

# Design of a Stabilizing and Switching Module for $\alpha$ -Helical Peptides



## **Dissertation**

zur Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

**Matthias Lipfert**

Kiel, 2017



## **Preface**



1. Gutachter: Prof. Dr. Frank D. Sönnichsen
  2. Gutachter: Prof. Dr. Axel Scheidig
- Tag der mündlichen Prüfung: 10.07.2017  
Zum Druck genehmigt:

gez. Prof. Dr. Natascha Oppelt  
Dekanin

Die vorliegende Arbeit wurde unter Anleitung von  
**Prof. Dr. Frank D. Sönnichsen**  
in der Zeit von Oktober 2013 bis Mai 2017  
am Otto Diels-Institut für Organische Chemie  
der Christian-Albrechts-Universität zu Kiel angefertigt.

---

### Eigenständigkeitserklärung

Hiermit erkläre ich, Matthias Lipfert, die vorliegende Dissertation selbstständig und ausschließlich unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel angefertigt zu haben. Inhalt und Form der Arbeit habe ich, abgesehen von Beratung durch meinen Betreuer, Prof. Dr. Frank D. Sönnichsen, eigenständig erarbeitet und verfasst. Bei der Entstehung der Arbeit wurden die Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft eingehalten. Weder die gesamte Arbeit noch Teile hiervon wurden an anderer Stelle im Rahmen eines Prüfungsverfahrens eingereicht. Diese Dissertation ist mein erster Promotionsversuch.

Teile von Projekten mit Kooperationspartnern wurden in Peer-Review-Zeitschriften publiziert. Diese wurden aufgrund der unterschiedlichen Thematik nicht mit in die Arbeit aufgenommen.

- (1) M. Schulz-Senft, M. Lipfert, A. Staubitz, Mechanopolymerchemie - Molekulare Wirkung durch Kraft, *Chem. Unserer Zeit* **2014**, *48*, 200-214.
- (2) J. Strueben, M. Lipfert, J. O. Springer, C. A. Gould, P. J. Gates, F. D. Sönnichsen, A. Staubitz, High-yield lithiation of azobenzenes by tin-lithium exchange, *Chemistry* **2015**, *21*, 11165-11173.





## Danksagung

An erster Stelle bedanke ich mich bei Frank Sönnichsen für die großartige Zeit. Die Möglichkeit der freien Entfaltung in der Beschäftigung mit meinem Thema und die gleichzeitige Unterstützung bei weiteren Projekten sind wirklich einmalig.

Die gute Atmosphäre in der Arbeitsgruppe wurde verstärkt durch die weiteren Mitglieder des Arbeitskreises. Insbesondere danke ich dabei Hauke Kobarg und Joana Gronow, die seit Beginn der Arbeit dabei waren. Weiterhin möchte ich mich bei den "neuen" Arbeitskreismitgliedern Nils Preußke und Katrin Bamberg bedanken, die nicht nur im Rahmen von Praktika und Abschlussarbeiten zu meinem Thema beigetragen haben, sondern dieses auch mit neuen Ideen und Ansätzen selbst weiterführen werden.

Einen großen Anteil an der besonderen Atmosphäre, sei es wissenschaftlicher oder privater Natur, hat auch zu jeder Zeit der Arbeitskreis um Anne Staubitz beigetragen. Vom wöchentlichen Frühstück bis zu dem einem oder anderen Feierabendbier. Insbesondere Jan Strüben, Jan-Ole Springer und Mathias Schulz sind dabei zu erwähnen.

Ein besonderer Dank gilt meinen Kooperationspartnern, die zum Gelingen meines Projektes einen wesentlichen Teil beigetragen haben. Dazu zählen Anne Müller (ja, ich weiß, jetzt Heitmann), Mathias Schulz, Jan Strüben (danke für den Linker) und Henrik Böhnke. Hinzu kommen Katrin Bamberg, Robert Eick, Nils Preußke und Adam Simpsons, die als Praktikanten (Bachelorarbeit, F2, F3, Masterarbeit RISE Student) mit zum Erfolg der Arbeit beigetragen haben.

Auch durfte ich mich an Projekten anderer beteiligen, die zum Teil zu Publikationen führten. Unter anderem mit M. Schulz und J. Strüben aus dem AK Staubitz und mit Julian Rudelt aus dem AK Verreet der Phytopathologie.

Ganz besonders möchte ich mich auch bei meinen Freunden bedanken, ohne die das Leben und Forschen in Kiel langweilig gewesen wäre. Alle aufzuzählen und in eine (gerechte) Reihenfolge zu bringen, ist unmöglich. Deswegen, danke für die Unterhaltung, die Musik, das Konzert, das Festival, den Abend,...

Zu guter Letzt danke ich meiner Familie, ohne deren Unterstützung die folgende Arbeit nicht möglich gewesen wäre.

"Maybe it meant something. Maybe not, in the long run, but no explanation, no mix of words or music or memories can touch that sense of knowing that you were there and alive in that corner of time and the world. Whatever it meant."

Hunter S. Thompson, *Fear and Loathing in Las Vegas*

## Kurzzusammenfassung

In dieser Arbeit wurde ein modulares Konzept zur Stabilisierung und Schaltung der Struktur  $\alpha$ -helikaler Proteine und Peptide untersucht. Zur Realisierung dieses Konzeptes haben wir ein Fusionsprotein entwickelt, welches sich aus zwei Teilen zusammensetzt. Der erste Teil besteht aus einem Protein mit einer stabilen helikalen Faltung und der Möglichkeit, eine photoresponsive Einheit zur Manipulation der Struktur einzuführen. Der zweite Teil wird von einer Peptidsequenz mit strukturabhängiger Aktivität gebildet. Das Schaltmodul basiert auf dem 20 Aminosäuren langen Trp-cage Miniprotein, dessen Struktur eine acht Aminosäuren lange N-terminalen  $\alpha$ -Helix enthält. Die Faltung wird durch hydrophobe Wechselwirkungen einer Tryptophan-Seitenkette mit mehreren C-terminalen Aminosäuren induziert, was zu einer hoch stabilen, tertiären Faltung führt. Zur Schaltung der Faltung haben wir in diese Struktur ein photoresponsives Molekül eingeführt. In dieser Arbeit wurde N-Hydroxysuccinimid-Ester als reaktive Gruppe und Azobenzol als Grundgerüst verwendet. Um diesen Schalter in den Trp-cage einzubinden, waren verschiedene Ersetzungen und Modifikationen in der Sequenz notwendig um eine selektive Verknüpfung zu gewährleisten. Die Verknüpfungsreaktion erweist sich als hochgradig selektiv und führt mit guter Ausbeute zu einfach aufzureinigendem Produkt. Die neue Trp-cage Sequenz und deren verlinkte Varianten wurden mit besonderem Fokus auf die Struktur und den Einfluss des Linkers auf ebenjene spektroskopisch charakterisiert.

Im zweiten Teil der Arbeit wurde eine Wiederholungseinheit des  $\alpha$ -helikalen Antigegefrierproteins „Type 1 HPLC6“ mit dem Trp-cage in einem Chimären-Konzept fusioniert. Zur Verknüpfung und Korrelation der Sekundärstrukturen beider Proteine wurden ihre Sequenzen überlappt, was zu einem Chimären-Protein führt. Dieses Protein lässt sich in drei Abschnitte einteilen. Die Sequenz des Antigegefrierproteins bildet den N-terminalen Abschnitt, der mittlere Abschnitt enthält Sequenzen beider Proteine, während der C-terminale Abschnitt aus der Trp-cage-Sequenz besteht. Nach diesem Prinzip ergeben sich vier Kombinationsmöglichkeiten unter Berücksichtigung der Orientierung der aktiven Seite des Antigegefrierproteins und der Tertiärstruktur des Trp-cages. Alle vier Varianten wurden zur Strukturbestimmung spektroskopisch charakterisiert. Darüber hinaus wurde die Aktivität der vier Chimären-Proteine mittels Beobachtung des Eiskristallwachstums in Gegenwart der Proteine bestimmt.



## Abstract

In this work, the concept of a modular approach to stabilize and switch the structure of  $\alpha$ -helical proteins and peptides was investigated. To achieve such, we designed a fusion protein which consists of two parts. The first part, an independent protein module, comprises a highly stable helical fold and the possibility to incorporate a light responsive unit for manipulation of the structure. The second part consists of a peptide sequence with a structure-dependent activity.

We used the 20 amino acid long Trp-cage miniprotein as basis for the stabilizing and switching module. Its tertiary structure consists of an eight amino acid long N-terminal  $\alpha$ -helix, whose fold is induced by encapsulation of the eponymous tryptophan side chain by several C-terminal residues. This results in a very stable tertiary fold of the module. In order to introduce switchability into the miniprotein, we cross-linked the Trp-cage with a photo responsive linker. In this work, we used N-hydroxysuccinimide esters (NHS) as reactive groups for the cross-linking, which are reactive towards primary amines (lysine residues), and azobenzene as the scaffold for a switchable linker. In order to apply this cross-linking method, several replacements and modifications in the Trp-cage sequence were required to ensure unambiguous linking. The cross-linking reaction proved to be highly selective and yielded easily purifiable protein product. The new Trp-cage and its cross-linked variants were spectroscopically characterized with primary focus on structure, stability and the influence of the linker on the fold.

In the second part of this work, one repeat of the  $\alpha$ -helical antifreeze protein AFP "Type 1 HPLC6" was fused with the Trp-cage module using a chimera strategy. In order to combine and correlate the secondary structure of the proteins, the sequences were overlapped, resulting in a chimera protein. This protein comprises at the N-terminus the sequence of the antifreeze protein, followed by a chimera region with a combination of both sequences, and is completed by the C-terminal sequence of the Trp-cage. The junction of the sequences is possible in four different ways altering the orientation of the active side of the antifreeze protein relative to the tertiary structure of the Trp-cage. All four variants were spectroscopically characterized in order to ascertain their structure. Furthermore, the activity of the four chimera proteins was determined by observing the ice crystal growth in presence of the proteins.



## Contents

<b>Preface</b>	<b>I</b>
<b>Kurzzusammenfassung</b>	<b>IX</b>
<b>Abstract</b>	<b>XI</b>
<b>Introduction and Objectives</b>	<b>1</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Antifreeze and Ice Binding Proteins . . . . .	5
1.1.1 Antifreeze Protein Type 1 . . . . .	9
1.2 The Trp-cage . . . . .	11
1.3 Switching at the Molecular Level . . . . .	13
1.3.1 Switching of Biopolymers . . . . .	15
<b>2 Motivation and Previous Work</b>	<b>17</b>
<b>3 Concept</b>	<b>21</b>
3.1 The Switch-Cage Module . . . . .	22
3.2 Concepts for Fusion . . . . .	23
<b>I The Switch-Cage Module</b>	<b>27</b>
<b>1 The Trp-Cage Fold</b>	<b>29</b>
1.1 The Folded Fraction . . . . .	30
1.2 The Unfolded Fractions . . . . .	34
<b>2 Link the Cysteine in the Trp-Cage with a Cross-Linker</b>	<b>41</b>
2.1 Introducing Cysteine Residues into Tc5b . . . . .	45
2.1.1 Tc5b-L2C-S20C . . . . .	48
2.2 Cross-Linking of Tc5b-L2C-S20C . . . . .	51
<b>3 Link the Lysine in the Trp-Cage with a Cross-Linker</b>	<b>59</b>

---

3.1	Trp-cage Variant for Cross-Linking Using NHS-Esters . . . . .	60
3.2	Cross-Linker with N-hydroxysuccinimide ester (NHS) Esters as Functional Group . . . . .	64
3.3	TcKKA Linked with the Lomant-Linker . . . . .	67
<b>4</b>	<b>TcKKA Linked with Azobenzene</b>	<b>73</b>
4.1	Properties of TcKKA-azo . . . . .	74
4.1.1	Photochromic Properties . . . . .	75
4.1.2	Structural Characterization of TcKKA-azo . . . . .	79
4.1.3	Structural Characterization of TcKKA-azo in <i>trans</i> Configuration	80
4.1.4	Structural Characterization of TcKKA-azo in the <i>cis</i> Configuration	83
4.2	Discussion of the Properties of TcKKA-azo . . . . .	86
<b>II</b>	<b>The AFP-Cage</b>	<b>89</b>
<b>1</b>	<b>Designing an AFP-Cage</b>	<b>91</b>
1.1	AFP-Tc-7 – DTASDAAAA-AAYTAWLA-DGGPSSGRPPPS . . . . .	96
1.2	AFP-Tc-6 – DTASDAAA-AAYLTWLA-DGGPSSGRPPPS . . . . .	104
1.3	AFP-Tc-5 – DTASDA-AAYAAWTA-DGGPSSGRPPPS . . . . .	107
1.4	AFP-Tc-4 – DTASD-AAYAAWLT-DGGPSSGRPPPS . . . . .	111
1.5	Discussion of the Fold at Neutral pH . . . . .	114
1.6	Activity of the AFP-Tc Chimera Proteins . . . . .	118
<b>III</b>	<b>Summary and Outlook</b>	<b>121</b>
<b>1</b>	<b>Linking and Switching</b>	<b>123</b>
<b>2</b>	<b>AFP-Cage Chimera Proteins</b>	<b>127</b>
<b>3</b>	<b>Other Fusion Partner and Possibilities</b>	<b>131</b>
<b>IV</b>	<b>Experimental Section</b>	<b>135</b>
<b>1</b>	<b>Solid Phase Peptide Synthesis</b>	<b>137</b>
1.1	Preparation of the Resin . . . . .	137
1.2	Coupling Cycle . . . . .	137
1.3	Removal of the Fmoc Protecting Group . . . . .	139
1.4	Cleavage and Deprotection of the Peptide . . . . .	139
1.5	HPLC and ESI-MS . . . . .	140
1.6	Sequence Difficulties . . . . .	142



---

<b>2</b>	<b>Procedure of Cross-Linking the Trp-Cage</b>	<b>145</b>
2.1	Cross-Linking with Activated Halides on Cys . . . . .	145
2.2	Cross-Linking with NHS Ester on Lys . . . . .	145
2.3	Linker Synthesis . . . . .	146
2.3.1	4,4'-Bis(2,5-dioxopyrrolodin-1yl)azobenzene . . . . .	146
<b>3</b>	<b>Used Methods and Instrumentation</b>	<b>147</b>
3.1	HPLC . . . . .	147
3.2	Mass Analysis . . . . .	147
3.3	Lyophilizer . . . . .	147
3.4	UV/vis Spectroscopy . . . . .	147
3.5	IR Spectroscopy . . . . .	148
3.6	Melting Points . . . . .	148
3.7	CD Spectroscopy . . . . .	148
3.8	NMR Spectroscopy . . . . .	148
3.8.1	Protein NMR . . . . .	149
3.8.2	Reference Shifts for the Trp-Cage Fold . . . . .	149
3.9	Ice Activity Assay . . . . .	150
3.10	Photo-Irradiation Equipment . . . . .	150
3.11	Molecular Modelling and MD Simulations . . . . .	151
<b>V</b>	<b>Appendix</b>	<b>153</b>
<b>1</b>	<b>Spectra and NMR Shifts of the Trp-cage Variants</b>	<b>155</b>
1.1	Tc5b and its Variants . . . . .	156
1.1.1	Tc5b . . . . .	156
1.1.2	Synthesis Details of Tc5b-L2C-S20C . . . . .	161
1.1.3	MALDI-ToF Spectra of Purified Cross-Linked Tc5b-L2C-S20C . . . . .	164
1.2	Tc10b and its Variants . . . . .	168
1.2.1	Tc10b . . . . .	168
1.2.2	Tc10b-S20C-21G . . . . .	172
1.2.3	Tc10b-A2K-S20C(S-tBu)-21G . . . . .	178
1.2.4	Tc10b-A2C(S-tBu)-S20K . . . . .	181
1.2.5	Tc10b-S20K . . . . .	184
1.2.6	TcKKA . . . . .	186
1.2.7	TcKKA-Lomant . . . . .	197
1.2.8	TcKKA-Azobenzene . . . . .	200
1.3	AFP-Tc Chimera Proteins . . . . .	204
1.3.1	15er Reference Peptide . . . . .	204

---

1.3.2	15erYW Reference Peptide . . . . .	209
1.3.3	AFP-Tc-4 . . . . .	214
1.3.4	AFP-Tc-5 . . . . .	239
1.3.5	AFP-Tc-5-KK . . . . .	264
1.3.6	AFP-Tc-7 . . . . .	268
1.3.7	AFP-Tc-6 . . . . .	279
1.4	4,4'-Bis(2,5-Dioxopyrrolodin-1-yl)Azobenzene . . . . .	292
<b>2</b>	<b>UV/vis-Spectra and Concentrations and Helical Content (CD)</b>	<b>293</b>
<b>3</b>	<b>Syntheses</b>	<b>297</b>
3.1	Tc10b . . . . .	297
<b>4</b>	<b>List of Chemicals</b>	<b>303</b>
	<b>List of Figures</b>	<b>303</b>
	<b>List of Schemes</b>	<b>311</b>
	<b>List of Tables</b>	<b>312</b>
	<b>Bibliography</b>	<b>317</b>
	<b>Abbreviations</b>	<b>323</b>

## **Introduction and Objectives**



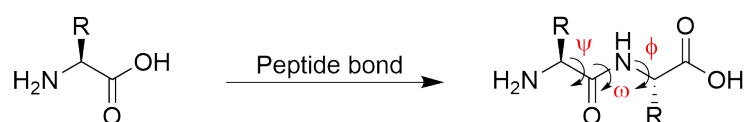
## 1. Introduction

Proteins are important modules for every living being. They play key roles in most process in cells. They may act as messenger,<sup>1</sup> storage,<sup>2</sup> or carrier (ion pumps<sup>3</sup>) molecules, catalyze chemical reactions (enzymes)<sup>4</sup> and are essential for the cells abilities to regulate its metabolism,<sup>5</sup> to communicate extracellularly<sup>6</sup> or to move.<sup>7</sup> This broad spectrum of functions can be fulfilled due to their great structural diversity. Proteins can be assessed at different structural levels, ranging from the primary up to the quaternary structure.

**Table 1.1.:** Categories for the 20 naturally occurring amino acids based on their side chains propensity to be in contact with water. The amino acids are listed with the three- and one-letter abbreviations.

Charged			Polar			Hydrophobic		
Name	Abbreviations		Name	Abbreviations		Name	Abbreviations	
Arginine	Arg	R	Asparagine	Asn	N	Glycine	Gly	G
Lysine	Lys	K	Glutamine	Gln	Q	Alanine	Ala	A
Aspartic acid	Asp	D	Histidine	His	H	Valine	Val	V
Glutamic acid	Glu	E	Serine	Ser	S	Leucine	Leu	L
			Threonine	Thr	T	Isoleucine	Ile	I
			Tyrosine	Tyr	Y	Phenylalanine	Phe	F
			Cysteine	Cys	C	Proline	Pro	P
			Methionine	Met	M			
			Tryptophan	Trp	W			

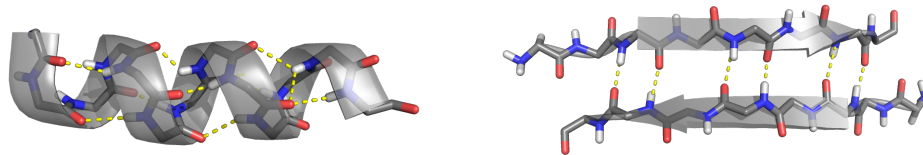
The primary structure is defined by the sequence, which is the order of the 20 naturally occurring amino acids. The amino acids differ in their side chains, ranging from a simple proton to an aromatic indole ring, i.e. glycine (Gly, G) and tryptophan (Trp, W). They may be grouped by the propensity of their side chains to interact with water. This results in charged (bearing a positive or negative charge), polar but uncharged, and hydrophobic groups (Table 1.1).



**Scheme 1.1.:** Scaffold of amino acids and the peptide bond with structure defining torsion angles  $\phi$ ,  $\psi$  and  $\omega$  highlighted in red.

The amino acids are covalently connected between the amine- and carboxyl function resulting in the peptide bond (Scheme 1.1) and form the primary structure of proteins. The lone pair of electrons located on the nitrogen is significantly delocalized, which gives the group a partial double bond character. The bond is defined by the torsion angle  $\omega$  between the H-N-C-O atoms. In most peptide bonds, a torsion angle of  $\omega = 180^\circ$

(*trans*) is preferred. An exception is the X-Pro peptide bond, in which the *trans* and the *cis* configuration is nearly equal in energy. The torsion angles  $\phi$  and  $\psi$  are flexible and enable the polypeptide backbone to adopt certain conformations.



**Figure 1.1.:** Exemplary secondary structure elements. Left:  $\alpha$ -helix. Right anti-parallel  $\beta$ -sheet. Both structures are represented in cartoon style with amide backbone in stick representation. Hydrogen bonds are displayed in yellow dashed lines.

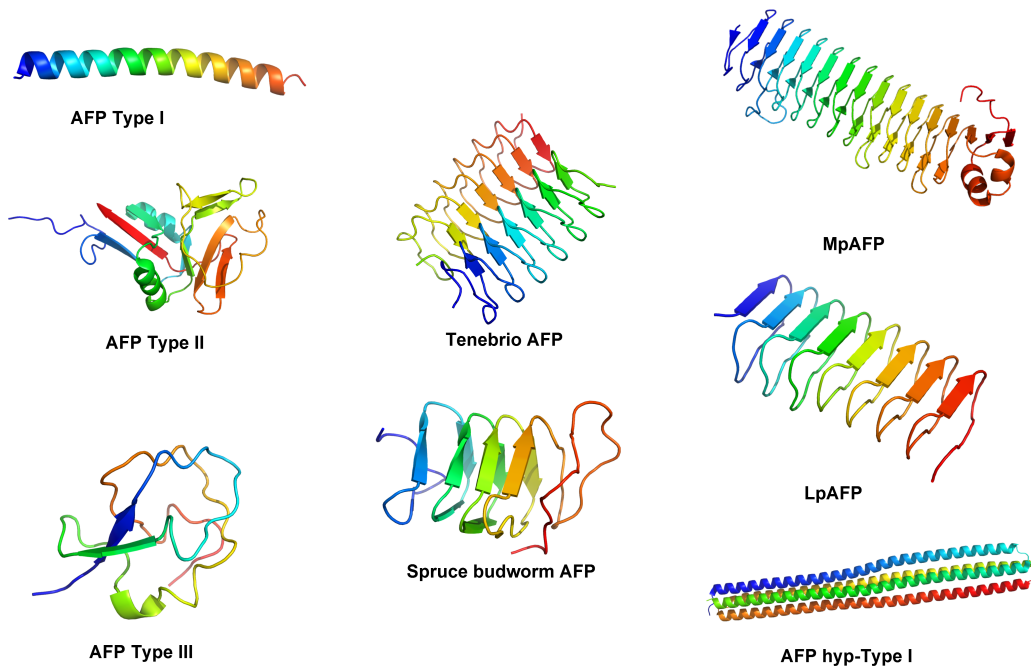
The primary structure can fold from flexible, disordered, a random coiled chain into locally regular conformations, driven by intramolecular hydrogen bonds. These "secondary structure elements" are defined by their hydrogen bonds between the backbone amide and carboxyl groups. The most common secondary structure elements are the  $\alpha$ -helix with 3.6 amino acids per turn and the  $\beta$ -strand (Figure 1.1). In the  $\alpha$ -helix, hydrogen bonds between carbonyl groups and amides in backbone with a gap of four are observed, which lead to a twisted backbone with torsion angles of  $\phi = -57^\circ$  and  $\psi = -47^\circ$ . In the  $\beta$ -sheet, hydrogen bonds are observed between two or more peptide chains. The  $\beta$ -sheet structure is defined by torsion angles for the backbone of  $\phi = -80^\circ$  and  $\psi = +150^\circ$ , which lead to a zig-zag structure of the backbone.<sup>8</sup>

The tertiary structure is the three dimensional arrangement of the polypeptide with one or more secondary structure elements. The three dimensional arrangement is predominantly driven by the hydrophobic effect involving the side chains of the amino acids. Under physiological conditions, solvent water molecules are not able to form hydrogen bonds to the hydrophobic side chains. As a result, stronger hydrogen bonds between the water molecules around the hydrophobic side chain are observed. The water molecules are ordered, which is an entropically unfavoured state. Folding of the protein into its tertiary structure leads to encapsulation of hydrophobic side chains of the protein. Thus, the hydrophobic regions are inaccessible to water and the tertiary structure is entropically driven by the hydrophobic effect.

The next and last structure level is based on the proteins ability to build up complexes with each other, either homo-multimers or hetero-multimers. This enlarged three dimensional arrangement of tertiary structured proteins is the quaternary structure and is facilitated by hydrophobic surface interactions as well as charged interactions.

## 1.1. Antifreeze and Ice Binding Proteins

A unique and interesting class of proteins are antifreeze proteins (AFPs) or ice binding proteins (IBPs). In the 1950s, P.F. SCHOLANDER suggested that Antarctic fish have antifreeze agents in their blood as an explanation for their ability to survive in sub-zero temperature habitats.<sup>9</sup> In the year 1969, A.L. DEVRIES screened the blood of ANTARCTIC NOTOTHENIOID and found that the concentration of sodium chloride, urea and free amino acids in the serum is indeed insufficient to explain the observed freezing point. They discovered a glycoprotein, the antifreeze glycoprotein (AFGP), in the ANTARCTIC NOTOTHENIOID fishes that accounts for an additional freezing point depression by 30%.<sup>10</sup>

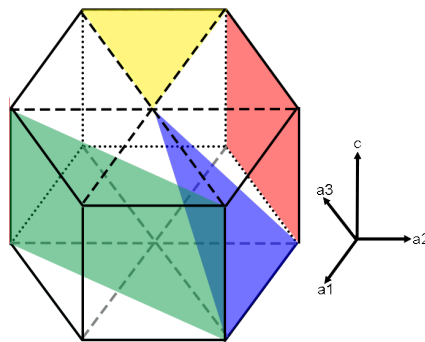


**Figure 1.2.:** Overview of the structural diversity of different AFPs obtained from multiple species.<sup>11–20</sup>

After this discovery, other AFPs have been detected and isolated not only from Antarctic fish, but in many different species, such as bacteria, fishes, insects, fungi.<sup>21</sup> They are solely protein based, i.e. do not contain carbohydrate groups. The independent evolution of AFP in this variety of organisms naturally leads to a broad structural diversity (Figure 1.2).<sup>11–20</sup> AFP structures ranges from a simple helix with repeated sequence (AFP Type 1) up to a globular lectin-type protein with united  $\alpha$ -,  $\beta$ -fold and a non repetitive sequence (AFP Type 2). Interestingly, AFPs have been identified with mul-

multiple repetitive sequence elements, which lead to enlarged sequences. These molecules fold as parallel  $\beta$ -helices (MpAFP, LpAFP). These proteins exhibit large, flat surfaces that are responsible for their hyperactivity.

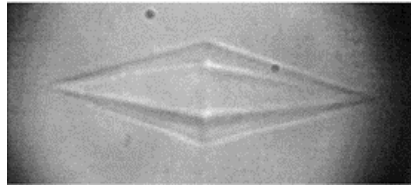
Despite the structural diversity, the mechanism of depressing the freezing point is assumed to be similar for all AFPs. AFPs possess the ability to bind specifically and selectively to the ice crystal's surface. This process is quasi-irreversible, which means that the AFPs do not desorb from the ice crystal surface once bound. The selective and characteristic binding to the ice crystal is enabled by a steric complementarity between the protein's and ice crystal's surface. To explain this ability, one has to look at the structure of water in its crystalline form.



**Figure 1.3.:** Different ice crystal surfaces, which AFP can interact with. Yellow represents a fraction of the basal plane (0001), red the primary prism plane ( $10\bar{1}0$ ), green the secondary prism plane ( $11\bar{2}0$ ) and blue the pyramidal plane ( $20\bar{2}1$ ).

Ice crystallizes under normal pressure as ice  $I_h$  in a hexagonal crystal form. In this form, a water molecule serves as proton donor and acceptor (two acceptors and two donors). Each water is surrounded by four water molecules in a tetrahedral arrangement. This results in a conventional hexagonal crystal cell described by a hexagonal axis system using three a-axis in a plane and a vertical c-axis. The crystal structure is described with the BRAVAIS-MILLER indices, which results i.e. for the basal plane in an index of 0001, as it is orthogonal to the three a-axis and intersects with the vertical c-axis. In the hexagonal ice crystal, several planes are present for an interaction with AFPs (colored planes, Figure 1.3). The unique ability of AFPs to bind to the ice crystal planes leads to a kinetic inhibition of the crystal's growth rate of the covered plane. Depending on the plane the AFP is bound to, the shape of the crystal is macroscopically changed.

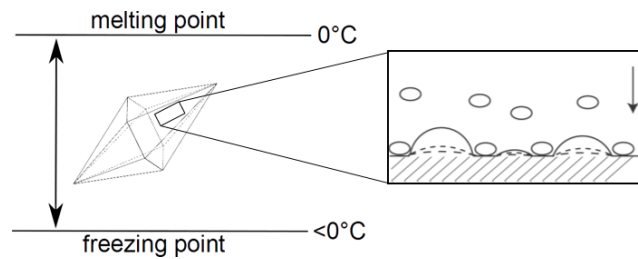




**Figure 1.4.:** The shape of an ice crystal in the presence of AFP Type 1. The AFP Type 1 binds to the pyramidal plane ( $20\bar{2}1$ ), resulting in an inhibition of the ice crystal's growth at these planes. The ice crystal has therefore a hexagonal bipyramidal geometry.

The first discovered AFGP binds to the six primary prism plane ( $10\bar{1}0$ , colored red).<sup>22</sup> This inhibits the growth of ice primarily of these planes to be the slowest. The macroscopic shape of the ice crystal is then that of a hexagonal cylinder. The large,  $\beta$ -helical AFPs (i.e. MpAFP) are able to bind in addition to the basal plane ( $0001$ ), resulting in crystals with hexagonal symmetries. The ability to bind to multiple surfaces (foremost the basal plane) results in a significant increase in their activity (hyperactive AFP), as the growth rate in the  $c$ -direction of the crystal is also inhibited. The AFP Type 1 has a surface which matches the pyramidal plane ( $20\bar{2}1$ ) of the ice crystal. This leads to an hexagonal bipyramidal shaped crystal geometry (Figure 1.4). Overall, the macroscopic shape of the ice crystal in the presence of AFP is therefore an ideal and simple measure to assess the activity of AFPs.

AFP also possess the ability to inhibit the recrystallization of ice crystals. In a system of constant temperature and pressure, larger ice crystals can grow at the expense of the smaller crystals. This effect is called OSTWALD ripening and describes the tendency of the system to optimize its energetic state. The small ice crystals exhibit a larger surface area to volume ratio and a higher curvature compared to large ice crystals. Depending on the curvature of the crystal's surface, the freezing point of the crystals vary. A large curvature leads to a locally increased freezing point, whereas a small curvature decreases the local freezing point relative to the planar ice-water interface that determines the global freezing point (KELVIN effect). Thus, in the re-crystallisation ripening small ice crystals melt while the larger crystals grow. In the presence of AFP, the crystals' growth rate is depressed, which stabilizes ice crystals and inhibit the ice crystal recrystallization.



**Figure 1.5.:** Right: Schematic view of the ice binding of AFP Type 1 to the  $20\bar{2}1$  ice surface. The proteins (sketched as ellipse) bind to the ice surface. The ice crystal can grow only in the free space not covered by the proteins (curved line). This resulting in a hexagonal bipyramidal crystal structure.

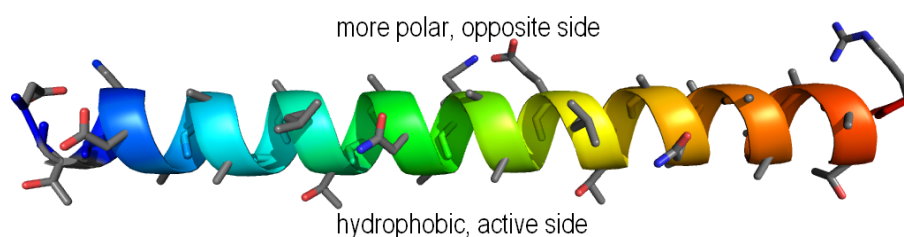
The same principle is the physical basis for the best-known activity of AFPs, the freezing point depression of its solutions.<sup>23,24</sup> Ice crystals slightly below the (colligative) freezing point in the presence of AFP are covered by the proteins (Figure 1.5). However, the protein covers only portions of the ice crystal's surface and the crystal can grow in-between the proteins. This leads to a curved interface of ice and water. At this interface, the freezing point is lowered, as it is thermodynamically disfavoured for water to adsorb to a curved ice plane rather a plain. If the local freezing point reaches the temperature of the solution, the crystal growth will stop. The same effect but in opposite direction is observed at temperatures above the freezing point. The protein is still bound to the ice crystal, thus the ice can melt only in the free space between the proteins. Melting leads to an inverse curvature and the increase of the ice crystal's melting point. The difference between the melting point and freezing point of the ice crystal is called thermal hysteresis (TH) and is used as the most distinctive characterization the activity of AFPs. The separation of the freezing and melting point of a solution is in contrast to the properties of a solution containing a solute, i.e. salt. In such systems, the freezing point and melting point are equal and the variation of the freezing point depends on the concentration of the solute and not on the type. Thus, the freezing point depression is one of the three known colligative properties. In contrast, solutions containing AFPs have separated freezing and melting points, which is a reflection of a kinetic growth inhibition keeping ice and solution from reaching the thermodynamic equilibrium. This is a meta stable state, and thus a non-colligative effect. Macroscopically, this effect is easily observed (under a microscope). At temperatures below the non-colligative freezing point, AFPs cannot inhibit the growth and the solution freezes completely. At temperatures above the melting point, any formed ice crystal will melt. Crystals however, will be stable at temperatures between the separated freezing and melting points. Ice crystals will neither grow nor melt in this thermal hysteresis window, they are meta stable.

