., Ek, K., Giorgio Righetti, P., Gianazza, E., et al., Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *J. Biochem. Biophys. Methods* 1982, 6, 317–339.
- [46] Tran, J.C., Doucette, A.A., Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem.* 2008, 80, 1568–1573.
- [47] Eeltink, S., Wouters, S., Does-Sousa, J.L., Svec, F., Advances in organic polymer-based monolithic column technology for high-resolution liquid chromatography-mass spectrometry profiling of antibodies, intact proteins, oligonucleotides, and peptides. *J. Chromatogr. A* 2017, 1498, 8–21.
- [48] Mohr, J., Swart, R., Samonig, M., Böhm, G., et al., High-efficiency nano- and micro-HPLC–high-resolution Orbitrap-MS platform for top-down proteomics. *Proteomics* 2010, 10, 3598–3609.
- [49] Shen, Y., Tolić, N., Piehowski, P.D., Shukla, A.K., et al., High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics. *J. Chromatogr. A* 2017, 1498, 99–110.
- [50] Catherman, A.D., Durbin, K.R., Ahlf, D.R., Early, B.P., et al., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol. Cell. Proteomics MCP* 2013, 12, 3465–3473.
- [51] Compton, P.D., Zamdborg, L., Thomas, P.M., Kelleher, N.L., On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* 2011, 83, 6868–6874.

- [52] Liu, X., Inbar, Y., Dorrestein, P.C., Wynne, C., et al., Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol. Cell. Proteomics MCP* 2010, 9, 2772–2782.
- [53] Durbin, K.R., Skinner, O.S., Fellers, R.T., Kelleher, N.L., Analyzing internal fragmentation of electrosprayed ubiquitin ions during beam-type collisional dissociation. *J. Am. Soc. Mass Spectrom.* 2015, 26, 782–787.
- [54] Tholey, A., Becker, A., Top-down proteomics for the analysis of proteolytic events - Methods, applications and perspectives. *Biochim. Biophys. Acta* 2017, 1864, 2191–2199.
- [55] Lottspeich, F., Engels, J.W. (Eds.), Bioanalytik, 3rd ed., Springer Spektrum, 2012.
- [56] van Deemter, J.J., Zuiderweg, F.J., Klinkenberg, A., Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography. *Chem. Eng. Sci.* 1956, 5, 271–289.
- [57] Kobayashi, H., Ikegami, T., Kimura, H., Hara, T., et al., Properties of monolithic silica columns for HPLC. *Anal. Sci. Int. J. Jpn. Soc. Anal. Chem.* 2006, 22, 491–501.
- [58] Petro, M., Svec, F., Gitsov, I., Fréchet, J.M.J., Molded Monolithic Rod of Macroporous Poly(styrene-co-divinylbenzene) as a Separation Medium for HPLC of Synthetic Polymers: “On-Column” Precipitation–Redissolution Chromatography as an Alternative to Size Exclusion Chromatography of Styrene Oligomers and Polymers. *Anal. Chem.* 1996, 68, 315–321.
- [59] Vlakh, E.G., Tennikova, T.B., Preparation of methacrylate monoliths. *J. Sep. Sci.* 2007, 30, 2801–2813.
- [60] Vissers, J.P.C., Blackburn, R.K., Moseley, M.A., A novel interface for variable flow nanoscale LC/MS/MS for improved proteome coverage. *J. Am. Soc. Mass Spectrom.* 2002, 13, 760–771.
- [61] Taylor, S.G., S, F.R., Disintegration of water drops in an electric field. *Proc R Soc Lond A* 1964, 280, 383–397.
- [62] Clegg, G.A., Dole, M., Molecular beams of macroions. 3. Zein and polyvinylpyrrolidone. *Biopolymers* 1971, 10, 821–826.
- [63] Wilm, M.S., Mann, M., Electrospray and Taylor-Cone theory, Dole’s beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Process.* 1994, 136, 167–180.
- [64] Iribarne, J.V., Thomson, B.A., On the evaporation of small ions from charged droplets. *J. Chem. Phys.* 1976, 64, 2287–2294.

- [65] Denisov, E., Damoc, E., Lange, O., Makarov, A., Orbitrap mass spectrometry with resolving powers above 1,000,000. *Int. J. Mass Spectrom.* 2012, 325–327, 80–85.
- [66] Shishkova, E., Hebert, A.S., Coon, J.J., Now, More Than Ever, Proteomics Needs Better Chromatography. *Cell Syst.* 2016, 3, 321–324.
- [67] Greenberg, R., Groves, M.L., Dower, H.J., Human beta-casein. Amino acid sequence and identification of phosphorylation sites. *J. Biol. Chem.* 1984, 259, 5132–5138.
- [68] Lam, H., Building and Searching Tandem Mass Spectral Libraries for Peptide Identification. *Mol. Cell. Proteomics MCP* 2011, 10.
- [69] Choi, H., Nesvizhskii, A.I., False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. *J. Proteome Res.* 2008, 7, 47–50.
- [70] Alves, G., Ogurtsov, A.Y., Yu, Y.-K., Assigning statistical significance to proteotypic peptides via database searches. *J. Proteomics* 2011, 74, 199–211.
- [71] Liu, X., Sirotkin, Y., Shen, Y., Anderson, G., et al., Protein identification using top-down. *Mol. Cell. Proteomics MCP* 2012, 11, M111.008524.
- [72] Wang, L.-X., Amin, M.N., Chemical and Chemoenzymatic Synthesis of Glycoproteins for Deciphering Functions. *Chem. Biol.* 2014, 21, 51–66.
- [73] Wopereis, S., Lefeber, D.J., Morava, E., Wevers, R.A., Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin. Chem.* 2006, 52, 574–600.
- [74] Hofsteenge, J., Mueller, D.R., de Beer, T., Loeffler, A., et al., New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* 1994, 33, 13524–13530.
- [75] Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., et al. (Eds.), Essentials of Glycobiology, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) 2009.
- [76] Owens, C.M., Poultry Meat Processing, Second Edition, CRC Press, 2010.
- [77] Vreeker, G.C.M., Wuhrer, M., Reversed-phase separation methods for glycan analysis. *Anal. Bioanal. Chem.* 2017, 409, 359–378.
- [78] Alley, W.R., Novotny, M.V., Structural Glycomic Analyses at High Sensitivity: A Decade of Progress. *Annu. Rev. Anal. Chem. Palo Alto Calif* 2013, 6, 237–265.

- [79] Pedrali, A., Tengattini, S., Marrubini, G., Bavaro, T., et al., Characterization of Intact Neo-Glycoproteins by Hydrophilic Interaction Liquid Chromatography. *Molecules* 2014, *19*, 9070–9088.
- [80] Leymarie, N., Zaia, J., Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal. Chem.* 2012, *84*, 3040–3048.
- [81] Harvey, D.J., Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Rev. Proteomics* 2005, *2*, 87–101.
- [82] Catalina, M.I., Koeleman, C.A.M., Deelder, A.M., Wuhler, M., Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun. Mass Spectrom. RCM* 2007, *21*, 1053–1061.
- [83] Songsrirote, K., Li, Z., Ashford, D., Bateman, A., et al., Development and application of mass spectrometric methods for the analysis of progranulin N-glycosylation. *J. Proteomics* 2010, *73*, 1479–1490.
- [84] Perry, D.C., Lehmann, M., Yokoyama, J.S., Karydas, A., et al., Progranulin mutations as risk factors for Alzheimer disease. *JAMA Neurol.* 2013, *70*, 774–778.
- [85] Minami, S.S., Min, S.-W., Krabbe, G., Wang, C., et al., Progranulin protects against amyloid β deposition and toxicity in Alzheimer's disease mouse models. *Nat. Med.* 2014, *20*, 1157–1164.
- [86] Goedert, M., Ghetti, B., Spillantini, M.G., Frontotemporal dementia: implications for understanding Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2012, *2*, a006254.
- [87] Cenik, B., Sephton, C.F., Cenik, B.K., Herz, J., et al., Progranulin: A Proteolytically Processed Protein at the Crossroads of Inflammation and Neurodegeneration. *J. Biol. Chem.* 2012, *287*, 32298–32306.
- [88] Plowman, G.D., Green, J.M., Neubauer, M.G., Buckley, S.D., et al., The epithelin precursor encodes two proteins with opposing activities on epithelial cell growth. *J. Biol. Chem.* 1992, *267*, 13073–13078.
- [89] Liao, L.M., Lallone, R.L., Seitz, R.S., Buznikov, A., et al., Identification of a Human Glioma-associated Growth Factor Gene, granulin, Using Differential Immuno-absorption. *Cancer Res.* 2000, *60*, 1353–1360.
- [90] Damme, P.V., Hoecke, A.V., Lambrechts, D., Vanacker, P., et al., Progranulin functions as a neurotrophic factor to regulate neurite outgrowth and enhance neuronal survival. *J Cell Biol* 2008, *181*, 37–41.

- [91] Becker, C., Kruse, M.-N., Slotty, K.A., Köhler, D., et al., Differences in the activation mechanism between the alpha and beta subunits of human meprin. *Biol. Chem.* 2003, 384, 825–831.
- [92] Becker-Pauly, C., Höwel, M., Walker, T., Vlad, A., et al., The alpha and beta subunits of the metalloprotease meprin are expressed in separate layers of human epidermis, revealing different functions in keratinocyte proliferation and differentiation. *J. Invest. Dermatol.* 2007, 127, 1115–1125.
- [93] Broder, C., Becker-Pauly, C., The metalloproteases meprin α and meprin β : unique enzymes in inflammation, neurodegeneration, cancer and fibrosis. *Biochem. J.* 2013, 450, 253–264.
- [94] Villa, J.P., Bertenshaw, G.P., Bond, J.S., Critical amino acids in the active site of meprin metalloproteinases for substrate and peptide bond specificity. *J. Biol. Chem.* 2003, 278, 42545–42550.
- [95] The, M., MacCoss, M.J., Noble, W.S., Käll, L., Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J. Am. Soc. Mass Spectrom.* 2016, 27, 1719–1727.
- [96] Robinson, N.E., Protein deamidation. *Proc. Natl. Acad. Sci. U. S. A.* 2002, 99, 5283–5288.
- [97] Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., et al., Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 2009, 6, 786–787.
- [98] Zappacosta, F., Pessi, A., Bianchi, E., Venturini, S., et al., Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. *Protein Sci. Publ. Protein Soc.* 1996, 5, 802–813.
- [99] Dokudovskaya, S., Williams, R., Devos, D., Sali, A., et al., Protease accessibility laddering: a proteomic tool for probing protein structure. *Struct. Lond. Engl.* 1993 2006, 14, 653–660.
- [100] Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., et al., Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* 2004, 51, 299–321.
- [101] Tolkatchev, D., Malik, S., Vinogradova, A., Wang, P., et al., Structure dissection of human progranulin identifies well-folded granulin/epithelin modules with unique functional activities. *Protein Sci. Publ. Protein Soc.* 2008, 17, 711–724.