### 1.1.1. Antifreeze Protein Type 1

The winter flounder *PSEUDOPLEURONECTES AMERICANUS* exhibits several different AFPs in its bodily fluids.<sup>25,26</sup> The best studied one is the 37 amino acid long  $\alpha$ -helical AFP Type 1 isoform HPLC6 (Table 1.2 and Figure 1.6). This AFP has served as model system for many years, due to its simple structure. The AFP Type 1 HPLC6 is able to bind to the pyramidal ice surface and inhibits the growth of the ice crystal completely. This results in a macroscopically crystal morphology of a hexagonal bipyramid with a characteristic c- to a-axis ratio of 2.2 : 1 and a TH of 1 °C.<sup>26</sup>

**Table 1.2.:** Sequence in one letter code of AFP Type 1 HPLC6.

<b>AFP Type 1 HPLC6</b>	D - TASDAAAAAAL - TAANAKAAAEL - TAANAAAAAAA - TAR-NH <sub>2</sub>
-------------------------	---



**Figure 1.6.:** Structure of AFP Type 1 HPLC6. The secondary structure is represented in ribbon cartoon style and colored in rainbow from N-(blue) to C-terminus (red). Amino acid side chains are represented in the stick style. Protein Data Bank (PDB) accession code: 1WFA.<sup>11</sup>

The AFP Type 1 HPLC6 is an alanine rich protein with a triplicated eleven residue motif: **Thr-Ala-Ala-X-Ala-X-X-Ala-Ala-X-X**. While X-positions can be any amino acid, these are most frequently also alanines.<sup>27</sup> The protein folds into a single, nearly perfectly straight  $\alpha$ -helix. This helix has only marginal stability and limited folding cooperativity, as the protein is 100% folded only near the freezing points, and almost linearly unfold with increase in temperature. In the folded conformation, the Thr residue and the Ala residues with a separation of  $i, i+4$  (with  $i$  the residue counting from) and  $i, i+8$  respective to Thr of the repetitive motif form on one side of the helix a contiguous hydrophobic surface, the active side of the protein<sup>28</sup> (Figure 1.6). The distance of the Thr residues match the spacing between oxygen atoms on the ice surface. This results in a preferred binding to the  $20\bar{2}1$  ice plane in  $11\bar{0}2$  direction, which has been experimentally determined and established.<sup>28,29</sup> The opposite side of the AFP is more hydrophilic and supports the protein's solubility. The sequence is terminated N- and C-terminally with an Asp and Arg, respectively, that serve as capping residues. At physiological conditions, their charged side chains fold back towards the backbone and

reduce an unfavoured interaction between the terminal charge and the dipole moment of the helix, which further stabilizes the folded conformation.<sup>27,30</sup>

As the AFP Type 1 HPLC6 exhibits a repetitive motif, studies to determine the influence of the length and helicity of the protein's activity have been performed. In these studies the sequence was shortened stepwise simultaneously from the N- and C-terminus.<sup>30</sup> It was shown that after shortening the sequence by three amino acids, the fold stability and activity of the truncated protein was similar to the full length protein. By truncation from 37 to 32 amino acids however, the helicity is decreased below 35%. This yields in a significant decrease of the TH by 60% to 0.4 °C. Further truncation to 28 amino acids (75% of the original number of residues) results in a helicity of 26.1%. For this protein, no TH was observed, but notably an ice shaping effect was observed, as a macroscopic hexagonal crystal morphology was formed. This length was taken to be the minimal required sequence length for the activity of AFP Type 1 HPLC6.

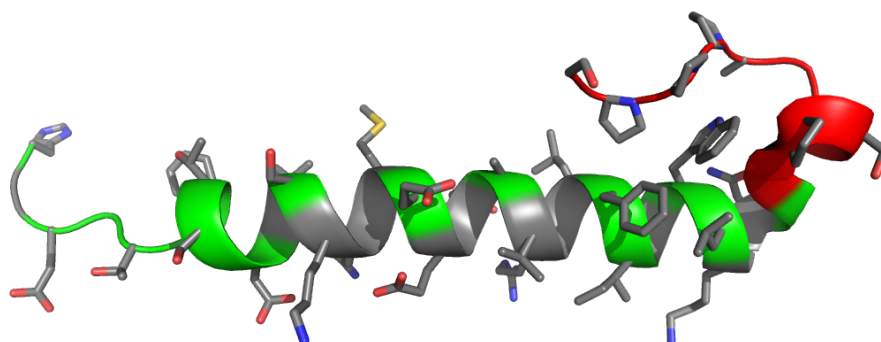
In a second approach, the activity of one eleven amino acids long ice binding motif was investigated.<sup>31</sup> In order to stabilize the structure, the N- and C-terminal capping residues were included (DTASD and TAR, respectively). Furthermore, a salt bridge was introduced for increased stability.<sup>32</sup> The orientation of the salt bridge was altered in favour of the helical dipole moment, resulting in two 15 amino acid long peptides (15KE and 15EK). Both peptides exhibit already a helical content of 62% and 58%, but a complete abolishing of TH and ice binding properties, as round, disk-shaped ice crystals were observed. In order to stabilize the structure of the peptide, the salt bridge of 15EK was replaced by a lactam-bridge forming a covalent bond (15EKlac). This significantly increased the helicity to 90%. The lactam-bridged 15EKlac displayed significant activity. Upon freezing, the crystal forms a hexagonal bipyramid shape. These results show that the activity is mainly structure and not length dependent.

## 1.2. The Trp-cage

From the oral secretion of the Gila monster (*HELODERMA SUSPECTUM*) the Exendin-4, a protein with a 50% sequence similarity to the helical glucagon-like-peptide-1 (GLP-1), was isolated.<sup>33–35</sup> Both proteins not only have sequential but also an activity similarity, i.e. binding to the GLP-1 receptor and therefore regulating insulin production (anti-diabetic actions).<sup>36</sup>

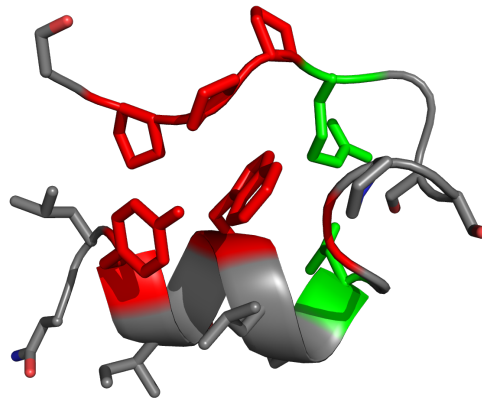
**Table 1.3.:** Sequence comparison of GLP-1 and Exendin-4 represented in one letter code. Similar amino acids are marked in bold and the binding side is underlined.

<b>GLP-1</b>	HDEFERHA <b>EGTFTSDVSSYLEGQA</b> <u>AKEFI</u> AWLVKGR-NH <sub>2</sub>
<b>Exendin-4</b>	<b>HGEGTFTSDLSKQMEEEA</b> <u>VRLFIEWL</u> KNGPSSGAPPPS-NH <sub>2</sub>



**Figure 1.7.:** Structure of the first Exendin-4 nuclear magnetic resonance (NMR)-structure (PDB: 1JRJ<sup>35</sup>). The secondary structure is represented in cartoon style and the amino acid side chains in stick style. Sequence similarities with GLP-1 are marked in green and C-terminal extension of Exendin-4 in red.

Comparison of the sequences reveals that Exendin-4 is extended at the C-terminus by ten and shortened at the N-terminus by six amino acids (Table 1.3). In contrast to the solely helical structure of GLP-1, the Exendin-4 structure exhibits a C-terminal extension consisting of a turn and a polyproline-II helix, resulting in a tertiary structured protein (Figure 1.7). The C-terminal extension of Exendin-4 encapsulates the present Trp side chain in the helix. This leads to a very stably folded C-terminus and stabilization of the adjacent helix, which interacts with the GLP-1 receptor.<sup>37</sup> The N-terminus itself shows significant fraying, which may be necessary for insertion through the membrane. Those structural features allow the Exendin-4 not only a longer half-life time in vitro, but also enable better permeation into the cell and a stronger binding to the GLP-receptor compared to GLP-1.<sup>38</sup> Therefore, it is commercially used as a therapeutic agent for diabetes type 2.



**Figure 1.8.:** Model of the Trp-cage NMR structures of the sequence Tc5b with the lowest energy (first model) (PDB accession code: 1L2Y<sup>39</sup>). The secondary structure is sketched in ribbon cartoon style for the protein backbone and stick style for amino acid side chains. The important amino acids for the cage fold are colored in red. The salt bridge between Asp at position 9 and Arg at position 16 is colored in green.

The significant increase in fold stability of the Exendin-4 motivated the group of J.W. NEIDIGH to look into the C-terminal folding motif. Based on the Exendin-4, they designed a 20 amino acid long peptide by N-terminal truncation, the Trp-cage construct Tc5b with the sequence NLYIQWLKDGGGRPPPS<sup>39</sup> (Figure 1.8). The resulting mini-protein is strikingly stably folded, but does not have an intrinsic function.<sup>40</sup>

In the miniprotein, the eponymous Trp side chain forms the inner core of a hydrophobic cage including the polyproline-II-helix, the turn and the Tyr (marked in red Figure 1.8). Furthermore, a salt bridge is present, stabilizing the tertiary arrangement between the polyproline-II helix and the  $\alpha$ -helix (marked in green).<sup>41,42</sup>

In a second round of sequence optimization, the different structural parts of the protein were optimized.<sup>43</sup> This study focused mainly on replacements in the helix. They demonstrated that the most important residues are the Trp (position 6) and Tyr (position 3). Tyr is necessary for shielding, but can be replaced by Leu and Phe with the expense of stability. All other residues in the helix can be replaced with limited destabilizing effect, but some even increase the protein's stability.

Studies with nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy showed that depending on the replacements in the sequence, the helical fold and cage fold can be separated and follow no stringent cooperative folding mechanism. If the thermal melting point ( $T_M$ ) differs more than 7 °C between the  $T_M$  of helix and cage, the folding is not cooperative, usually indicating an intrinsically stabilized helix and an intermediate between the folded and unfolded state of the Trp-cage.

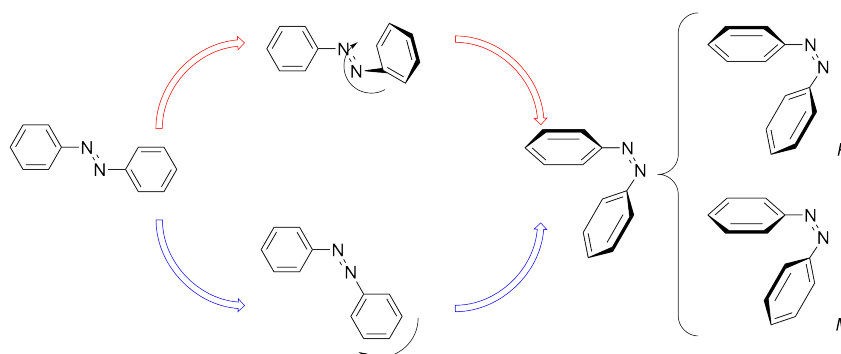
In addition to the optimization of the 20 amino acids, studies of N-terminal extensions

of the Trp-cage were performed.<sup>44</sup> An N-terminal insertion of six Ala leads to a stabilization of the cage fold and notably, to a propagation of the N-terminal helical fold to include the inserted Ala. In conclusion, the Trp-cage cannot only be modified within its N-terminal helix, but also the existing helical fold propagates into N-terminal extensions. Since the first publication, the Trp-cage protein has been used for molecular dynamics (MD) simulations.<sup>45</sup> Due to its unique and stable fold, it serves as a model system for calculations aimed at optimizing force field parameters for protein simulations and in calculations for assessing fold properties and folding mechanism pathways.<sup>46–49</sup>

### 1.3. Switching at the Molecular Level

Controlled motion or switching of properties can be found for many objects of daily use, ranging from a simple light switch up to complex machines. Transferring the principle of switching to a molecular level leads to the family of molecular switches. For a molecular switch, a molecule has to fulfill several characteristics. First, a switch requires two different electronic states, which differ in their properties, for example in their conformation or configuration.<sup>50</sup> These states have to be convertible into each other. Furthermore, a switch has to respond to external stimuli (light, electricity, magnetism, mechanical force), whereas the two states should not respond to the exact same stimulus. Referring to light responsive molecular switches, the excitation wavelengths of each conformer have to be separated. Another characteristic is the life-time of the conformers; optimally both should be thermodynamically stable.

Three exemplary molecular scaffolds, which fulfill these characteristics, are azobenzene and spiropyran.<sup>51,52</sup> Both systems respond to light as external stimulus and exhibit two conformations, which differ in their length among others. An advantage of these systems are the easy accessible chemical modification and therefore possible tuning of the photochromic properties. Foremost, this allows the covalent incorporation in multiple materials. Since azobenzene was mainly used in this work, its characteristics are discussed in more detail.



**Scheme 1.2.:** Mechanism of the *trans* to *cis* isomerization of azobenzene. Top: rotational mechanism around the "activated" N=N bond; Bottom: mechanism of inversion at N-atom. Right: *P* and *M* enantiomer of the *cis* azobenzene.

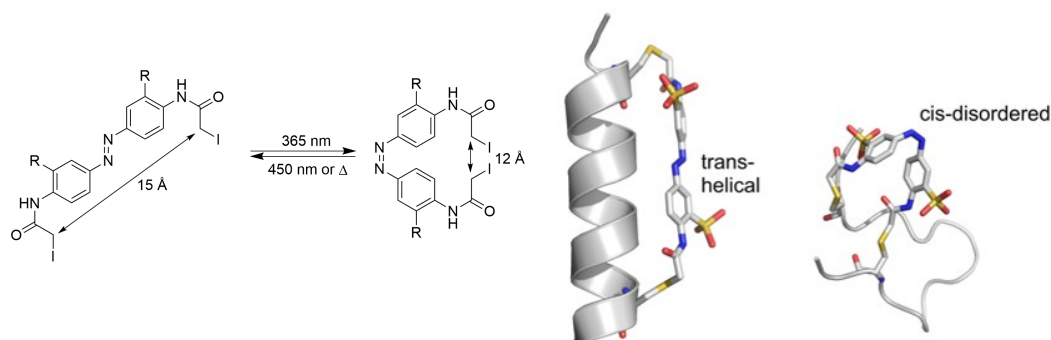
Azobenzene was discovered in the year 1937<sup>53</sup> and ever since has been used due to its favourable thermal stability<sup>54</sup> and photochromic properties, i.e. nearly no photo bleaching.<sup>55</sup> Upon irradiation with ultra violet (UV) light at  $\lambda = 365\text{ nm}$ , the molecule can isomerize from the thermodynamic stable *trans* to the meta-stable *cis* form. From the *cis* configuration, azobenzene can either thermodynamically relax back or can be converted by irradiation with  $\lambda = 440\text{ nm}$  to the *trans* state.<sup>56</sup> Although the mechanism is still discussed in the literature, two possibilities are commonly accepted for both photochemical conversions: an out-of-the-plane rotational mechanism<sup>57</sup> around the activated azo group and an in-plane inversion mechanism at the nitrogen (Scheme 1.2).<sup>58</sup> The switching process from the *trans* to *cis* configuration is bi-directional. Without constraints, both possible *cis* enantiomers (*P* or *M*) are energetically equal. However, by incorporating azobenzene in a chiral cyclic compound, the switching process may become uni-directional upon irradiation and one enantiomer is favoured.<sup>59</sup> This results in an additional helicoidal chirality of these compounds.

These photochromic properties of the azobenzene lead to a high potential in photo-mechanical applications. Incorporated in materials, the different properties of the azobenzene can be used. In particular, the controlled motion is of great interest. Combined with a liquid crystalline polymer, the characteristics of the bulk material can be altered upon switching. Furthermore, a macroscopic effect in shape and length can be observed. This can be used for artificial muscles or 3D motors.<sup>60</sup>



### 1.3.1. Switching of Biopolymers

A special application of molecular switching is the structural control of proteins. With the manipulation of structure, the related activity can be regulated. For that purpose, a switchable molecule can be introduced in the biopolymers. A currently employed strategy is the incorporation of bi-functionalized azobenzene on the amino acid's side chains.<sup>61</sup> A major advantage of this strategy is the easy transfer to larger systems, i.e. to the  $\sim 24$  kDa big trans membrane protein Cadherin controlling the ability of  $Ca^{2+}$  dependent adhesion<sup>62</sup> or the photo control of the 60 kDa SHAKER K ion channel for potassium.<sup>63</sup>



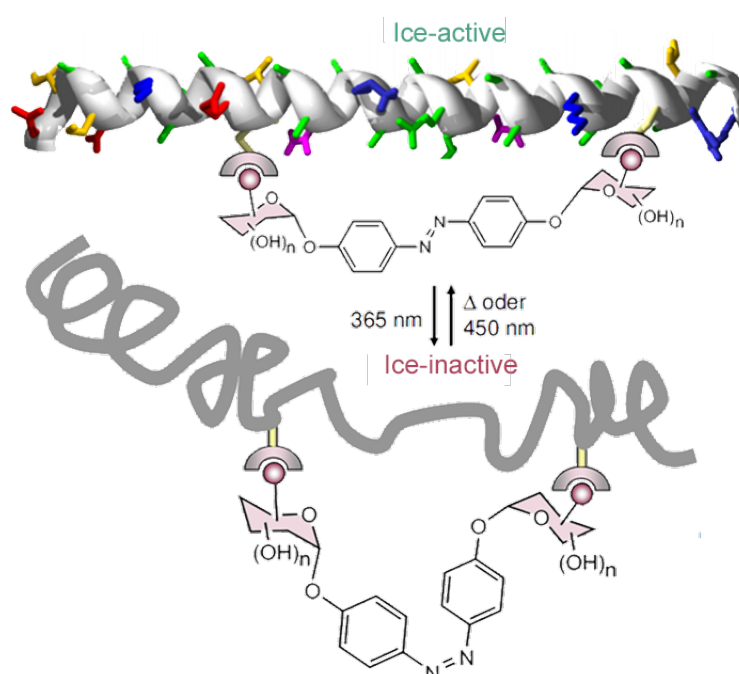
**Figure 1.9.:** Concept of G. A. WOOLLEY of azobenzene cross-linked to side chains. Left: structure of iodoacetamide functionalized azobenzene derivative. Effective length difference of connection points is  $\sim 3$  Å. Right: Peptide linked with iodoacetamide azobenzene via Cys side chains. With azobenzene in the *trans* configuration, it stabilizes the secondary structure and in the *cis* configuration it destabilizes the  $\alpha$ -helix. Reprint with permission by the "Royal Society of Chemistry".<sup>64</sup>

Depending on the employed concept, the azobenzene can be incorporated in a manner that the distance of the connection points matches the length of the *trans* or *cis* isomer. In this state of the linker, the fold is induced and stabilized. The counterpart does not match the distance (either too long or short) and therefore destabilizes the secondary structure. This concept is reported for numerous  $\alpha$ -helical peptides (Figure 1.9).<sup>64–66</sup> With different substitution patterns of the azobenzene, the life-time of the *cis* isomer can be tailored to various applications, and thus the life-time of the fold. Additional advantage of a switchable protein is the ability to gain new insights in the folding process and manipulation of the folding dynamics.<sup>65</sup> This allows not only being able to switch between a folded and unfolded or disrupted state, but also between two functional states.



## 2. Motivation and Previous Work

The overall objective of this work is the transfer of the WOOLLEY-type photo-switching of protein conformation to a unique functional protein, the  $\alpha$ -helical AFP Type 1. Photo-control of the AFP structure should control its activity. While one can envision a number of possible new applications of an AFP with switchable function, the ability to modify the protein's activity with high spatial and temporal resolution will be a potent tool to study the still controversial mechanism of binding of AFP to the ice surface.



**Figure 2.1.:** Concept of a photo-switchable AFP Type 1 by H. KOBARG. The AFP Type 1 HPLC6 is cross-linked with a glyco-azobenzene (in collaboration with A. MUELLER). With azobenzene in *trans* configuration, the connection points match the distance of a helical fold (top) of the AFP. Thus, the protein is active and can bind to the ice surface. In the *cis* configuration of the linker, the structure of the AFP is disrupted. This leads to reduction its ice binding ability (bottom).

H. KOBARG, in collaboration with A. MUELLER (Lindhorst group), started this approach in our group.<sup>67</sup> The original WOOLLEY-type cross-linker comprises an azobenzene moiety as photochrome and iodoacetamides as reactive groups. (Figure 2.1). In H. KOBARG's design, additional carbohydrates are inserted for improved solubility of the linker. This modification consequently increases the length of the linker, matching the spacing between the residues  $i, i + 21$  in contrast to the  $i, i + 7$  spacing for the original WOOLLEY-linker. In the AFP sequence (Type 1 HPLC6), Cys residues were incorpo-

rated in appropriate positions. The exact positions were chosen so that the linker does not sterically interfere with the protein's active site, i.e. the linker is placed opposite to the ice binding side in the folded protein. With the azobenzene linker in the *trans* configuration, the AFP helix can form and possibly may even be stabilized by the linkage and should enable interaction of the protein with the ice surface. Upon irradiation, the azobenzene linker isomerizes to its *cis*-configuration leading to a shortening of the maximal linker length. Therefore the fold should be disrupted as the reduced linker length is incompatible with 21 amino acids in six helical turns, resulting in a reduction or abolishment of the AFP activity.

The implementation of the approach turned out to be very challenging in several aspects. First, the synthesis of the AFP by Fluorenylmethyloxycarbonyl (Fmoc) based solid phase peptide synthesis (SPPS) was difficult. The AFP sequence comprises a large number of non-polar Ala residues and several long poly-alanine stretches occur. The achieved synthesis yield for the wild type AFP (Type 1 HPLC6) was below 20% (H. KOBARG, unpublished). Incorporating Cys residues in the chosen positions for the cross-linking with the azobenzene decreased the yield of the respective protein even further. In addition, the purification of these crude proteins was difficult due to very close retention times of product and by-product in reverse-phase (RP) high pressure liquid chromatography (HPLC).

The cross-linking reaction at the thiol function of the cysteine residue created additional problems. The reaction requires a reductant to prevent cysteine oxidation and disulfides formation. Unfortunately, the chosen reducing agent tris-(2-carboxyethyl)-phosphin (TCEP) reacted with the cross-linker. In addition, the cross-linking reaction itself was not selective. Both complications led to various side products that rendered the product purification impossible.<sup>67</sup>

An additional unfavourable observation is that the obtained cross-linked AFP products had unexpectedly low solubility in water. Once dissolved, solutions were not stable and precipitations were observed over time. Possible reasons for this behavior could be a combination of the hydrophobicities of protein sequence and cross-linker, as well impurities of the cross-linking reaction.

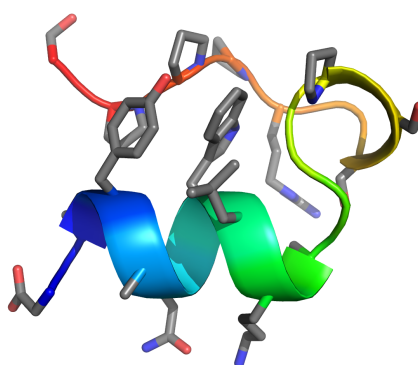
Owing to the large number and severity complications encountered, this approach to generate switchable AFPs was not pursued any further. Principally, the problems in the Cys cross-linking reaction appeared solvable (for example by choosing a different reducing agent, conducting of the experiment under inert atmosphere and at high dilution of both reactants). Similarly, the hydrophilicity of the linker can be further increased (albeit attempts to incorporate the sulfonic-acid variant of the WOOLLEY linker as published failed<sup>68</sup>). However, incorporation of further carbohydrate units yield linkers with intrinsic flexibility and dimensions which are incompatible with the requirements for effective switching and the length of the folded AFPs and are thus not applicable to this example.

These optimizations require a great deal of work and possibly needed to be repeated for every AFPs. Not only are these efforts individually significant, the results are expected to depend on each other, and foremost on the sequence of the specific AFP used. Thus, to a large extent, these optimizations would have to be repeated for each AFP peptide linker combination. But foremost, these efforts would not solve the two other encountered significant problems, the high hydrophobicity/low solubility of the AFP that is a main factor for the observed aggregation, and the synthetic difficulties to produce pure AFP by SPPS in a sufficient yield. These are intrinsic to the AFP Type 1 sequence, and cannot be changed in this sequence without affecting the protein's activity. This realization resulted in the generation of a completely new approach, in which the AFP is coupled with a second protein unit that is easy to synthesize, has a high intrinsic solubility, and can be modified to be a switchable module. Fusion protein approaches are widely known in biochemistry, e.g. green fluorescent protein (GFP) to render proteins' fluorescence detectable,<sup>69</sup> fusion proteins to enhance recombinant bacterial expression (small ubiquitin-like modifier (SUMO),<sup>70</sup> thioredoxin<sup>71</sup>). The challenge in our idea and application, however, is to couple the structure of the AFP to the structural state of the fusion partner, if one wishes to switch the AFP activity by affecting the fold of the fusion partner. Our search for solutions to this problem lead to the identification of the Trp-Cage mini-protein as the promising fusion protein.



### 3. Concept

In the following work, a new concept for stabilizing and switching  $\alpha$ -helical peptides is presented. The concept follows a modular approach. In this approach, the functional sequence, the  $\alpha$ -helical protein, is not directly modified to achieve photo-control. Rather, a protein module is attached, which comprises the light responsive unit. Provided that the structural state of the protein module can be coupled with that of the  $\alpha$ -helical protein, this approach has the advantage that the switchable module needs to be developed (designed and optimized) only once. Ideally, the sequence of the functional  $\alpha$ -helical proteins is not or only minimally altered, thus the approach is sequence independent and can be applied to any helical peptide with a function of interest simply by attaching the switching module. In other words, the modular approach separates function and switching, and thus facilitates their independent optimization.

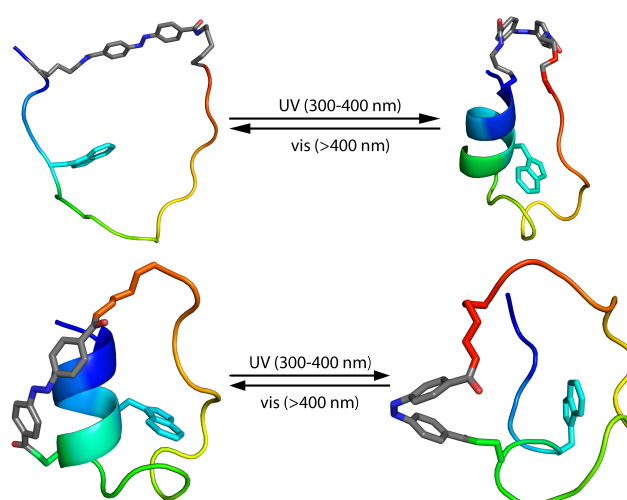


**Figure 3.1.:** Model of the Trp-cage variant Tc10b. The secondary structure is presented in cartoon style and colored in rainbow, starting at the N-terminus in blue and ending at the C-terminus in red. The side chains are presented in stick style in colored according to the element (carbon in grey, nitrogen in blue and oxygen in red).

The Trp-cage mini-protein fulfills the demands on such a module (Figure 3.1). Its tertiary structure consists of an eight residue N-terminal helix, which is induced by the encapsulation of the eponymous Trp side chain by the C-terminal loop. Owing to the tertiary structure, the Trp-cage exhibits an unusually high stability and is the smallest folded protein so far. The Trp-cage is known to be highly soluble. It has an unbiased sequence, no extended hydrophobic sequence stretches and is known to be easily synthesized by SPPS. The protein does not have an intrinsic function, and thus can be converted into an otherwise functionless switching module.

### 3.1. The Switch-Cage Module

The switching of the structure of the Trp-cage can be implemented in several ways. The underlying strategy is the disturbance of the Trp-cage folding nucleus with an associated solvent exposure of the Trp side chain. The approach should lead to an unfolding of the protein, including the N-terminal  $\alpha$ -helix. The dimensions of the Trp-cage in relation to those of the linker allows several anchor points for the linker to fulfill the strategy. It is possible to attach the  $i, i+7$  WOOLLEY-linker to the helix to switch the secondary structure. However, switching of the tertiary structure is possible. From the dimensions of the Trp-cage, anchor points in the C-terminal Trp-cage loop are also possible, however, most residues in this sequence are essential for the fold and cannot be altered. Nevertheless, the C-terminal residue (position 20) is ideally suited as attachment point, since it is known to accommodate other residues without effect on the fold and, foremost, is in a crucial position in the tertiary structure. With the second attachment point either at the beginning or the end of the N-terminal helix, two different implementations of the strategy are obtained, in which a maximum geometric change on the Pro-Pro-Pro (position 17-19) sequence in the C-terminal loop is exerted.



**Figure 3.2.:** Two different concepts for a switchable Trp-cage. Top: an azobenzene at a push position between the N- and C-terminus; Bottom: an azobenzene at a pull position between the C-terminus and the C-terminal end of the helix.

In the approach called "push-approach", the Trp-cage fold is disturbed by the azobenzene in the *trans* configuration, as the linker length does not match the distance between the anchor residues (Figure 3.2). In the opposite "pull approach", the azobenzene stabilizes the fold in its *trans* configuration, as the linker length matches with the distance between the anchor residues.

Both concept have the advantage that only one amino acid in the helix is needed as



anchor residues, where again residue replacements are well tolerated. The pull-approach has been implemented and studied by N. PREUSSKE during his Masters-thesis, while the push approach has been pursued in this study.

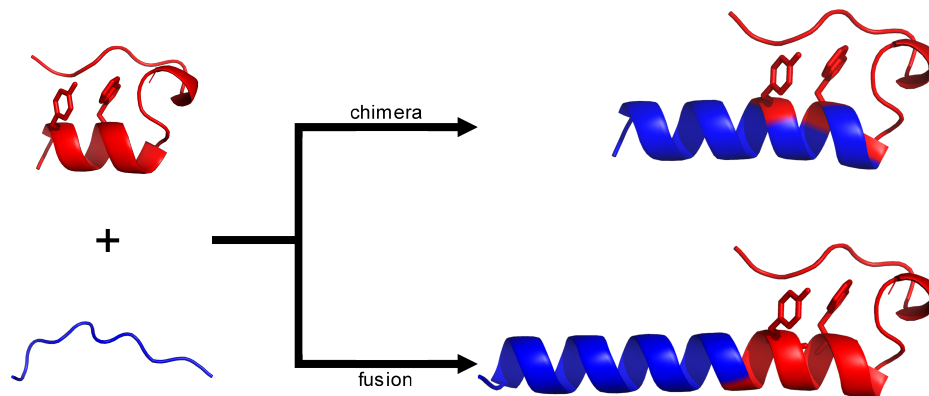
### 3.2. Concepts for Fusion

In a modular approach, two proteins are joined leading to a fusion protein. The fusion concept is well known and utilized in particular with recombinant technology to enhance bacterial expression. In this concept, spacer residues are added between the proteins modules to ensure that the protein domains behave independently and do not affect their fold and activities. This is achieved by adding spacers either sequentially in the primary sequence or *via* side chains. Thus, the fusion protein exhibits independent folding domains and activity.

In our approach, however, we also wish to achieve synergistic properties in form of a structural correlation or dependencies of the fold between the two joined modules. We aim to use the properties of the Trp-cage (stability and solubility) and the biological activity of peptides with  $\alpha$ -helical propensity, resulting in a fusion protein with a stabilized fold of the biological active peptide.

The fold of  $\alpha$ -helices is known to be cooperative.<sup>72</sup> Thus, by adding residues to an existing helical structure, hydrogen bonds between the peptide groups  $i, i + 4$  are established, leading to a propagation of the existing helix conformation to the added residue backbone conformation. As existing helix or helix nucleation point, the Trp-cage module can be used. The Trp-cage exhibits in its tertiary structure a remarkably stable, non-fraying N-terminal helix. Previous studies with the Trp-cage protein showed that the  $\alpha$ -helical fold, induced by the Trp-cage, is propagated to six, N-terminally added Ala residues leading to a stable  $\alpha$ -helix.<sup>44</sup>

The junction between the Trp-cage module and the biological active peptide with  $\alpha$ -helical propensity can be accomplished in two different strategies.



**Figure 3.3.:** Concept for combining the Trp-cage with a biological active peptide sequence in two alternative approaches: a) chimera (top) and b) direct fusion strategy (bottom). The secondary structure is represented in cartoon ribbon style. The Trp-cage residues are marked in red and the residues from the biological peptide in blue.

The simplest strategy is the N-terminal elongation of the Trp-cage sequence with the peptide (Figure 3.3), thus a direct fusion. On the one hand, the Trp-cage fold should not be influenced by this extension and able to fold in its tertiary structure. On the other hand, the Trp-cage helix is stable and provides a structural nucleation point, thus this helix is expected to be cooperatively propagated into the sequence of the fused peptide. This should lead to an induced and stabilized helical fold of the biological active peptide. The advantage of this strategy is that neither the peptide nor the Trp-cage sequence need any further modifications. In addition, spacer residues between the Trp-cage and the peptide can be inserted to control the orientation between the Trp-cage's tertiary structure and the peptide's helix. This method is suitable for all peptides with helical propensity.

The second strategy is to create a chimeric fusion protein. The idea is to impart properties from each protein into the resulting chimera. In contrast to the direct fusion strategy, the sequence of the biological active peptide and the N-terminal helix of the Trp-cage coincide, resulting in a reduction of the overall peptide length. To realize this strategy, one needs to carefully consider for each residue in the chimera region, which amino acid from which part of the sequence has to be included. From the Trp-cage perspective the two aromatic residues of the Trp-cage Tyr and Trp are needed in the appropriate positions in the helix. The other positions in the helix can either be the original Trp-cage sequence or the biological active peptide, depending whether the position is important for structure or activity of the AFP. At the C-terminal end of the chimera protein, the Trp-cage fold is completed by addition of 12 residues (the cage loop, including the polyproline-II-helix). The resulting chimera protein exhibits the tertiary structure of the Trp-cage with an elongated helix. Thus, the protein can be divided in three parts,

the N-terminal extension (biological active peptide sequence), chimera helix (Trp-cage and biological active peptide sequence) and the cage loop (Trp-cage sequence). The advantage of this strategy is that the overall protein length can be significantly reduced. However, the relative orientation of the peptides' active side to the tertiary fold of Trp-cage has to be considered carefully, as both should not interfere with the each other. This method is suitable for all peptides, which activity is located on a surface side of the protein helix.

Hereinafter, the results for the Switch-cage module and fusion are presented separately in two parts. In the first part, the Trp-cage fold is discussed in detail and the results for introducing the ability of switching the Trp-cage are presented and discussed. In the second part, the fusion of the Trp-cage with the AFP Type 1 HPLC6 sequence is presented and discussed.



**Part I.**

**The Switch-Cage Module**



## 1. The Trp-Cage Fold

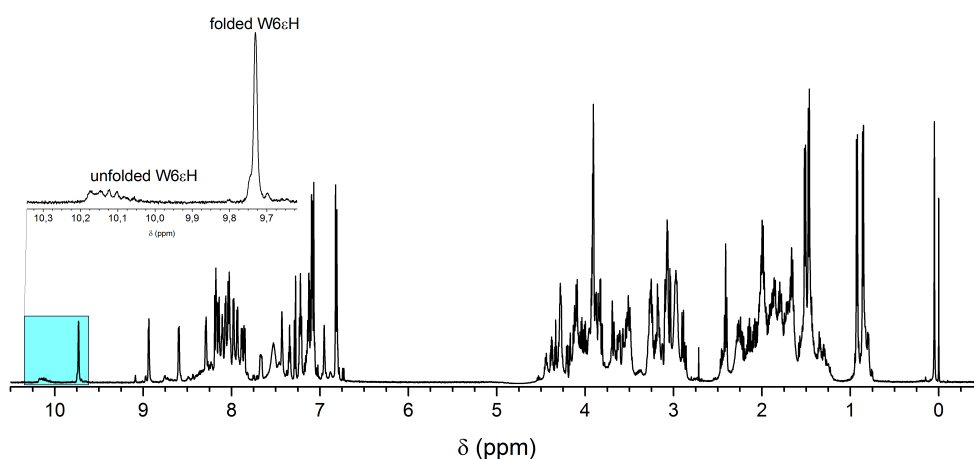
The Trp-cage exhibits a unique and unusually stable fold for its small size. Since its identification, numerous studies have aimed to optimize the fold stability, analyzing almost every amino acid in the sequence for its effect.<sup>39,41,43,73,74</sup> Thus, many variants (> 40) have been described and are well characterized.

**Table I.1.1.:** Numbered Tc10b sequence in one letter code.

Numbering	1	6	11	16
Tc10b	DAYAQ	WLKDG	GPSSG	RPPPS

We selected the Trp-cage variant Tc10b (Table I.1.1) as starting molecule and reference. Tc10b is a very stable variant with a thermal unfolding mid-point (melting point,  $T_M$ ) of  $56^\circ\text{C}$  at pH 7 and  $T_M = 41^\circ\text{C}$  at pH 2.5. The pH dependency of the fold stability originates in the presence of a salt bridge between residue 9 and 16 at neutral pH. According to literature, Tc10b folds cooperatively and is 99.5% folded at 280 K and pH 7.<sup>43</sup>

Tc10b was synthesized by Fmoc solid phase peptide synthesis and purified by HPLC chromatography. The protein was characterized by NMR spectroscopy.



**Figure I.1.1.:**  $^1\text{H}$ -NMR spectrum of Tc10b at 298 K and pH 3 with enlargement of the indole proton region.

The  $^1\text{H}$  NMR spectrum at pH 3 and 298 K is characteristic for the presence of a folded peptide and is typical for a Trp-cage. The eponymous Trp side chain at the center of the structure is the driving force for folding. Its well separated indole proton W6

at the downfield edge of the spectrum is a strong indicator for structure. The signal of W6H $\epsilon$  is known to shift upfield. Here, W6H $\epsilon$  is observed at a chemical shift ( $\delta$ ) of 9.75 ppm, indeed significantly shifted to high field compared to the random coil shift of 10.22 ppm.<sup>75</sup> A second good indicator for the presence of a folded protein is the chemical shift dispersion of amide and aliphatic resonances. While this not obvious from the shown <sup>1</sup>H spectrum, NMR-assignments using 2D-MR experiments determine structure based differences between resonance of amino acids of the same type. Notably, all resonance frequencies match well the reported values for Tc10b. A closer look at the indole proton region in the <sup>1</sup>H NMR spectrum of Tc10b (inset, Figure 1.1.1) however reveals a general and severe problem. Besides the mentioned upfield shifted signal of W6H $\epsilon$  of the folded Trp-cage, multiple additional and partially broad W6H $\epsilon$  signals above 10 ppm are observed. The intensity ratio of all the resonances relative to the folded signal is 1 : 3. As these resonances originate from multiple partially or complete unfolded Trp-cage sequences, a significant proportion of the protein ( $\approx 25\%$ ) is incorrectly folded. Therefore, it was very surprising that these resonances have not been presented and discussed in detail in the literature on the Trp-cage before. Only exception are two studies that briefly mention the observation of a “second form”.<sup>39,41</sup> Principally, the resonances could originate from by-products of the synthesis as sequence deletions or early terminations would be expected to be unfolded. This possibility, however, can be ruled out, as the synthesized protein was demonstrated to be of high purity (HPLC-electron spray ionization (ESI)-mass spectrometry (MS),  $> 95\%$ ). Thus, these resonances must originate from alternative conformations of the protein. However, these must interconvert with the folded protein only slowly on the NMR chemical shift time scale in order to yield separate sets of resonances in the NMR spectrum. Slow conformational exchange transitions are rare in proteins and principally not expected in a fast and cooperatively folding, small protein domain. We contribute these conformations to the presence of four proline amino acids in the Trp-cage sequence (as will be discussed in detail below), as X-Pro sequences are known to undergo a slow *trans/cis*-peptide bond isomerization under specific circumstances.

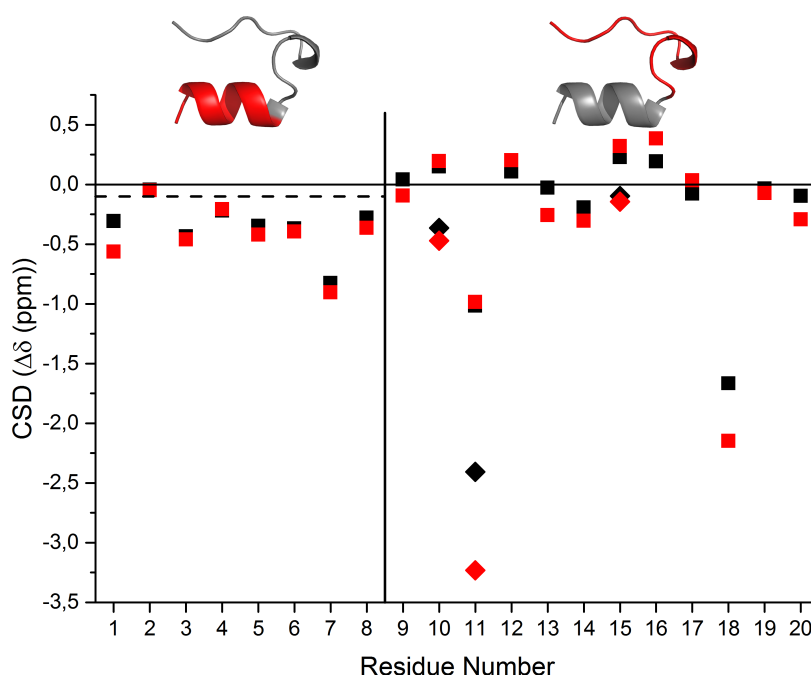
### 1.1. The Folded Fraction

The secondary structure of folded proteins can easily be assessed by analyzing the chemical shifts of H $\alpha$  protons. These undergo typical upfield or downfield changes depending on the backbone conformation of the respective amino acid. A straight forward method for interpretation was developed by D. WISHART.<sup>76,77</sup>

$$CSD = \delta_{obs} - \delta_{ref} \quad (1.1.1)$$



In this, the deviation of the observed chemical shift of the  $H\alpha$  from the reference random coil shift is calculated (Equation I.1.1). The resulting chemical shift deviation (CSD) has been shown to be secondary structure dependent. Thus, the CSD is often termed the secondary chemical shift. Negative CSD values ( $< 0.1 \text{ ppm}$ ) are indicative of helical backbone conformation. Conversely, positive values ( $> 0.1 \text{ ppm}$ ) indicate a  $\beta$ -conformation for the respective residue. Thus, a plot of the CSD against the residue number facilitates the determination and a presentation of the secondary structure present in proteins.



**Figure I.1.2.:** Chemical shift deviation plot of  $H\alpha$ -atoms in Tc10b at 298 K at pH 3 (black) and pH 7 (red). The CSD of the  $H\alpha$  is represented as squares. The deviations for the second  $H\alpha'$  protons in glycines are depicted as diamonds. The dashed line represents the cutoff for the presence of a helix. The expected fold is presented above the plot in ribbon presentation. The vertical solid line separates the sequence region of the helix (residue 1-8) to the left from that of the Trp-cage loop (residue 9-20) on the right.

The CSD plot is shown for Tc10b at 298 K and pH 3 and pH 7 (Figure I.1.2). The Trp-cage fold generally can be divided in two parts: the N-terminal helix from residue 1 to 8 and a C-terminal cage loop that completes the fold (9-20).

The chemical shift deviation of the N-terminal residues from 1 to 8 of Tc10b is continuously negative. Except for residue number 2, the limit for a helical structure is exceeded for all residues. Notably, the CSD for residue 7 is particularly negative, which is typical for the Trp-cage. Likely, that position experiences an aromatic-ring current shift caused by the indole side chain of the preceding amino acid W6.

From residue 9 onwards, nearly all residues show CSD values larger than the significance

limit of 0.1 ppm. Both positive and negative deviations are observed, further it is notable that the sign of the CSD frequently changes between neighboring residues. This behavior indicates the absence of classical (regular) secondary structure elements (no sequence stretches of three or more residues with identical backbone conformation), while the presence and magnitude of CSD manifest the presence of structure and absence of conformational flexibility. This behavior is fully consistent with the loop structure from residue 9-15 and the polyproline-II helix from residue 17-19 present in this region of the Trp-cage. Remarkably, the CSD values for residue Gly 11 and Pro 18 are strongly negative. Again, this observation can be attributed to the ring-current effect of the aromatic indole ring system of the Trp side chain. These CSDs are therefore also a strong indicator for the close proximity between the respective residues and the Trp side chain. They reflect the encapsulation of the Trp side chain by Gly 11 and Pro 18 that is typical of the Trp-cage tertiary structure. Thus, the CSDs of these two residues are excellent markers for the Trp-cage fold.

A comparison of the CSD plots of Tc10b at pH 3 and 7 shows that the CSDs are nearly identical. The region from residue 1 to 8 exhibits only minor changes in the CSD values, indicating a similar stability of the  $\alpha$ -helix. The cage loop also exhibits a similar pattern at both pH values. However, the magnitude of  $H\alpha$  CSD of the marker residues Gly 11 and Pro 18 is significantly larger at pH 7. This indicates that the cage fold stability is higher at neutral pH, owing to the formation of a salt bridge between the oppositely charged residues Asp 9 and Arg 16.

The observed chemical shifts for Tc10b at pH 3 and pH 7 are in good agreement with the chemical shifts reported in the literature.<sup>43</sup> For a complete unbiased assessment of the tertiary structure of each analyzed variant, it is advantageous to not only analyze the  $H\alpha$  protons, but also all additional side chain protons that are affected characteristically in the Trp-cage fold. A method for this was established by BARUA et al.<sup>43</sup> They selected for the representation of the helix the sum of the  $H\alpha$ -CSD from residue 2-8 (sum of the chemical shift deviation ( $\Delta\delta$ )). In order to describe the cage loop, protons are selected, which show the most influence by the ring current of the Trp side chain. These protons are the  $H\alpha$  of position 7, the  $H\alpha'$  of Gly (11), the  $H\alpha$  and  $H\beta'$  of Pro (18), and the  $H\delta$  and  $H\delta'$  of Pro (19). The CSD of these protons were summed up and serve as reference for the cage loop. They decided to separate the helix and the cage loop, as variants of the Trp-cage can possess an intrinsic stable helix, and thus not a stringent cooperative folding between the helix and the cage loop.

**Table I.1.2.:** CSD for selected protons in the structure representing the cage fold of Tc10b at 298 K at pH 3 and pH 7.

Cage Loop	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
pH 3	-0.83	-2.41	-1.67	-1.02	-0.34	-0.53	<b>-6.79</b>	<b>71</b>	
pH 7	-0.90	-3.23	-2.15	-1.51	-0.44	-0.63	<b>-8.87</b>	<b>92</b>	
Helix	A2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	K8 $\alpha$		
	$\Delta\delta(ppm)$							$\Sigma(ppm)$	$\chi_F(\%)$
pH 3	-0.045	-0.434	-0.22	-0.35	-0.37	-0.83	-0.28	<b>-2.52</b>	<b>84</b>
pH 7	-0.05	-0.46	-0.21	-0.42	-0.40	-0.90	-0.36	<b>-2.80</b>	<b>94</b>

Application of this method for the Trp-cage Tc10b at pH 3 and 298 K results in a sum of the CSD for the helix of  $\Delta\delta_{helix} = -2.52 ppm$  and for the cage loop of  $\Delta\delta_{cage} = -6.79 ppm$ , respectively. At pH 7, the sum for the helix is  $\Delta\delta_{helix} = -2.80 ppm$  and for the cage loop to  $\Delta\delta_{cage} = -8.87 ppm$ .

A comparison of the sums between the two pH values confirm the general observations in the CSD plot. The helix shows only a small stabilization upon an increase in pH, the cage loop show significant changes, reflecting the location of the stabilizing salt bridges. The observed chemical shifts and the discussed CSDs (shift changes versus an unstructured reference point) have been interpreted with the underlying assumption that an equilibrium between folded and unfolded conformation of a protein exists and that the interconversion between these conformations is fast on the NMR time scale. Under these conditions, the measured chemical shift is averaged over folded and unfolded conformation or, more precisely, all conformations present. This is in contrast to the slower *cis/trans* proline conformational isomerizations that yield separate sets of resonances due to the lack of signal averaging. See discussion in section below. Thus, with the knowledge of chemical shifts for the folded state of the protein and the random coiled reference shifts for the unfolded state, the position of the folding equilibrium, i.e. fraction folded ( $\chi_F$ ), of a protein can be determined. The published shifts of Tc10b at pH 7 and 280 K, where Tc10b is 99.5% folded, can thus be used in conjunction with the method by BARUA to determine the  $\chi_F$  of the helix and the cage loop (Equation I.1.2).<sup>43</sup>

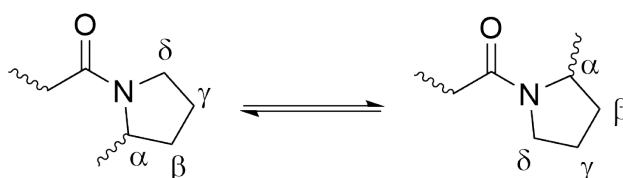
$$\chi_F = \left( 0.995 * \frac{CSD_{obs}}{CSD_{ref}} \right) \quad (I.1.2)$$

Calculation of the fraction folded for Tc10b at 298 K yields in a value for  $\chi_F^{helix} = 84\%$  and  $\chi_F^{cage} = 71\%$  at pH 3 and  $\chi_F^{helix} = 94\%$  and  $\chi_F^{cage} = 92\%$  at pH 7 and 298 K. This method is used in the following chapters for comparing the fold stabilities of all Trp-cage variants that were synthesized in this project.

## 1.2. The Unfolded Fractions

The folding dynamics of the Trp-cage is more complicated than expected based on present literature results that suggest a completely cooperative two state folding mechanism. In the previous section, the chemical shift dependence of proton resonances in the Trp-cage was interpreted in terms of an existing fast equilibrium between the folded Trp-cage with an unfolded population that yields one averaged NMR resonance for both states. This behavior is in agreement with the expectations for a two-state, cooperative the folding mechanism of such a small protein folding unit and all literature results for the Trp-cage. Our NMR results however reveal the presence of multiple conformational equilibria that occur on distinct time-scales. These additional conformational equilibria are manifested in the presence of multiple resonances for the W6H $\epsilon$  proton. Their respective chemical shifts indicate that the resonances originate from partially folded or completely unfolded conformations.

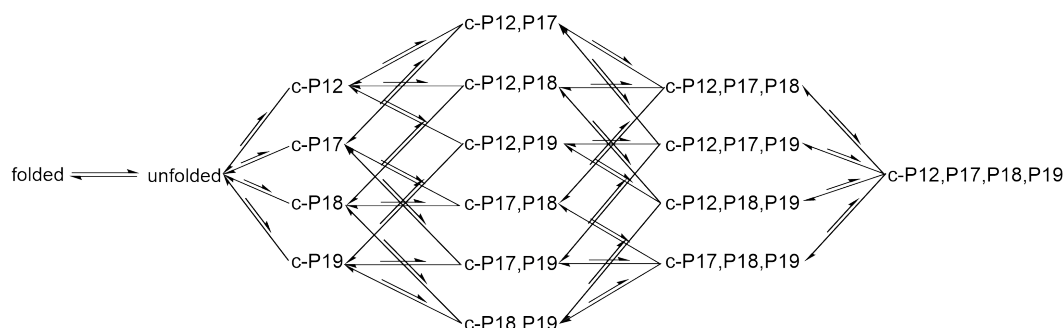
Multiple conformations in the unfolded state, in particular slow processes that lead to separate resonances in the NMR spectrum, are unexpected for a small protein. All conformational conversions are expected to occur on the fast NMR time scale and should be included in the single, fast-average signal. In this disordered or "random coil" state, interactions between atoms are nearly completely absent and energetic barriers between conformers are small.



**Scheme I.1.1.:** Scaffold of X-Pro. The peptide bond of X-Pro can isomerize between *cis* and *trans*.

An exception is the peptide bond. The peptide bond has a partially double bond character, thus it can either be in *trans* configuration ( $\omega = 0^\circ$ ) or in the *cis* configuration ( $\omega = 180^\circ$ ). Usually, the energy difference between the two configurations is high and energetically, the *trans* configuration favoured (with an approximate ratio of 1 : 1000). However, the X-Pro peptide bond is known to undergo a *trans/cis* isomerization, as the energy differences between the two configurations is significantly smaller as the C $\alpha$  and C $\delta$  offer nearly equal steric repulsion to the preceding carbonyl group (Scheme I.1.1). Thus, about 2 – 80% of *cis* X-Pro and an exchange rate of the order of  $2.5 \cdot 10^{-3} \text{ sec}^{-1}$  have been observed in unstructured peptides containing a single Pro residue when no tertiary interactions shift the equilibrium to one side.<sup>78–81</sup> For the reason

of latter, *cis/trans*-Pro isomers are often referred to as conformers rather than strictly configuration isomers.

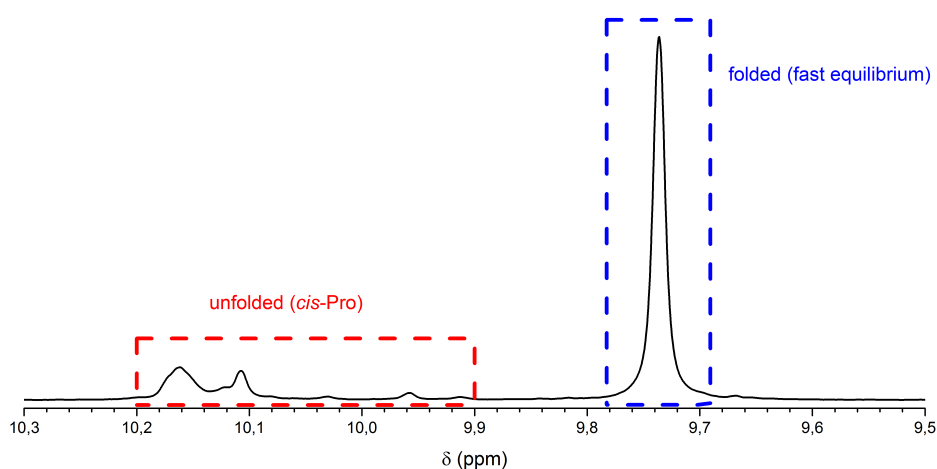


**Scheme I.1.2.:** Schematic view of the folding and unfolding equilibrium with the downstream equilibrium of the *cis*-Pro isomerization of the Trp-cage. In the unfolded, all *trans* state of the Trp-cage, the X-Pro peptide bond can undergo a isomerization to the *cis* configuration (indicated by c-PXX, with XX the number of the Pro in the sequence), resulting in four different conformers. These conformers can also undergo a isomerization, which leads to a total number of 15 unfolded conformations with at least one X-Pro peptide bond in *cis*.

All four Pro peptide bonds in the folded Trp-cage are in the *trans* configuration. In the unfolded, all *trans* state of the fast equilibrium, the X-Pro peptide bond can undergo an isomerization from the *trans* to the *cis* configuration. With a X-Pro peptide bond in *cis* configuration, the Trp-cage cannot refold correctly and this conformation is trapped and extracted from the fast equilibrium. The Trp-cage sequence contains four Pro residues, whereas each can undergo a isomerization to the *cis* configuration. In this state, with a single X-Pro in *cis* configuration, the three remaining X-Pro can also isomerize (Scheme I.1.2), leading to a total number of 15 possible conformations. In order to estimate the population of unfolded conformations with at least one X-Pro in the *cis* configuration, an average possibility of 5% *cis* for each X-Pro is assumed. With this assumption, the probability for one X-Pro in the *cis* configuration in the Trp-cage accumulates to 20%. The overall probability, including all four X-Pro and combinations, is determined to be  $\approx 23\%$ . However, the total amount present in a folding equilibrium and thus visible in the NMR, is also dependent on the population in the unfolded, all *trans* conformation. Even at the thermal mid-point (equilibrium between folded and unfolded), a probability of 5% per X-Pro isomerization leads to a observable ratio of 90 : 10 (folded and unfolded *trans* configuration to *cis*-Pro unfolded). However, with an increased average probability of 20% per X-Pro isomerization, ratios of  $\approx 60 : 40$  are possible at the thermal mid-point.

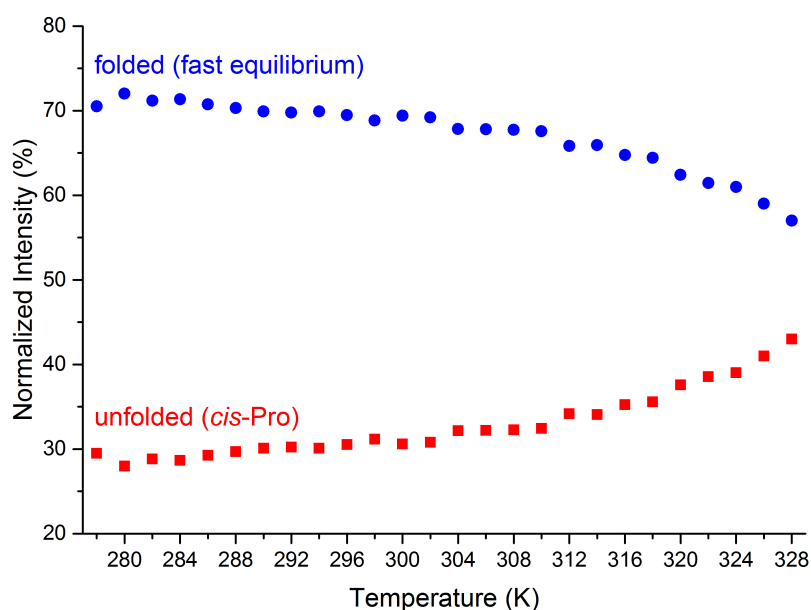
In order to prove this interpretation of the observed multiple unfolded conformations in the NMR spectrum of Trp-cage (inset, Figure I.1.1) being accumulated configurations with *cis* X-Pro peptide bonds, two approaches were pursued.

The first approach exploits the fact that owing to the limited stability, proteins unfold with increasing temperature. Following our hypotheses, an increase in unfolded equilibrium population should result and also lead to an increase in *cis*-unfolded conformers. Thus, the intensity of resonances for unfolded conformers should grow at the expense of the intensity of the folded conformer. This was tested for practical reasons not with Tc10b, but a closely related variant of similar stability, our variant TcKKA (described in detail in section 3.1). Its sequence **Ac-DKYAQLADGGPSSGRPPPK** differs from that of Tc10b in the amino acids highlighted in bold.



**Figure I.1.3.:**  $^1\text{H}$  NMR spectrum of the region of the W6H $\epsilon$  proton from 9.5 to 10.3 ppm of TcKKA at pH 3 and 298 K. The sharp resonance at 9.75 ppm is the fast equilibrium signal and its integral is taken as all *trans* configuration (blue box). The multiple signals between 9.9 and 10.2 ppm are the *cis*-Pro unfolded signals and the area integral represents the *cis*-unfolded.

A series of  $^1\text{H}$  NMR spectra of TcKKA from 278 K to 328 K in steps of 2 K at pH 3 was recorded and the ratio of the fast equilibrium resonances and the *cis*-Pro unfolded resonances determined. For this, the W6H $\epsilon$  resonances are particularly suitable, since the signals of the fast, all *trans* equilibrium and the *cis*-Pro conformations are well separated (Figure I.1.3). As a result, the integral of the upfield shifted W6H $\epsilon$  resonance represents the fast equilibrium signal of the all *trans* conformations (folded and unfolded) and the overall intensity of the *cis*-Pro unfolded population is represented by the integral of the area between 9.9 and 10.3 ppm. The sum of both integrals was normalized to 100 in order to remove intensity differences originating from NH proton exchange with the solvent.



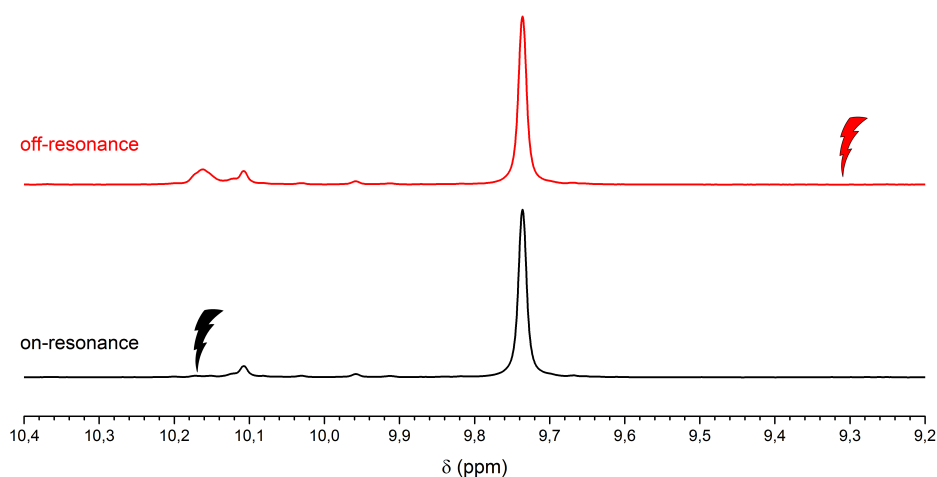
**Figure I.1.4.:** Normalized intensity of the folded all *trans* (fast equilibrium, blue circles) and unfolded (*cis*-Pro, red squares) conformations against temperature. The integral of the folded conformation was determined for the upfield shifted W6H $\epsilon$  below 10 ppm and represents the Trp-cage fold in fast equilibrium of the folding and unfolding process. For the representation of the unfolded *cis*-Pro conformations, the integral of all W6H $\epsilon$  signals between 9.9 and 10.2 ppm is summed up.

A plot of the normalized intensity of the fast, all *trans* equilibrium (blue) and the unfolded *cis*-Pro conformations (red) against the temperature highlights the relationship between the conformations (Figure I.1.4). At low temperatures (278-286 K), the ratio between the fast, all *trans* equilibrium and *cis*-Pro unfolded conformations is approximately constant ( $\approx 3 : 1$ ). From 286 K on, the intensity of the fast equilibrium indole signal decreases (all *trans* conformation), whereas the *cis*-Pro conformations increase. This shows that upon thermal denaturation of the Trp-cage, the total amount of *cis*-Pro unfolded conformations present in the sample also increases. Foremost, the fold stability of the protein is in relation to the amount of *cis*-Pro unfolded conformations, and thus it can be utilized as an additional stability criteria for new Trp-cage variants.

The second approach attempted to spectroscopically prove the direct exchange between the all *trans* conformation and the *cis*-Pro unfolded conformations. For this, a variant of a nuclear Overhauser effect (NOE) experiment was performed, the so called transferred NOE (tr-NOE). This type of experiment is usually used for the investigation of ligand-protein interactions.<sup>82</sup> In this experiment, resonances of the protein are selectively irradiated and saturated. Due to spin diffusion, this saturation is transferred

and leads to the saturation of all protein resonances. In the case of binding a ligand, this magnetization is also transferred from the protein to the ligand, thus proving association of ligand and protein.

A particular advantage of this approach is that the saturation transfer can similarly be observed in an exchange situation on a slow time scale. With this experiment, processes involving two or more different states, such as isomerization, can be studied. For this application, two requirements are needed. First, the states have to differ significantly in their chemical shifts for a selective irradiation. Second, the exchange rate must be faster than the relaxation of the magnetization. For the Trp-cage, both requirements are clearly fulfilled for the W6H $\epsilon$  proton.

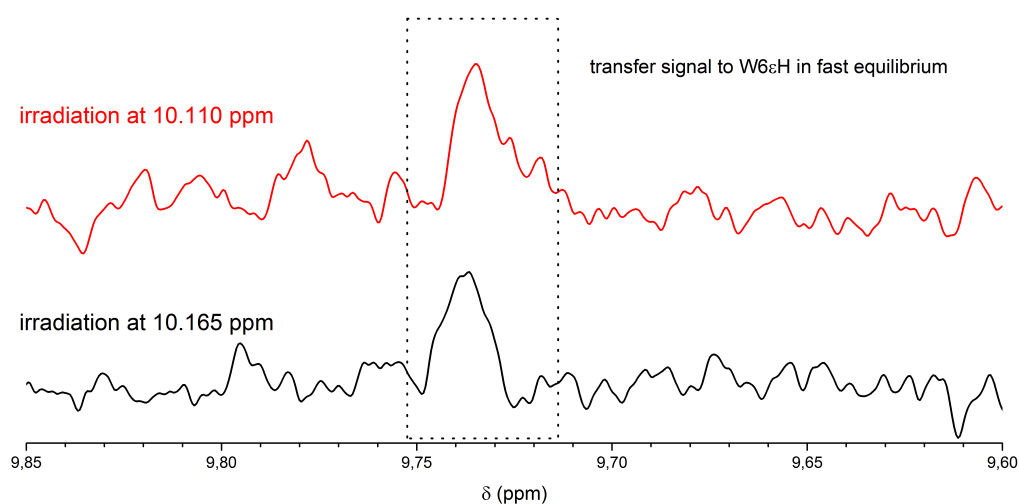


**Figure I.1.5.:** Enlarged W6H $\epsilon$  region of the 1D steady state tr-NOE of TcKKA at 298 K and pH 3 between  $\delta = 9.2$  and  $10.4$  ppm. In the black spectrum the *cis*-Pro W6H $\epsilon$  at  $10.165$  ppm (black mark) was selectively irradiated and is the *on-resonance* spectrum. The red spectrum is the *off-resonance* reference spectrum with an irradiation  $9.300$  ppm (red mark).

The experiment was performed as steady state version with a exchange time of  $4$  s. Two separate spectra are recorded, one *on-resonance*- and one *off-resonance*-saturated spectrum. In the *on-resonance* irradiation spectrum, one resonance of W6H $\epsilon$  of the *cis*-Pro conformations was selectively irradiated and saturated (black spectrum, Figure I.1.5). During a delay, that serves as exchange time, the saturated conformation is able to isomerize to the all *trans* state, which is included in the fast equilibrium signal of W6H $\epsilon$ . The saturation of this *cis*-Pro conformation is thereby "transferred" and the fast equilibrium signal of W6H $\epsilon$  should be partially saturated, i.e. reducing its signal. In the *off-resonance* spectrum, a second part of the spectrum was selectively irradiated with equal power. In order to reduce artefacts of imperfect selectivity, such as unwanted

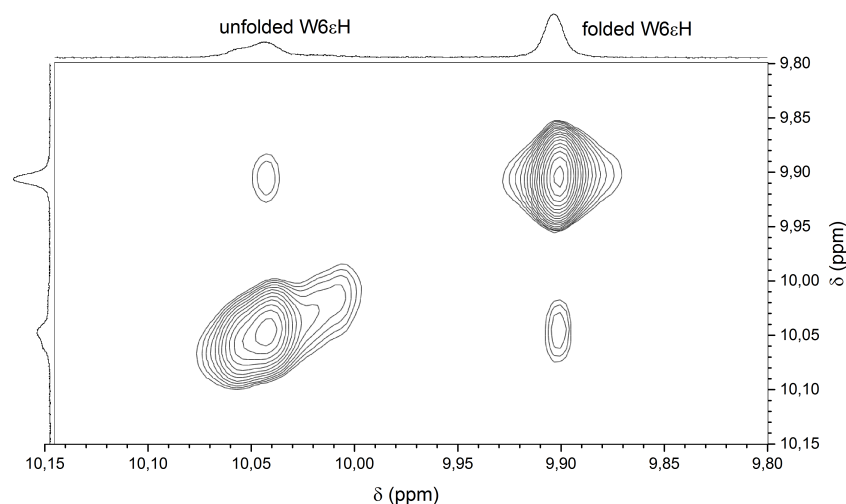


direct saturation of the fast equilibrium indole signal, the *off-resonance* spectrum was recorded with irradiation equally distanced from the signal of the folded W6H $\epsilon$ . For better visibility of the transferred magnetization, a difference spectrum between the *off-* and *on-resonance* was calculated that shows a resonance line proportional to the saturation transfer.