- [102] Ren, D., Julka, S., Inerowicz, H.D., Regnier, F.E., Enrichment of cysteine-containing peptides from tryptic digests using a quaternary amine tag. *Anal. Chem.* 2004, 76, 4522–4530.
- [103] Qiao, X., Wang, R., Yan, H., Wang, T., et al., Development of a novel imidazolium-based aromatic quaternary ammonium tag: synthesis and application to the efficient analysis of cysteinyl-peptides by mass spectrometry. *Rapid Commun. Mass Spectrom. RCM* 2014, 28, 256–264.
- [104] Grice, E.A., Kong, H.H., Renaud, G., Young, A.C., et al., A diversity profile of the human skin microbiota. *Genome Res.* 2008, 18, 1043–1050.
- [105] Archer, G.L., Staphylococcus aureus: a well-armed pathogen. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 1998, 26, 1179–1181.
- [106] Jarvis, W.R., Martone, W.J., Predominant pathogens in hospital infections. *J. Antimicrob. Chemother.* 1992, 29 Suppl A, 19–24.
- [107] Hiramatsu, K., Katayama, Y., Matsuo, M., Sasaki, T., et al., Multi-drug-resistant Staphylococcus aureus and future chemotherapy. *J. Infect. Chemother. Off. J. Jpn. Soc. Chemother.* 2014, 20, 593–601.
- [108] Lyu, W., Curtis, A.R., Sunkara, L.T., Zhang, G., Transcriptional Regulation of Antimicrobial Host Defense Peptides. *Curr. Protein Pept. Sci.* 2015, 16, 672–679.
- [109] Nuti, R., Goud, N.S., Saraswati, A.P., Alvala, R., et al., Antimicrobial Peptides: A Promising Therapeutic Strategy in Tackling Antimicrobial Resistance. *Curr. Med. Chem.* 2017, 24, 4303–4314.
- [110] Rishi, P., Singh, A.P., Arora, S., Garg, N., et al., Revisiting eukaryotic anti-infective biotherapeutics. *Crit. Rev. Microbiol.* 2014, 40, 281–292.
- [111] Jung, S., Mysliwy, J., Spudy, B., Lorenzen, I., et al., Human beta-defensin 2 and beta-defensin 3 chimeric peptides reveal the structural basis of the pathogen specificity of their parent molecules. *Antimicrob. Agents Chemother.* 2011, 55, 954–960.
- [112] Bechinger, B., Gorr, S.-U., Antimicrobial Peptides: Mechanisms of Action and Resistance. *J. Dent. Res.* 2017, 96, 254–260.
- [113] Schitteck, B., Hipfel, R., Sauer, B., Bauer, J., et al., Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nat. Immunol.* 2001, 2, 1133–1137.

- [114] Hansmann, B., Schröder, J.-M., Gerstel, U., Skin-Derived C-Terminal Filaggrin-2 Fragments Are *Pseudomonas aeruginosa*-Directed Antimicrobials Targeting Bacterial Replication. *PLoS Pathog.* 2015, *11*, e1005159.
- [115] Gläser, R., Harder, J., Lange, H., Bartels, J., et al., Antimicrobial psoriasin (S100A7) protects human skin from *Escherichia coli* infection. *Nat. Immunol.* 2005, *6*, 57–64.
- [116] Harder, J., Schroder, J.-M., RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *J. Biol. Chem.* 2002, *277*, 46779–46784.
- [117] Jones, D.E., Bevins, C.L., Paneth cells of the human small intestine express an antimicrobial peptide gene. *J. Biol. Chem.* 1992, *267*, 23216–23225.
- [118] Jones, D.E., Bevins, C.L., Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptides in host defense of the human bowel. *FEBS Lett.* 1993, *315*, 187–192.
- [119] Diamond, G., Zasloff, M., Eck, H., Brasseur, M., et al., Tracheal antimicrobial peptide, a cysteine-rich peptide from mammalian tracheal mucosa: peptide isolation and cloning of a cDNA. *Proc. Natl. Acad. Sci. U. S. A.* 1991, *88*, 3952–3956.
- [120] Selsted, M.E., Tang, Y.Q., Morris, W.L., McGuire, P.A., et al., Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J. Biol. Chem.* 1993, *268*, 6641–6648.
- [121] Mena, K.D., Gerba, C.P., Risk assessment of *Pseudomonas aeruginosa* in water. *Rev. Environ. Contam. Toxicol.* 2009, *201*, 71–115.
- [122] Oliver, A., Mulet, X., López-Causapé, C., Juan, C., The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* 2015, *21–22*, 41–59.
- [123] MacPherson, P., Valentine, K., Chadderton, V., Dardamissis, E., et al., An Outbreak of *Pseudomonas Aeruginosa* Infection Linked to a “Black Friday” Piercing Event. *PLoS Curr.* 2017, *9*.
- [124] Tümmler, B., Bosshammer, J., Breitenstein, S., Brockhausen, I., et al., Infections with *Pseudomonas aeruginosa* in patients with cystic fibrosis. *Behring Inst. Mitt.* 1997, 249–255.
- [125] P. Tam, J., Rong Wu, C., Liu, W., Wen Zhang, J., Disulfide bond formation in peptides by dimethyl sulfoxide. Scope and applications. *J. Am. Chem. Soc. - J AM CHEM SOC* 1991, *113*.

- [126] Broussard, J.A., Yang, R., Huang, C., Nathamgari, S.S.P., et al., The desmoplakin-intermediate filament linkage regulates cell mechanics. *Mol. Biol. Cell* 2017, 28, 3156–3164.
- [127] Dick, L.W., Kim, C., Qiu, D., Cheng, K.-C., Determination of the origin of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. *Biotechnol. Bioeng.* 2007, 97, 544–553.
- [128] Baechle, D., Flad, T., Cansier, A., Steffen, H., et al., Cathepsin D is present in human eccrine sweat and involved in the postsecretory processing of the antimicrobial peptide DCD-1L. *J. Biol. Chem.* 2006, 281, 5406–5415.
- [129] Brophy, M.B., Nolan, E.M., Manganese and microbial pathogenesis: sequestration by the Mammalian immune system and utilization by microorganisms. *ACS Chem. Biol.* 2015, 10, 641–651.
- [130] Sharma, S., Sahoo, N., Bhunia, A., Antimicrobial Peptides and their Pore/Ion Channel Properties in Neutralization of Pathogenic Microbes. *Curr. Top. Med. Chem.* 2016, 16, 46–53.
- [131] Bijlmakers, M.-J., Marsh, M., The on-off story of protein palmitoylation. *Trends Cell Biol.* 2003, 13, 32–42.
- [132] Sternlicht, M.D., Werb, Z., How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* 2001, 17, 463–516.
- [133] Resing, K.A., Thulin, C., Whiting, K., al-Alawi, N., et al., Characterization of profilaggrin endoproteinase 1. A regulated cytoplasmic endoproteinase of epidermis. *J. Biol. Chem.* 1995, 270, 28193–28198.
- [134] Eissa, A., Diamandis, E.P., Human tissue kallikreins as promiscuous modulators of homeostatic skin barrier functions. *Biol. Chem.* 2008, 389, 669–680.
- [135] Yamasaki, K., Schaubert, J., Coda, A., Lin, H., et al., Kallikrein-mediated proteolysis regulates the antimicrobial effects of cathelicidins in skin. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 2006, 20, 2068–2080.
- [136] Latendorf, T., Gerstel, U., Wu, Z., Bartels, J., et al., Cationic Intrinsically Disordered Antimicrobial Peptides (CIDAMPs) Represent a New Paradigm of Innate Defense with a Potential for Novel Anti-Infectives. *Sci. Rep.* 2019, 9, 3331.
- [137] Murphy, E.C., Mohanty, T., Frick, I.-M., FAF and SufA: proteins of *Finnegoldia magna* that modulate the antibacterial activity of histones. *J. Innate Immun.* 2014, 6, 394–404.

- [138] Fischer, E.H., Phosphorylase and the origin of reversible protein phosphorylation. *Biol. Chem.* 2010, *391*, 131–137.
- [139] Fischer, E.H., Cellular regulation by protein phosphorylation. *Biochem. Biophys. Res. Commun.* 2013, *430*, 865–867.
- [140] Stig, Å., Cysteamine S-Phosphoric Acid. *Acta Chem. Scand.* 1960, *14*, 1980–1984.
- [141] Stig, Å., Further Studies on S-Substituted Phosphorothioic Acids. Mixed Lithiumsodium Salts of S-(1-Carboxyethyl) Phosphorothioic Acid and S-(2-Carboxyethyl) Phosphorothioic Acid. *Acta Chem. Scand.* 1961, *15*, 575–582.
- [142] Fuhs, S.R., Hunter, T., pHisphorylation: the emergence of histidine phosphorylation as a reversible regulatory modification. *Curr. Opin. Cell Biol.* 2017, *45*, 8–16.
- [143] Stadtman, T.C., Emerging awareness of the critical roles of S-phosphocysteine and selenophosphate in biological systems. *BioFactors Oxf. Engl.* 1994, *4*, 181–185.
- [144] Kübler, D., Pyerin, W., Bill, O., Hotz, A., et al., Evidence for ecto-protein kinase activity that phosphorylates Kemptide in a cyclic AMP-dependent mode. *J. Biol. Chem.* 1989, *264*, 14549–14555.
- [145] Rowan, F.C., Richards, M., Bibby, R.A., Thompson, A., et al., Insights into Aurora-A kinase activation using unnatural amino acids incorporated by chemical modification. *ACS Chem. Biol.* 2013, *8*, 2184–2191.
- [146] Chalker, J.M., Lercher, L., Rose, N.R., Schofield, C.J., et al., Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chem. Int. Ed Engl.* 2012, *51*, 1835–1839.
- [147] Ge, G., Greenspan, D.S., Developmental roles of the BMP1/TLD metalloproteinases. *Birth Defects Res. Part C Embryo Today Rev.* 2006, *78*, 47–68.
- [148] Mac Sweeney, A., Gil-Parrado, S., Vinzenz, D., Bernardi, A., et al., Structural basis for the substrate specificity of bone morphogenetic protein 1/tolloid-like metalloproteases. *J. Mol. Biol.* 2008, *384*, 228–239.
- [149] Hung, C.-W., Koudelka, T., Anastasi, C., Becker, A., et al., Characterization of post-translational modifications in full-length human BMP-1 confirms the presence of a rare vicinal disulfide linkage in the catalytic domain and highlights novel features of the EGF domain. *J. Proteomics* 2016, *138*, 136–145.
- [150] Smyth, D.G., Stein, W.H., Moore, S., The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. *J. Biol. Chem.* 1963, *238*, 227–234.

- [151] García Alonso, J.I., Rodríguez-González, P., González-Gago, A., González-Antuña, A., Determination of the uncertainties in the theoretical mass isotopomer distribution of molecules. *Anal. Chim. Acta* 2010, 664, 68–76.