**Figure I.1.6.:** Enlargement of the 1D steady state tr-nuclear Overhauser effect spectroscopy (NOESY) difference spectra of W6H $\epsilon$  in fast equilibrium between 9.60 and 9.85 ppm with on-resonance irradiation at 10.165 (black spectrum) and 10.110 ppm (red spectrum). The signal at 9.75 ppm (dashed box) depicts the transferred signal from the respective *cis*-Pro species into the all-*trans* fast equilibrium resonances.

Two *cis*-Pro unfolded conformations with resonances at 10.110 and 10.165 ppm were chosen for the irradiation in the tr-NOE experiment, as they exhibit the highest intensity and therefore the best chance to detect a transfer to the W6H $\epsilon$  in the fast equilibrium. In both difference spectra, a signal at the resonance of W6H $\epsilon$  in the fast equilibrium is observed (Figure I.1.6). This proves that from both *cis*-Pro unfolded conformations magnetization is transferred to the all *trans* conformation. These conformations are in exchange with the fast equilibrium signal. Thus, proving a slow process in the protein folding. This is in excellent agreement with the above described hypothesis of unfolded conformations with an X-Pro peptide bond in *cis* configuration.

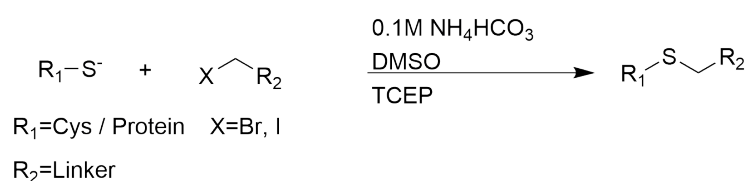


**Figure I.1.7.:**  $^1\text{H},^1\text{H}$  NOESY spectrum of our Trp-cage variant TcKKA at 328 K and pH 3. Enlarged region of the W6H $\epsilon$  between  $\delta = 9.80$  and  $10.15$  ppm. The diagonal peaks correspond to the peaks in the 1D NMR and the cross peak between  $\delta = 9.89$  ppm and  $10.04$  ppm to the exchange between the fast equilibrium signal and the unfolded *cis*-Pro conformation.

With increase in temperature, not only the total amount of unfolded conformations increase but also the exchange rate between the *trans* and *cis* configuration of the X-Pro peptide bond. This leads to the possibility to perform a  $^1\text{H},^1\text{H}$  NOESY experiment at elevated temperature in order to show the exchange between the different conformations. The  $^1\text{H},^1\text{H}$  NOESY experiment is usually used to determine short distances through space in molecules. In this experiment, short mixing times ( $\approx 200$  ms) are used in which magnetization between spins is transferred *via* dipolar coupling through space. However, chemical exchange processes can also be observed, as they similarly transfer magnetization from one resonance (or conformation) to another. In the recorded  $^1\text{H},^1\text{H}$  NOESY spectrum, exchange cross peaks are observed between the W6H $\epsilon$  of the fast equilibrium signal and unfolded *cis*-Pro signal (Figure I.1.7). The exchange cross peak above the diagonal (top, left signal) describes the exchange from the *cis*-Pro unfolded conformations to the all *trans* conformation and the cross peak below the exchange from the all *trans* conformation to the *cis*-Pro unfolded conformations. All performed experiments prove that a slow equilibrium is present in the Trp-cage folding, leading to additional unfolded species, which are kinetically trapped and extracted from the fast equilibrium between the folded and unfolded conformation.

## 2. Link the Cysteine in the Trp-Cage with a Cross-Linker

The azobenzene linker introduced by WOOLLEY has been successfully used in numerous studies to switch the conformation of proteins. It comprises the azobenzene moiety and two iodoacetamide functional groups. This group facilitates the chemoselective covalent attachment to Cys- side chains of the proteins (Scheme I.2.1).



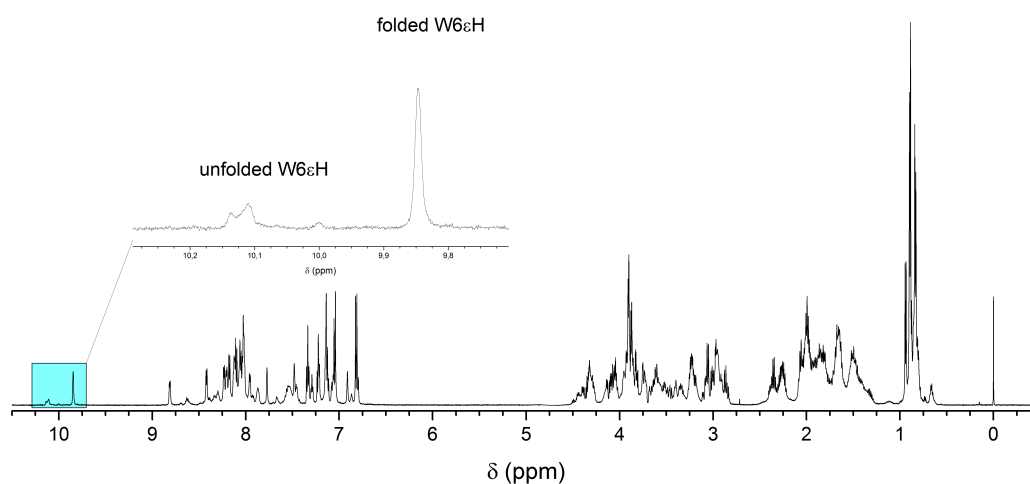
**Scheme I.2.1.:** Reaction scheme of cross-linking of Cys with iodo or bromo functionalized linker.

The iodide in the acetamide group represents an activated halide. Principally, the reaction can also be performed with bromides in equivalent positions, i.e. bromo acetamides or bromide in benzylic positions. The linking reaction is carried out under alkaline conditions. As the Cys's thiol group easily oxidizes and forms disulfide bridges under these conditions, the presence of a reductant such as TCEP is required.

**Table I.2.1.:** The sequences of Trp-cage variants Tc5b and Tc10b presented in one letter code, which are used in work. The differences between these variants are marked in bold.

<b>Tc5b</b>	<b>N</b> LYIQWLKDGGPSSGRPPPS
<b>Tc10b</b>	<b>D</b> AYAQLWLKDGGPSSGRPPPS

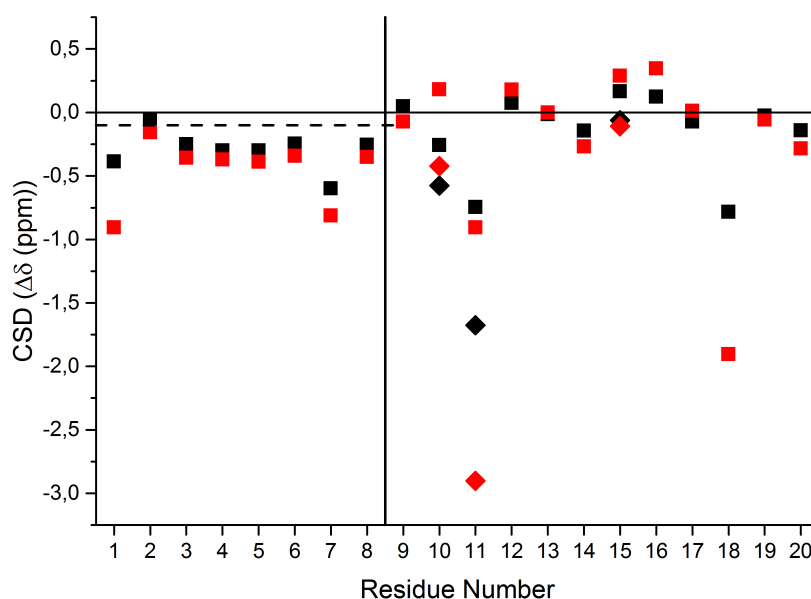
The Trp-cage Tc5b, a variant possessing average stability, was chosen for the following exploratory experiments (Table I.2.1). Tc5b and our reference variant Tc10b share 85% sequence identity. Only three amino acids in the first turn of the N-terminal helix differ between these two proteins. The sequence changes from Asp-Ala-Tyr-Ala in Tc10b to Asn-Leu-Tyr-Ile in Tc5b, resulting in a destabilization of the fold by 14 °C ( $T_M$  Tc5b: of 42 °C , Tc10b: 56 °C).<sup>43</sup>



**Figure I.2.1.:**  $^1\text{H}$  NMR spectrum of Tc5b at 298 K and pH 3 with enlargement of the indole proton region.

Tc5b shows a typical  $^1\text{H}$  NMR spectrum at 298 K and pH 3 for a folded Trp-cage. The signal of the encapsulated indole W6H $\epsilon$  is observed at  $\delta = 9.85 \text{ ppm}$ . Similar to other Trp-cage variants, unfolded *cis*-Pro conformations lead to clearly visible signals of the indole proton above  $10 \text{ ppm}$ . The ratio of the resonance intensities of the fast folding equilibrium and the *cis*-Pro unfolded conformations is 69 : 31, which is comparable to Tc10b.

At pH 8, i.e. the conditions of the linking reaction, the signal of the fast equilibrium folded indole shifts upfield by  $0.10 \text{ ppm}$  to  $\delta = 9.75 \text{ ppm}$ . The interpretation of this shift average yields a fraction folded of 93%, the ratio of the fast equilibrium signal to *cis*-Pro unfolded signals changes to 81 : 19. Both values suggest the presence of a more stable fold at neutral pH.



**Figure I.2.2.:** CSD plot of the  $H\alpha$  atoms of Tc5b at pH 3 (black) and pH 8 (red). The  $H\alpha$  CSD is represented as square and the  $H\alpha'$  (Gly) as diamond. The dashed line represents the cut-off for the presence of a helix. The solid line separates the N-terminal helix (residue 1-8) from those of the Trp-cage loop (residue 9-20).

The  $H\alpha$  CSD plot confirms that Tc5b exhibits a cage fold (Figure I.2.2). Expectedly, from residue 1 to 8 the chemical shift deviation is negative and confirms the helical secondary structure. From residue 9 to the C-terminal end, the CSD shows the typical distribution for the loop in the folded Trp-cage at both pH values. The protein is characterized as folded with a higher stability at neutral pH, as the deviations for the  $H\alpha$  protons increase in magnitude.

To determine the stability of the Trp-cage based on the fast equilibrium of folded and unfolded conformation (with all X-Pro in *trans* configuration), the chemical shifts of characteristic protons in the tertiary fold were assessed again by the method of BARUA. The sum of the CSDs of characteristic protons was calculated as described for Tc10b and the folded fraction determined (Table I.2.2).

**Table I.2.2.:** CSDs for selected protons in the cage structure representing the cage and helical fold of Tc5b at 298 K (pH 3 and pH 8).

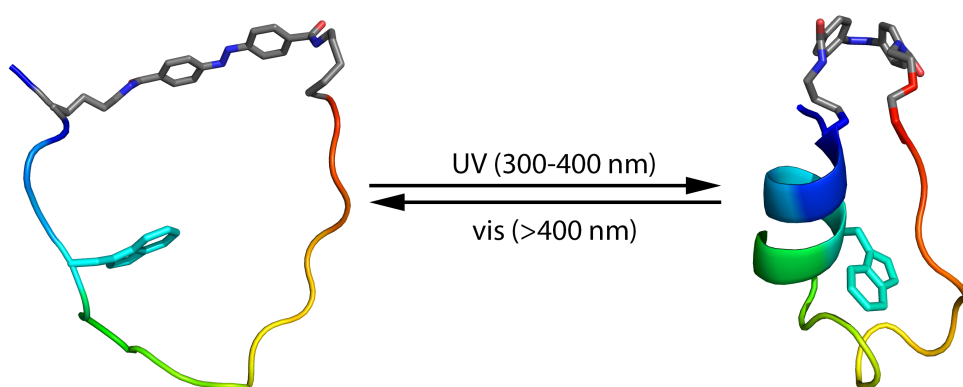
Cage Loop	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
pH 3	-0.60	-1.68	-0.78	-0.79	-0.19	-0.35	<b>-4.39</b>	<b>46</b>	
pH 8	-0.81	-2.90	-1.90	-1.27	-0.38	-0.56	<b>-7.82</b>	<b>81</b>	
Helix	L2 $\alpha$	Y3 $\alpha$	I4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	K8 $\alpha$		
	$\Delta\delta(ppm)$							$\Sigma(ppm)$	$\chi_F(\%)$
pH 3	-0.05	-0.25	-0.30	-0.30	-0.25	-0.60	-0.26	<b>-2.01</b>	<b>67</b>
pH 8	-0.17	-0.36	-0.37	-0.39	-0.34	-0.81	-0.35	<b>-2.79</b>	<b>93</b>

Notably, the fractions folded of the helix and the cage loop differ significantly by 20%, which is unexpected for a cooperative folding. Structurally interpreted, the secondary structure (helix) of the Trp-cage is intrinsically stabilized and compared to the tertiary structure (cage loop). As a result, the Trp-cage can exhibit a helical propensity in the absence of the tertiary structure. A criterion for an intrinsically stable helix, and thus not an exclusively cooperative folding mechanism, are the differences in the fraction folded and the  $T_M$  of helix and cage. If the difference of the latter is smaller than 7 °C, the protein folds cooperatively.

The fraction folded of Tc5b at neutral pH is determined to be  $\chi_F^{cage} = 81\%$  and  $\chi_F^{helix} = 93\%$ . A comparison between Tc5b and Tc10b shows that under conditions of the cross-linking reaction, the Trp-cage variant Tc5b exhibits a similar stability, making this protein a suitable variant for reaction tests and optimizations.

## 2.1. Introducing Cysteine Residues into Tc5b

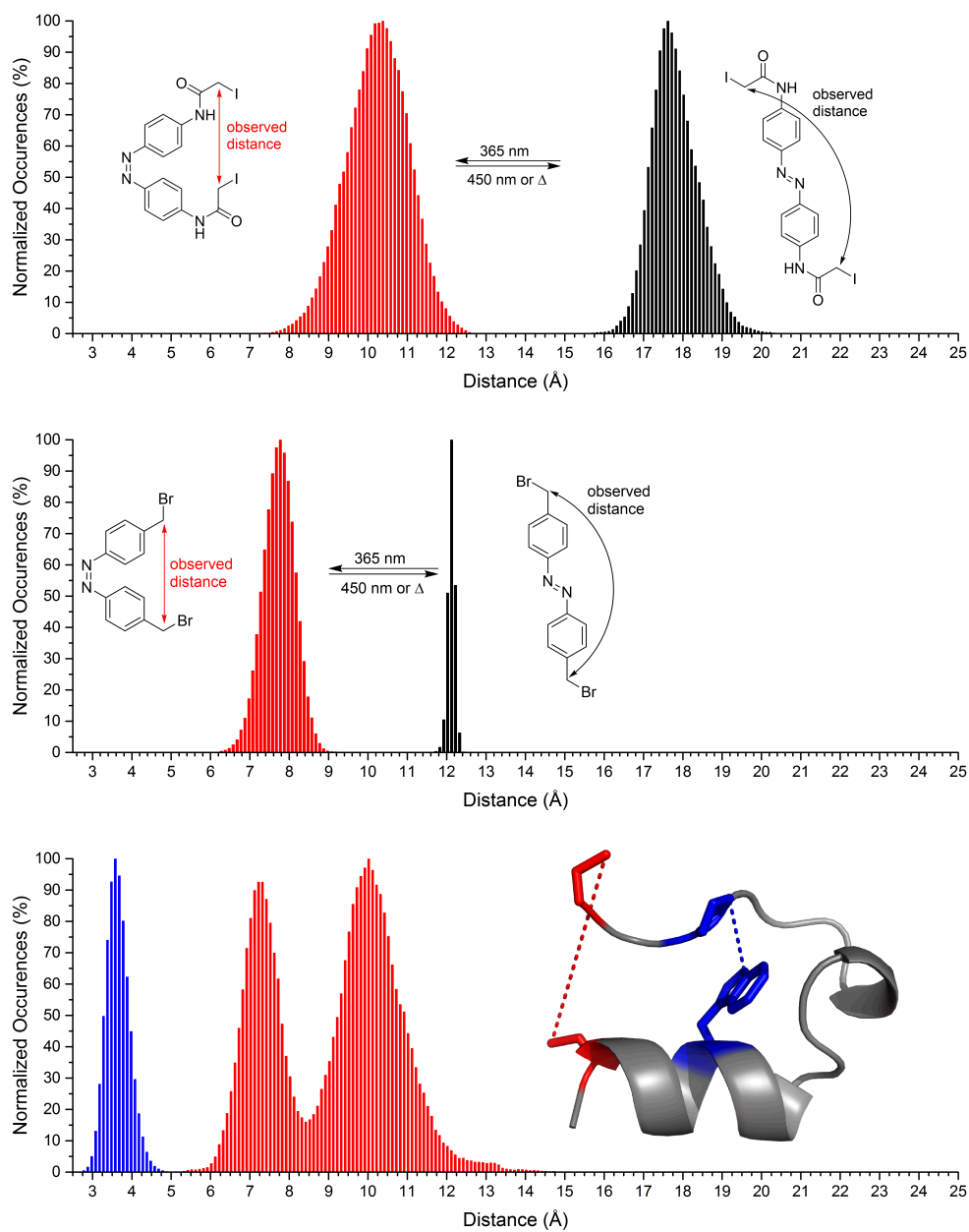
The Trp-cage does not have any native Cys residues in its sequence. Therefore, the only sequence alterations required for protein cross linking are the incorporation of two cysteine amino acids. The best positions for their incorporation depend on the steric properties of the chosen linker (mainly its length) and the employed concept for effective switching of the fold. Here, we pursue the push-concept with the linkers 4,4'-iodoacetamide-azobenzene (WOOLLEY linker) and 4,4'-dibromomethyl azobenzene.



**Figure I.2.3.:** Concept for a switchable Trp-cage with azobenzene in a push position between the N- and C-terminus. The azobenzene linker destabilizes or disrupts the tertiary structure of the Trp-cage in its *trans* configuration (left) and stabilizes structure in its *cis* configuration (right).

In this concept, the azobenzene in the *cis*-configuration should match the distance between the two anchor points, i.e. the two Cys-residues, in the folded conformation (Figure I.2.3). With azobenzene in the *trans* conformation, the attachment points are pushed apart, locally disturbing the structure of the protein. Ideally, this should lead to an exposure of the Trp side chain to the solvent. As the fold of the Trp-cage is hydrophobically driven by encapsulation of this side chain, its exposure should lead also to a complete unfolding of the protein.

In order to identify the best positions for introducing of the two Cys residues, the length of the linker and distance between possible connection points in the protein were determined in MD simulations of the two linkers.

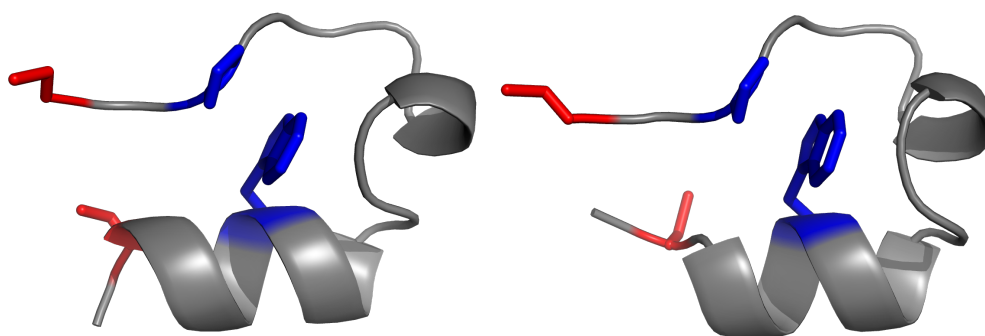


**Figure I.2.4.:** Top: Distance distribution of 4,4'-diiodoacetamide azobenzene (WOOLLEY linker) between C-6 and C-6' for *cis* (red) and *trans* (black). Middle: Distance distribution 4,4'-dibromomethyl azobenzene between C-5 and C-5' for *cis* (red) and *trans* (black). Bottom: Distance distribution of Cys side chain γS of Tc5b-L2C-S20C (red) and between W6-Hε and P18-Hβ' as control (blue).



MD simulations of the WOOLLEY linker show an effective distance change between C-6 and C-6' in the *trans* and *cis* isomer of 7 Å (17.5 Å in *trans* and 10.5 Å in *cis* (Figure I.2.4, detailed information of the simulations can be found in the Experimental section 3.11). 4,4'-dibromomethyl azobenzene shows a higher rigidity and shorter distances for the *cis* (7.6 Å) and for the *trans* (12.1 Å) isomer. This results in an effective length change of 3.5 Å.

In order to determine positions for the Cys residues in Tc5b, the published structure was scanned for possible connection points matching the length of the linker. The best positions were identified to be position 2 in the helix and position 20 at the C-terminus for both linker lengths. Thus, MD simulations of a Trp-cage variant with the two Cys residues in the appropriate positions were also performed and the distance between the Cys side chains analyzed. During the simulation, the distance between W6-H $\epsilon$  and P18-H $\beta'$  was monitored as control for the folded state of the protein (Figure I.2.4). The distance distribution between the  $\gamma$ S atoms of Cys at position 2 and 20 reveals that two maxima are present. One distribution maximum is centered at 7.5 Å and the other one at 10.0 Å.



**Figure I.2.5.:** The two major conformations of Tc5b-L2C-S20C observed in the MD simulation. The Trp side chain is encapsulated during the simulation, but at the N-terminus and C-terminus flexibility was observed. This is a result of interactions between the Cys at position 2 and the C-terminus. This yields in a different conformation of the first two N-terminal amino acids of the helix and thereby two minimum structures.

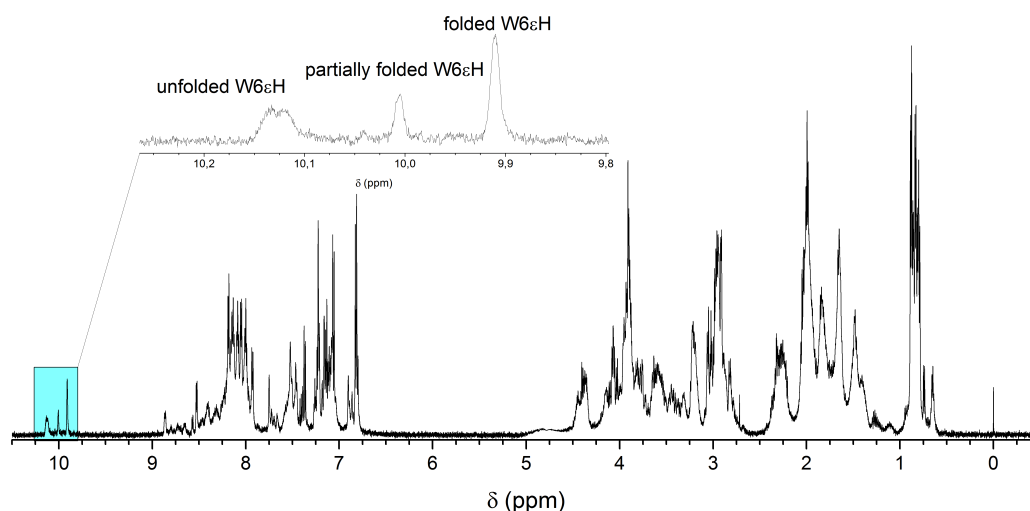
At first, the two maxima (rather than one) in the distance distribution of side chains in the Trp-cage appears surprising. An analysis of the simulated Trp-cage conformations reveals their origin. At the beginning of the simulation, the helix is folded up to the N-terminus (left, Figure I.2.5) with a backbone torsion angle of  $\phi = -60^\circ$  and  $\psi = -35^\circ$  of Cys at position 2. This results in shorter distances between the side

chains (centered 7.5 Å). During the second half of the simulation, the position 2 torsion angles change ( $\phi = -150^\circ$  and  $\psi = +36^\circ$ ). While this results in a partially unfolding of the first N-terminal loop of the helix, the tertiary structure of the cage remains folded (right, Figure I.2.5). The increased distance and flexibility lead to the second maximum with a average distance of 10.0 Å between the two Cys residues.

Each maximum in the distance distribution of the connection residues in the Trp-cage matches the length of one photochromic linker in its *cis* conformation: 4,4'-dibromomethyl azobenzene with the maximum at 7.5 Å and the WOOLLEY-linker with the maximum at 10.0 Å. Furthermore, neither linker matches these distances in their respective *trans*-configuration, and thus should fulfill the requirement for efficient switching.

### 2.1.1. Tc5b-L2C-S20C

The double Cys replacement of the Trp-cage Tc5b at position 2 and 20 was synthesized (by S. ROTHEMUND, IZKF Leipzig, with detailed synthesis protocols of all variants documented in the Appendix, subsection 1.1.2) and characterized by NMR spectroscopy.

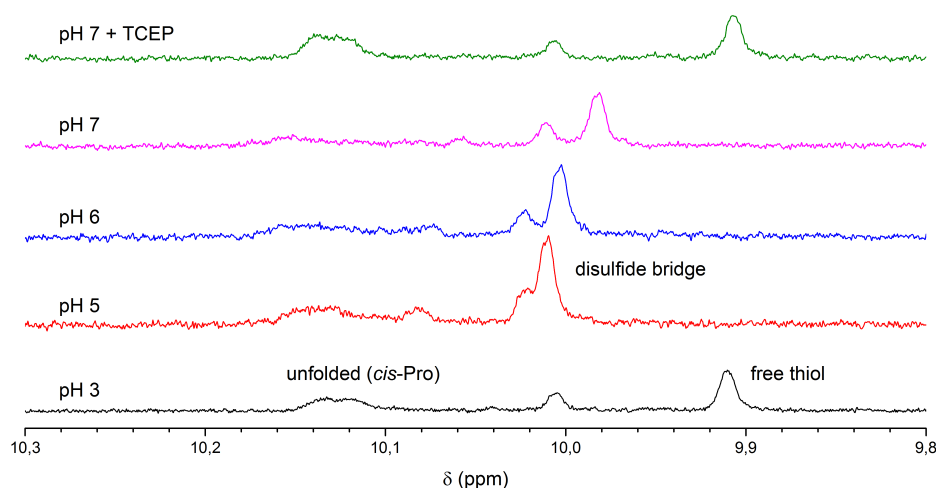


**Figure I.2.6.:**  $^1\text{H}$  NMR spectrum of Tc5b-L2C-S20C at 298 K and pH 3 with magnification of indole proton region.

The  $^1\text{H}$  NMR spectrum of Tc5b-L2C-S20C at 298 K and pH 3 indicates a destabilizing effect of the Cys replacements (Figure I.2.6). The signal of the encapsulated indole proton W6H $\epsilon$  is observed at  $\delta = 9.91$  ppm, suggesting a higher fraction unfolded in the fast equilibrium population. Also, unfolded *cis*-Pro conformations are visible above 10 ppm. Surprisingly, another defined, sharp signal for the W6H $\epsilon$  is observed. Its chemical shift

of  $\delta = 10.01 \text{ ppm}$  is in-between the folded conformation and the unfolded *cis*-Pro conformations, thus it is unclear what nature this species has. Because of its visibility in the NMR spectrum, it must be a species in slow exchange with the other conformers. Thus, the resonance may also arise from an unusual *cis*-Pro conformer, for example a partially folded one. Alternatively, the signal could originate in a side reaction of the Cys side chain, an intra- and/or inter-molecular disulfide-bridge formation between the two Cys residues. Disulfide bonds are known to occur in aqueous solution at neutral pH (or above), when solvent accessible thiol groups are oxidized by oxygen present in the solvent.

We first tested for this side reaction by changing the protein solution conditions and by monitoring changes by  $^1\text{H}$  NMR spectra (Figure I.2.7).



**Figure I.2.7.:**  $^1\text{H}$  NMR spectra of Tc5b-L2C-S20C at 298 K and different pH values. Expansion of indole proton region between  $\delta = 9.8$  and  $10.3 \text{ ppm}$ . Bottom to top: Tc5b-L2C-S20C at pH 3 (black), pH 5 (red), pH 6 (blue), pH 7 (magenta) and pH 7 with addition of TCEP (green). The label *free thiol* describes the correctly folded protein in fast equilibrium.

At pH 3, the three different signals of the indole proton W6H $\epsilon$  are observed as discussed above. The novel, unexplained signal is clearly evident at  $\delta = 10.01 \text{ ppm}$ . By increasing the pH to 5, the fast equilibrium signal (free thiol) at pH 3 vanishes and new signals with high intensities are observed at  $\delta = 10.02 \text{ ppm}$ . By further increase of the pH, the three NMR spectra remain unchanged except for slight chemical shift changes of all resonances. This could suggest that already after the first pH change to 5 all protein reacted and oxidized. Thus, we added the reductant tris-(2-carboxyethyl)-phosphine (TCEP) to the solution. The observed spectrum is identical to the initial spectrum at pH 3, despite the difference in pH. This experiment series proves that disulfide oxidation does indeed occur rapidly at pH 5, but is fully reversible upon addition of reducing

agent. Second, it proves that the sharp resonance at 10.01 *ppm* persist in the presence of reducing agent. This shows, that the additional, partially unfolded signal cannot be attributed to a formation of intra or inter molecular disulfide bridges.

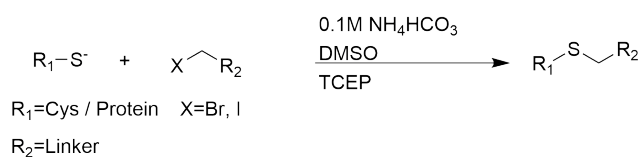
In order to test for the alternative explanation for the resonance at 10.01 *ppm*, we pursued an amino acid replacement study using analogs aimed at detecting whether steric interactions of the Cys residues could cause the observation of novel partially folded *cis*-Pro Trp-cage conformation. In the Trp-cage variant Tc10b-21Gly, we introduced Cys at position 20, replacing the native Ser residue. This, from a structural point conservative replacement, leads to a Trp-cage variant that is determined to be correctly folded (for complete data see Appendix, subsection 1.2.2). By the additional introduction of Cys at position 2, a protein variant was obtained that yields a nearly identical <sup>1</sup>H NMR spectrum compared to that of Tc5b-L2C-S20C (Figure I.2.6), in particular showing the additional sharp signal with shift of 10.01 *ppm*, in between the resonances of folded and folded conformations of the Trp-cage.

In conclusion, these experiments revealed that the partially folded conformation is not originated from a formation of disulfide bridges. The replacement study shows that the replacement to Cys at position 2 in the helix leads to the partially folded conformation and is attributed to a specific *cis*-Pro isomerization. As in the helix no Pro residues are present, an influence to the isomerization of the Pro residues at the C-terminus (17-19), due to a close proximity of the Cys side chain, is most likely. This hypotheses is in accordance to observations made in Part II.

Although this variant exhibits a complex fold with two stable folded conformations, it was used for cross-linking, with the aim of an associated simplification of the fold.

## 2.2. Cross-Linking of Tc5b-L2C-S20C

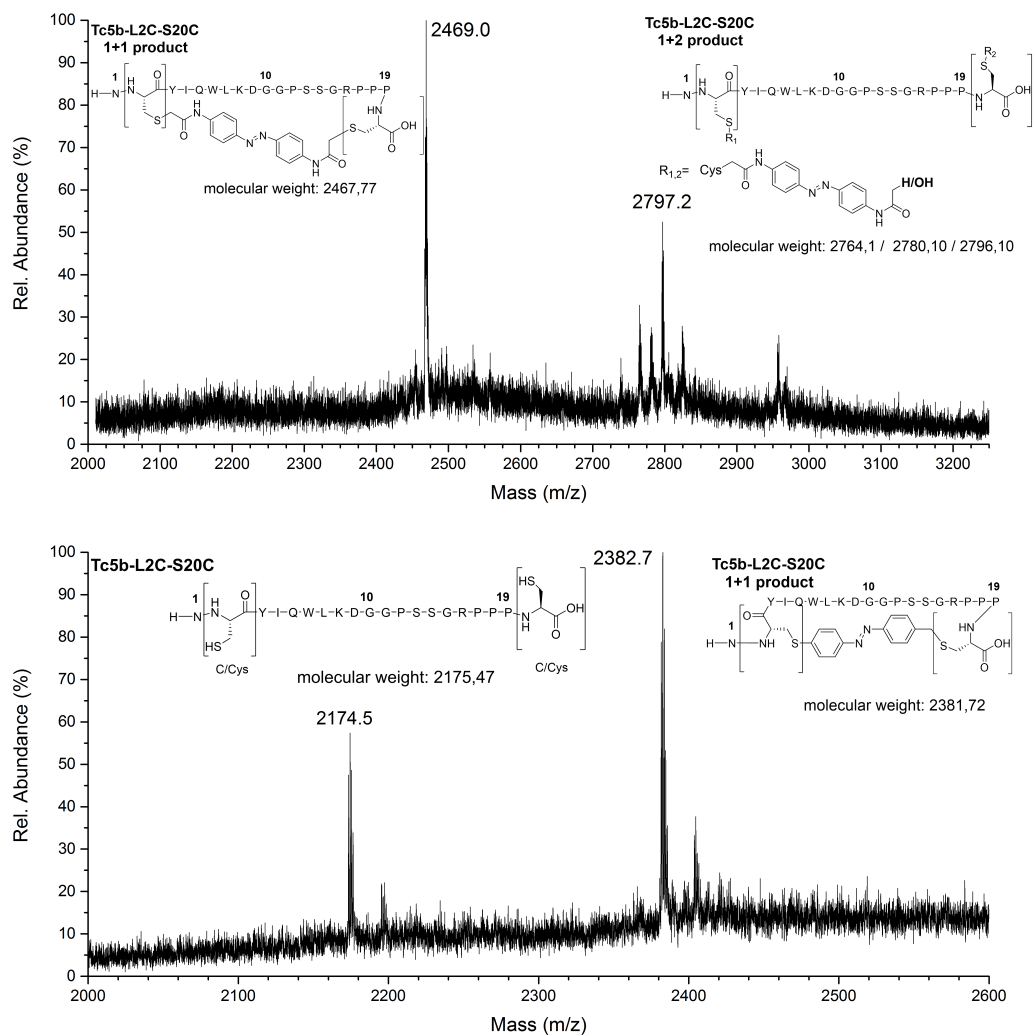
The linking of Tc5b-L2C-S20C with the two photochromic cross-linkers was performed following a procedure by KUMITA and coworkers,<sup>83</sup> which is described in detail in the Experimental Section (section 2.1). The WOOLLEY-linker was provided by H. KOBARG and 4,4'-dibromomethyl azobenzene by J. STRÜBEN. The reaction was monitored using HPLC, matrix assisted laser desorption ionization (MALDI) mass spectrometry with a time of flight (ToF) mass separator, and NMR spectroscopy.