10.Anhang

a. Chromatogramme und Massenspektren der Progranulin-Versuche

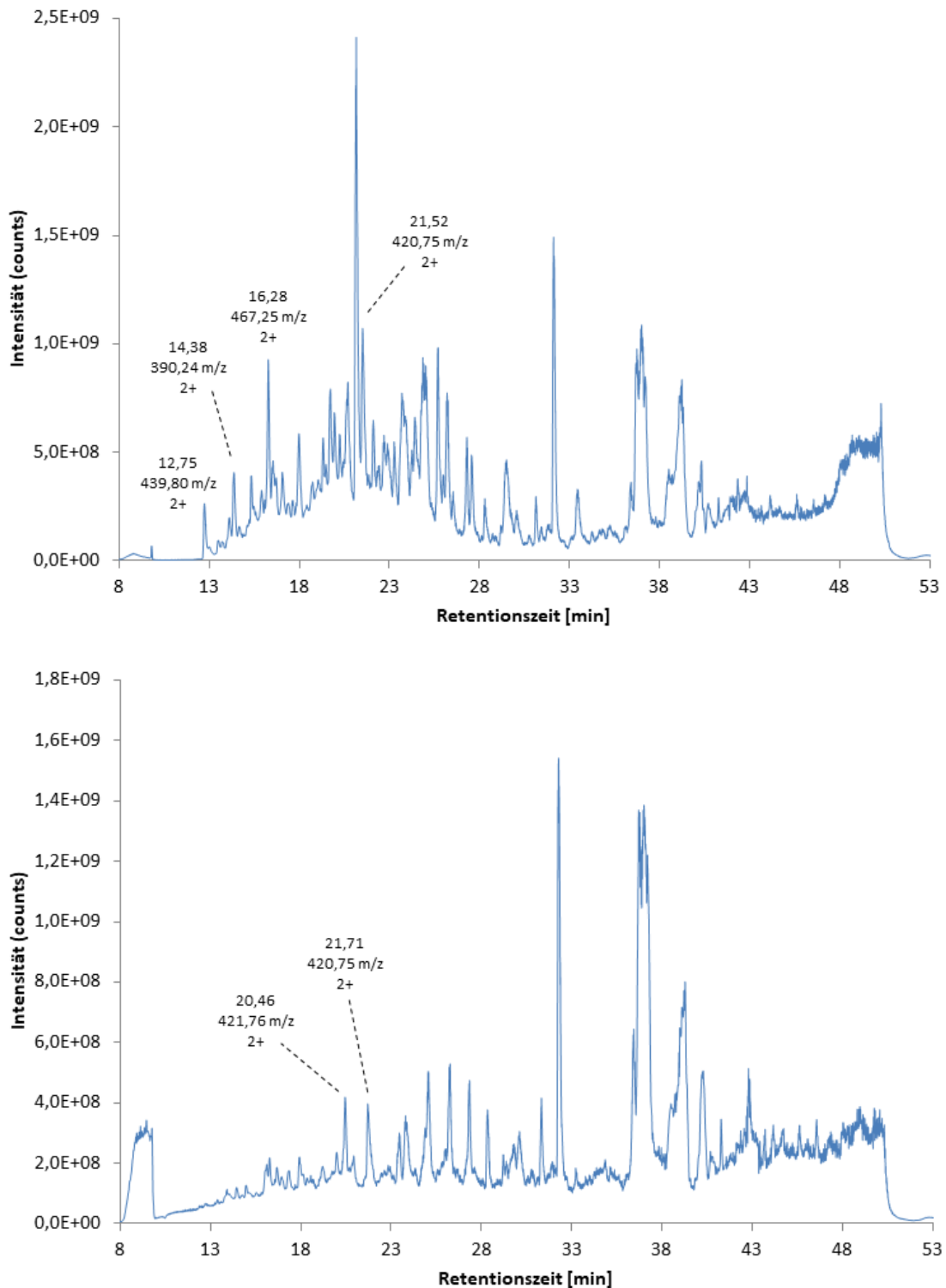


Abbildung 47 TIC des 4-stündigen Meprin- α - (oben) und Meprin- β -Verdaus (unten)(Bottom-Up) Für den Bottom-Up Ansatz ergeben sich ähnliche Chromatogramme, wie sie schon nach 30 min beobachtet wurden: Eine Vielzahl von Signalen ist sichtbar, von welchen aber wenige, bzw. keine dem Progranulin zugeordnet werden konnten. Es sei angemerkt, dass auch hier wieder nur nach Peptiden gesucht wurde, welche eine

Dimethylierung am N-Terminus trugen, um sie als von Meprin- α generiert deklarieren zu können. Auch hier wurden viele Störsignale beobachtet, die nicht von Peptiden herrühren, z.B. bei 32 und 37 min.

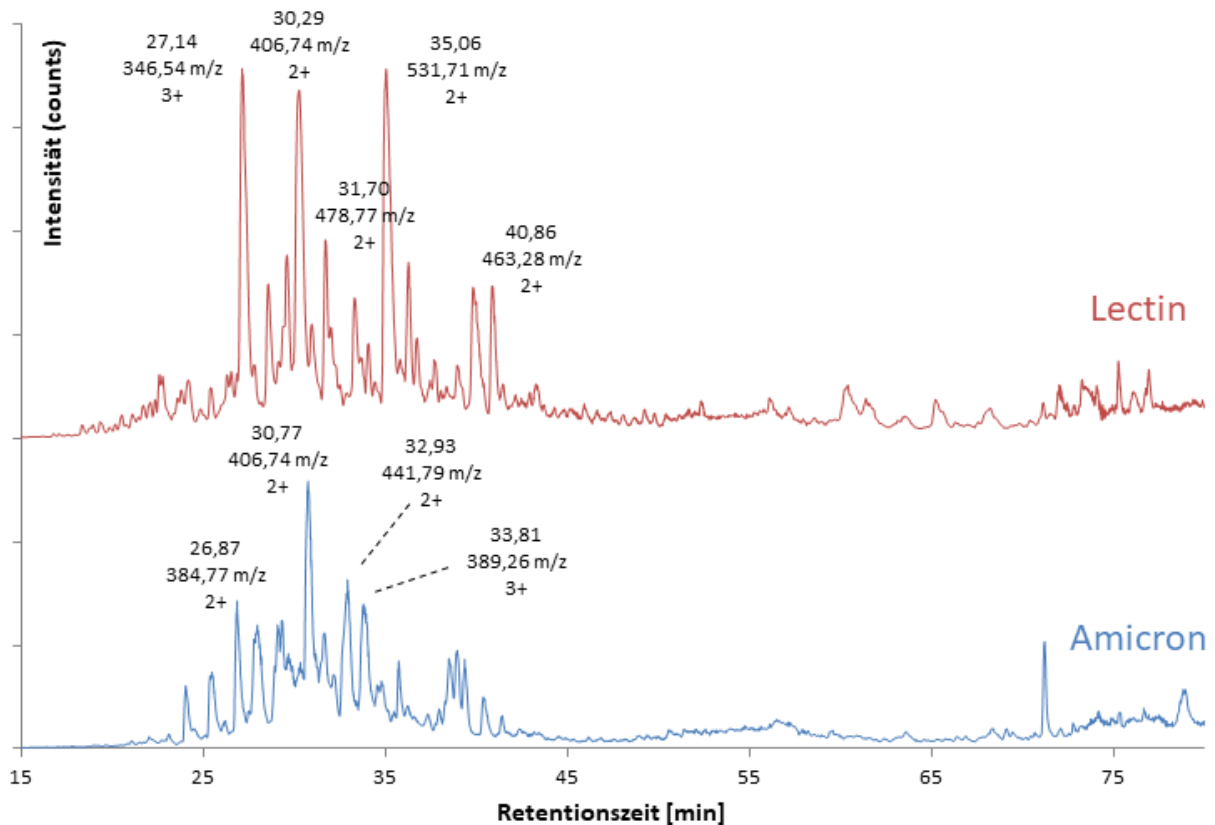


Abbildung 48 TIC des 4-stündigen Meprin- α -Verdaus (Top-Down) Nach der längeren Inkubationszeit zeigen sich im Vergleich zum 30-minütigen Verdau (**Abbildung 17**) die gleichen Signale, nur mit anderen relativen Intensitäten (z.B. 346,54 m/z bei etwa 27 min oder 463,28 m/z bei ca. 41 min). Auch hier eluieren alle Peptide zwischen etwa 25 und 45 min Retentionszeit.

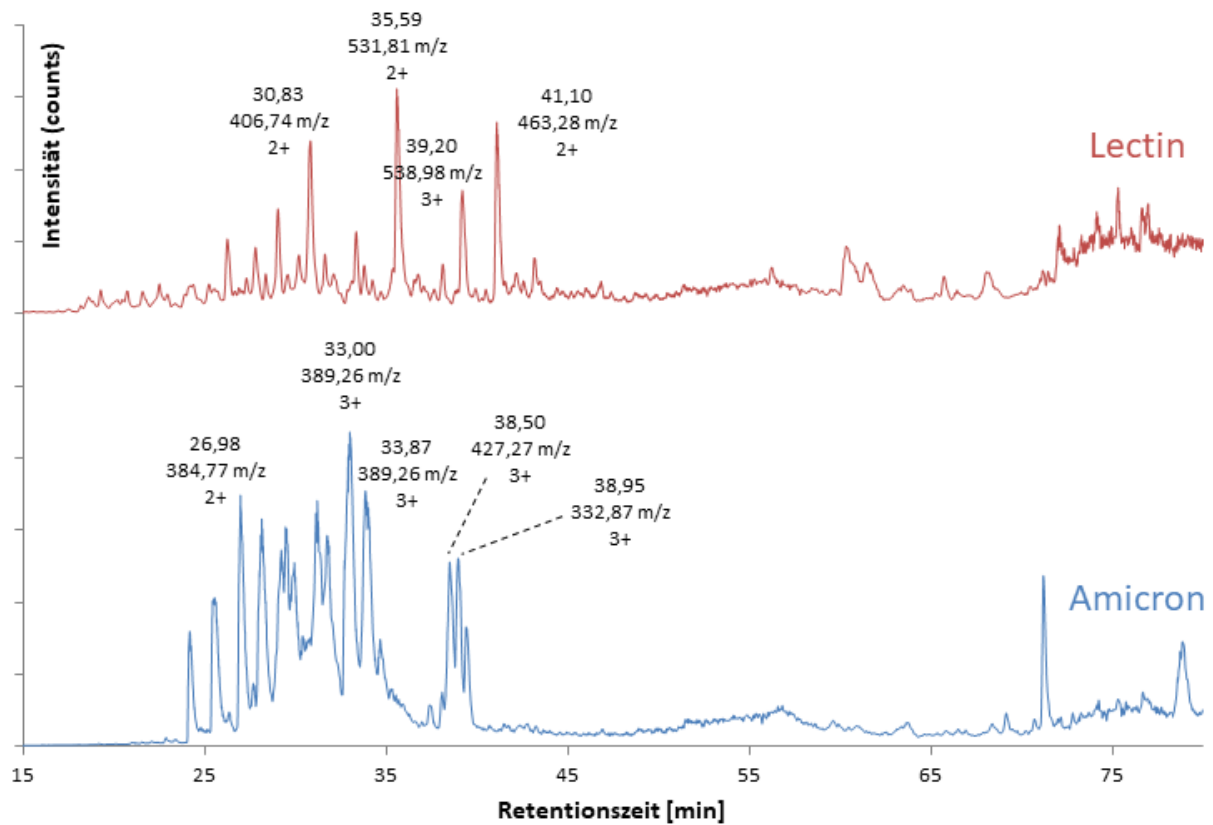


Abbildung 49 TIC des 4-stündigen Meprin- β -Verdaus (Top-Down) Da Meprin- β nachgewiesenermaßen über eine höhere Selektivität verfügt, zeigen auch die Chromatogramme weniger Signale (vgl. **Abbildung 48**). Dass die Spezifität der beiden Proteasen jedoch ähnlich ist, zeigt sich im Auftreten von Peptiden, die für beide Enzyme beobachtet wurden (z.B. 406,74 m/z bei 30,29, bzw 30,83 min). Da diese jedoch nicht identifiziert wurden, kann diese These nicht zweifelsfrei belegt werden.

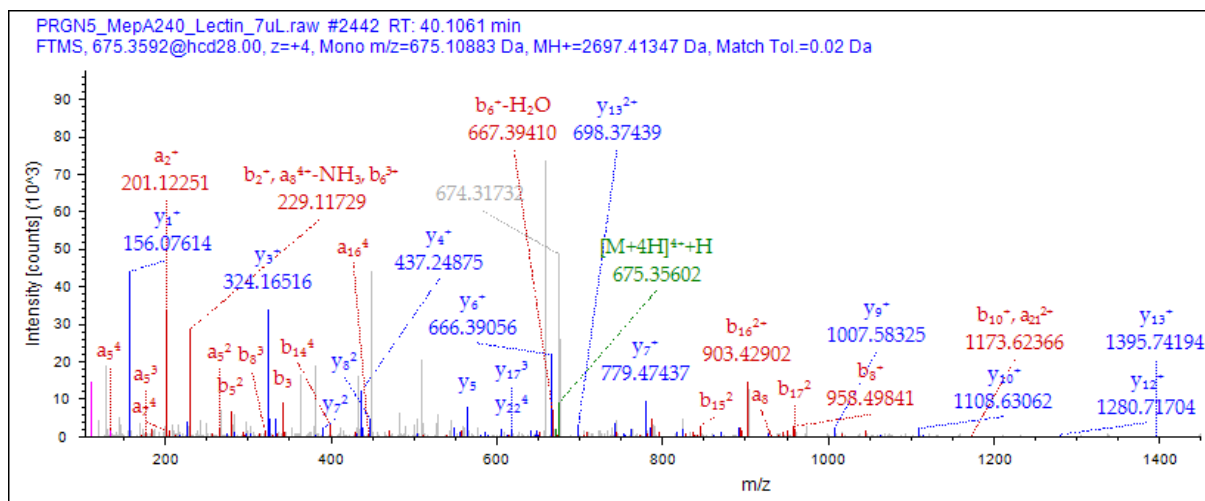


Abbildung 50 MS/MS-Spektrum des PGRN 254-277 (LC-ESI MS, positiver Ionenmodus, Vorläufer: $[M+4H]^{4+} = 675,109 \text{ m/z}$) Das Spektrum zeigt wenige intensive Fragmentsignale. Dazu zählen vor allem die kürzeren y-Fragmente. Neben b- und y-Ionen wurden auch eine Reihe von a-Ionen detektiert. Die Massenabweichung des Vorläufers betrug 0,38 ppm.

Tabelle 12 Identifizierte Fragment-Ionen für das in Abbildung 50 gezeigte Peptid Wie schon aus dem Massenspektrum ersichtlich, wurden für die Sequenz vor allem y-Ionen detektiert. Dieser Umstand könnte auf die beiden basischen Aminosäuren Histidin und Lysin am Peptid-C-Terminus zurückzuführen sein.

#1	Imm.	a ⁺	a ²⁺	a ³⁺	a ⁴⁺	b ⁺	b ²⁺	b ³⁺	b ⁴⁺	Seq.
1	88,039	88,039	44,523	30,018	22,765	116,034	58,521	39,350	29,764	D
2	86,096	201,123	101,065	67,713	51,036	229,118	115,063	77,044	58,035	L
3	86,096	314,207	157,607	105,407	79,307	342,202	171,605	114,739	86,306	I
4	101,071	442,266	221,637	148,094	111,322	470,261	235,634	157,425	118,321	Q
5	60,044	529,298	265,153	177,104	133,080	557,293	279,150	186,436	140,079	S
6	101,107	657,393	329,200	219,803	165,104	685,388	343,198	229,134	172,102	K
7	133,043	817,424	409,215	273,146	205,111	845,419	423,213	282,478	212,110	C-Carbamidom.
8	86,096	930,508	465,758	310,841	233,382	958,503	479,755	320,172	240,381	L
9	60,044	1017,540	509,274	339,851	255,140	1045,535	523,271	349,183	262,139	S
10	101,107	1145,635	573,321	382,550	287,164	1173,630	587,318	391,881	294,163	K
11	102,055	1274,677	637,842	425,564	319,425	1302,672	651,840	434,896	326,424	E
12	88,039	1389,704	695,356	463,906	348,182	1417,699	709,353	473,238	355,180	N-Deamidated
13	44,049	1460,741	730,874	487,585	365,941	1488,736	744,872	496,917	372,940	A
14	74,060	1561,789	781,398	521,268	391,203	1589,784	795,396	530,600	398,201	T
15	74,060	1662,837	831,922	554,950	416,465	1690,832	845,919	564,282	423,463	T
16	88,039	1777,864	889,435	593,293	445,221	1805,859	903,433	602,624	452,220	D
17	86,096	1890,948	945,978	630,987	473,492	1918,943	959,975	640,319	480,491	L
18	86,096	2004,032	1002,520	668,682	501,763	2032,027	1016,517	678,014	508,762	L
19	74,060	2105,079	1053,043	702,365	527,025	2133,074	1067,041	711,696	534,024	T
20	101,107	2233,174	1117,091	745,063	559,049	2261,169	1131,088	754,395	566,048	K
21	86,096	2346,258	1173,633	782,758	587,320	2374,253	1187,630	792,089	594,319	L
22	70,065	2443,311	1222,159	815,109	611,583	2471,306	1236,157	824,440	618,582	P
23	44,049	2514,348	1257,678	838,788	629,343	2542,343	1271,675	848,119	636,341	A
24	110,071									H

Seq.	y ⁺	y ²⁺	y ³⁺	y ⁴⁺	#2
D					24
L	2582,386	1291,697	861,467	646,352	23
I	2469,302	1235,155	823,772	618,081	22
Q	2356,218	1178,612	786,077	589,810	21
S	2228,159	1114,583	743,391	557,795	20
K	2141,127	1071,067	714,381	536,037	19
C-Carbamidom.	2013,032	1007,020	671,682	504,013	18
L	1853,001	927,004	618,339	464,006	17
S	1739,917	870,462	580,644	435,735	16
K	1652,885	826,946	551,633	413,977	15
E	1524,790	762,899	508,935	381,953	14
N-Deamidated	1395,748	698,378	465,921	349,692	13
A	1280,721	640,864	427,578	320,936	12
T	1209,684	605,346	403,899	303,176	11
T	1108,636	554,822	370,217	277,914	10
D	1007,588	504,298	336,534	252,653	9
L	892,561	446,784	298,192	223,896	8
L	779,477	390,242	260,497	195,625	7
T	666,393	333,700	222,803	167,354	6
K	565,346	283,176	189,120	142,092	5
L	437,251	219,129	146,422	110,068	4
P	324,167	162,587	108,727	81,797	3
A	227,114	114,061	76,376	57,534	2
H	156,077	78,542	52,697	39,775	1

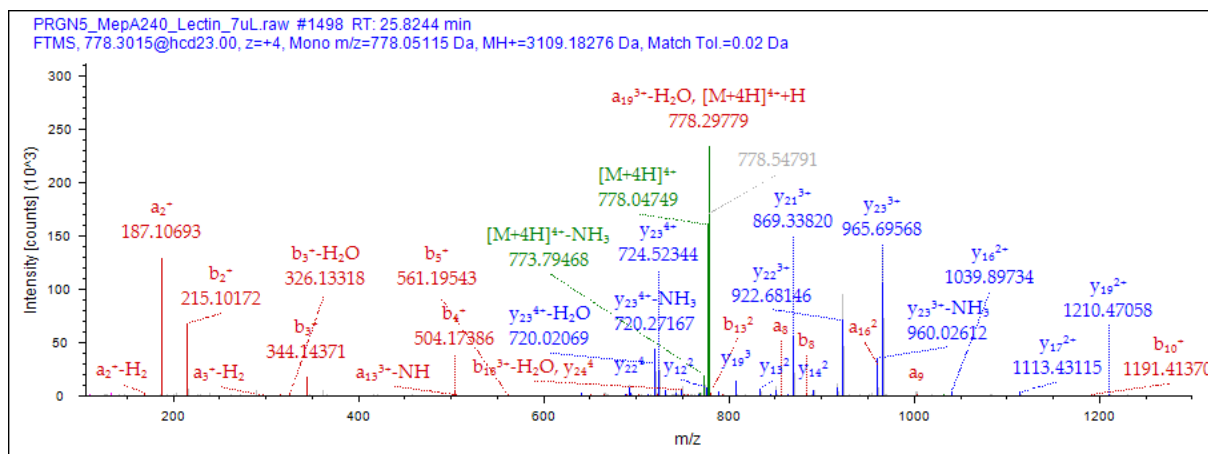


Abbildung 51 MS/MS-Spektrum des PGRN 518-542 (LC-ESI MS, positiver Ionenmodus, Vorläuferion: $[M+4H]^{4+} = 778,051 \text{ m/z}$) Das Spektrum wird von Fragmenten dominiert, die aus einer Dissoziation der Bindung zwischen Val-519 und Glu-520 sowie Cys-521 und Gly-522 entstanden sind. Die höchste Intensität aller Ionen zeigt das Vorläuferion. Dieser Umstand ist auf die im Vergleich zu **Abbildung 50** niedrigere NCE von 23 % zurückzuführen. Die Massenabweichung des Vorläuferions betrug 1,80 ppm.

Tabelle 13 Identifizierte Fragment-Ionen für das in **Abbildung 51** gezeigte Peptid Aufgrund der niedrigen NCE von 23 % wurden für die Größe des Peptids verhältnismäßig wenige Fragmente detektiert. Die wenigen intensiven Signale gehören zu den Fragmenten der o.a. Bindungsbrüche, nämlich a_2^+ , b_2^+ , y_{21}^{3+} sowie y_{23}^{3+} und y_{23}^{4+} .

#	Imm.	a^+	a^{2+}	a^{3+}	a^{4+}	b^+	b^{2+}	b^{3+}	b^{4+}	Seq.
1	88,039	88,039	44,523	30,018	22,765	116,034	58,521	39,350	29,764	D
2	72,081	187,108	94,058	63,041	47,532	215,103	108,055	72,372	54,531	V
3	102,05 5	316,150	158,579	106,055	79,793	344,145	172,576	115,387	86,792	E
4	133,04 3	476,181	238,594	159,399	119,801	504,176	252,592	168,730	126,79 9	C- Carbamidom
5	30,034	533,202	267,105	178,406	134,056	561,197	281,102	187,737	141,05 5	G
6	102,05 5	662,245	331,626	221,420	166,317	690,240	345,624	230,751	173,31 5	E
7	30,034	719,266	360,137	240,427	180,572	747,261	374,134	249,759	187,57 1	G
8	110,07 1	856,325	428,666	286,113	214,837	884,320	442,664	295,445	221,83 6	H
9	120,08 1	1003,394	502,201	335,136	251,604	1031,38 9	516,198	344,468	258,60 3	F
10	133,04 3	1163,424	582,216	388,480	291,612	1191,41 9	596,213	397,811	298,61 0	C- Carbamidom
11	110,07 1	1300,483	650,745	434,166	325,876	1328,47 8	664,743	443,498	332,87 5	H
12	88,039	1415,510	708,259	472,508	354,633	1443,50 5	722,256	481,840	361,63 2	D
13	88,039	1530,537	765,772	510,851	383,390	1558,53 2	779,770	520,182	390,38 9	N- Deamidated
14	101,07 1	1658,596	829,802	553,537	415,404	1686,59 1	843,799	562,868	422,40 3	Q
15	74,060	1759,644	880,325	587,219	440,666	1787,63 8	894,323	596,551	447,66 5	T
16	133,04 3	1919,674	960,341	640,563	480,674	1947,66 9	974,338	649,895	487,67 3	C- Carbamidom
17	133,04 3	2079,705	1040,356	693,906	520,682	2107,70 0	1054,35 4	703,238	527,68 0	C- Carbamidom
18	129,11 3	2235,806	1118,407	745,940	559,707	2263,80 1	1132,40 4	755,272	566,70 6	R