**Scheme I.2.2.:** Reaction scheme of cross-linking of Cys with iodo- or bromo-functionalized linker. The activated alkyl halogen can react in a nucleophilic substitution reaction with the deprotonated thiol to form a thio-ether.

The cross-linking reaction with the WOOLLEY-linker or 4,4'-dibromomethyl azobenzene is a non-selective reaction (Scheme I.2.2). Therefore, not only in the 1 + 1 product (Trp-cage + linker), but also a 1 + 2 (Trp-cage + two linker molecules) with a hydrolyzed reactive position of the linker or oligomeric products (2 + 2 and other combinations) can be expected.

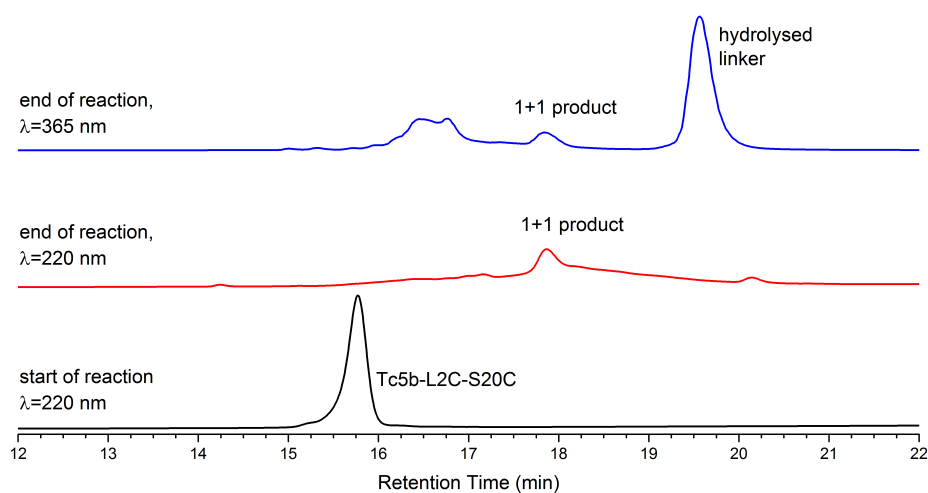
The cross-linking of the Trp-cage with these cross-linker turned out to be more complicated than the literature suggested. Therefore, a great effort was put into optimizing the reaction by varying solvent mixtures and conditions. However, the results were almost identical, thus the results of a standard procedure is presented.



**Figure I.2.8.:** Top: MALDI-ToF spectrum of the crude reaction solution of Tc5b-L2C-S20C with the WOOLLEY linker.  $m/z$ :  $[M + H]^+$  calculated for  $[Tc5b - L2C - S20C - (azo) + H]^+$  2468.1, found 2469.0 and  $[M + H]^+$  calculated for  $[Tc5b - L2C - S20C - (azo)_2(hydrolyzed) + H]^+$  2797.1, found 2797.2. Bottom: MALDI-ToF spectrum of the crude reaction solution of Tc5b-L2C-S20C with the 4,4'-dibromomethyl azobenzene.  $m/z$ :  $[M + H]^+$  calculated for  $[Tc5b - L2C - S20C + H]^+$  2176.5, found 2474.5 and  $[M + H]^+$  calculated for  $[Tc5b - L2C - S20C - (azo) + H]^+$  2382.7, found 2382.7.

With respect to the multiple cross-linking possibilities, at low reactant concentrations the reaction should primarily yield the cyclized 1 + 1 product in, since the cyclization reaction is an intramolecular ring closure. Therefore, the 1 + 1 product is favoured compared to the products of one protein with two linker (1 + 2 product) or two proteins with one linker (2 + 1 product). The MALDI-ToF spectrum of the reaction solution with the WOOLLEY-linker indicates a complete conversion of Tc5b-L2C-S20C. Mainly the  $m/z$  peak for the desired cross-linked 1 + 1 product is observed. Beside this, peaks of one protein with two cross-linkers in different hydrolyzed states (1 + 2 products) are observed with significant intensities (Figure I.2.8).

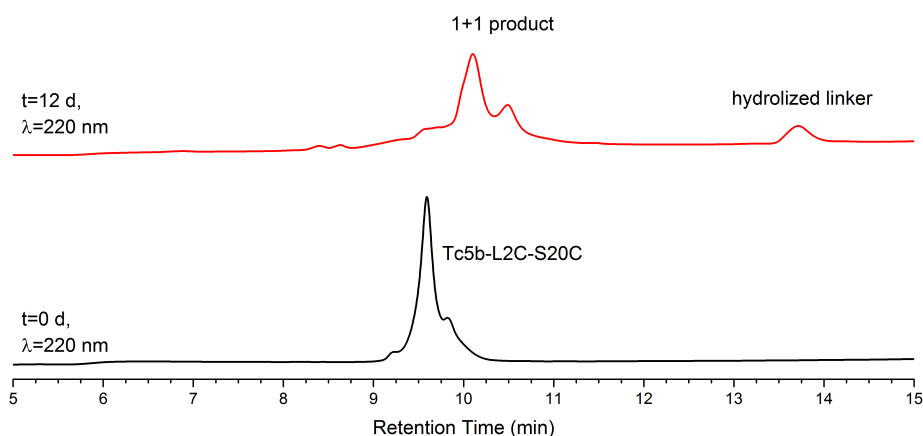
In contrast, the cross-linking reaction with 4,4'-dibromomethyl azobenzene shows an incomplete conversion and peaks of doubly linked protein (1 + 2) are not observed. The reduced reaction rate and the enhanced selectivity can be fully explained by the reduced solubility of 4,4'-dibromomethyl azobenzene compared to the WOOLLEY-linker. The linker shows only limited solubility in the reaction solution and mainly precipitates, significantly reducing the concentration of the linker available for the reaction in the solution.



**Figure I.2.9.:** Progress of the linking reaction between Tc5b-L2C-S20C and the WOOLLEY linker was monitored by HPLC. From bottom to top: Tc5b-L2C-S20C (black chromatogram), reaction solution with a detection wavelength of 220 nm (red chromatogram) and 365 nm (blue chromatogram).

Both cyclization reactions were also followed by HPLC. The HPLC profile of the reaction mixture Trp-cage and WOOLLEY-linker after reaction completion (Figure I.2.9) shows one main product peak with a retention time ( $t_R$ ) of 18 min. This was assigned to the desired 1 + 1 cyclized product based on its MALDI spectrum (see Appendix, Figure V.1.7). The reaction is complete (> 90%) as a peak of the starting material ( $t_R = 15.5$  min) was not observed. A clear peak for a 1 + 2 cross-linked product is

also not apparent, despite its detection in the mixture characterized by MALDI (Figure I.2.8). However, very broad peaks between 14 to 17 *min* and a peak at 20 *min* are noticeable. These originate from unreacted, hydrolyzed linker or linker side products. This is highlighted by measuring a subsequent HPLC profile obtained under nearly identical conditions, which was recorded with a detection wavelength ( $\lambda$ ) of 365 *nm*. At this wavelength, the aromatic linkers have a significant higher absorption coefficient. The profile confirms that the assigned product peak is linked with an aromatic linker. Further, the broad signals sharpen and the signal at 20 *min* shows an increased intensity. Thus, the peaks are assigned to side products of the aromatic linker.



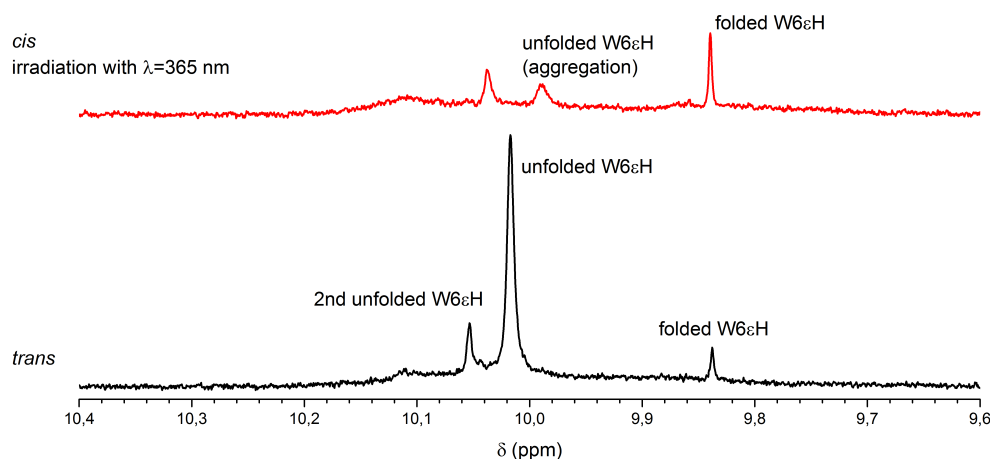
**Figure I.2.10.:** Reaction progression of Tc5b-L2C-S20C and 4,4'-dibromomethyl azobenzene monitored by HPLC. From bottom to top HPLC chromatograms of the reaction mixture after  $t = 0 d$  (black chromatogram), 12 *d* (red chromatogram) with a detection wavelength of  $\lambda = 220 nm$ .

The reaction progress of Tc5b-L2C-S20C with 4,4'-dibromomethyl azobenzene was similarly monitored by analytical HPLC with a detection wavelength  $\lambda = 220 nm$  (Figure I.2.10). Besides the peak of Tc5b-L2C-S20C at  $t_R = 9.5 min$ , two minor peaks at  $t_R = 9.2 min$  and  $t_R = 9.9 min$  were observed in the beginning of the reaction (black chromatogram). Those peaks could not be identified with the equipment at that time. After 12 days of reaction time (red chromatogram), most of the protein had reacted and three peaks were apparent ( $t_R = 10.0 min$ ,  $t_R = 10.5 min$  and  $t_R = 14.0 min$ ). The fraction with  $t_R = 10.0 min$  contains the correct cross-linked product. For the peak at  $t_R = 10.5 min$  unambiguous assignment was not possible, but judged by the retention time a multiple cross-linked side product, such as 1 + 2 or 2 + 2, is a plausible origin. The third peak ( $t_R = 14.0 min$ ) is assigned to the unreacted linker in its hydrolyzed state as discussed for the other linker.



The cross-linked proteins were purified and further characterized by NMR spectroscopy. For both cross-linked Trp-cages, a low signal to noise ratio of the spectra is observed, as only limited product quantities were available owing to the low yield ( $< 20\%$ ) of the cyclization reaction. It is also clearly noticeable that the resonances of the protein are broadened, which reduces the signal to noise ratio even further. These effects are significantly more pronounced for the cross-linked Trp-cage with the WOOLLEY-linker. Its  $^1\text{H}$  NMR spectrum can be found in the Appendix (Figure V.1.8).

The NMR spectrum of Tc5b-L2C-S20C linked with 4,4'-dibromomethyl azobenzene has a significantly better signal to noise ratio. This can be attributed to the increased selectivity of the cross-linking reaction, and thus higher yield ( $\approx 40\%$ ) and higher concentration of the sample. Similar to the cross-linked Tc5b-L2C-S20C with WOOLLEY-linker, line-broadening is observed for the protein resonances (full spectra can be found in the Appendix, Figure V.1.11). In order to show the effect of the cross-linker to the Trp-cage fold in the *trans* or *cis* configuration, the focus is set to the indole proton region of W6H $\epsilon$  between  $\delta = 9.5$  and  $10.4$  ppm.



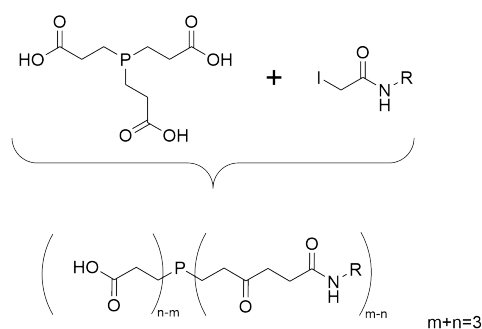
**Figure I.2.11.:** Enlarged  $^1\text{H}$  NMR spectra between  $9.5$  and  $10.4$  ppm (indole proton region) of Tc5b-L2C-S20C cross-linked with 4,4'-dibromomethyl azobenzene (top) in *trans* (black) and *cis* (red) at  $298$  K and  $\text{pH } 3$ .

In the magnified indole proton region of W6H $\epsilon$  of the  $^1\text{H}$  NMR spectrum with the linker in the *trans* configuration, three sharp signals are observed (Figure I.2.11). The signal  $\delta = 10.02$  ppm is assigned to W6H $\epsilon$  of the Trp-cage in an unfolded conformation judged by its chemical shift. The clear signal at  $\delta = 10.05$  ppm is assigned to a second unfolded conformation of the Trp-cage. Judged by its chemical shift and intensity its occurrence can be explained by a single *cis*-Pro unfolded conformation. The third signal at  $9.84$  ppm is assigned to a W6H $\epsilon$  of the Trp-cage in its folded conformation. A folded Trp-cage signal should only be observable if the sample contains already *cis*-azobenzene.

This could have happened during the purification process, since several steps cannot be performed exclusively in the dark.

Upon irradiation at  $\lambda = 365 \text{ nm}$ , the azobenzene is switched from the *trans* to the *cis* configuration. This results in a significant decrease of intensity of the indole signal of the unfolded Trp-cage at  $\delta = 10.02 \text{ ppm}$ . Simultaneously, the intensity of the folded signal at  $9.84 \text{ ppm}$  increases, indicating an encapsulation of the Trp side chain, and thus a refolding with the linker in the *cis*-configuration. However, the two signals of the unfolded Trp-cage not only decrease in intensity, but also experience an upfield shift and a line broadening. Furthermore, a significant signal loss is observed. Both observations lead to the conclusion that the protein sample aggregated. Therefore, no further experiments and assignments were possible.

The results of the cross-linking with the two azobenzene linker reveal that a slower but more selective cross-linking reaction is of advantage, regarding the purification and therefore yield. The changes in the  $^1\text{H}$  NMR spectra of azobenzene in the *trans* or *cis* configuration indicate that the protein is influenced differently by each state. The chemical shift of the W6H $\epsilon$  in both states is in good agreement with our aim to destabilize the Trp-cage fold in the *trans* configuration of the linker and stabilizing its fold in its *cis* configuration.



**Scheme 1.2.3.:** Possible side reactions of TCEP (top, left) with the reactive group of the WOODLEY linker.

The aggregation of both cross-linked variants can be a result of the side-products of the linker with the reducing agent TCEP (Scheme 1.2.3). These by-products are soluble in water and due to their variety not fully separable by purification by RP-HPLC. Evidence for their presence in the protein samples are prominent signals of the linker in the NMR spectra (spectra in the Appendix, Figure V.1.8, Figure V.1.11). Changes in solvents (to methanol or dimethylsulfoxide, data not shown) or additional purification steps could not diminish the linker contamination nor reduce the aggregation of the protein product.

This hypothesis was later substantiated by experiments performed by H. KOBARG.<sup>67</sup> After an upgrade of the HPLC with a simultaneous ESI-MS detection (see Appendix

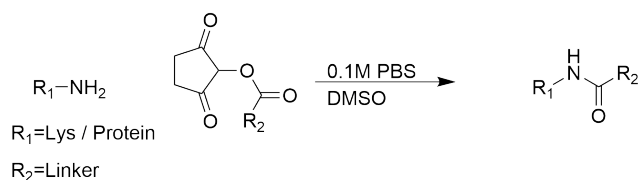
section 3.2 for more details), additional peaks of the HPLC chromatograms were observed and assigned. He showed various side reactions of the linker, i.e. hydrolysis, reactions with reducing agent TCEP, leading to water soluble azobenzene. These species exhibit similar retention times on the HPLC as the desired product. Therefore, the purified samples still contained unreacted azobenzene, possibly causing the observed aggregation.

Overall, these experiments suggest that control of the Trp-cage fold using this approach with azobenzene cross-linker is principally possible. However, significant improvements regarding the properties of the linking reaction, e.g. yield, elimination of soluble linker by-products, ideally avoiding the use of a reductant, altogether, are necessary. Such improvements are expected to also improve the protein behavior and reduce the sample aggregation tendency.



### 3. Link the Lysine in the Trp-Cage with a Cross-Linker

The results of cross-linking the Trp-cage between Cys residues revealed several important observations. On the positive side, the selectivity of the cross-linking reaction can be tuned by the solubility of the linker. With limited solubility the reaction is slower, but significantly lower fractions of side products were observed. On the negative side, we learned that the use of one specific reductant caused multiple problems. The linker reacted with the reductant, leading to water soluble side-products that cause significant difficulties in the purification. The major drawback of using Cys residues for cross-linking is the occurrence of multiple (partially) folded species of the Trp-cage.



**Scheme I.3.1.:** Reaction scheme of cross-linking of Lys with N-hydroxysuccinimide ester functionalized linker. The activated ester can react with primary amines to form a peptide bond.

In order to overcome the substantial problems, we decided to pursue an alternative cross-linking chemistry. A promising possibility appeared to be the linking chemistry based on N-hydroxysuccinimide esters NHS (Scheme I.3.1). These esters are known to react exclusively with primary amines, therefore lysine is the only reactive amino acid.<sup>84</sup> A particular advantage using this chemistry is that the reaction proceeds without any additional reagents, which largely reduces the possibility of side reactions.

For implementing this chemistry, we intended to use a more stable Trp-cage variant than Tc05b to reduce the effect of possible unfavourable steric interactions upon including the long Lys side chain. Therefore, the studies are based on the Trp-cage variant Tc10b with a  $T_M$  of 56 °C, elevated by 14 °C compared to Tc5b.

### 3.1. Trp-cage Variant for Cross-Linking Using NHS-Esters

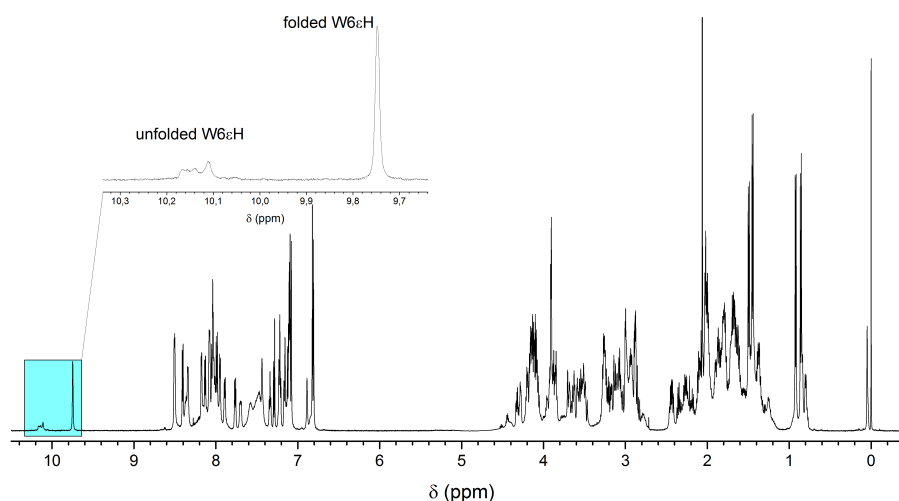
Active esters are reactive towards primary amines. In order to apply this chemistry to the Trp-cage, several replacements are necessary in the Tc10b sequence to incorporate an azobenzene linker.

**Table I.3.1.:** Trp-cage sequence TcKKA for NHS based cross-linking. Replacements in the sequence of Tc10b are marked in bold.

<b>Tc10b</b>	DAYAQLKDGGPSSGRPPPS
<b>Tc-KKA</b>	<b>Ac</b> -DKYAQL <b>AD</b> GGPSSGRPP <b>PK</b>

First, the N-terminus has to be acetylated as it presents a primary amin functionality that would compete for the linker. Secondly, a native Lys residue at position 8 needs to be replaced. We chose Ala of Lys, as this replacement is known to be fold stabilizing.<sup>43</sup> The connection points for the cross-linking were kept identical to our previous attempt and Lys residues were introduced at position 2 and 20 to yield the variant TcKKA (Table I.3.1).

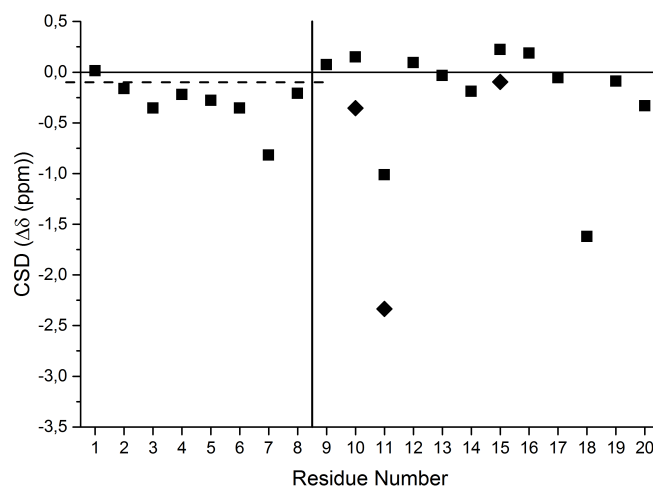
The synthesis of Tc10b-Ac-A2K-K8A-S20K (TcKKA) is described in detail in the Appendix (Figure V.1.21, Table V.3.5). In the following, the folding properties of this Trp-cage variant are discussed.



**Figure I.3.1.:**  $^1\text{H}$  NMR spectrum of TcKKA at 298 K and pH 3. Magnification of indole proton region.

The  $^1\text{H}$  NMR spectrum (Figure I.3.1) indicates a well folded Trp-cage protein. The signal of the indole proton W6 $\epsilon$  in the fast folding equilibrium state is observed at  $\delta = 9.75$  ppm. The ratio between the fast equilibrium signal and the *cis*-Pro unfolded indole signals is determined to be 69 : 31. Both observations are comparable to Tc10b

and are clear evidence that all replacements are well accepted. Notably, the Lys residue at position 2 does not cause alternative (partially) folded conformations in this variant.



**Figure I.3.2.:** CSD plot of the  $H\alpha$  atoms of TcKKA at pH 3.  $H\alpha$  as black squares and  $H\alpha'$  diamonds. The dashed line represents the cutoff for the presence of a helix. The vertical solid line separates the CSD of the helix (residue 1-8) to the left from those of the Trp-cage loop (residue 9-20) on the right.

The  $H\alpha$ -CSD plot confirms that the variant TcKKA exhibits a Trp-cage fold (Figure I.3.2). As expected, the chemical shift deviations from residue 1 to 8 are negative. This is evidence for the helical secondary structure. From residue 9 to the C-terminal end, the CSDs show the expected distribution for the cage loop in the folded Trp-cage. The characteristic high deviations for Gly (11)  $H\alpha'$  and Pro (18)  $H\alpha$ , which are additionally influenced by the ring current of the encapsulated aromatic indole, are strong evidence for the correct tertiary structure.

**Table I.3.2.:** CSDs for selected protons in the cage structure representing the cage fold of TcKKA at 298 K at pH 3.

Cage Loop	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
pH 3	-0.82	-2.34	-1.62	-0.83	-0.34	-0.54	<b>-6.48</b>	<b>67</b>	
pH 7.9	-0.96	-3.46	-2.31	-1.64	-0.45	-0.68	<b>-9.49</b>	<b>99</b>	
Helix	K2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$		
	$\Delta\delta(ppm)$							$\Sigma(ppm)$	$\chi_F(\%)$
pH 3	-0.16	-0.35	-0.22	-0.28	-0.36	-0.82	-0.21	<b>-2.4</b>	<b>80</b>
pH 7.9	-0.24	-0.46	-0.20	-0.37	-0.42	-0.96	-0.30	<b>-2.95</b>	<b>99</b>

The determination of the stability of the Trp-cage's fast equilibrium folding at pH 3 and 298 K is accomplished again by evaluation of the CSDs of characteristic protons grouped according to the tertiary structure (Table I.3.2). This results in  $\chi_F^{cage} = 67\%$  for the

cage loop and  $\chi_F^{helix} = 80\%$  for the helix. Both values are in the same range as for Tc10b at similar conditions.

Furthermore, the fraction folded was determined for TcKKA at 274 K and pH 7.9. At these conditions, the fraction folded for both, helix and cage loop, are determined to be  $> 99\%$  and the highest values we have seen among all our protein variants. The protein under these conditions is completely folded and very stable, which confirms that all replacements and modifications are well accepted.

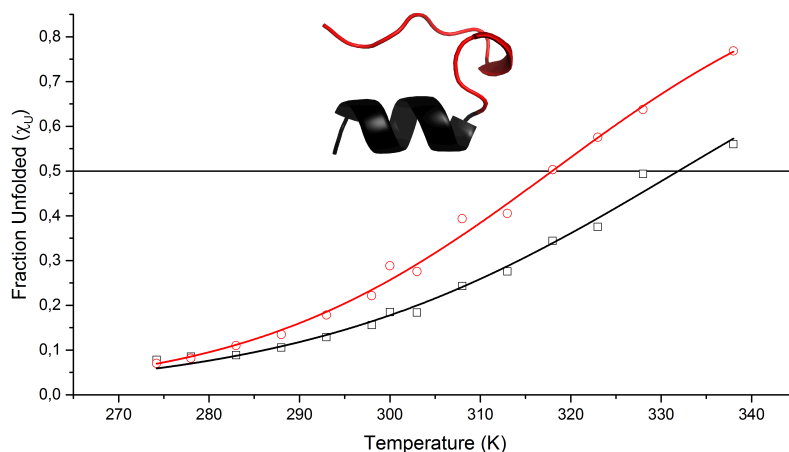
The orientations and interactions of the side chains of the introduced Lys residues are of importance for the cyclization reaction. A measure for the environment of the Lys side chains can be the chemical shift of the Lys  $\zeta$ H amino group. These chemical shifts for the introduced Lys (2) and Lys (20) residues differ in the  $^1\text{H}$  NMR spectrum (observed  $\delta_{K2\zeta} = 7.574 \text{ ppm}$  and  $\delta_{K20\zeta} = 7.473 \text{ ppm}$ ) by  $0.1 \text{ ppm}$ . This indicates a difference in their environment. One could speculate that the higher shift for Lys (2)  $\zeta$ H originates in stronger hydrogen bonding of the amino function, for example with the free C-terminal carboxy function. However, in protic solvents, multiple effects have an influence on this shift (competing hydrogen bonds with the solvent, pH and temperature effects) so that an unambiguous interpretation is difficult. Support for an interaction between the Lys (2) side chain and the C-terminus can be derived from  $^1\text{H}, ^1\text{H}$  NOESY NMR experiments. In this experiment, close proximity of protons can be detected. The spectra  $^1\text{H}, ^1\text{H}$  NOESY spectra of TcKKA show cross peaks between the Lys (2)  $\text{H}\beta$  and the  $\text{H}\delta$  of Pro at position 19. These two residues are in close contact, which would be expected for a hydrogen bonded, conformationally restricted Lys side chain.

The stability of the fold is also characterized by the  $T_M$ , the temperature of the half point of protein denaturation. In order to determine the  $T_M$ , a series of 2D  $^1\text{H}, ^1\text{H}$  total correlation spectroscopy (TOCSY) experiments between 274.2 K and 358 K were recorded and the representative shifts for the helix and cage loop were determined and evaluated. For each temperature, the fraction unfolded ( $\chi_U = 1 - \chi_F$ ) was calculated. With increasing temperature, multiple forms, line-broadening and overlapping with the residual water signal occurred. Foremost, with increasing temperature the folding dynamics facilitates the presence of *trans/cis* Pro and the unfolded *cis*-Pro conformations gain intensity and are more prominent in the sample compared to the conformation in the fast folding equilibrium. As consequence, the chemical shifts could only be determined unambiguously up to 338 K. Noticeably, upon cooling TcKKA from 328 K to 298 K, the protein showed full reversibility of the fold.

$$y = \left( \frac{A_1 - A_2}{1 + e^{(T-T_m)/dT}} \right) + A_2 \quad (1.3.1)$$



The fraction unfolded in the fast equilibrium  $\chi_U$  was plotted against the temperature (Figure I.3.3) and a constrained Boltzmann fit (Equation I.3.1) was performed (Equation I.3.1). The lower and upper limit  $A_1$  and  $A_2$  were set to 0 or 1, respectively.



**Figure I.3.3.:** The fraction unfolded  $\chi_U$  plotted against the temperature for N-terminal  $\alpha$ -helix (black squares) and the cage loop (red circles), both with a constrained Boltzmann fit (black and red line).

The plot of the fraction unfolded against the temperature (Figure I.3.3) shows that at low temperature both, the secondary structure (helix) and the tertiary structure (cage loop), are  $> 90\%$  folded. However, with increase in temperature we observe a small slope for the thermal denaturation. Notably, the cage loop unfolds faster compared to the helix, which leads to a lower melting point by  $> 10^\circ\text{C}$ . Both observations, the small slope and the difference in the  $T_M$ , show that the folding of TcKKA is not exclusively cooperative.

**Table I.3.3.:** Melting points ( $T_M$ ) of the helix and the cage loop of TcKKA, Tc10b and Tc10b-K8A at acidic conditions (pH 3 for TcKKA and pH 2.5 for Tc10b and Tc10b-K8A, NEIDIGH<sup>43</sup>).

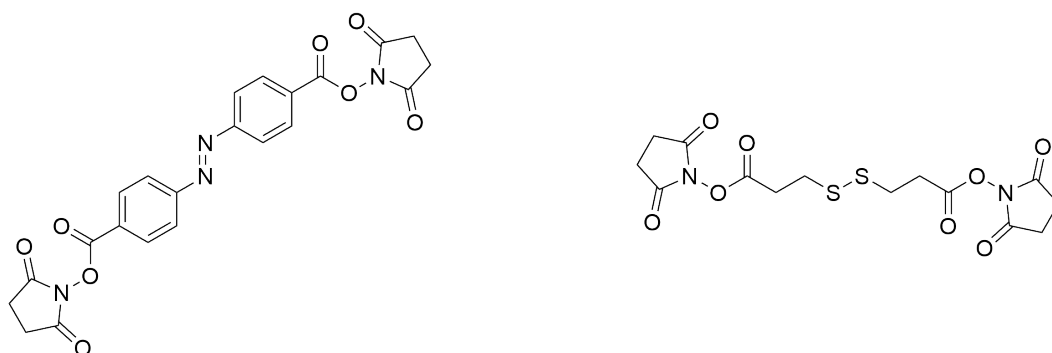
Trp-cage	$T_M(\text{helix}) (K(^{\circ}\text{C}))$	$T_M(\text{cage loop}) (K(^{\circ}\text{C}))$
TcKKA	$332 \pm 0.8 (59 \pm 0.8)$	$318 \pm 0.5 (45 \pm 0.5)$
Tc10b	317 (44)	312 (39)
Tc10b-K8A	325 (52)	319 (46)

Comparing the  $T_M$  of TcKKA with ( $332\text{ K}$  for the helix and  $318\text{ K}$  for the cage loop) to those of Tc10b and the variant Tc10b-K8A<sup>43</sup> shows that the replacements and modifications lead to a stabilization. The helix is markedly stabilized by  $15^\circ\text{C}$  and  $7^\circ\text{C}$  compared to Tc10b and Tc10b-K8A, respectively. The  $T_M$  value of the loop is unaltered compared

to Tc10b-K8A, but elevated by 6 °C compared to Tc10b. This shows that our modified Trp-cage exhibits a very stable fold. We therefore used this new Trp-cage variant for establishing a new cross-linking method described in the following.

### 3.2. Cross-Linker with NHS Esters as Functional Group

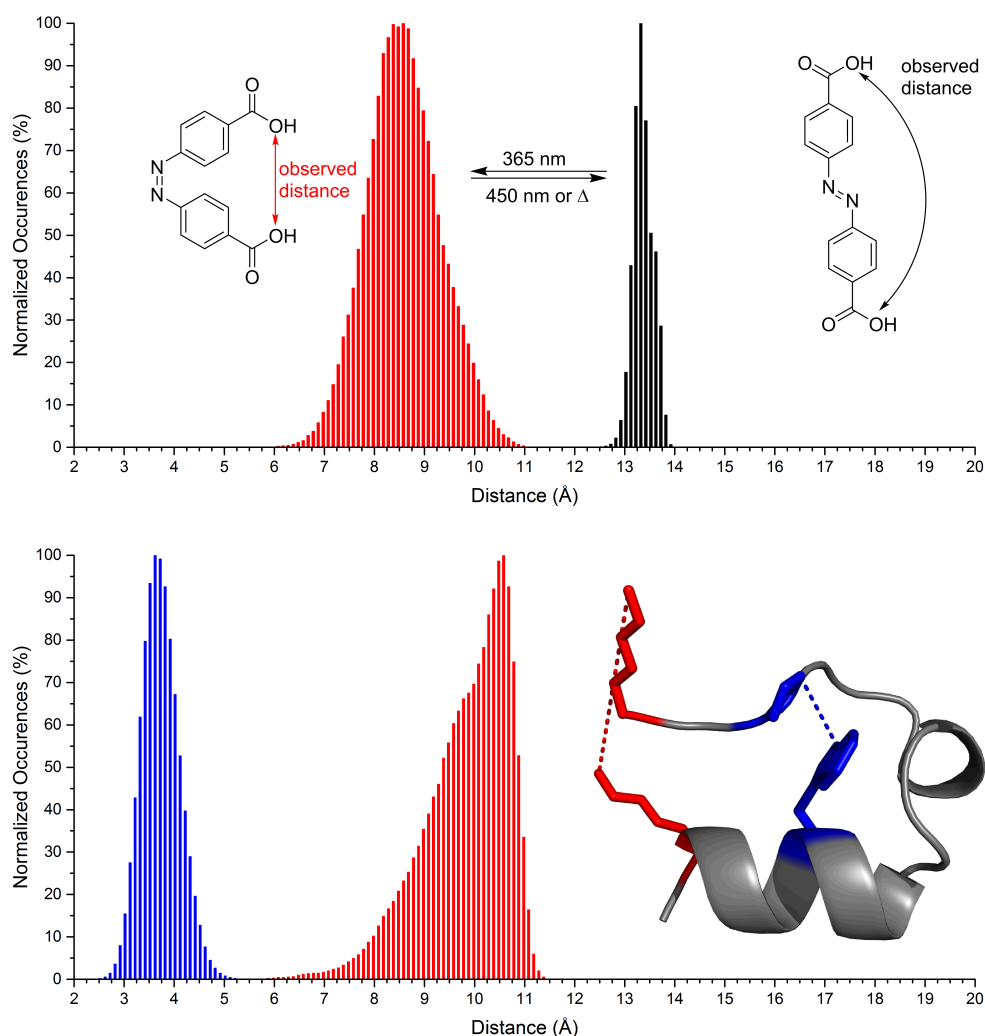
Cross-linking the Trp-cage variant TcKKA requires a linker with a selective reactivity towards primary amines, such as the Lys side chain. A functional group, which meets this criteria is the N-hydroxysuccinimide ester.



**Scheme 1.3.2.:** Scaffold of an azobenzene based photochromic linker with NHS functionalized reactive groups for cross-linking (left) and LOMANT'S REAGENT (right).

Based on the azobenzene scaffold, a photo-switchable cross-linker with NHS groups was designed and synthesized (see Experimental Section subsection 2.3.1 for details) for cross-linking with TcKKA. In order to verify the design of the switchable cross-linker, the length and the flexibility of the linker and Trp-cage side chains were determined by MD simulations of 4,4'-dibenzoic acid azobenzene and the TcKKA.

The linker dithiobis(succinimidylpropionate) is known as LOMANT'S REAGENT (right structure, Scheme 1.3.2). It comprises two NHS ester functionalities and is commercially available. This linker exhibits no photochromic switching properties and is fully flexible. The spacer arm has a length of eight atoms, resulting in a maximum span of 12 Å.<sup>85</sup> Furthermore, this linker has a disulfide bridge, which can be *opened* by reduction (with common reductants, such as TCEP) and *closed* by oxidization (i.e. at high pH and atmospheric oxygen). LOMANT'S REAGENT was chosen primarily to test and optimize the reaction conditions for cross-linking the Trp-cage TcKKA. Moreover, it will reveal possible sterical hindrances in the protein or the linker interferes with the Trp-cage structure.



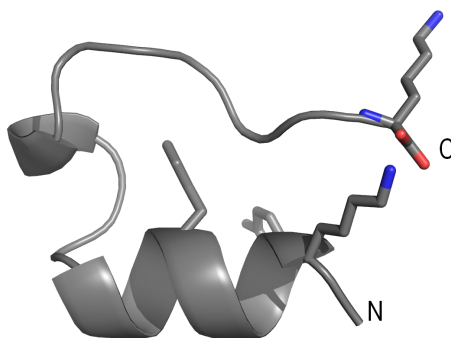
**Figure I.3.4.:** Top: Distribution the distance between *O-6* and *O-6'* in MD calculations of 4,4'-dibenzoic acid azobenzene in the *cis* (red) and *trans* (black) configuration. Bottom: Distance distribution of Lys side chain N $\epsilon$  of Tc10b-A2K-K8A-S20K (TcKKA) (red) and the distance distribution between Trp 6 H $\epsilon$  and Pro 18 H $\beta'$  (blue) as control for the presence of the intact fold of the cage.

MD simulations were employed to assess the available distance between the reactive groups of the azobenzene cross-linker (top, Figure I.3.4). In the *trans* configuration, the linker has a narrow distribution around an average length of 13.3 Å. The distance distribution of the *cis* configuration is significantly wider and centered at 8.5 Å. The photoisomerization is thus expected to result in an effective change of distance of 4.8 Å.

A similar calculation was performed for the Trp-Cage focusing on the distance of the reacting amino groups of Lys (2) and (20). The short distance of Pro 18 H $\beta'$  and Trp 6 H $\epsilon$  was monitored as a control for the folded state of the protein. The distance

distribution between the two  $N_{\epsilon}$  nitrogen atoms of the Lys residue at position 2 and 20 shows a distribution maximum at a distance of 10.5 Å. Unexpectedly for two long, flexible and terminal side chains, the distribution has an asymmetric shape and a shoulder (at 9.5 Å, bottom, Figure I.3.4), which suggest that the mobility of one (or both) side chains is constrained and has two major conformations.

The distance distribution of the respective atoms in the *cis*-linker and the protein almost completely overlap, the size of the linker matches and should be compatible and support the folded Trp-cage state. On the other hand, due to the completely separate distance distribution, the *trans*-linker is incompatible with the folded state of the Trp-cage.



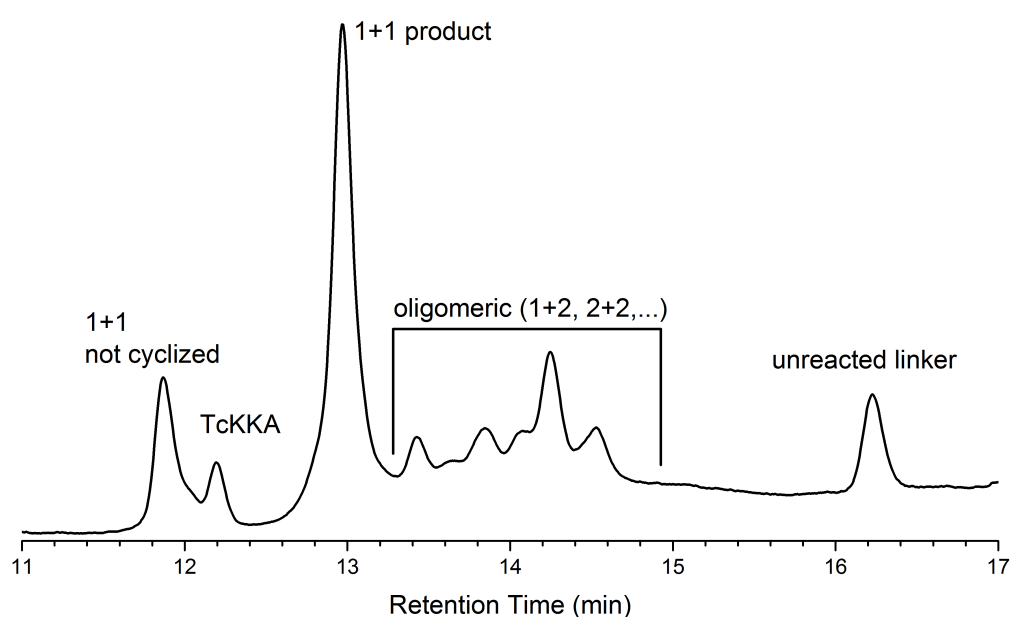
**Figure I.3.5.:** Representative structure of Tc10b-A2K-K8A-S20K (TcKKA) of the MD simulation with highlighted Lys (2 and 20) Tyr (3) and Trp (6) in stick style. Carbon is colored in grey, nitrogen in blue and oxygen in red. The secondary structure is represented in cartoon style.

The unusual distribution shape for the protein prompted us to analyze the conformations of the calculated structures and specifically the mobility of the Lys side chains in more detail. Lys at position 20 (Figure I.3.5) does not experience any restraints and can move freely. The amino function of the Lys side chain in the helix at position 2 frequently interacts with the free, negatively charged C-terminal carboxy group in a preferred, relative defined conformation. On the remaining structures, the side chain conformations are very variable (flexibility), pointing away from the C-terminus. As a result, the distance distribution between the  $N_{\epsilon}$  of both Lys side chains is shaped asymmetrically.

A putative interaction between the C-terminus and the Lys (2) side chain was already suggested by the NMR data (discussed above). Thus, the MD simulation and NMR experiments support each other.

### 3.3. TcKKA Linked with the Lomant-Linker

The Trp-cage TcKKA was cross-linked with LOMANT'S REAGENT and purified as described in the Experimental Section (section 2.2). The reaction was performed in a 1:1 mixture of buffered water<sup>a</sup> and dimethyl sulfoxide (DMSO). The Trp-cage and LOMANT'S REAGENT are completely soluble in this solvent mixture, making the reaction fast but not highly selective. Therefore, exclusive formation of the cyclic 1 + 1 product cannot be expected, but also multiply linked species should form.



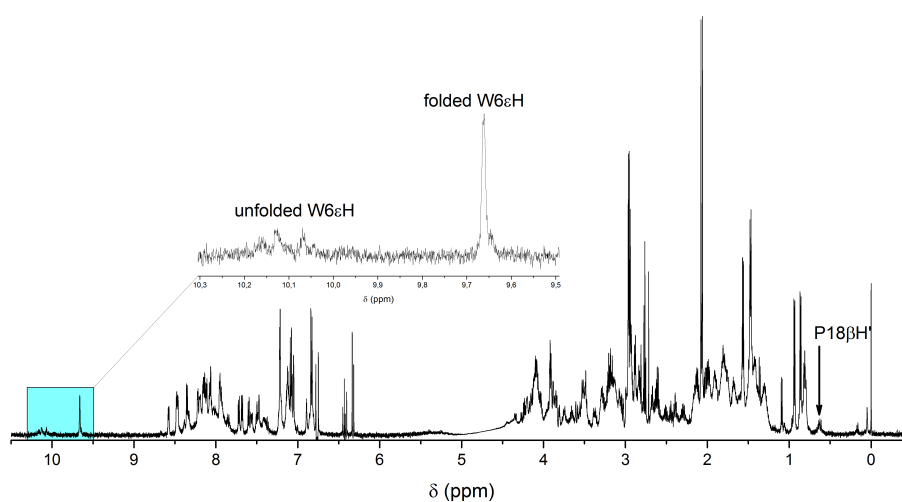
**Figure I.3.6.:** HPLC chromatogram of the reaction mixture of TcKKA with LOMANT'S REAGENT prior purification with a detection wavelength  $\lambda = 220 \text{ nm}$

The reaction progress was monitored by HPLC (Figure I.3.6). The profile shows multiple peaks as expected. The main peak at  $t_R = 13.0 \text{ min}$  is assigned to the cyclic 1 + 1 product based on its ESI-MS spectrum (see Appendix, (Figure V.1.23)). It is noticeable that the product is well separated from the starting materials ( $t_R = 12.2 \text{ min}$  (TcKKA) and  $t_R = 16.2 \text{ min}$  unreacted linker) and from all side products. Most of the side product peaks can also be assigned, such as doubly linked Trp-cage or oligomers of Trp-cage with the linker.

A specific advantage of LOMANT'S REAGENT as cross-linker is its disulfide bridge, as it can be reduced and cleaved. The free thiols can also be re-oxidized, if done under very dilute conditions of the protein, the disulfide bridge reforms selectively intra-molecularly, and all oligomeric species are converted into the desired 1 + 1 product.

<sup>a</sup>pH 7.2, 20 mM Phosphate buffered saline (PBS) and 1 M NaCl

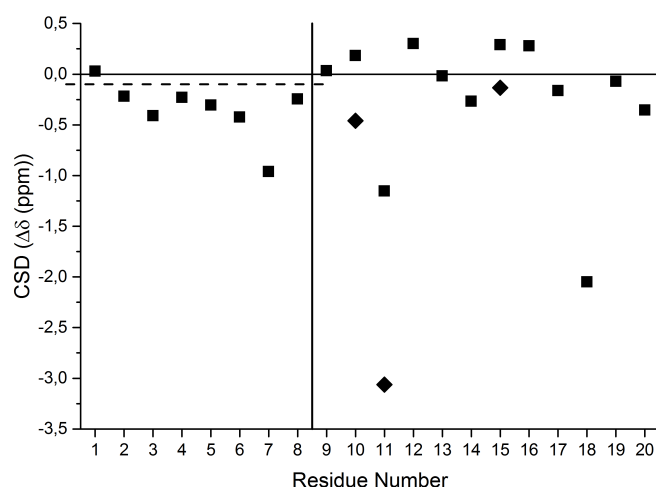
The (1 + 1) LOMANT-linked TcKKA (TcKKA-LOMANT) was purified and characterized by NMR spectroscopy at 298 K and pH 3 (Figure I.3.7).



**Figure I.3.7.:**  $^1\text{H}$ -NMR spectrum of TcKKA linked with LOMANT'S REAGENT at 298 K and pH 3. Magnification of indole proton region.

The  $^1\text{H}$  NMR spectrum (Figure I.3.7) of TcKKA-LOMANT shows characteristics of a folded Trp-cage and is comparable to the unlinked TcKKA. The signal of the indole ( $\text{W6H}\epsilon$ ) of the folded Trp-cage is observed at  $\delta = 9.66 \text{ ppm}$ . In comparison to the unlinked Trp-cage it is shifted upfield by  $0.09 \text{ ppm}$ . Further, the resonance for Pro (18)  $\text{H}\beta'$  is clearly visible with a strong upfield shift ( $\delta = 0.62 \text{ ppm}$ ). The position of these resonances indicate a stabilizing effect of the cross-linker on the protein fold.

Resonances for the *cis*-Pro unfolded species are also observed in the NMR spectrum. These appear comparatively sharp, broad resonances are not apparent. This is likely due to the low signal-to-noise ratio of the data, which also prevents the determination of a *cis*-Pro unfolded to folded  $\text{W6H}\epsilon$  intensity ratio.



**Figure I.3.8.:** CSD plot for the  $H\alpha$ -atoms of TcKKA linked with LOMANT'S REAGENT at pH 3. The  $H\alpha$  are represented as squares and  $H\alpha'$  (Gly) as diamonds. The dashed line is the cut-off for the presence of a helix. The solid line separates the N-terminal helix (residues 1-8) from those of the cage loop (residue 9-20).

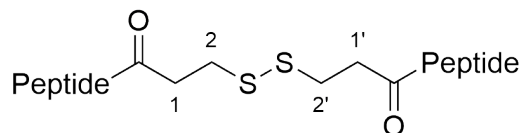
Evaluating the secondary chemical shifts for the  $H\alpha$  (CSD plot) of TcKKA-LOMANT confirms the first impression of the  $^1H$  NMR spectrum (Figure I.3.8). From residue 2 to 8, the deviations of the  $H\alpha$  atoms of the linked Trp-cage are negative and exceed the limit for the presence of a helical secondary structure. From residue 9 on, the typical pattern for the C-terminal Trp-cage loop is observed. In particular, the  $H\alpha$  derivations of Gly (11) ( $\delta\Delta = -3.063 ppm$ ) Pro (18) ( $\delta\Delta = -2.051 ppm$ ) are significantly upfield shifted compared to the unlinked TcKKA. This is a clear evidence that the fold stability is significantly increased.

**Table I.3.4.:** CSDs for selected protons representing the fold of TcKKA-LOMANT at 298 K at pH 3.

Cage Loop	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
	-0.96	-3.06	-2.05	-1.28	-0.55	-0.73	<b>-8.64</b>	<b>90</b>	
Helix	K2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$		
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
	-0.22	-0.41	-0.23	-0.32	-0.43	-0.96	-0.24	<b>-2.81</b>	<b>94</b>

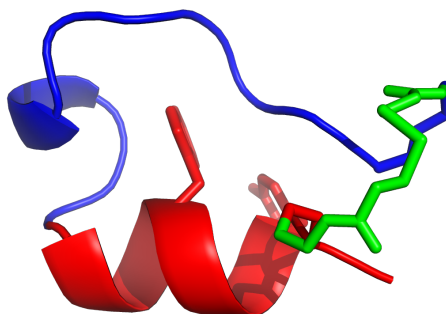
Analysis of the sums of the CSDs support the observation of a stabilizing effect (Table I.3.4). This is particularly reflected in the fraction folded for the cage loop and the helix, as they are significantly increased. In comparison to the unlinked TcKKA ( $\chi_F^{cage} = 67\%$  and  $\chi_F^{helix} = 80\%$ ), the cross-linking results in a significant stabilization

of both structural parts in the Trp-cage.



**Scheme I.3.3.:** Numbered scaffold of LOMANT'S REAGENT.

In order to investigate possible interactions of the linker and Trp-cage, the chemical shifts of the linker were evaluated. The proton bearing ethylene groups of the LOMANT linker exhibit for each group (dashed and not dashed protons) only one set of signals (H-1,1':  $\delta = 2.95$  ppm; H-2,2':  $\delta = 2.76$ ). The observed chemical shifts are observed at in the expected region and exhibit coupling constants usually observed for flexible alkyl chains. This can only occur, when the chemical environment is equal and no interactions of the linker with the protein exist. Support for this is found by analysis of  $^1\text{H}, ^1\text{H}$  NOESY experiments. For the spacer, close contacts to the protein's residues are not observed.



**Figure I.3.9.:** Representative model of TcKKA-LOMANT of MD simulations. The secondary structure is represented in cartoon style and the side chain of Tyr and Trp as well the LOMANT-linker in stick style. The helix is colored in red, the C-terminal cage loop in blue and the LOMANT-linker in green.

Using MD simulations of TcKKA-LOMANT, a model for the cross-linked protein was created (Figure I.3.9). The protein conformation was stable over the entire trajectory showing no deleterious effect of the flexible linker on the protein stability. The linker is positioned near the termini and placed well separated from any protein side chain. It is located on the opposite side of the protein, relative to the tyrosine side chain and is



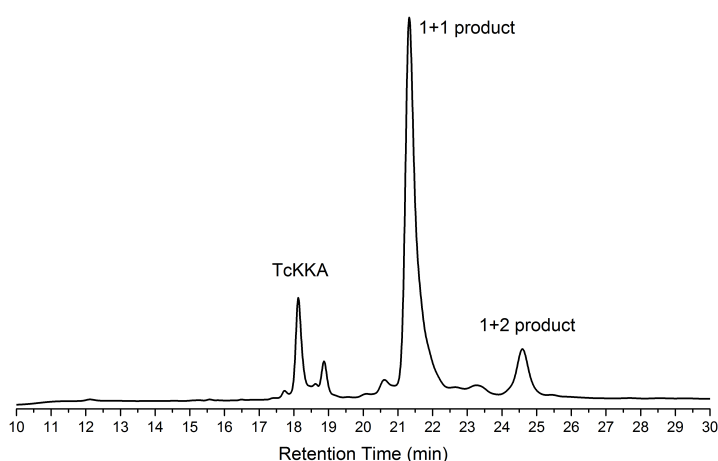
distanced over 5 Å away from the Trp side chain and other. This position is consistent with the NMR data.

The results display that the Trp-cage can be selectively cross-linked using NHS ester. The linker does not interfere with the protein fold. Foremost, the results show that the flexible linker even leads to a significant stabilization of the Trp-cage fold.



## 4. TcKKA Linked with Azobenzene

Following the promising results of the cross-linking with LOMANT'S REAGENT, TcKKA was linked and purified with the NHS functionalized azobenzene under similar conditions. In contrast to LOMANT'S REAGENT, the NHS functionalized azobenzene possesses a limited solubility and precipitated in the reaction solvent (a 1:1 buffered water<sup>a</sup>/DMSO mixture). Owing to the limited concentration of the linker in solution, the reaction rate is very slow resulting in total reaction time of 12 weeks. However, this leads to a highly selective cross-linking, because at low protein and linker concentration, the second step of the cyclization reaction occurs significantly faster intra-molecularly than inter-molecularly. This step determines whether a cyclized 1 + 1 product or a 2 + 1 protein-linker product is obtained.



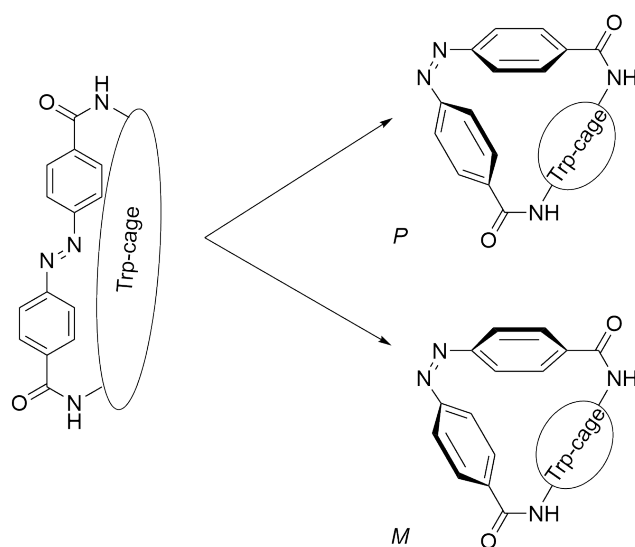
**Figure I.4.1.:** HPLC chromatogram of the reaction mixture of TcKKA with azobenzene prior purification with a detection wavelength of 220 nm.

In the HPLC-chromatogram of the crude reaction mixture prior to purification, three major peaks are observed (Figure I.4.1). The peak at  $t_R = 18.0 \text{ min}$  is assigned to the protein starting material based on ESI-MS data. In the highest peak, the desired 1 + 1 cyclized product TcKKA-azo is observed ( $t_R(\text{TcKKA} - \text{azo}) = 21.5 \text{ min}$ ) and in the peak at  $t_R = 24.5 \text{ min}$  the doubly linked 1 + 2 side product is detected. This profile is significantly simplified compared to the previous cross-linking attempts with azobenzene. The 1 + 1 product is well separated from the side products and therefore easily purified.

<sup>a</sup>pH 7.2, 20 mM PBS and 1 M NaCl

### 4.1. Properties of TcKKA-azo

For a protein, which is cross-linked with a photo-switchable linker, several aspects need to be considered for its characterization. On the one hand, the photochromic properties, such as switching efficiency, reversibility and relaxation, of the linker in the protein environment are of interest, as they characterize the linker and provide first insights of interactions between protein and linker.

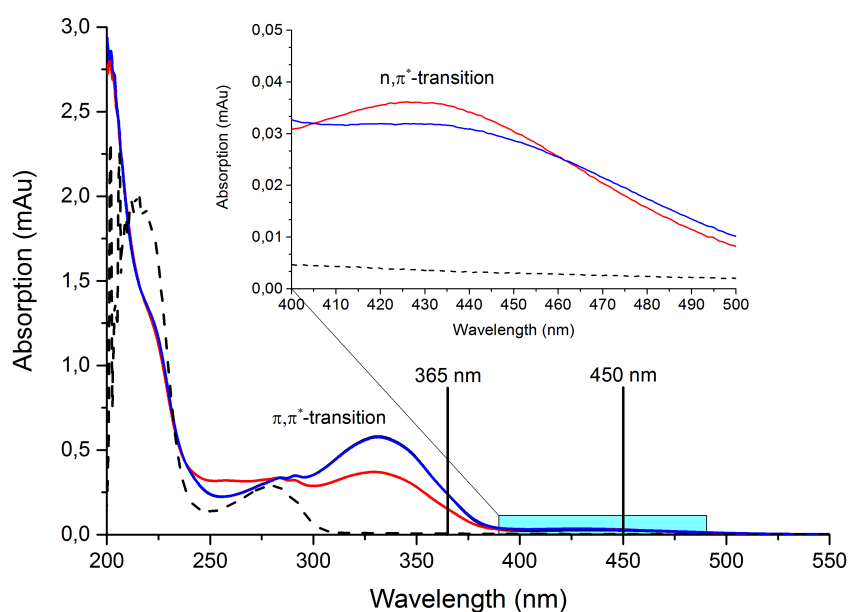


**Scheme I.4.1.:** Scaffold of the switching of azobenzene linked to the Trp-cage. The *cis*-TcKKA-azo (right) can exhibit an additional helical chirality (*P* or *M*), induced by the linker.

On the other hand, the structural characteristics of the protein (and the linker) display the influence of the cross-linker in its different states of *trans*- and *cis*-configuration on the protein structure. In this context, the stereochemical aspects of the linker are also important. The azobenzene cross-linker can induce additional chirality in its *cis* configuration, as it exhibits two enantiomeric forms (Scheme I.4.1). While in azobenzene compounds without constraints both enantiomers are energetically equal, in constrained (i.e. cyclized) azobenzene compounds one enantiomer may be energetically preferred. In the cross-linked Trp-cage *cis*-TcKKA-azo, the protein moiety is chiral. Thus, the additional (*P* or *M*) chirality of the azobenzene causes the formation of two diastereomers. In the presence of (steric) constraints, the energy of the two diastereomers may differ, as well as the kinetics of the photoisomerization rates to the *cis*- or *trans*-form, leading to different populations of the diastereomers.

### 4.1.1. Photochromic Properties

The purified TcKKA-azo was first characterized by ultra violet and visible light (UV/vis) spectroscopy (Figure I.4.2). In addition to the expected absorption at  $\lambda \approx 200 \text{ nm}$  of the peptide bond and the absorption bands between  $\lambda = 250 \text{ nm}$  and  $\lambda = 300 \text{ nm}$  (aromatic residues Tyr and Trp), bands are observed above  $\lambda = 300 \text{ nm}$ . For the azobenzene linker, two absorption bands are expected, the intense  $\pi, \pi^*$ -transition in the UV region and the significantly weaker  $n, \pi^*$ -transition in the visible region.<sup>86</sup> Differences between the *trans* and the *cis* isomer can be observed for both transitions. In the *trans* configuration, the  $\pi, \pi^*$ -transition is very intense, whereas the forbidden  $n, \pi^*$ -transition is much weaker. In the *cis* configuration, the  $\pi, \pi^*$ -transition is shifted to shorter wavelengths and much weaker compared to the *trans*. In contrast, the  $n, \pi^*$ -transition is allowed, resulting in an increased intensity.

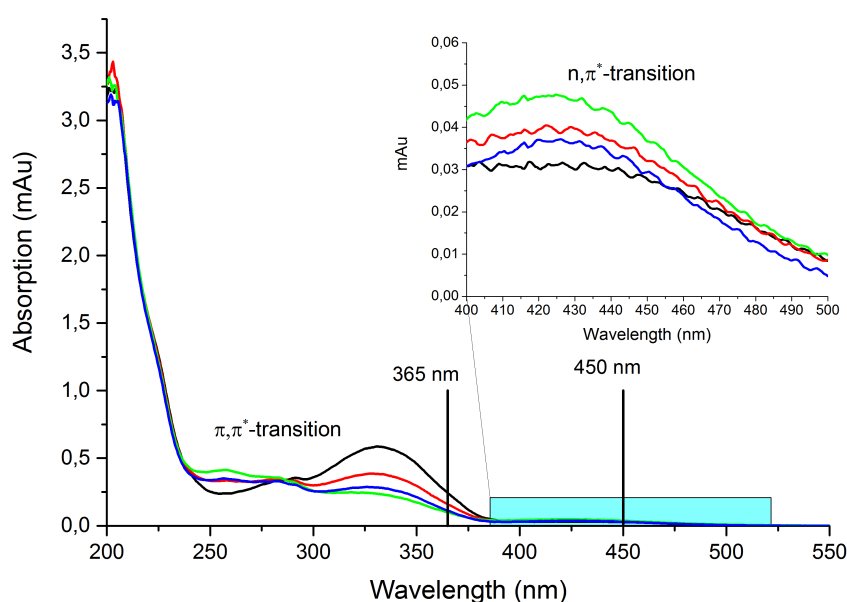


**Figure I.4.2.:** UV/vis spectra of TcKKA-azo after irradiation to the PSS at  $\lambda = 450 \text{ nm}$  (blue spectrum) and  $\lambda = 365 \text{ nm}$  (red spectrum) at 298 K and pH 3 and as comparison the UV/vis spectrum of TcKKA-unlinked at pH 7.9 (dashed black line) with a magnification between 225 nm to 525 nm.

For TcKKA-azo, UV/vis spectra were recorded after irradiation to the *trans*- photo stationary state (PSS) (no changes in the spectra observed) with an excitation wavelength of 450 nm (90% *trans* isomer). The commonly used excitation wavelength of 365 nm was used to reach the PSS of the *cis* isomer (40% *cis* isomer) (Figure I.4.2). In the spectrum of the PSS at 450 nm, the intense  $\pi, \pi^*$ -transition band is observed at  $\lambda = 332 \text{ nm}$  and the weak  $n, \pi^*$ -transition band at  $\lambda \approx 450 \text{ nm}$ . Upon irradiation to the PSS at

365 nm, the intensity of  $\pi, \pi^*$ -transition band diminishes and at  $\lambda = 432$  nm a band is clearly observed. This absorption band is assigned to the  $n, \pi^*$ -transition band of the *cis*-isomer. It is notable that the shift is caused by the increased population of the *cis*  $n, \pi^*$  band.

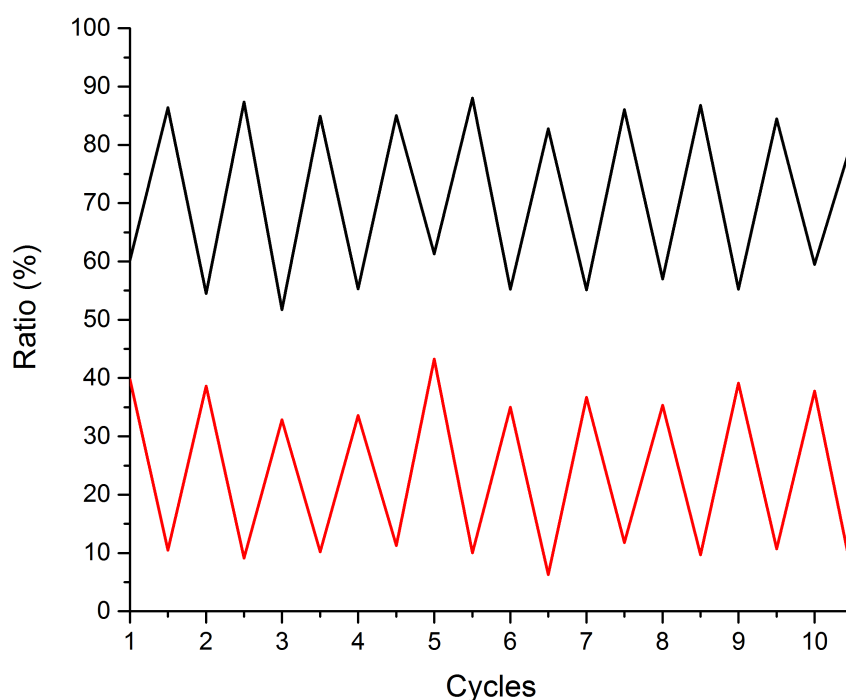
However, comparison of the maximum at  $\lambda = 332$  nm in the UV/vis spectrum after irradiation at a wavelength of 450 nm and the irradiation at  $\lambda = 365$  nm to the PSS suggests that the excitation wavelength may not be efficient and need further optimization to increase the *trans* to *cis* isomerization.



**Figure I.4.3.:** UV/vis spectra of TcKKA linked with azobenzene at pH 3. The spectra were recorded in the PSS of TcKKA-azo after irradiation at  $\lambda = 450$  nm (black spectrum), at  $\lambda = 365$  nm (red spectrum), at  $\lambda = 340$  nm (blue spectrum) and at  $\lambda = 350$  nm (green spectrum) with a magnification between the  $\lambda = 225$  and  $\lambda = 525$  nm.

In order to improve the switching efficiency of the *trans* to the *cis* isomer, the excitation wavelength of 365 nm was successively changed towards the maximum of the *trans*- $\pi, \pi^*$  transition and the results followed by UV/vis spectroscopy. Adjustments of the wavelength were accomplished by using a fluorescence spectrometer as it allows a continuous change in the excitation wavelengths. In Figure I.4.3, UV/vis spectra of TcKKA-azo in the PSS after irradiation with different wavelengths ( $\lambda = 450$  nm, 365 nm, 350 nm and 340 nm) are shown. After irradiation with  $\lambda = 450$  nm, the azobenzene is converted to 90% into the *trans* configuration (determined by  $^1\text{H}$  NMR spectroscopy). Subsequent irradiation at the commonly used wavelength of 365 nm switches the azobenzene to the

*cis*-isomer. The conversion at this wavelength is determined to be only 40%. By shifting the excitation wavelength near to the absorption maximum of the  $\pi, \pi^*$ -transition of the *trans* isomer to 340 nm, the efficiency of the *trans* to *cis* isomerization is significantly increased. However, optimal efficiency is achieved by irradiation at  $\lambda = 350 \text{ nm}$ . At this wavelength, 30% *trans* configuration and 70% *cis* configuration are obtained in the PSS.



**Figure I.4.4.:** Ratio of *trans* (black) and *cis* (red) linked TcKKA-azo. At integer values, the sample was irradiated with  $\lambda = 365 \text{ nm}$  and in between with  $\lambda = 450 \text{ nm}$ .

Focusing on reversibility of the switching process, ten cycles with irradiation at  $\lambda = 365 \text{ nm}$  for the *trans* to *cis* isomerization and irradiation at  $\lambda = 450 \text{ nm}$  for the *cis* to *trans* isomerization were performed and followed by  $^1\text{H}$  NMR. The ratio of TcKKA-azo with the linker in the *trans* and *cis* configuration was determined using the integral values of the indole proton W6H $\epsilon$ , which exhibits different chemical shifts depending on the configuration of the linker. The sum of both integrals was normalized to 100 (Figure I.4.4). As visualized in the plot, TcKKA-azo can be irradiated multiple times without signs of aggregation, fatigue or photo-bleaching.

Furthermore, the thermal relaxation behavior of TcKKA-azo from its *cis* to *trans* configuration was investigated. Usually, azobenzene relaxes from *cis* to *trans* thermally.<sup>87</sup> The rate depends on ring substitutions, solvent and temperature.