19	88,039	2350,833	1175,920	784,282	588,464	2378,82 8	1189,91 8	793,614	595,46 2	D
20	87,055	2464,876	1232,942	822,297	616,974	2492,87 1	1246,93 9	831,628	623,97 3	N
21	129,11 3	2620,977	1310,992	874,330	656,000	2648,97 2	1324,99 0	883,662	662,99 8	R
22	101,07 1	2749,035	1375,021	917,017	688,014	2777,03 0	1389,01 9	926,348	695,01 3	Q
23	30,034	2806,057	1403,532	936,024	702,270	2834,05 2	1417,53 0	945,355	709,26 8	G
24	159,09 2	2992,136	1496,572	998,050	748,790	3020,13 1	1510,56 9	1007,38 2	755,78 8	W
25	44,049									A

Seq.	y ⁺	y ²⁺	y ³⁺	y ⁴⁺	#2
D					25
V	2994,152	1497,580	998,722	749,293	24
E	2895,083	1448,045	965,699	724,526	23
C-Carbamidom.	2766,041	1383,524	922,685	692,266	22
G	2606,010	1303,509	869,342	652,258	21
E	2548,989	1274,998	850,334	638,003	20
G	2419,946	1210,477	807,320	605,742	19
H	2362,925	1181,966	788,313	591,487	18
F	2225,866	1113,437	742,627	557,222	17
C-Carbamidom.	2078,797	1039,902	693,604	520,455	16
H	1918,767	959,887	640,260	480,447	15
D	1781,708	891,358	594,574	446,182	14
N-Deamidated	1666,681	833,844	556,232	417,426	13
Q	1551,654	776,331	517,890	388,669	12
T	1423,595	712,301	475,203	356,654	11
C-Carbamidom.	1322,548	661,777	441,521	331,392	10
C-Carbamidom.	1162,517	581,762	388,177	291,385	9
R	1002,486	501,747	334,834	251,377	8
D	846,385	423,696	282,800	212,352	7
N	731,358	366,183	244,458	183,595	6
R	617,315	309,161	206,443	155,084	5
Q	461,214	231,111	154,410	116,059	4
G	333,156	167,082	111,723	84,044	3
W	276,134	138,571	92,716	69,789	2
A	90,055	45,531	30,690	23,269	1

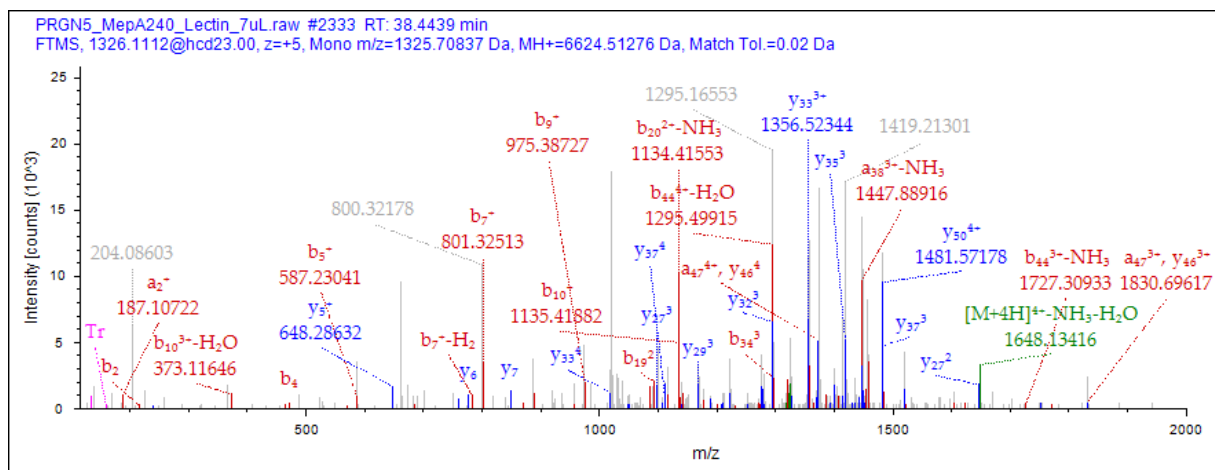


Abbildung 52 MS/MS-Spektrum des PGRN 363-418 (LC-ESI MS, positiver Ionenmodus, Vorläuferion: $[M+5H]^{5+} = 1325,708$ m/z) Das Spektrum zeigt, mit Ausnahme von b_7^+ , nur wenige intensive Signale, die zu kurzen Peptidionen gehören. Dafür konnten viele größere Fragmente mit zwei, drei und sogar vier Ladungsträgern beobachtet werden. Die Massenabweichung des Vorläuferions betrug 1,30 ppm.

Tabelle 14 Identifizierte Fragment-Ionen für das in Abbildung 52 gezeigte Peptid Wie von einem Peptid mit C-terminalem Arginin zu erwarten, konnten die intensivsten Signale hauptsächlich y-Ionen zugeordnet werden.

#1	Imm.	a ⁺	a ²⁺	a ³⁺	a ⁴⁺	a ⁵⁺	Seq.
1	88,039	88,039	44,523	30,018	22,765	18,414	D
2	72,081	187,108	94,058	63,041	47,532	38,227	V
3	70,065	284,160	142,584	95,392	71,796	57,638	P
4	133,043	444,191	222,599	148,735	111,803	89,644	C-Carbamidom.
5	88,039	559,218	280,113	187,078	140,560	112,649	D
6	88,039	674,245	337,626	225,420	169,317	135,655	N-Deamidated
7	72,081	773,313	387,160	258,443	194,084	155,469	V
8	60,044	860,345	430,676	287,453	215,842	172,875	S
9	60,044	947,377	474,192	316,464	237,600	190,281	S
10	133,043	1107,408	554,208	369,808	277,607	222,287	C-Carbamidom.
11	70,065	1204,461	602,734	402,158	301,871	241,698	P
12	60,044	1291,493	646,250	431,169	323,629	259,104	S
13	60,044	1378,525	689,766	460,180	345,387	276,511	S
14	88,039	1493,552	747,280	498,522	374,143	299,516	D
15	74,060	1594,600	797,803	532,205	399,405	319,726	T
16	133,043	1754,630	877,819	585,548	439,413	351,732	C-Carbamidom.
17	133,043	1914,661	957,834	638,892	479,421	383,738	C-Carbamidom.
18	101,071	2042,719	1021,863	681,578	511,435	409,350	Q
19	86,096	2155,804	1078,405	719,273	539,706	431,967	L
20	74,060	2256,851	1128,929	752,955	564,968	452,176	T
21	60,044	2343,883	1172,445	781,966	586,726	469,582	S
22	30,034	2400,905	1200,956	800,973	600,982	480,987	G
23	102,055	2529,947	1265,477	843,987	633,242	506,795	E
24	159,092	2716,027	1358,517	906,014	679,762	544,011	W
25	30,034	2773,048	1387,028	925,021	694,017	555,415	G
26	133,043	2933,079	1467,043	978,364	734,025	587,422	C-Carbamidom.

27	133,043	3093,109	1547,058	1031,708	774,033	619,428	C-Carbamidom.
28	70,065	3190,162	1595,585	1064,059	798,296	638,838	P
29	86,096	3303,246	1652,127	1101,754	826,567	661,455	I
30	70,065	3400,299	1700,653	1134,105	850,830	680,866	P
31	102,055	3529,342	1765,174	1177,119	883,091	706,674	E
32	44,049	3600,379	1800,693	1200,798	900,850	720,882	A
33	72,081	3699,447	1850,227	1233,821	925,617	740,695	V
34	133,043	3859,478	1930,243	1287,164	965,625	772,701	C-Carbamidom.
35	133,043	4019,508	2010,258	1340,508	1005,633	804,707	C-Carbamidom.
36	60,044	4106,540	2053,774	1369,518	1027,391	822,114	S
37	88,039	4221,567	2111,287	1407,861	1056,147	845,119	D
38	110,071	4358,626	2179,817	1453,547	1090,412	872,531	H
39	101,071	4486,685	2243,846	1496,233	1122,427	898,143	Q
40	110,071	4623,744	2312,376	1541,919	1156,691	925,555	H
41	133,043	4783,774	2392,391	1595,263	1196,699	957,561	C-Carbamidom.
42	133,043	4943,805	2472,406	1648,607	1236,707	989,567	C-Carbamidom.
43	70,065	5040,858	2520,933	1680,957	1260,970	1008,977	P
44	101,071	5168,916	2584,962	1723,644	1292,985	1034,589	Q
45	30,034	5225,938	2613,473	1742,651	1307,240	1045,993	G
46	136,076	5389,001	2695,004	1797,005	1348,006	1078,606	Y
47	74,060	5490,049	2745,528	1830,688	1373,268	1098,816	T
48	133,043	5650,079	2825,543	1884,031	1413,275	1130,822	C-Carbamidom.
49	72,081	5749,148	2875,078	1917,054	1438,042	1150,635	V
50	44,049	5820,185	2910,596	1940,733	1455,802	1164,843	A
51	102,055	5949,228	2975,117	1983,747	1488,062	1190,651	E
52	30,034	6006,249	3003,628	2002,755	1502,318	1202,056	G
53	101,071	6134,308	3067,657	2045,441	1534,332	1227,667	Q
54	133,043	6294,338	3147,673	2098,784	1574,340	1259,673	C-Carbamidom.
55	101,071	6422,397	3211,702	2141,470	1606,355	1285,285	Q
56	129,113						R

#1	b ⁺	b ²⁺	b ³⁺	b ⁴⁺	b ⁵⁺	Seq.
1	116,034	58,521	39,350	29,764	24,013	D
2	215,103	108,055	72,372	54,531	43,826	V
3	312,155	156,581	104,723	78,794	63,237	P
4	472,186	236,597	158,067	118,802	95,243	C-Carbamidom.
5	587,213	294,110	196,409	147,559	118,248	D
6	702,240	351,624	234,752	176,315	141,254	N-Deamidated
7	801,308	401,158	267,774	201,083	161,067	V
8	888,340	444,674	296,785	222,841	178,474	S
9	975,372	488,190	325,796	244,599	195,880	S
10	1135,403	568,205	379,139	284,606	227,886	C-Carbamidom.
11	1232,456	616,732	411,490	308,869	247,297	P
12	1319,488	660,248	440,501	330,627	264,703	S

13	1406,520	703,764	469,511	352,385	282,110	S
14	1521,547	761,277	507,854	381,142	305,115	D
15	1622,594	811,801	541,536	406,404	325,325	T
16	1782,625	891,816	594,880	446,412	357,331	C-Carbamidom.
17	1942,656	971,832	648,223	486,419	389,337	C-Carbamidom.
18	2070,714	1035,861	690,910	518,434	414,949	Q
19	2183,798	1092,403	728,604	546,705	437,566	L
20	2284,846	1142,927	762,287	571,967	457,775	T
21	2371,878	1186,443	791,298	593,725	475,181	S
22	2428,900	1214,953	810,305	607,980	486,586	G
23	2557,942	1279,475	853,319	640,241	512,394	E
24	2744,022	1372,514	915,345	686,761	549,610	W
25	2801,043	1401,025	934,353	701,016	561,014	G
26	2961,074	1481,040	987,696	741,024	593,021	C-Carbamidom.
27	3121,104	1561,056	1041,040	781,032	625,027	C-Carbamidom.
28	3218,157	1609,582	1073,391	805,295	644,437	P
29	3331,241	1666,124	1111,085	833,566	667,054	I
30	3428,294	1714,651	1143,436	857,829	686,465	P
31	3557,336	1779,172	1186,450	890,090	712,273	E
32	3628,374	1814,690	1210,129	907,849	726,481	A
33	3727,442	1864,225	1243,152	932,616	746,294	V
34	3887,473	1944,240	1296,496	972,624	778,300	C-Carbamidom.
35	4047,503	2024,255	1349,839	1012,631	810,306	C-Carbamidom.
36	4134,535	2067,771	1378,850	1034,389	827,713	S
37	4249,562	2125,285	1417,192	1063,146	850,718	D
38	4386,621	2193,814	1462,879	1097,411	878,130	H
39	4514,680	2257,844	1505,565	1129,425	903,742	Q
40	4651,739	2326,373	1551,251	1163,690	931,154	H
41	4811,769	2406,388	1604,595	1203,698	963,160	C-Carbamidom.
42	4971,800	2486,404	1657,938	1243,705	995,166	C-Carbamidom.
43	5068,853	2534,930	1690,289	1267,969	1014,576	P
44	5196,911	2598,959	1732,975	1299,983	1040,188	Q
45	5253,933	2627,470	1751,982	1314,239	1051,592	G
46	5416,996	2709,002	1806,337	1355,004	1084,205	Y
47	5518,044	2759,526	1840,019	1380,266	1104,415	T
48	5678,074	2839,541	1893,363	1420,274	1136,421	C-Carbamidom.
49	5777,143	2889,075	1926,386	1445,041	1156,234	V
50	5848,180	2924,594	1950,065	1462,800	1170,442	A
51	5977,223	2989,115	1993,079	1495,061	1196,250	E
52	6034,244	3017,626	2012,086	1509,316	1207,655	G
53	6162,303	3081,655	2054,772	1541,331	1233,266	Q
54	6322,333	3161,670	2108,116	1581,339	1265,272	C-Carbamidom.
55	6450,392	3225,700	2150,802	1613,353	1290,884	Q
56						R

Seq.	y ⁺	y ²⁺	y ³⁺	y ⁴⁺	y ⁵⁺	#2
D						56
V	6509,477	3255,242	2170,497	1628,125	1302,701	55
P	6410,408	3205,708	2137,474	1603,357	1282,887	54
C-Carbamidom.	6313,355	3157,181	2105,123	1579,094	1263,477	53
D	6153,325	3077,166	2051,780	1539,087	1231,471	52
N-Deamidated	6038,298	3019,653	2013,437	1510,330	1208,465	51
V	5923,271	2962,139	1975,095	1481,573	1185,460	50
S	5824,202	2912,605	1942,072	1456,806	1165,646	49
S	5737,170	2869,089	1913,062	1435,048	1148,240	48
C-Carbamidom.	5650,138	2825,573	1884,051	1413,290	1130,833	47
P	5490,108	2745,557	1830,707	1373,282	1098,827	46
S	5393,055	2697,031	1798,357	1349,019	1079,417	45
S	5306,023	2653,515	1769,346	1327,261	1062,010	44
D	5218,991	2609,999	1740,335	1305,503	1044,604	43
T	5103,964	2552,486	1701,993	1276,746	1021,599	42
C-Carbamidom.	5002,916	2501,962	1668,310	1251,485	1001,389	41
C-Carbamidom.	4842,886	2421,946	1614,967	1211,477	969,383	40
Q	4682,855	2341,931	1561,623	1171,469	937,377	39
L	4554,796	2277,902	1518,937	1139,455	911,765	38
T	4441,712	2221,360	1481,242	1111,184	889,148	37
S	4340,665	2170,836	1447,560	1085,922	868,939	36
G	4253,633	2127,320	1418,549	1064,164	851,532	35
E	4196,611	2098,809	1399,542	1049,908	840,128	34
W	4067,569	2034,288	1356,528	1017,648	814,320	33
G	3881,489	1941,248	1294,501	971,128	777,104	32
C-Carbamidom.	3824,468	1912,738	1275,494	956,872	765,699	31
C-Carbamidom.	3664,437	1832,722	1222,151	916,865	733,693	30
P	3504,406	1752,707	1168,807	876,857	701,687	29
I	3407,354	1704,181	1136,456	852,594	682,277	28
P	3294,270	1647,638	1098,761	824,323	659,660	27
E	3197,217	1599,112	1066,410	800,060	640,249	26
A	3068,174	1534,591	1023,396	767,799	614,441	25
V	2997,137	1499,072	999,717	750,040	600,233	24
C-Carbamidom.	2898,069	1449,538	966,694	725,273	580,420	23
C-Carbamidom.	2738,038	1369,523	913,351	685,265	548,413	22
S	2578,007	1289,507	860,007	645,257	516,407	21
D	2490,975	1245,991	830,997	623,499	499,001	20
H	2375,949	1188,478	792,654	594,743	475,996	19
Q	2238,890	1119,948	746,968	560,478	448,584	18
H	2110,831	1055,919	704,282	528,463	422,972	17
C-Carbamidom.	1973,772	987,390	658,596	494,198	395,560	16
C-Carbamidom.	1813,741	907,374	605,252	454,191	363,554	15
P	1653,711	827,359	551,908	414,183	331,548	14
Q	1556,658	778,833	519,558	389,920	312,137	13

G	1428,599	714,803	476,871	357,905	286,526	12
Y	1371,578	686,293	457,864	343,650	275,121	11
T	1208,515	604,761	403,510	302,884	242,509	10
C-Carbamidom.	1107,467	554,237	369,827	277,622	222,299	9
V	947,436	474,222	316,484	237,615	190,293	8
A	848,368	424,688	283,461	212,847	170,479	7
E	777,331	389,169	259,782	195,088	156,272	6
G	648,288	324,648	216,768	162,828	130,463	5
Q	591,267	296,137	197,760	148,572	119,059	4
C-Carbamidom.	463,208	232,108	155,074	116,558	93,447	3
Q	303,178	152,092	101,731	76,550	61,441	2
R	175,119	88,063	59,045	44,535	35,830	1

18	133,043	1996,809	998,908	666,275	499,958	2024,804	1012,906	675,606	506,956	C-Carbamidom.
19	133,043	2156,840	1078,924	719,618	539,965	2184,835	1092,921	728,950	546,964	C-Carbamidom.
20	102,055	2285,882	1143,445	762,632	572,226	2313,877	1157,442	771,964	579,225	E
21	86,096	2398,967	1199,987	800,327	600,497	2426,961	1213,984	809,659	607,496	L
22	70,065	2496,019	1248,513	832,678	624,760	2524,014	1262,511	842,010	631,759	P
23	60,044	2583,051	1292,029	861,689	646,518	2611,046	1306,027	871,020	653,517	S
24	30,034	2640,073	1320,540	880,696	660,774	2668,068	1334,537	890,027	667,772	G
25	101,107	2768,168	1384,588	923,394	692,797	2796,163	1398,585	932,726	699,796	K
26	136,076	2931,231	1466,119	977,749	733,563	2959,226	1480,117	987,080	740,562	Y
27	30,034	2988,253	1494,630	996,756	747,819	3016,247	1508,627	1006,087	754,817	G
28	133,043	3148,283	1574,645	1050,099	787,826	3176,278	1588,643	1059,431	794,825	C-Carbamidom.
29	133,043	3308,314	1654,661	1103,443	827,834	3336,309	1668,658	1112,774	834,833	C-Carbamidom.
30	70,065	3405,367	1703,187	1135,794	852,097	3433,362	1717,184	1145,125	859,096	P
31	104,053	3536,407	1768,707	1179,474	884,857	3564,402	1782,705	1188,806	891,856	M
32	70,065	3633,460	1817,234	1211,825	909,120	3661,455	1831,231	1221,156	916,119	P
33	88,039									N-Deamidated

Seq.	y ⁺	y ²⁺	y ³⁺	y ⁴⁺	#2
S					33
S	3707,460	1854,234	1236,492	927,621	32
V	3620,428	1810,718	1207,481	905,863	31
M	3521,360	1761,184	1174,458	881,095	30
C-Carbamidom.	3390,319	1695,663	1130,778	848,335	29
P	3230,289	1615,648	1077,434	808,328	28
D	3133,236	1567,122	1045,083	784,064	27
A	3018,209	1509,608	1006,741	755,308	26
R	2947,172	1474,090	983,062	737,548	25
S	2791,071	1396,039	931,028	698,523	24
R	2704,039	1352,523	902,018	676,765	23
C-Carbamidom.	2547,938	1274,472	849,984	637,740	22
P	2387,907	1194,457	796,641	597,732	21

D	2290,854	1145,931	764,290	573,469	20
G	2175,827	1088,417	725,947	544,712	19
S	2118,806	1059,907	706,940	530,457	18
T	2031,774	1016,391	677,929	508,699	17
C-Carbamidom.	1930,726	965,867	644,247	483,437	16
C-Carbamidom.	1770,695	885,851	590,903	443,429	15
E	1610,665	805,836	537,560	403,422	14
L	1481,622	741,315	494,546	371,161	13
P	1368,538	684,773	456,851	342,890	12
S	1271,485	636,246	424,500	318,627	11
G	1184,453	592,730	395,489	296,869	10
K	1127,432	564,220	376,482	282,613	9
Y	999,337	500,172	333,784	250,590	8
G	836,274	418,640	279,429	209,824	7
C-Carbamidom.	779,252	390,130	260,422	195,568	6
C-Carbamidom.	619,221	310,114	207,079	155,561	5
P	459,191	230,099	153,735	115,553	4
M	362,138	181,573	121,384	91,290	3
P	231,098	116,052	77,704	58,530	2
N-Deamidated	134,045	67,526	45,353	34,267	1

b. Hornerin-Peptide

Im Folgenden sind alle identifizierten Sequenzen aufgeführt. Die z.T. große Überlappung erklärt die geringe Sequenzabdeckung von etwas über 40 %.

340	350	360	370			
<u>YSYSRGHYES</u>	<u>GSGQTSGFGQ</u>	<u>HESGSGQSSG</u>	<u>YSKHGSGSGH</u>			
SRGHYES	GSGQTSGFGQ	HESGSGQSSG	Y			
SRGHYES	GSGQTSGFGQ	HESGSGQSSG				
ES	GSGQTSGFGQ	HESGSGQSSG	Y			
410	420	430	440	450	460	
<u>GQHESASRHS</u>	<u>SGRGQHSSGS</u>	<u>GQSPGHGQRG</u>	<u>SGSGQSPSSG</u>	<u>QHGTGFGRSS</u>	<u>SSGPYVSGSG</u>	
HS	SGRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGRSS	SSGPY	
HS	SGRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGRSS	SSGP	
HS	SGRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGR		
HS	SGRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFG		
	SGRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGR		
	GRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFG		
	GRGQHSSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGF		
	SSGS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFGR		
	GS	GQSPGHGQRG	SGSGQSPSSG	QHGTGFG		
500	510	520	530	540	550	
<u>HGOHGSRSGO</u>	<u>SSRGERQGSS</u>	<u>AGSSSSYGQH</u>	<u>GSGSRQSLGH</u>	<u>SRHSGSGSQS</u>	<u>PSPSRGRHES</u>	
SGQ	SSRGERQGSS	AGSSSSYGQH	GSGSRQSLGH	S	HGSGSGQS	PSPSRGRHES
	GERQGSS	AGSSSSYGQH	GSGSRQSLGH	S		
	GERQGSS	AGSSSSYGQH	GSGSRQSL			
560	570					
<u>GSRQSSSYGP</u>	<u>HGYGSGRSSS</u>					
GSRQSSSYGP	HG					
680	690	700	710	720	730	
<u>HSSSYGOHGS</u>	<u>GSGWSSSNGP</u>	<u>HGSVSGQSSG</u>	<u>FGHKSGSGQS</u>	<u>SGYSQHGSGS</u>	<u>SHSSGYRKHG</u>	
GQHGS	GSGWSSSNGP	HGSVSGQSSG	FGHKSGSGQS	SG		
	SSSNGP	HGSVSGQSSG	FGHKSGSGQS	SGYSQHGSGS	SHSSGY	
	SSSNGP	HGSVSGQSSG	FGHKSGSGQS	SGY		
	SSSNGP	HGSVSGQSSG	FGHKSGSGQS	SG		
	SSNGP	HGSVSGQSSG	FGHKSGSGQS	SGYS		
	SSNGP	HGSVSGQSSG	FGHKSGSGQS	SGY		
	SNGP	HGSVSGQSSG	FGHKSGSGQS	SGY		
	NGP	HGSVSGQSSG	FGHKSGSGQS	SGY		
		SSG	FGHKSGSGQS	SGYSQHGSGS	SHSSGY	