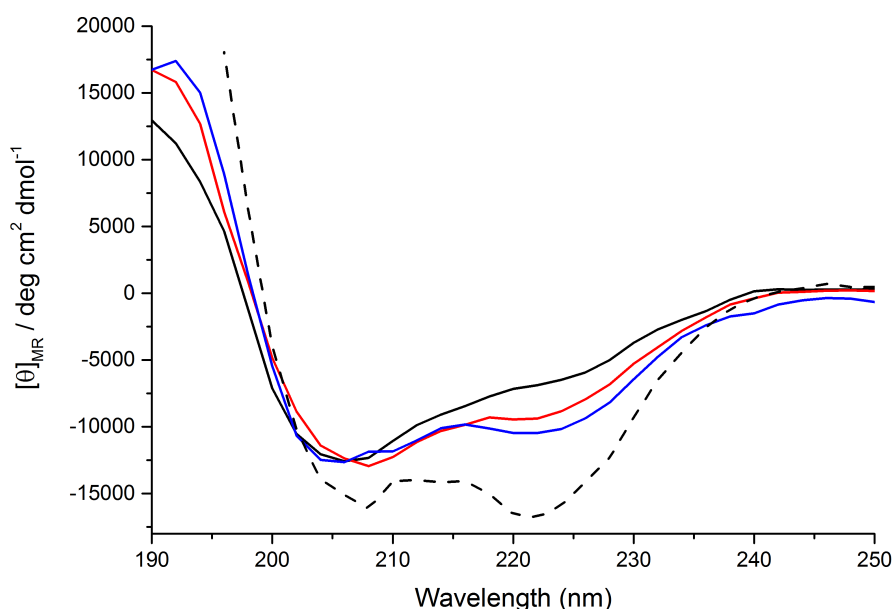
To determine the thermal back relaxation rate, NMR spectra were recorded and the intensity differences over time evaluated. For our cross-linked variant TcKAA-azo in the *cis* configuration, we did not observe any intensity changes over the course of 14 days at 1 °C. This reveals that TcKKA-azo exhibit a very long or absent thermal back relaxation rate. Only by irradiation with  $\lambda = 450\text{ nm}$ , a conversion to *trans* configuration was possible. This property is contrary to observations by N. PREUSSKE during his Masters thesis. He followed the pull approach of Trp-cage conformational switching, in which the azobenzene stabilizes in its *trans* configuration the protein structure and disrupts the fold in the *cis* configuration. In this environment, the *cis*-azobenzene cross-linker exhibits a thermal back relaxation of  $\approx 18\text{ h}$ .

The combined photochromic properties of TcKKA-azo indicate that the azobenzene moiety is influenced by the protein fold of the Trp-cage. Thus, a detailed evaluation of the structural properties of the linker and, foremost, protein fold will allow a delineation of possible interactions between the azobenzene and the Trp-cage in the different configurations of the linker.



#### 4.1.2. Structural Characterization of TcKKA-azo

For a first, overall assessment of the effect of the *cis/trans* configuration switch of azobenzene on the Trp-cage fold, CD spectra were recorded at the PSS with irradiation at  $\lambda = 450 \text{ nm}$ ,  $\lambda = 365 \text{ nm}$  and  $\lambda = 350 \text{ nm}$  (Figure I.4.5). As comparison, unlinked TcKKA at  $1^\circ\text{C}$  in  $0.1 \text{ M NH}_4\text{HCO}_3$  buffer at pH 7.9 is also shown. In the CD spectrum of TcKKA, two minima are observed at  $\lambda = 208 \text{ nm}$  and  $\lambda = 222 \text{ nm}$  and a steep ascent towards lower wavelengths ( $< 200 \text{ nm}$ ). These three characteristics are typical for proteins with helical content, and the shape and magnitude of the absorption curve is consistent with a folded Trp-cage variant.



**Figure I.4.5.:** CD spectra of TcKKA linked with azobenzene at pH 3 and 298 K after irradiation at different wavelengths. After irradiation at a wavelength of  $450 \text{ nm}$  (black), at  $365 \text{ nm}$  (red) and  $350 \text{ nm}$  (blue) and as comparison the CD spectrum of TcKKA-unlinked at  $274 \text{ K}$  and pH 7.9.

The cross-linked TcKKA-azo with azobenzene predominantly in its *trans* configuration (PSS<sub>450</sub>) exhibits a spectrum with a minimum at  $\lambda = 208 \text{ nm}$ , whereas the minimum at a wavelength of  $222 \text{ nm}$  vanished. The ascent towards lower wavelengths is reduced significantly. Both observations are interpreted as a reduction of (helical) structure in the protein.

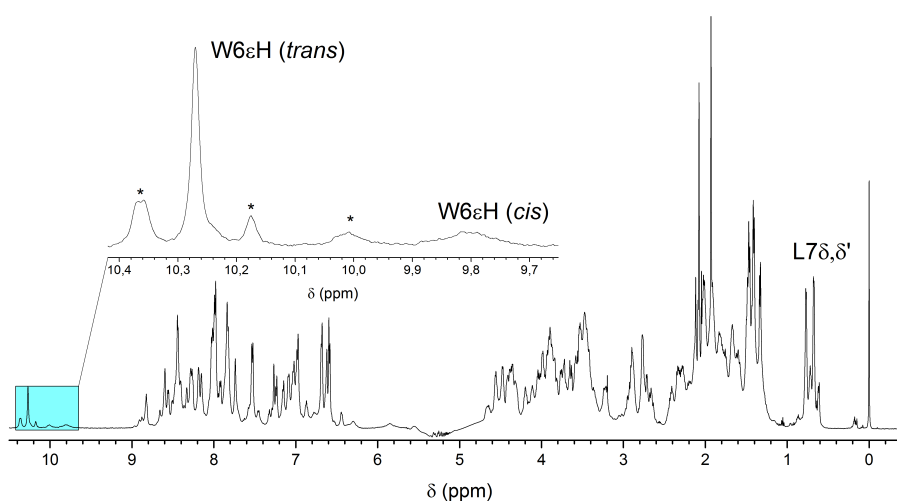
Upon irradiation at wavelengths of  $365 \text{ nm}$  or  $350 \text{ nm}$ , the azobenzene switches from its *trans* to the *cis* configuration. With increasing fraction of *cis*-azobenzene in the sample (40% and 70%, respectively), the minimum at a wavelength at  $222 \text{ nm}$  is present and the

ascent to lower wavelengths is increased. Both observations demonstrate an increase in helicity of the sample and suggest that in the *cis*-azobenzene state the Trp-cage refolds towards its native structure. This is a first evidence for structural effects of the azobenzene on the Trp-cage.

However, since in CD spectroscopy the curves of the mixture components *trans*- and *cis*-TcKKA-azo are averaged, we analyzed the PSS samples by NMR experiments performed at 298 K and pH 3, as in NMR the two isomers lead to a distinct set of resonances.

#### 4.1.3. Structural Characterization of TcKKA-azo in *trans* Configuration

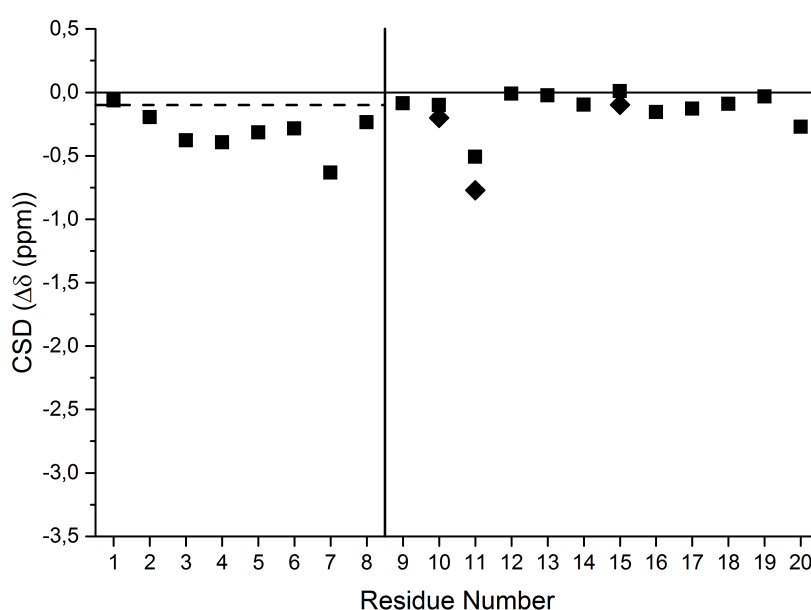
The NMR spectra for TcKKA-azo to characterize the *trans* configuration were recorded at the PSS after irradiation at a wavelength of 450 nm. In this state, 90% *trans* azobenzene are present as determined by the ratio of the integral values of W6H $\epsilon$ . Thereby, the *trans* state is represented by the intensity of resonance lines between 10.2 ppm and 9.9 ppm and the *cis* state by the intensity of the resonance at 9.8 ppm.



**Figure I.4.6.:**  $^1\text{H}$  NMR spectrum of TcKKA linked with azobenzene in the *trans* state after irradiation at  $\lambda = 450\text{ nm}$  at 298 K and pH 3 with magnification of the indole proton region. Signals marked with an asterisk are unfolded W6H $\epsilon$  with Pro in *cis* configuration.

The  $^1\text{H}$  NMR spectrum (Figure I.4.6) displays mainly an unfolded Trp-cage. The indole proton resonance W6H $\epsilon$  of the fast equilibrium folding is shifted downfield to  $\delta = 10.11\text{ ppm}$ , indicating a deshielding of the Trp side chain. Furthermore, additional unfolded indole signals, both downfield and upfield shifted relative to the main signal are observed. Those signals are again assigned to unfolded conformations with *cis*-X-Pro peptide bonds. Their summed intensity is more prominent in comparison to the respective intensity sum in the unlinked, folded TcKKA. This is in good agreement with the

expectation that the probability of the *trans* / *cis* Pro isomerization is reflective of the unfolded all-*trans* equilibrium unfolded population, thus depends on the fold stability. A second major change is the upfield shift of the L7  $\delta$  and  $\delta'$  methyl groups in *trans* TcKKA-azo, indicating a deshielding of the helical side of the Trp side chain relative to unlinked TcKKA. Furthermore, it is observed that no additional prominent aromatic resonances are present in contrast to the cross-linking experiments at Cys residues. This confirms the successful purification and separation from the hydrolyzed linker.



**Figure I.4.7.:** CSD plot of the H $\alpha$  atoms for TcKKA-azo at pH 3 with the azobenzene in its *trans* configuration. The deviation of H $\alpha$  are represented as squares and H $\alpha'$  (Gly) as diamonds. The dashed line represents the cut-off for a present helix. The vertical solid line separates the N-terminal helix (residue 1-8) and the Trp-cage loop (residue 9-20).

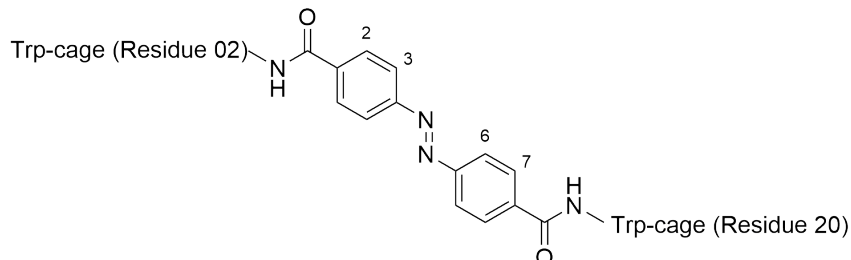
The chemical shifts of the H $\alpha$  atoms for TcKKA-azo were evaluated (Figure I.4.7) for the major of *trans*-azobenzene protein conformation. It is observed that the chemical shift deviations of the N-terminal helix up to residue 8 are continuously negative and exceed the helical limit. This shows that this region remains in a helical conformation and does not unfold as expected. The magnitude of the CSDs even suggest an average helical amount comparable to the helix in the folded Trp-cage. Based on the CSDs from residue 9 onwards, it is however clear that the tertiary fold of the Trp-cage is disrupted as all CSDs are small in their magnitude, including the characteristic deviations of Gly (11) and Pro (18).

**Table I.4.1.:** CSDs for selected protons in the cage structure representing the cage fold and the helix of TcKKA-azo with the linker in its *trans* configuration at 298 K at pH 3.

Cage Loop	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
	-0.63	-0.77	-0.09	-0.00	-0.04	-0.12	<b>-1.58</b>	<b>17</b>	
Helix	K2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$		
	$\Delta\delta(ppm)$							$\Sigma(ppm)$	$\chi_F(\%)$
	-0.20	-0.38	-0.39	-0.33	-0.29	-0.63	-0.24	<b>-2.45</b>	<b>82</b>

The fraction folded determined by the sum of the CSD of fold-characteristic protons for *trans*-TcKKA-azo quantify this observation (Table I.4.1). While the fraction unfolded of the cage loop is only  $\chi_F^{cage} = 17\%$ , the fraction folded for the helix remained nearly unchanged with  $\chi_F^{helix} = 82\%$ , which is comparable to the unlinked TcKKA ( $\chi_F^{helix} = 80\%$ ).

These results display that the azobenzene linker in the *trans* state is able to disrupt the tertiary structure of the Trp-cage, resulting in an unstructured cage loop (residue 9-20). Surprisingly, the secondary structure element, the helix, remains folded, implying the presence of a fold stabilizing effect on the helix, possibly by of the linker in its *trans* state.

**Scheme I.4.2.:** Numbered scaffold of the azobenzene incorporated in the Trp-cage.

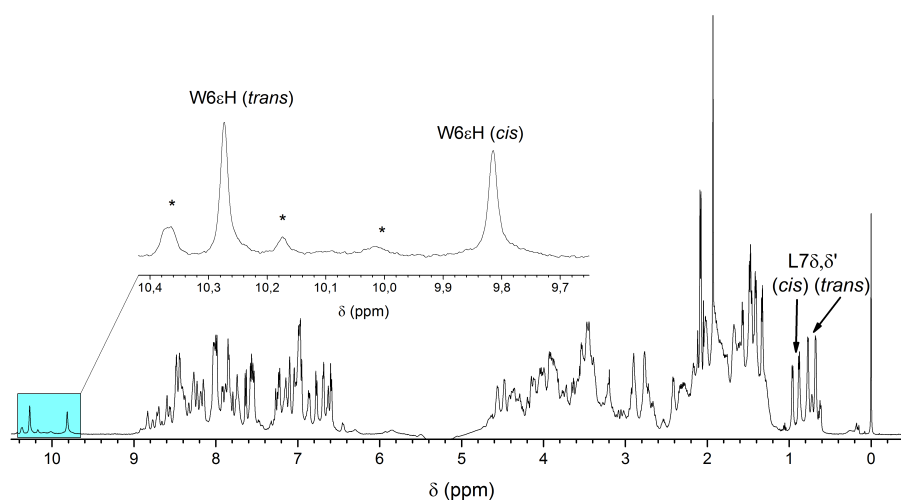
In order to explain the unexpected high helical content after the successful unfolding of the tertiary structure of the Trp-cage, the linker's proton resonances in the *trans* state were evaluated. We focused on the linker properties in the NMR spectra and noticed the incorporated azobenzene exhibits unexpected chemical shifts. In *trans*, the phenyl ring linked to Lys 2 exhibits chemical shifts of  $\delta = 7.99$  (*H*-2,2') and 7.87 (*H*-3,3') *ppm*. The second ring linked to Lys 20 shows chemical shifts of  $\delta = 7.37$  (*H*-6,6') and 7.65 (*H*-7,7') *ppm*. The different chemical shifts are indicative for a different chemical environment for the two phenyl rings in the "supposedly unfolded" state and implies an interaction or influence between linker and protein and the presence of residual helical structure.

Support for an interaction between the azobenzene linker and the protein is drawn from  $^1H, ^1H$  NOESY spectroscopy, in which close contacts between protons are observed. In

this spectrum, we observed for the protons of the phenyl ring linked to residue Lys (2) NOE cross peaks between  $H-2,2'$  and the aromatic side chains, specifically to Y3  $H\delta/H\delta'$  and W6  $\zeta H'$  and  $\eta H$ . In contrast, the second ring exhibits no NOE contacts with proteins in the Trp-cage protein. These observations support the different chemical shifts and can be interpreted as a partial shielding of the aromatic residues of the protein by the linker, which implies the presence of residual structure in the unfolded TcKKA state.

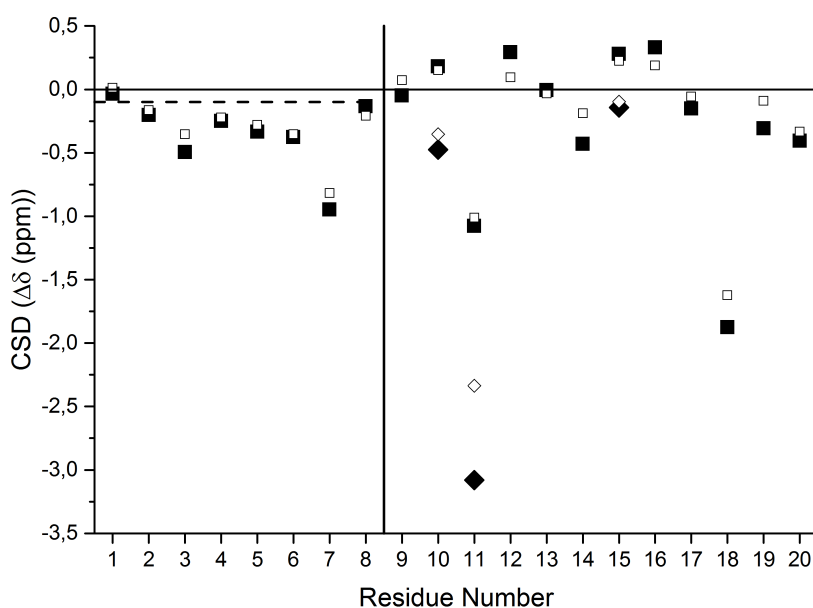
#### 4.1.4. Structural Characterization of TcKKA-azo in the *cis* Configuration

By irradiating the sample at a wavelength of  $365\text{ nm}$ , the azobenzene linker isomerizes from the *trans* to *cis* configuration and yields  $\approx 40\%$  *cis*-TcKKA-azo. By comparison of the spectroscopic data for TcKKA-azo in its *trans* configuration, the resonances of the *cis* configuration were unambiguously assigned by 2D NMR spectroscopy.



**Figure 1.4.8.:**  $^1\text{H}$  NMR spectrum of TcKKA linked with azobenzene in the *cis* state at 298 K and pH 3 with magnification of indole proton region. Signals marked with an asterisk are  $W6\epsilon H$  signals of the unfolded Trp-cage with Pro in *cis*.

In the  $^1\text{H}$  NMR spectrum, a prominent indole proton resonance at  $\delta = 9.79\text{ ppm}$  is observed (Figure 1.4.8). Furthermore, the resonances for L7  $\delta$  and  $\delta'$  methyl groups exhibit typical chemical shifts, comparable with the folded, unlinked Trp-cage. Both observations can be interpreted as a shielding of the Trp side chain and therefore, the presence of the Trp-cage's tertiary structure.



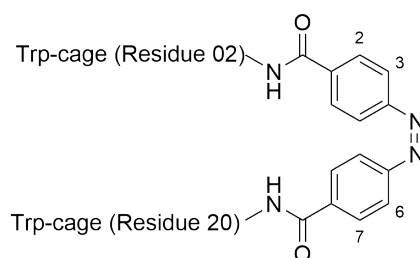
**Figure I.4.9.:** CSD plot of the  $H\alpha$  atoms of TcKKA-azo at pH 3 with the linker in its *cis* configuration. Deviation of the  $H\alpha$  are represented as squares and  $H\alpha'$  as diamonds. The dashed line represents the cut-off for a present helix. The vertical solid line separates the N-terminal helix from the cage loop. As comparison, the CSD of the  $H\alpha$  atoms of the unlinked TcKKA are presented as hollow squares and diamonds.

The CSDs for TcKKA-azo with the linker in its *cis* configuration (Figure I.4.9, filled squares and diamonds) show that the N-terminal part of the Trp-cage until residue 8 exhibits a helical structure. From residue 9 onwards, the typical distribution for a folded Trp-cage loop is observed. Especially Gly (11)  $H\alpha'$  and Pro (18)  $H\alpha$  exhibit a large CSD shift, indicating the presence of the protein's tertiary structure. A comparison with the CSDs of the unlinked TcKKA (hollow squares and diamonds) reveal that not only the expected distribution is observed (evidence for a fully and correctly folded Trp-cage motif), but also that the linker has a stabilizing effect on the fold of Trp-cage leading to increased CSDs in TcKKA-azo.

**Table I.4.2.:** CSDs for selected protons in the cage structure representing the cage loop and helix of TcKKA-azo in *cis* at 298 K at pH 3.

<b>Cage Loop</b>	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$		
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$
	-0.95	-3.08	-1.88	-1.40	-0.47	-0.70	<b>-8.47</b>	<b>88</b>
<b>Helix</b>	K2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$	
	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$
	-0.20	-0.50	-0.25	-0.34	-0.38	-0.95	-0.13	<b>-2.75</b> <b>92</b>

The analysis of the sum of the CSD for representative protons for the Trp-cage fold leads to fractions folded for the Trp-cage loop and the helix (Table I.4.2). The fraction folded of *cis*-TcKKA-azo is determined to be  $\chi_F^{helix} = 92\%$  for the helix and  $\chi_F^{cage} = 88\%$  for the cage loop. These values support the prior interpretations and observations, showing a cross-linked protein with expected and stabilized tertiary structure.



**Scheme I.4.3.:** Numbered scaffold of the azobenzene in *cis* incorporated in the Trp-cage.

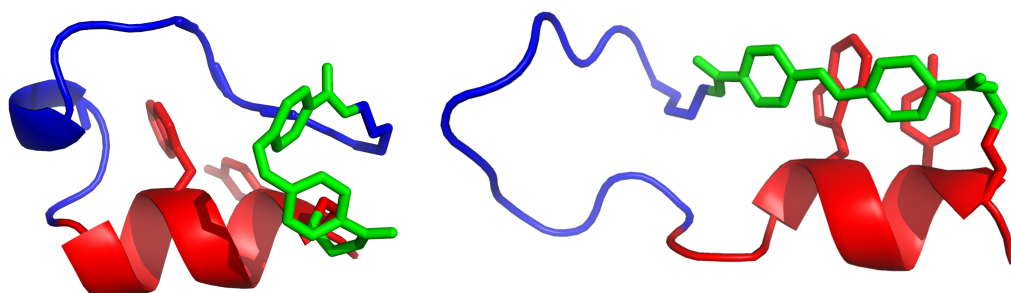
In order to identify possible interactions of the linker in its *cis* configuration and the protein, the chemical shifts of the linker were determined and evaluated. The phenyl ring close to residue 2 exhibits chemical shifts of  $\delta = 7.58$  (*H*-2,2') and 6.82 (*H*-3,3') *ppm* and the second ring linked to residue 20 has chemical shifts of  $\delta = 7.64$  (*H*-6,6') and 6.93 (*H*-7,7') *ppm*. As both the chemical shift of both phenyl rings is almost identical, a similar chemical environment can be expected for them. Furthermore, we did not observe short distance contacts in the  $^1\text{H}, ^1\text{H}$  NOESY spectrum between the protons of linker and the protein. Both aspects, the chemical shifts and the absence of short distances, suggest a flexibility of the connection between the linker and the protein (Lys residues) and an absence of a preferred or fixed conformation of the linker.

The azobenzene cross-linker in its *cis* configuration exhibit two enantiomeric forms (*P* and *M*) in chiral and cyclized compounds, in which one resulting diastereomer may be energetically preferred. This leads to an additional helicoidal chirality. Indications for the presence of such chirality can principally be observed in NMR and CD spectra. In the NMR spectra, one would expect two different sets of resonances for each diastereomer. With CD spectroscopy, the absorption in the UV region (300 – 500 *nm*) of azobenzene can be utilized to detect and determine the chirality of the linker, as the diastereomers absorb left- or right polarized light differently. If both diastereomers are equally present, the effect will cancel out and a flat line is expected. We recorded such CD spectra for our cross-linked Trp-cage, but due to the high concentration needed for such experiment, a reliable measurement could not be achieved with the available protein amount. In the NMR spectra of *cis*-TcKKA-azo, only one set of signals is observed. This suggests that either only one diastereomer is present or both diastereomers (*P* or *M*) exhibit similar

resonances for all protons. The latter explanation is only possible if the chiral elements (protein and linker) are far apart and do not influence each other. A requirement for this in the *cis*-TcKKA-azo is flexibility of the side chains of the anchor residues (Lys 2 and 20). In the presented data, we did not observe any interaction between the linker and the protein, thus the "flexibility" argument is sufficient to explain the lack of observation of resonances from a second diastereomer in the NMR.

## 4.2. Discussion of the Properties of TcKKA-azo

The combined spectroscopic data (chemical shifts and the observations in the  $^1\text{H}, ^1\text{H}$ NOESY spectra) of TcKKA-azo were used to create models of *cis*- and *trans*-TcKKA-azo.



**Figure I.4.10.:** Possible models of TcKKA-azo with the azobenzene linker in *cis* (left) and *trans* configuration (right) based on the results of the spectroscopic data. The secondary structure is represented in cartoon style and the side chains of Tyr and Trp as well as the azobenzene in stick style. The helix is colored in red, the C-terminal loop in blue and the azobenzene in green.

The evaluation of the combined data for *cis*-TcKKA-azo suggests a simple model for the cross-linked protein (left, Figure I.4.10). The azobenzene linker connected to flexible side chains (Lys residues), is located on the side of the protein. This leads to no direct interactions between the protein and the linker. In this state, the N-terminal helix (blue) and the C-terminal cage loop (red) are folded and stabilized.

The spectroscopic data of *trans*-TcKKA-azo lead to a model, in which the cage loop (colored blue) is unfolded and flexible (right, Figure I.4.10). Thus, the Trp side chain is unprotected and exposed to the solvent. While we had expected that this leads to a complete unfolding of the protein helix, the linker instead interacts with the aromatic residues and partially shields the Trp-side chain. Furthermore, this is sterically possible to occur well in the presence of a helical backbone structure, thus stabilizes the secondary structure and prevents helix unfolding. Therefore, the expected unfolded state is not completely unfolded as directly evidenced by observed NOEs, which reveal a specific positioning between the linker and the Trp aromatic side chain. This also leads



to ring current shift of the phenyl rings of the linker, caused by the aromatic Trp side chain, explaining the difference in the chemical shifts of the two ring systems in the linker.

The suggested model of *cis*-TcKKA-azo without direct interactions between the linker and the folded protein is surprising with respect to the absence of thermal back relaxation of the linker from the *cis* to the *trans* configuration. This is a remarkable observation, as the Trp-cage fold exhibits a thermodynamic stability at 280 K ( $\Delta G_U^{280K}$ ) of approximately  $12 \text{ kJ/mol}$  ( $5.5 \text{ kJ/mol}$  at 298 K).<sup>43</sup> In contrast, the *trans* isomer of the azobenzene is more stable (about  $50 \text{ kJ/mol}$ ) than the *cis* isomer and the barrier for a *cis* to *trans* isomerization is about  $100 \text{ kJ/mol}$ .<sup>86</sup> Thus, usually a thermal back relaxation to the ground state (*trans*) is observed for azobenzene and would be similarly expected in the context of the protein. The absence of back relaxation demonstrates that in our cross-linked Trp-cage, the photochromic properties and the stability of the two isomers of the azobenzene linker are strongly influenced by the protein, even if we cannot identify the cause in the model.

The putative interaction between the aromatic rings of the linker and protein in the *trans*-TcKKA-azo provides a good explanation for the yet another unexpected observation of the blue shift of the best excitation wavelength for the isomerization from *trans* to the *cis* configuration. Nevertheless, it was proven that the tertiary structure of the Trp-cage can be influenced differently by the *cis* or *trans* configuration of the linker. Although the helix remains folded in both configurations of the linker, differences in the convolution or a kink in the helix are likely to be present in the *trans* configuration.

In conclusion, we showed that the synthesis of cross-linked protein using NHS esters as reactive groups and Lys side chains as connection points proved to be superior to the commonly used cross-linking chemistry on Cys residues. The cross-linked Trp-cage TcKKA-azo is fully soluble in the *cis* and *trans* configuration and has no tendency of aggregation. By irradiation with light, TcKKA-azo can be switched between the *trans* and the *cis* state with an efficiency of 70% and 90%, respectively. In the *cis* states of the cross-linker, the protein is folded with an enhanced stabilization compared to the unlinked Trp-cage. While in the *trans* configuration the tertiary structure of the protein is disordered, a helical fold was observed due to interactions between the linker and the protein. However, the protein also influences the properties of the linker, as we did not observe a thermal back relaxation from the *cis*-TcKKA-azo to the *trans*-TcKKA-azo. Both properties, the structural and photo-chemical, correlate and have to be optimized to increase the efficiency of the Trp-cage switching module.



**Part II.**

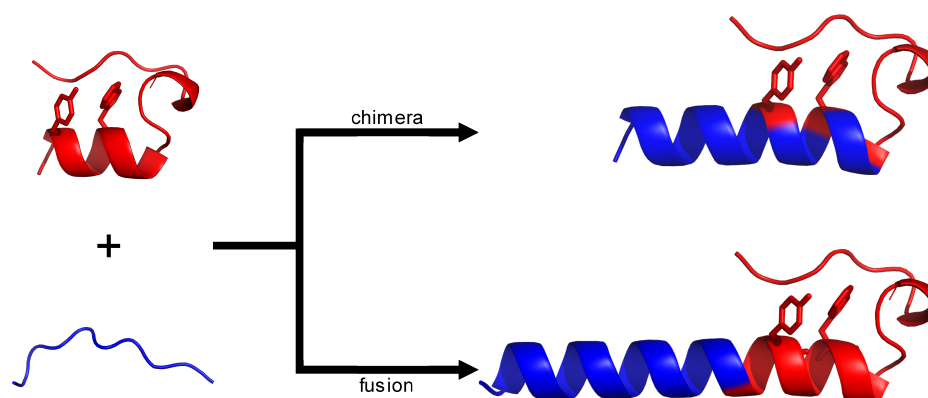
**The AFP-Cage**



## 1. Designing an AFP-Cage

The activity and structure of peptides correlate strongly. For small peptide sequences, however, the structural propensity is limited. Consequently, these peptides exhibit conformational flexibility that reduces the presence of the active conformations. Stabilizing such sequences is therefore of great interest. Numerous studies with the aim of stabilization have been published, using approaches such as the introduction of hydrogen bond surrogates,<sup>88</sup> lactam bridges,<sup>89</sup> or other rigid linkages between side chains. Disadvantages of such modifications are often low yields of the additionally required chemical reaction steps and their incompatibility with large-scale production procedures.

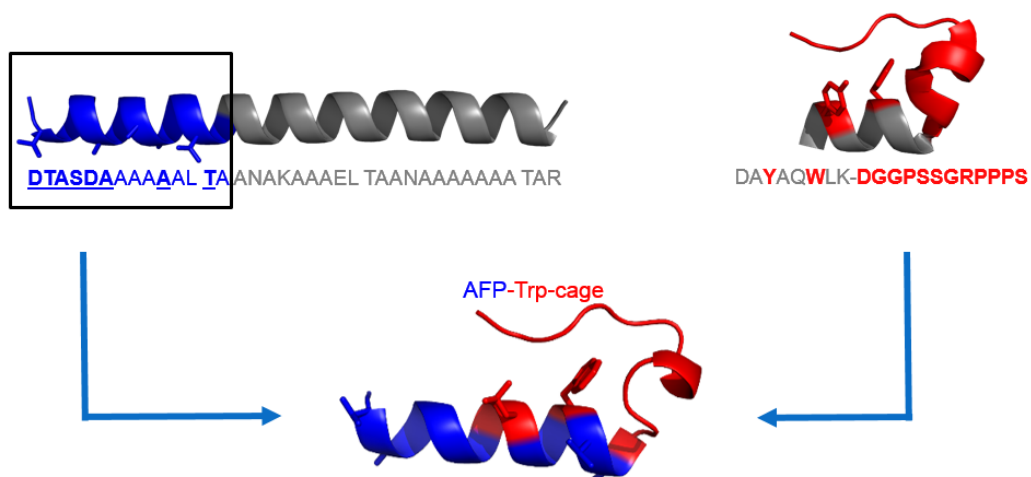
A method that uses standard synthetic methods, does not require chemical modification steps, and foremost leads to a fully natural peptide is of high interest. In the following chapters, we present a strategy that utilizes the stability of the Trp-cage folding motif and confers its fold to short peptide sequences without an intrinsic stable structure. Attaching the sequence of a biological peptide with helical propensity to the N-terminus of the Trp-cage should lead to propagation of the existing helical fold into the peptide sequence.



**Figure II.1.1.:** Concept for combining the Trp-cage with a biological active peptide sequence in two alternative approaches: a) chimera (top) and b) fusion approach (bottom). The secondary structure is represented in cartoon ribbon style. The Trp-cage residues are marked in red and the residues from the biological peptide in blue.

The sequences can be merged in two different ways, either by a direct fusion or a chimera approach (Figure II.1.1). In the fusion approach, the existing Trp-cage helix is N-terminally extended with the peptide's sequence. For varying the orientation of the active side of the peptide relative to the Trp-cage loop, additional residues with helical conformation can be inserted between the Trp-cage helix and the peptide.