740	750	760	770	780	
<u>SRSGQSSRSE</u>	<u>QHGSSSGLSS</u>	<u>SYGQHSGSGH</u>	<u>QSSGHGRQGS</u>	<u>SGSHSPSRVR</u>	
SGQSSRSE	QHGSSSGLSS	SYGQHSGSGH	QSSGHGRQGS	SGSHSPSRVR	
SGQSSRSE	QHGSSSGLSS	SYGQHSGSGH	QSSGHGR		
SE	QHGSSSGLSS	SYGQHSGSGH	QSSGHGRQGS	SGSHSPS	

920	930	940	950	960	970
<u>SGSGRSSSSG</u>	<u>RHSGSGSQSS</u>	<u>GFGHKSSSGQ</u>	<u>SSGYTQHGSG</u>	<u>SGHSSSYEQH</u>	<u>GSRSGQSSRS</u>
SSSSG	RHSGSGSQSS	GFGHKSSSGQ	SSGYTQHGSG	SGHSSSYEQH	GS
SSSSG	RHSGSGSQSS	GFGHKSSSGQ	SSGYTQHG		
SSSSG	RHSGSGSQSS	GFGHKSSSGQ	SSGYTQH		
SSSSG	RHSGSGSQSS	GFGHKSSSGQ	SSGYT		
SSSSG	RHSGSGSQSS	GFGHKSSSGQ	SSGY		
SSG	RHSGSGSQSS	GFGHKSSSGQ	SSGYTQH		

980	990	1000	1010	1020	1030
<u>EOHGSSSGSS</u>	<u>SSYGQHGSGS</u>	<u>RQSLGHGQHG</u>	<u>SGSGQSPSPS</u>	<u>RGRHGSGSGQ</u>	<u>SSSYGPyRSG</u>
HGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS		
HGSSSGSS	SSYGQHGSGS	RQSL			
		QSLGHGQHG	SGSGQSPSPS	RGRHGSGSGQ	SSSY
		GHGQHG	SGSGQSPSPS	RGRHGSGSGQ	SSSYGPY
		GHGQHG	SGSGQSPSPS	RGRHGSGSGQ	SSSYGP
		HGQHG	SGSGQSPSPS	RGRHGSGSGQ	SSSYGPY
		HG	SGSGQSPSPS	RGRHGSGSGQ	SSSYGPY
		G	SGSGQSPSPS	RGRHGSGSGQ	SSSYGPY
			SGSGQSPSPS	RGRHGSGSGQ	SSSYGPY
			SGSGQSPSPS	RGRHGSGSGQ	SSSYGP

1040	1050	1060	1070	1080
<u>SGWSSSRGPY</u>	<u>ESGSGHSSGL</u>	<u>GHRESRSGQS</u>	<u>SGYGQHGSSS</u>	<u>GHSSTHGQHG</u>
RGPY	ESGSGHSSGL	GHRESR		
	ESGSGHSSGL	GHRESRSGQS	SGYGQHGSSS	GHSSTHGQHG

1140	1150	1160	1170	1180
<u>RGSGSRQSPS</u>	<u>YGRHGSGSGR</u>	<u>SSSSGQHGSG</u>	<u>LGESSGFGHH</u>	<u>ESSSGQSSSY</u>
QSPS	YGRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
QSPS	YGRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	E
QSPS	YGRHGSGSGR	SSSSGQHGSG	LGE	
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSS
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQS
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSG
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESS
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ES
	GRHGSGSGR	SSSSGQHGSG	LGESSGFGHH	E
	RHGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	HGSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	GS GSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS

1160	1170	1180	1190	1200
SSSSGQHGS	LGESSGFGHH	ESSSGQSSSY	SOHGSGSGHS	SGYGQHGS

SSSSGQHGS	LGESSGFGHH	ESSSGQSSSY	SOHGSGSGHS	SGYGQHGS
SSSSGQHGS	LGESSGFGHH	ESSSGQSSSY		
SSSSGQHGS	LGESSGFGHH	ESSSGQSSS		
SSSSGQHGS	LGESSGFGHH	ESSSGQSS		
SSSSGQHGS	LGESSGFGHH	ESSSGQS		
SSSSGQHGS	LGESSGFGHH	ESSSGQ		
SSSSGQHGS	LGESSGFGHH	ESSSG		
SSSSGQHGS	LGESSGFGHH	ESSS		
SSSSGQHGS	LGESSGFGHH	ESS		
SSSSGQHGS	LGESSGFGHH	ES		
SSSSGQHGS	LGESSGFGHH	E		
SSSSGQHGS	LGESSGFGHH			
SSGQHGS	LGESSGFGHH	ESSSGQSSS		
SGQHGS	LGESSGFGHH	ESSSGQSSS		
GQHGS	LGESSGFGHH	ESSSGQSSS		
HGS	LGESSGFGHH	ESSSGQSSS		

1250	1260	1270	1280	1290	1300
HSPSRGRHS	GLGHSSSHGQ	HSGSGSRSS	RGPYESRSGH	SSVFGQHESG	SGHSSAYSQH

GRHS	GLGHSSSHGQ	HSGSGSRSS	RGPYE		
GRHS	GLGHSSSHGQ	HSGSGSRSS	RGPY		
HGS	GLGHSSSHGQ	HSGSGSRSS	RGPYE		
HGS	GLGHSSSHGQ	HSGSGSRSS	RGPY		
GS	GLGHSSSHGQ	HSGSGSRSS	RGPY		
	GLGHSSSHGQ	HSGSGSRSS	RGPY		
			ESRSGH	SSVFGQHESG	SGHSSAY
			ESRSGH	SSVFGQHESG	SGHSSA
			ESRSGH	SSVFGQHESG	SGHS
			ESRSGH	SSVFGQHESG	SGH
			SGH	SSVFGQHESG	SGHSSAY
			SGH	SSVFGQHESG	SGHSSA

1360	1370	1380	1390	1400	1410
<u>RHGAGSGOSP</u>	<u>SRGRHGS GSG</u>	<u>HSSSYGQHGS</u>	<u>GSGWSSSSGR</u>	<u>HGSGSGQSSG</u>	<u>FGHHESSWO</u>

SP	SRGRHGS GSG	HSSSYGQHGS	GSGWS		
	RGRHGS GSG	HSSSYGQHGS	GSGW		
	GRHGS GSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
	GRHGS GSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
	GRHGS GSG	HSSSYGQHGS	GSGWSSSSGR	HG	
	GRHGS GSG	HSSSYGQHGS	GSGWSSSSGR		
	GRHGS GSG	HSSSYGQHGS	GSGWS		
	GRHGS GSG	HSSSYGQHGS	GSGW		
	HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
	HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
	HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
	HGSGSG	HSSSYGQHGS	GSGW		
		GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
		GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
		GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
		GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
		GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHH
		GS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
			SSSSGR	HGSGSGQSSG	FGHHESSWQ
			SSSSGR	HGSGSGQSSG	FGHHESSW
			SGR	HGSGSGQSSG	FGHHESSW
			GR	HGSGSGQSSG	FGHHESSWQ
				HGSGSGQSSG	FGHHESSW

1440	1450	1460	1470	1480	1490
<u>GSRSGOSSRG</u>	<u>ERHGSSSGSS</u>	<u>SSYGQHGSGS</u>	<u>RQSLGHGQHG</u>	<u>SGSGQSPSPS</u>	<u>RGRHGS GSGQ</u>

G	ERHGSSSGSS	SSYGQHGSGS	RQSL		
	HGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS	
	HGSSSGSS	SSYGQHGSGS	RQSL		

1470	1480	1490	1500
<u>RQSLGHGQHG</u>	<u>SGSGQSPSPS</u>	<u>RGRHGS GSGQ</u>	<u>SSSYSPYGSG</u>

QSLGHGQHG	SGSGQSPSPS	RGRHGS GSGQ	SSSY
GHGQHG	SGSGQSPSPS	RGRHGS GSGQ	SSSYSPY

1610	1620	1630	1640	1650
<u>RGSGSRQSPS</u>	<u>YGRHSGSGSR</u>	<u>SSSSGQHGSG</u>	<u>LGESSGFGHH</u>	<u>ESSSGQSSSY</u>

QSPS	YGRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
QSPS	YGRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	E
QSPS	YGRHSGSGSR	SSSSGQHGSG	LGE	
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSS
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQS
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSG
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESS
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ES
	GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	E
	RHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	HSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
	GSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
		SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY
		SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
		SSSSGQHGSG	LGESSGFGHH	ESSSGQSS
		SSSSGQHGSG	LGESSGFGHH	ESSSGQS
		SSSSGQHGSG	LGESSGFGHH	ESSSGQ
		SSSSGQHGSG	LGESSGFGHH	ESSSG
		SSSSGQHGSG	LGESSGFGHH	ESSS
		SSSSGQHGSG	LGESSGFGHH	ESS
		SSSSGQHGSG	LGESSGFGHH	ES
		SSSSGQHGSG	LGESSGFGHH	E
		SSSSGQHGSG	LGESSGFGHH	
		SSGQHGSG	LGESSGFGHH	ESSSGQSSS
		SGQHGSG	LGESSGFGHH	ESSSGQSSS
		GQHGSG	LGESSGFGHH	ESSSGQSSS
		HGSG	LGESSGFGHH	ESSSGQSSS

1630	1640	1650	1660	1670
<u>SSSSGQHGSG</u>	<u>LGESSGFGHH</u>	<u>ESSSGQSSSY</u>	<u>SOHGSGSGHS</u>	<u>SGYGQHGSR</u>

SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY	SOHGSGSGHS	SGYGQHGS
------------	------------	------------	------------	----------

1720	1730	1740	1750	1760	1770
<u>QSPSRGRHGS</u>	<u>GLGHSSSHGQ</u>	<u>HGSGSGRSSH</u>	<u>RGPYESRSGH</u>	<u>SSVFGQHESG</u>	<u>SGHSSAYSQH</u>

GRHGS	GLGHSSSHGQ	HGSGSGRSSH	RGPYE		
GRHGS	GLGHSSSHGQ	HGSGSGRSSH	RGPY		
HGS	GLGHSSSHGQ	HGSGSGRSSH	RGPYE		
HGS	GLGHSSSHGQ	HGSGSGRSSH	RGPY		
GS	GLGHSSSHGQ	HGSGSGRSSH	RGPY		
	GLGHSSSHGQ	HGSGSGRSSH	RGPY		
			ESRSGH	SSVFGQHESG	SGHSSAY
			ESRSGH	SSVFGQHESG	SGHSSA
			ESRSGH	SSVFGQHESG	SGHS
			ESRSGH	SSVFGQHESG	SGH
			SGH	SSVFGQHESG	SGHSSAY
			SGH	SSVFGQHESG	SGHSSA

1820	1830	1840	1850	1860
<u>SGSSQSSSYG</u>	<u>QQGSGSGQSP</u>	<u>SRGRHSGSGG</u>	<u>HSSSYGQHGS</u>	<u>GSGWSSSSGR</u>

G	QQGSGSGQSP	SRGRHSGSGG	HSSSYGQHGS	GSGWSSSSGR
G	QQGSGSGQSP	SRGRHSGSGG	HSSSYGQHGS	GSGWS
	SGSGQSP	SRGRHSGSGG	HSSSYGQHGS	GSGWSSSSGR
	SGSGQSP	SRGRHSGSGG	HSSSYGQHGS	GSGWSS
	SP	SRGRHSGSGG	HSSSYGQHGS	GSGWS

1840	1850	1860	1870	1880
<u>SRGRHSGSGG</u>	<u>HSSSYGQHGS</u>	<u>GSGWSSSSGR</u>	<u>HGSGSGQSSG</u>	<u>FGHHESSWO</u>

RGRHSGSGG	HSSSYGQHGS	GSGW		
GRHSGSGG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
GRHSGSGG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
GRHSGSGG	HSSSYGQHGS	GSGWSSSSGR	HG	
GRHSGSGG	HSSSYGQHGS	GSGWSSSSGR		
GRHSGSGG	HSSSYGQHGS	GSGWS		
GRHSGSGG	HSSSYGQHGS	GSGW		
HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
HGSGSG	HSSSYGQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
HGSGSG	HSSSYGQHGS	GSGW		
	GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
	GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHES
	GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
	GQHGS	GSGWSSSSGR	HGSGSGQSSG	FGHH
	GS	GSGWSSSSGR	HGSGSGQSSG	FGHHE
		SSSSGR	HGSGSGQSSG	FGHHESSWQ
		SSSSGR	HGSGSGQSSG	FGHHESSW
		SGR	HGSGSGQSSG	FGHHESSW
		GR	HGSGSGQSSG	FGHHESSWQ
			HGSGSGQSSG	FGHHESSW

1920	1930	1940	1950	1960	1970
<u>EOHGSSSGSS</u>	<u>SSYGQHGSGS</u>	<u>RQSLGHGQHG</u>	<u>SGSGQSPSPS</u>	<u>RGRHSGSGSQ</u>	<u>SSSYGPGYSG</u>

HGSSSGSS	SSYGQHGSGS	RQSLGHGQHG	SGSGQSPSPS		
HGSSSGSS	SSYGQHGSGS	RQSL			
		QSLGHGQHG	SGSGQSPSPS	RGRHSGSGSQ	SSSY
		GHGQHG	SGSGQSPSPS	RGRHSGSGSQ	SSSYGPY
		GHGQHG	SGSGQSPSPS	RGRHSGSGSQ	SSSYGP
		HGQHG	SGSGQSPSPS	RGRHSGSGSQ	SSSYGPY
		HG	SGSGQSPSPS	RGRHSGSGSQ	SSSYGPY
		G	SGSGQSPSPS	RGRHSGSGSQ	SSSYGPY
			SGSGQSPSPS	RGRHSGSGSQ	SSSYGPY
			SGSGQSPSPS	RGRHSGSGSQ	SSSYGP

1980	1990	2000	2010	2020
<u>SGWSSSRGPY</u>	<u>ESGSGHSSGL</u>	<u>GHRESRSGQS</u>	<u>SGYGQHGSSS</u>	<u>GHSSTHGQHG</u>

RGPY	ESGSGHSSGL	GHRESR		
	ESGSGHSSGL	GHRESRSGQS	SGYGQHGSSS	GHSSTHGQHG

2220 2230 2240 2250
RGPYESRSGH SSVFGQHESG SGHSSAYSQH GSGSGHFCSQ

ESRSGH SSVFGQHESG SGHSSAY
 ESRSGH SSVFGQHESG SGHSSA
 ESRSGH SSVFGQHESG SGHS
 ESRSGH SSVFGQHESG SGH
 SGH SSVFGQHESG SGHSSAY
 SGH SSVFGQHESG SGHSSA

2290 2300 2310 2320 2330
SGSSQSSSYG OQSGSGSQSP SRGRHGSQSG HSSSYGQHGS GSGWSSSSSGR

G OQSGSGSQSP SRGRHGSQSG HSSSYGQHGS GSGWSSSSSGR
 G OQSGSGSQSP SRGRHGSQSG HSSSYGQHGS GSGWS

2300 2310 2320 2330 2340 2350
OQSGSGSQSP SRGRHGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHESSWO

SGSQSP SRGRHGSQSG HSSSYGQHGS GSGWSSSSSGR
 SSGSQSP SRGRHGSQSG HSSSYGQHGS GSGWSS
 SP SRGRHGSQSG HSSSYGQHGS GSGWS
 RGRHGSQSG HSSSYGQHGS GSGWS
 GRHGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHESS
 GRHGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHE
 GRHGSQSG HSSSYGQHGS GSGWSSSSSGR HG
 GRHGSQSG HSSSYGQHGS GSGWSSSSSGR
 GRHGSQSG HSSSYGQHGS GSGWS
 GRHGSQSG HSSSYGQHGS GSGW
 HGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHESS
 HGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHESS
 HGSQSG HSSSYGQHGS GSGWSSSSSGR HGSQSGQSSG FGHHE
 HGSQSG HSSSYGQHGS GSGW
 QHGS GSGWSSSSSGR HGSQSGQSSG FGHHESS
 QHGS GSGWSSSSSGR HGSQSGQSSG FGHHESS
 QHGS GSGWSSSSSGR HGSQSGQSSG FGHHE
 QHGS GSGWSSSSSGR HGSQSGQSSG FGH
 GS GSGWSSSSSGR HGSQSGQSSG FGHHE
 SSSSGR HGSQSGQSSG FGHHESSWO
 SSSSGR HGSQSGQSSG FGHHESSW
 SGR HGSQSGQSSG FGHHESSW
 GR HGSQSGQSSG FGHHESSWO
 HGSQSGQSSG FGHHESSW

2380 2390 2400 2410 2420
GSRSGQSSRG ERHGSSSGSS SSYGQHGSGS RQSLGHGQHG SSGSQSPSPS

G ERHGSSSGSS SSYGQHGSGS RQSLGHGQHG SSGSQSPSPS
 HGSSSGSS SSYGQHGSGS RQSLGHGQHG SSGSQSPSPS
 HGSSSGSS SSYGQHGSGS RQSL

2410 2420 2430 2440
RQSLGHGQHG SSGSQSPSPS RGRHGSQSGQ SSSYSPYGS

QSLGHGQHG SSGSQSPSPS RGRHGSQSGQ SSSY
 GHGQHG SSGSQSPSPS RGRHGSQSGQ SSSYSPY

2450	2460	2470	2480	2490
<u>SGWSSSRGPY</u>	<u>ESGSGHSSGL</u>	<u>GHRESRSGQS</u>	<u>SGYGQHGSSS</u>	<u>GHSSTHGQHG</u>

RGPY	ESGSGHSSGL	GHRESR		
	ESGSGHSSGL	GHRESRSGQS	SGYGQHGSSS	GHSSTHGQHG

2550	2560	2570	2580	2590
<u>RGSGSRQSPS</u>	<u>YGRHSGSGSR</u>	<u>SSSSGQHGSG</u>	<u>LGESSGFGHH</u>	<u>ESSSGQSSSY</u>

QSPS	YGRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS
QSPS	YGRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	E
QSPS	YGRHSGSGSR	SSSSGQHGSG	LGE	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSS	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQS	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSG	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESS	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ES	
GRHSGSGSR	SSSSGQHGSG	LGESSGFGHH	E	
RHSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS	
HGSGSGSR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS	
GSGSGR	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS	
	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY	
	SSSSGQHGSG	LGESSGFGHH	ESSSGQSSS	
	SSSSGQHGSG	LGESSGFGHH	ESSSGQSS	
	SSSSGQHGSG	LGESSGFGHH	ESSSGQS	
	SSSSGQHGSG	LGESSGFGHH	ESSSGQ	
	SSSSGQHGSG	LGESSGFGHH	ESSSG	
	SSSSGQHGSG	LGESSGFGHH	ESSS	
	SSSSGQHGSG	LGESSGFGHH	ESS	
	SSSSGQHGSG	LGESSGFGHH	ES	
	SSSSGQHGSG	LGESSGFGHH	E	
	SSSSGQHGSG	LGESSGFGHH		
	SSGQHGSG	LGESSGFGHH	ESSSGQSSS	
	SGQHGSG	LGESSGFGHH	ESSSGQSSS	
	GQHGSG	LGESSGFGHH	ESSSGQSSS	
	HGSG	LGESSGFGHH	ESSSGQSSS	

2570	2580	2590	2600	2610
<u>SSSSGQHGSG</u>	<u>LGESSGFGHH</u>	<u>ESSSGQSSSY</u>	<u>SQHSGSGGHS</u>	<u>SGYGQHGSR</u>

SSSSGQHGSG	LGESSGFGHH	ESSSGQSSSY	SQHSGSGGHS	SGYGQHGSR
------------	------------	------------	------------	-----------

2660	2670	2680	2690
<u>QSPSRGRHGS</u>	<u>GLGHSSSHGQ</u>	<u>HGSGSGRSSH</u>	<u>RGPYESRLGH</u>

GRHGS	GLGHSSSHGQ	HGSGSGRSSH	RGPYE
GRHGS	GLGHSSSHGQ	HGSGSGRSSH	RGPY
HGS	GLGHSSSHGQ	HGSGSGRSSH	RGPYE
HGS	GLGHSSSHGQ	HGSGSGRSSH	RGPY
GS	GLGHSSSHGQ	HGSGSGRSSH	RGPY
	GLGHSSSHGQ	HGSGSGRSSH	RGPY

11. Erklärung zum Eigenanteil der Dissertation und Publikationen

Hiermit versichere ich an Eides statt, dass ich meine Dissertation – abgesehen von der Beratung meiner wissenschaftlichen Betreuer – nach Form und Inhalt eigenständig angefertigt habe. Nicht eigens durchgeführte Versuche oder Ergebnisse dergleichen, die dem Verständnis der Ausrichtung der praktischen Arbeiten dienen, sind im Text eindeutig gekennzeichnet. Ich habe außer den in der Arbeit aufgeführten keine weiteren Hilfsmittel verwendet. Die vorliegende Arbeit lag bisher keiner anderen akademischen Institution als Dissertation vor – weder in Teilen noch im Ganzen. Die Arbeit entstand unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft. Bis zum Tage der Vorlage dieser Arbeit wurde mir kein akademischer Grad entzogen.

Kiel, den

Alexander Becker

Liste der in den Promotionssemestern entstandenen Publikationen

T. Latendorf, U. Gerstel, Z. Wu, J. Bartels, A. Becker, A. Tholey, J.-M. Schröder, Cationic Intrinsically Disordered Antimicrobial Peptides (CIDAMPs) Represent a New Paradigm of Innate Defense with a Potential for Novel Anti-Infectives, *Sci Rep* 2019, 9:3331.

U. Gerstel, T. Latendorf, J. Bartels, A. Becker, A. Tholey, J.-M. Schröder, Hornerin contains a Linked Series of Ribosome-Targeting Peptide Antibiotics, *Sci Rep* 2018, 8:16158.

A. Tholey, A. Becker, Top-down proteomics for the analysis of proteolytic events – Methods, applications and perspectives, *Biochim Biophys Acta Mol Cell Res.* 2017, 1864, 2191-2199.

C. W. Hung, T. Koudelka, C. Anastasi, A. Becker, C. Moali, A. Tholey, Characterization of post-translational modifications in full-length human BMP-1 confirms the presence of a rare vicinal disulfide linkage in the catalytic domain and highlights novel features of the EGF domain, *J Proteomics* 2016, 138, 136-145.