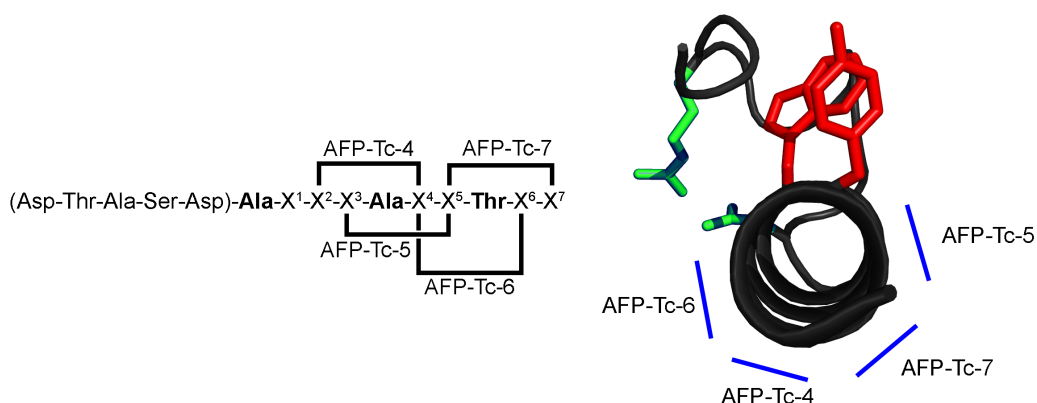
An alternative, more complex strategy is the chimera approach, which is discussed in detail. In this approach, the proteins are not sequentially fused, but the two helical turns of the Trp-cage helix are overlapped with the two C-terminal turns of the biological active peptide. In the Trp-cage helix, the primarily important amino acids for the fold are only the eponymous Trp and Tyr. These interacting side chains need to be placed with a separation of three residues and are located on the same side of the helix. Thus, focusing now on the biological active peptide, only two changes are required in the implementation of this approach: First, the Trp-cage residues Tyr and Trp need to replace residues three and six in the peptide, counting from its C-terminus. Second, the cage loop of twelve residues needs to be attached to the peptide's C-terminus. Because of the mixed primary sequence in the overlapping region, where residues originate from different protein sequences, this sequence-overlapping fusion method is called "chimera" approach. Its advantage is that the chimera protein length is reduced by eight residues compared to the fusion approach. Furthermore, it may better stabilize the helical fold of the biologically active peptide, as the cage stabilization is within the peptide sequence region. On the other hand, the inserted Trp-cage residues and loop structure may sterically interfere with the peptide's biological activity. Thus, the approach can be used only with peptides, where the residues for activity are in regions of the helix not sterically affected (covered) by the Trp-cage loop. An example for the latter are AFPs, as their activity (the ice adsorption surface) is located solely on one side of the helix.



**Figure II.1.2.:** Concept for the design of a chimera protein of a Trp-cage with one repeat of the AFP Type 1 HPLC6. The underlined residues in the AFP sequence are the capping motif and the essential amino acids for the ice-binding. The secondary structure of the backbone is sketched in ribbon cartoon style. The sequence of AFP Type 1 HPLC6 is colored in blue and the Trp-cage in red. The backbone of the resulting AFP-Trp-cage chimera protein (AFP-Tc) is colored by the sequence either in blue for the AFP or red for the Trp-cage residues.

The AFP Type 1 HPLC6 is an alanine rich  $\alpha$ -helical member of the diverse antifreeze protein family. Its active side is formed by an eleven amino acid sequence repeat consisting of **Thr**-(Ala)-X-X-**Ala**-X-X-X-**Ala**-X-X. The X-positions can be any amino acid, whereas Ala is the most frequent amino acid type.<sup>27,90</sup> AFP Type 1 HPLC6 exhibits three repeats of this motif in its sequence (Figure II.1.2, top left). The Thr- and two Ala-residues (highlighted in bold) are essential for the protein's ice activity, as they are located on one side of the helix and form the hydrophobic surface that binds to the ice surface. Terminal sequence variations serve to cap and stabilize the terminal helical turns by hydrogen bonds and to minimize the helical dipole effect under physiological conditions. At the C-terminus, the capping motif consists of an Arg residue, whose positively charged side chain can partially compensate the negatively charged C-terminus and the negative dipole moment of the helix. At the N-terminus, the sequence is terminated by an Asp residue, which reduces the effect of a charge repulsion between the N-terminal charge and the positive side of the helix-dipole moment.

In our pursued chimera approach, we chose to join the sequences of the first ice binding repeat including the N-terminal capping structure of AFP HPLC6 with the Trp-cage motif (Figure II.1.2). In the chosen motif of the AFP, the two essential aromatic residues for the Trp-cage folding motif, Tyr and Trp, can be inserted at any position, that is not essential for the activity (Figure II.1.3). Whereby, the Trp and the Tyr residue have to be inserted in the sequence with a spacing of  $i, i + 3$ .



**Figure II.1.3.:** Left: Sequence pattern of the used ice binding motif including the N-terminal capping sequence (in brackets). The positions marked in bold are essential and the positions marked as  $X^n$  are arbitrary for the activity and can be replaced. Right: View along the helix (from N to C) of the chimera protein with marked orientations (blue lines) of the ice binding surface. The aromatic residues Trp and Tyr (marked in red) and the salt bridge between Asp-Arg (marked in green) are conserved.

Considering the free positions in the chosen ice binding motif and the appropriate separation of the two aromatic residues, four different variants are possible (Figure II.1.3, left). For example, the insertion of the Tyr and Trp residue at the position  $X^2$  and  $X^4$ , respectively, leads to the shortest AFP-Tc chimera AFP-Tc-4 (Figure II.1.3), whereas the longest chimera AFP-Tc-7 results from the insertion of the Tyr and Trp residue at the positions  $X^5$  and  $X^7$ , respectively. The AFP-Tc chimera numbering scheme reflects to the position  $X^n$  of the Trp residue.

The resulting four different AFP-Trp-cage chimera proteins differ primarily in the location of the ice binding surface relative to the cage loop (Figure II.1.3). With the assumption of an ideal  $\alpha$ -helix with 3.6 amino acids per turn, the relative orientation of the ice binding side in relation to the Trp residue is  $+200^\circ$  for AFP-Tc-4,  $+100^\circ$  for AFP-Tc-5,  $-100^\circ$  for AFP-Tc-6 and  $-200^\circ$  for AFP-Tc-7. Thus, the steric position of the ice binding side of the chimeras AFP-Tc-4 and AFP-Tc-7 is almost opposite the C-terminal cage loop. Consequently, we expect no sterical interference of the Trp-cage residues tertiary structure with the AFP activity. In AFP-Tc-5 and AFP-Tc-6, the location of the ice binding surfaces of AFP-Tc-5 and AFP-Tc-6 are in proximity to the cage loop and salt bridge. It is expected that interactions in the AFP's activity can be observed, due to putative steric interference of the aromatic residues of the Trp-cage and the salt bridge in the variants AFP-Tc-5 and AFP-Tc-6, respectively. Although, AFP-Tc-4 and AFP-Tc-7 appeared to be the most promising design, we decided to synthesize all four variants in order to experimentally determine the best chimera and best location of the binding side.

**Table II.1.1.:** Sequences of the four different AFP-Tc- $X^n$  variants. Replacements within the Trp-cage helix are marked in bold.

Name	Sequence		
	Helical Extension	Chimera Helix	Cage Loop
<b>Tc10b</b>		DAYAQWLK	DGGPSSGRPPPS
<b>AFP-Tc-4</b>	DTASD	<b>AA<del>Y</del>AAWLT</b>	DGGPSSGRPPPS
<b>AFP-Tc-5</b>	DTASDA	<b>AA<del>Y</del>AAWTA</b>	DGGPSSGRPPPS
<b>AFP-Tc-6</b>	DTASDAAA	<b>AA<del>Y</del>LTWLA</b>	DGGPSSGRPPPS
<b>AFP-Tc-7</b>	DTASDAAAA	<b>AA<del>Y</del>TAWLA</b>	DGGPSSGRPPPS

In the following, the different sequence regions of the chimera are named the "helical extension" (AFP sequence), the "chimera helix" (combination of AFP and Trp-cage sequence) and the "cage loop" (Trp-cage sequence) (Table II.1.1). The chimera sequences are numbered based on the original Trp-cage sequence, starting at residue 1 of the Tc10b Trp-cage, which is equivalent to the beginning of the chimera helix sequence and position 9 as the start of the cage loop. The AFP sequence precedes, its residues are numbered from the N-terminus starting with 1'.

The four designed AFP-Tc proteins were synthesized by S. ROTHEMUND (IZKF,



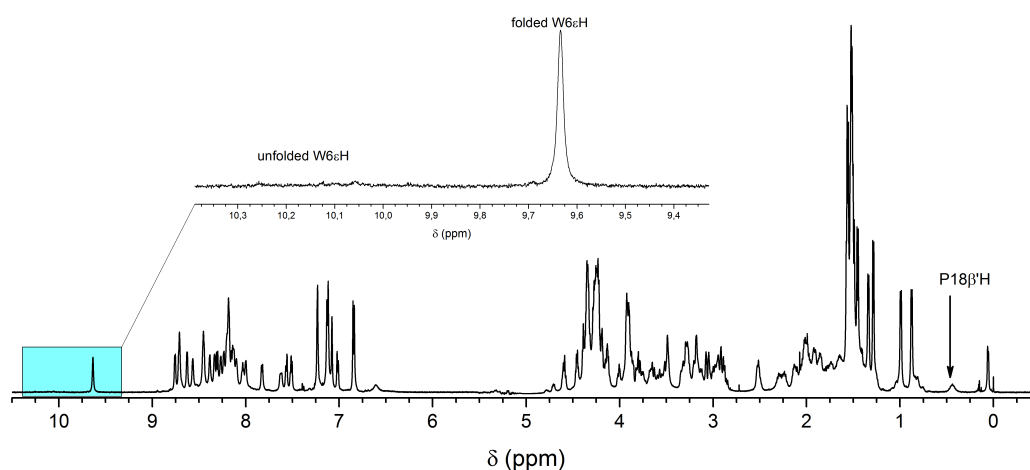
Leipzig). Protocols of the syntheses can be found in the Appendix (section 1.3). In the following chapters, the results of the spectroscopic characterizations of all AFP-Tc chimera proteins are presented. The characterizations were performed by NMR spectroscopy at pH 3 and 7.9 (0.1 M  $\text{NH}_4\text{HCO}_3$  buffer) at 298 K and 274 K. Additionally, CD spectra for each chimera protein were recorded at 274 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer and ice activity assays were performed.

### 1.1. AFP-Tc-7 – DTASDAAAA-AAYTAWLA-DGGPSSGRPPPS

AFP-Tc-7 is with 29 amino acids the longest member of the AFP-Tc chimera family. The two aromatic residues Tyr and Trp are located in the positions X<sup>5</sup> and X<sup>7</sup>, respectively (Table II.1.2). The helical extension includes nine amino acids formed by the five residue long capping sequence Asp-Thr-Ala-Ser-Asp of AFP Type 1 HPLC6 and four Ala residues. The helical sequence is terminated with the C-terminal cage loop.

**Table II.1.2.:** Sequence of AFP-Tc-7 and Tc10b. Replacements within the Trp-cage helix are marked in bold.

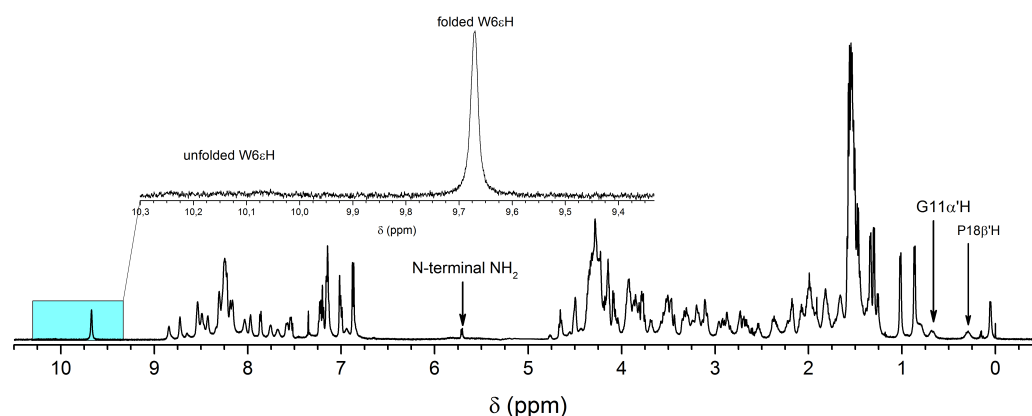
Numbering	1'	1	9
<b>AFP-Tc-7</b>	DTASDAAAA-	<b>AAYTAWLA</b>	-DGGPSSGRPPPS
<b>Tc10b</b>		DAYAQWLK-	DGGPSSGRPPPS
<b>AFP motif</b>	DTASDAXXX-	AXXTXXXXA	



**Figure II.1.4.:** <sup>1</sup>H NMR spectrum of AFP-Tc-7 at 274 K and pH 3 with enlargement of the indole proton region around 10 ppm.

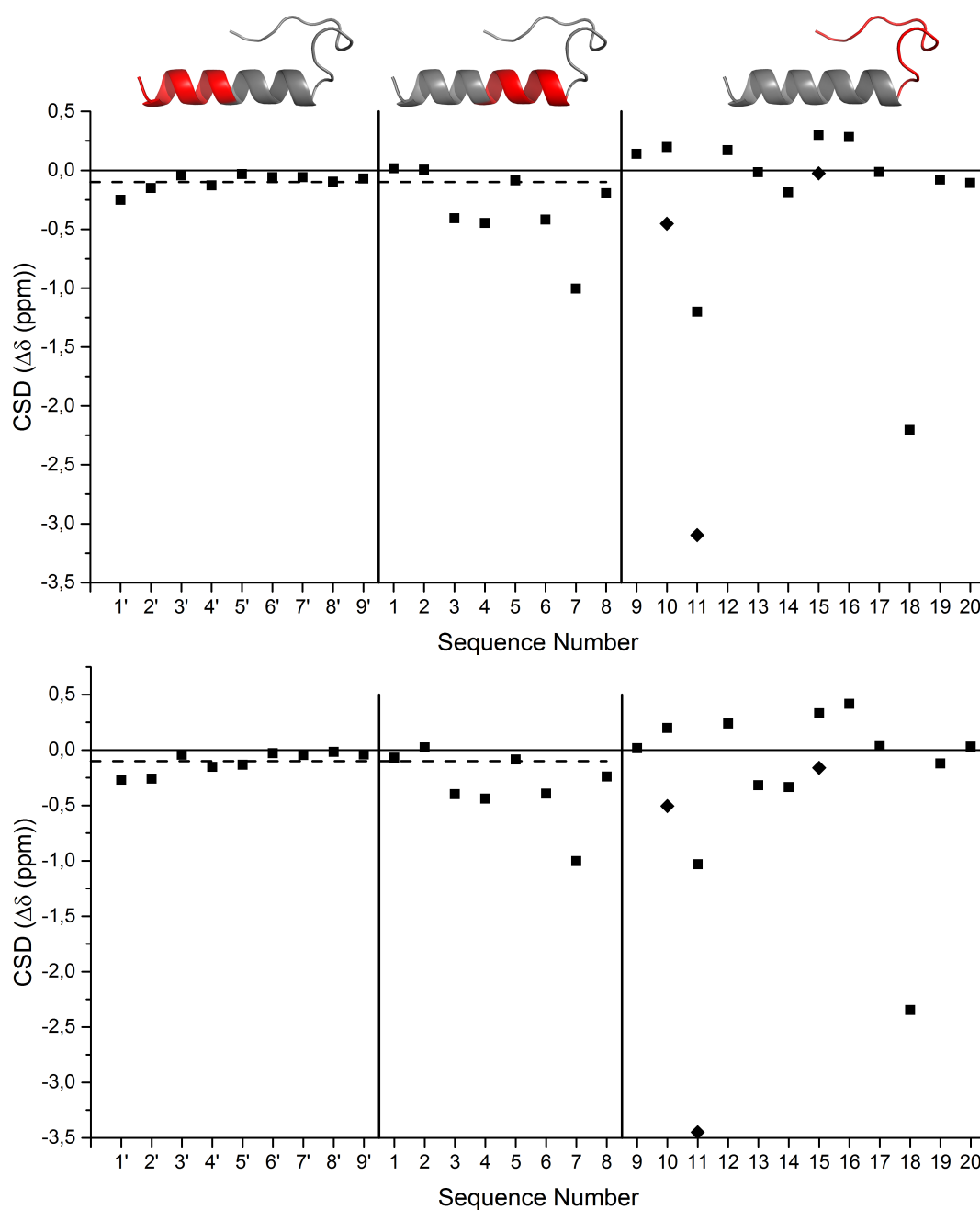
The <sup>1</sup>H NMR spectrum at 274 K and pH 3 is characteristic for a folded Trp-cage (Figure II.1.4), as signified by the unique upfield chemical shift ( $\delta = 9.64 \text{ ppm}$ ) of the indole proton W6H<sub>e</sub> in the fast folding equilibrium. A similar fold-significant observation is the Pro (18) H $\beta'$  resonance at the distinct chemical shift of  $\delta = 0.44 \text{ ppm}$ . Relative to its intrinsic shift, this represents an upfield shift by  $\approx 1.6 \text{ ppm}$  and is explained by the atoms position close to and directly above the aromatic ring of the Trp side chain in the folded protein. Above  $\delta = 10 \text{ ppm}$ , signals of *cis*-Pro unfolded conformations in slow equilibrium are detected. Integration of the fast equilibrium signal and the *cis*-Pro unfolded signals leads to a ratio of 91 : 9. These observations strongly confirm an

existing and stable cage fold.



**Figure II.1.5.:**  $^1\text{H}$  NMR spectrum of AFP-Tc-7 at 274 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer with enlargement of indole proton region around 10 ppm.

At pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer, the spectrum has a very similar appearance. The intensity ratio between the indole resonance  $\text{W6H}\epsilon$  of the *cis*-Pro unfolded conformations above 10 ppm and the equilibrium-folded conformation is with 92 : 8 not altered. The chemical shift of  $\text{W6H}\epsilon$  in the fast folding equilibrium is at  $\delta = 9.67$  ppm. Markedly, beside the Pro (18)  $\text{H}\beta'$  resonance at  $\delta = 0.30$  ppm, another strongly shifted resonance at  $\delta = 0.68$  ppm is clearly visible and is assigned to the Gly (11)  $\text{H}\alpha'$  proton. All three chemical shifts are indicative for the folded protein conformation. In particular, the increased upfield shift of Gly (11)  $\text{H}\alpha'$  and Pro (18)  $\text{H}\beta'$  relative to pH 3 is an expression of a higher stability of AFP-Tc-7 at pH 7.9.



**Figure II.1.6.:** CSD plot of  $H\alpha$ -atoms in AFP-Tc-7 pH 3 (top) and pH 7.9 (bottom) at 274 K. The chemical shift deviation of the  $H\alpha$  is represented as black squares and the deviation of  $H\alpha'$  (Gly) as black diamonds. The dashed line at  $y = -0.1$  ppm represents the cutoff for the presence of a helix. The expected fold is presented above the plot in the ribbon presentation. The vertical solid line between residue 9' and 1 separates the CSD of the N-terminal extension (residues 1'-9') and the solid line between residue 8 and 9 the Trp-cage helix (residues 1-8) from those of the Trp-cage loop (residues 9-20) on the right.

For the purpose of clarity, the fold of the AFP-Tc chimeras is analyzed with a focus on its three parts: the C-terminal cage loop from residue 9-20, the chimera helix from residue 1-8 and the helical extension indicated by dashed numbers (1'-9').

The chemical shift deviations for the cage loop exhibit the unique distribution as observed for other Trp-cage variants described in Part I. In particular, the Gly (11) H $\alpha'$  and Pro (18) H $\alpha$  atoms as representatives for the cage loop, and thus for the tertiary structure exhibit the unusual large negative shift deviation as an effect of a ring-current of the aromatic indole ring system of the Trp side chain.

The CSDs for the chimera helix (residue 1-8) are continuously negative. Except for the first two residues, the limit for a helical structure is exceeded. Notably, the deviations for Tyr (3) and Trp (6) and their neighboring amino acids (i+1) are particularly negative, which is typical for the tertiary Trp-cage structure. The CSDs for the H $\alpha$  atoms in the helical extension (residue 1' to 9') are also continuously negative. However, the limit for an  $\alpha$ -helix is not exceeded for all residues.

The comparison of the CSD plot at pH of 3 and 7.9 shows that the CSD plots are largely alike. Minor differences are observed for the CSDs in the cage loop (residue 9-20), indicating a slight stabilization of the fold at pH 7.9. This is attributed to the intramolecular salt bridge at neutral pH, which stabilizes the cage loop. Both parts of the helix, the chimera helix and the helical extension, exhibit only minor changes in the CSDs, indicating a similar stability of the protein at both pH values.

**Table II.1.3.:** CSDs for selected protons in the chimera AFP-Tc-7 representing the fold of the cage loop and chimera helix at 274 K and pH 3 and 7.9.

<b>Cage Loop</b>	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$			
<b>pH</b>	$\Delta\delta(ppm)$						$\Sigma(ppm)$	$\chi_F(\%)$	
3	-1.00	-3.10	-2.21	-1.47	-0.48	-0.64	<b>-8.89</b>	<b>92</b>	
7.9	-1.00	-3.45	-2.35	-1.61	-0.50	-0.64	<b>-9.55</b>	<b>99</b>	
<b>Chimera Helix</b>	A2 $\alpha$	Y3 $\alpha$	T4 $\alpha$	A5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$		
<b>pH</b>	$\Delta\delta(ppm)$							$\Sigma(ppm)$	$\chi_F(\%)$
3	0.01	-0.41	-0.45	-0.09	-0.42	-1.00	-0.19	<b>-2.55</b>	<b>85</b>
7.9	0.02	-0.40	-0.44	-0.08	-0.39	-1.00	-0.24	<b>-2.54</b>	<b>85</b>

The quantification of the fast equilibrium folding population was performed as described in Part 1. The sums of the CSDs were calculated for the cage loop and the chimera helix, respectively. For the cage loop, the CSDs of Gly (11) H $\alpha'$ , Pro (18) H $\alpha$  and H $\beta'$ , Pro (19) H $\delta$  and H $\delta'$  and additionally of the H $\alpha$  of the residue at position 7 were taken into account (Table II.1.3). For representation of the chimera helix, the sum of the H $\alpha$  CSDs from residues 2-8 is calculated. Evaluation of the sums for AFP-Tc-7 at pH 3 leads to  $\Delta\delta_{cage} = -8.89 ppm$  for the C-terminal cage loop (residues 9-20) and  $\Delta\delta_{helix} = -2.55 ppm$  for the chimera helix (residues 2-8).

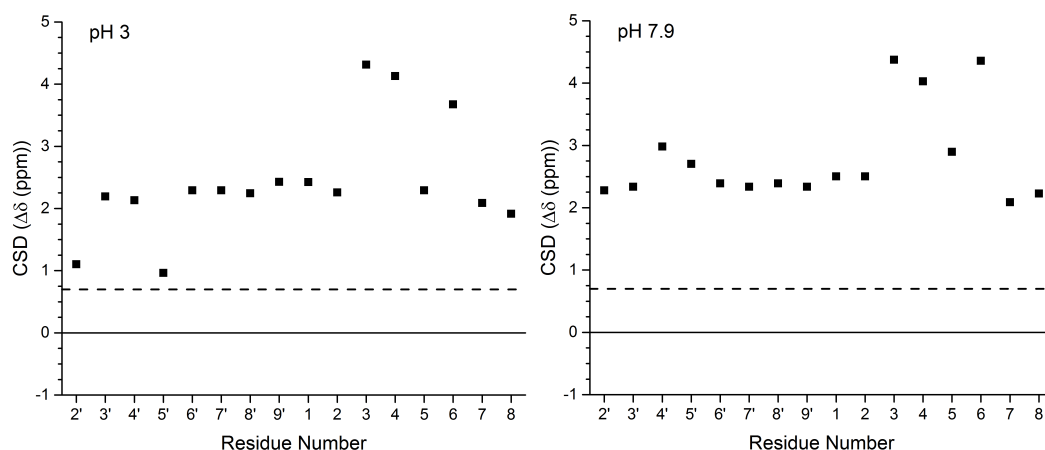
These sums set in relation to the reference CSDs of J. W. NEIDIGH's Trp-cage variant

Tc10b<sup>43</sup> yield a fraction folded in the fast equilibrium  $\chi_F$  as described in Part I, which results for AFP-Tc-7 at pH 3 in a fraction folded for the cage loop of  $\chi_F^{cage} = 92\%$  and for the chimera helix of  $\chi_F^{chim.} = 85\%$ , respectively. At pH 7.9, the fraction folded for the cage loop is determined to  $\chi_F^{cage} = 99\%$  and for the chimera helix to  $\chi_F^{chim.} = 85\%$ .

These fraction folded are appear paradox. The fraction folded for the cage loop with  $\chi_F^{cage} = 99\%$  is completely consistent with the discussed <sup>1</sup>H NMR spectrum. Thus, a fraction folded of only 85% would suggest flexibility (and significantly lower stability). Structurally, this seems impossible as any helical unfolding necessarily leads to an abolishment of all tertiary interactions, thus to the complete flexibility of the loop. Therefore the folding measure for the loop can never be higher than that for the helix. The fraction folded for the helix based on proton chemical shift deviations ( $H\alpha$  CSD) seem to underestimate the stability. This is indeed the case, as shown by comparison of the chemical shifts with those of the parent AFP Type 1 HPLC6 sequence measured at nearly identical conditions.<sup>91</sup> The published HPLC6  $H\alpha$  shifts for the first five residues are identical to our measured values in AFP-Tc-7. Further, in both proteins it is noticeable that the CSD for Ala residues at best reaches and never exceeds the cut-off value of  $-0.1$  ppm. As AFP Type 1 HPLC6 has been shown to be 100% folded at these conditions (cf. its CD spectra<sup>91</sup>), the same implies for AFP-TC-7. Accordingly, the  $H\alpha$  CSD is not a reliable measure of helical content and stability in alanine-rich proteins. An alternative structural measure that may be used is the <sup>13</sup>C $\alpha$  chemical shift, as it also experiences typical upfield or downfield changes depending on the backbone conformation of the respective amino acid.

$$CSD(^{13}C\alpha) = \delta_{obs} - \delta_{ref} \quad (II.1.1)$$

A method for structural interpretation of the <sup>13</sup>C $\alpha$  shifts was established by D. WISHART,<sup>92</sup> in which the observed chemical shifts are subtracted from reference random coil shifts, resulting in the CSDs (Equation II.1.1). If the absolute value of the CSD is larger than  $0.7$  ppm, the presence of a secondary structure is indicated. In contrast to the  $H\alpha$  CSD, a positive deviation of the chemical shift indicates  $\alpha$ -helicity.



**Figure II.1.7.:**  $^{13}\text{C}_\alpha$  CSD plot of the helical region from residue 2' to 8 for AFP-Tc-7 pH 3 (left) and pH 7.9 (right) at 274 K. The horizontal dashed line represents the limit for  $\alpha$ -helicity.

The chemical shift deviations of the  $^{13}\text{C}_\alpha$  in the helical region of AFP-Tc-7 from residue 2' to 8 were evaluated at 274 K and pH 3 and 7.9 (Figure II.1.7). AFP-Tc-7 exhibits consistently positive shift deviations for all residues in the helix at pH 3. All residues exceed the helical limit, indicating the presence of a continuous helicity. Remarkable are the large downfield CSDs of Trp and Tyr residues. Both are essential in the cage fold, and thus not only affected by the secondary structure, but in particular by the tertiary Trp-cage fold. The Thr residue at position 4 also shows an unusually high deviation, which is attributed to an effect of the tertiary Trp-cage structure. In contrast, the CSDs for the residues 2' and 5' are significantly lower compared to the other residues, which might indicate N-terminal fraying.

At pH 7.9, the chemical shift deviations are mostly comparable to pH 3. All residues exhibit a large positive deviation that exceed the helical limit. Noticeable is the change of the CSDs for the residues at position 2' and 5' to larger values. This indicates that the N-terminus of the helix stabilizes, which is attributed to the capping effect of the Asp at position 1'.

$$\overline{\Delta\delta} = \frac{\sum_n(CSD)}{n} \quad (\text{II.1.2})$$

To establish a method to quantify the helical fold and enabling comparisons between the chimeras, the chemical shift deviations of the  $^{13}\text{C}_\alpha$  of the entire helix were summed up and averaged (Equation II.1.2). Furthermore, the helix was divided into the chimera (residues 1-8) and the helical extension (residues 1'-9'). For each, the  $^{13}\text{C}_\alpha$  CSD were summed up and averaged. The observed average chemical shift deviations  $\overline{\Delta\delta}$  for AFP-Tc-7 at pH 7.9 for the complete helix, chimera helix and helical extension are the highest

observed values for all AFP-Tc chimeras. Moreover, CD spectroscopy determines this variant to be  $\geq 100\%$  helical overall (shown below). Consistently, the H $\alpha$  CSDs established the region of the cage loop to be 99% folded and a comparison of the chemical shifts with those of the AFP Type 1 HPLC6 indicates a  $\geq 99\%$  folded conformation in this variant. Thus, we decided to use the  $^{13}\text{C}\alpha$  shifts of AFP-Tc-7 at pH 7.9 and 274 K as a reference for a 99.5% folded AFP-Tc state.

$$\chi_F = 0.995 * \frac{\overline{\Delta\delta}_{obs}}{\overline{\Delta\delta}_{ref}} \quad (\text{II.1.3})$$

With the reference  $\overline{\Delta\delta}$  from AFP-Tc-7 at pH 7.9 and 274 K, the fraction folded for the complete helix ( $\chi_F^{helix}$ ), chimera helix ( $\chi_F^{chim.}$ ) and helical extension ( $\chi_F^{exten.}$ ) can be determined according to Equation II.1.3.

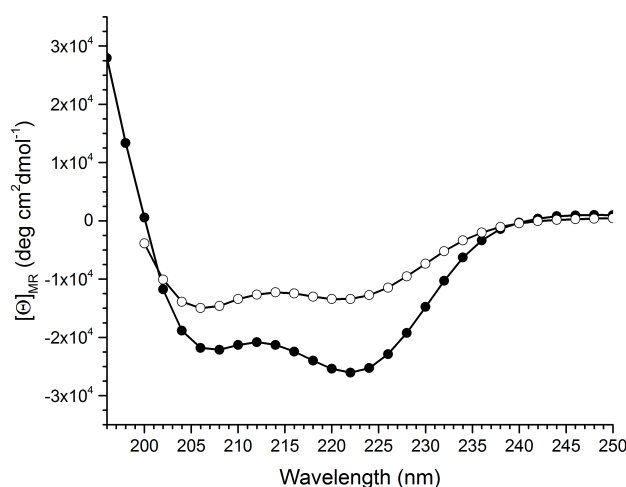
**Table II.1.4.:** Average chemical deviations ( $\overline{\Delta\delta}$ , determined by the  $^{13}\text{C}\alpha$  shifts) and fraction folded ( $\chi_F$ ) for the AFP-Tc-7 at 274 K and pH 3 and 7.9 for the complete helix, the chimera helix and the helical extension. The values of AFP-Tc-7 at pH 7.9 are the reference values and marked in italic.

pH	Helix		Chimera Helix		Helical Extension	
	$\overline{\Delta\delta}$ (ppm)	$\chi_F^{helix}$ (%)	$\overline{\Delta\delta}$ (ppm)	$\chi_F^{chim.}$ (%)	$\overline{\Delta\delta}$ (ppm)	$\chi_F^{exten.}$ (%)
3	2.4	87	2.9	92	2.0	79
7.9	<i>2.8</i>	<i>99.5</i>	<i>3.1</i>	<i>99.5</i>	<i>2.5</i>	<i>99.5</i>

The fraction folded of the entire helix for AFP-Tc-7 at pH 3 displays a decrease of the helicity (Table II.1.4). By comparison the fraction folded for the chimera helix and the helical extension it becomes clear that the decrease in helicity of the complete helix is attributed to the helical extension ( $\chi_F^{exten.} = 79\%$ ). The fraction folded of the chimera helix is in good agreement with the cage loop's. Both exhibit a fraction folded of 92%.

In addition to the NMR characterizations, CD spectra were recorded. In CD spectroscopy, the UV absorption of the amide in the peptide bond is utilized. In general, two transitions are responsible for the CD signal, the  $\pi\pi^*$  transition at  $\lambda = 222\text{ nm}$  and  $n\pi^*$  at  $\lambda = 208\text{ nm}$ . Depending on the chirality of the sample, the absorption of left- or right polarized light differs. The difference spectrum allows a global determination of the predominant fold of the protein, averaged over all residues.





**Figure II.1.8.:** CD spectra of Tc10b (hollow circles) and AFP-Tc-7 (filled circles) recorded at 274 K and pH 7.9. Plotted is the wavelength against the mean residue molar ellipticity.

AFP-Tc-7 and for comparison Tc10b were characterized by CD spectroscopy at 274 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  (Figure II.1.8). Both proteins exhibit CD-curves with similar shape. Specifically, the two minima at wavelengths of 208 nm and 222 nm are characteristic for helical structure. The latter minimum is usually used for quantification of helical content.

$$[\theta]_{222} = \left(1 - \frac{1}{N_r}\right) (-44,000 + 250T) \quad (\text{II.1.4})$$

Following significant prior work, P. LOU and coworkers developed an updated method to calculate the reference mean residue ellipticity at  $\lambda = 222 \text{ nm}$   $[\theta]_{MR}$  for a 100% folded helix (Equation II.1.4, with  $N_r$ =number of residues,  $T$ =temperature in  $^\circ\text{C}$ ).<sup>93</sup> This reference value is required to convert the measured mean residue ellipticity for a given protein into a helical content following Equation II.1.5.

$$\text{Helical content} = 100 * \frac{[\theta]_{MR}(\text{obs})}{[\theta]_{MR}(\text{ref})} \quad (\text{II.1.5})$$

With a measured  $[\theta]_{MR}(\text{obs}) = -26,054 \text{ deg cm}^2 \text{ dmol}^{-1}$  we obtain a helical content of 62% in AFP-Tc-7. Assuming a 100% stable helix, this implies that  $\approx 18$  residues contribute to the helical content. This agrees well with the expected 17 residues long helix in AFP-Tc-7. The difference of one amino acid can easily be explained by further, minor contributions of the loop residues to the  $[\theta]_{MR}(\text{obs})$ . Most likely it originates in a contribution of the C-terminal residue of the helix (D9), which has a helical  $\phi$  backbone

angle. However, we excluded this residue from our stated expectation values based on its non helical CSD value.

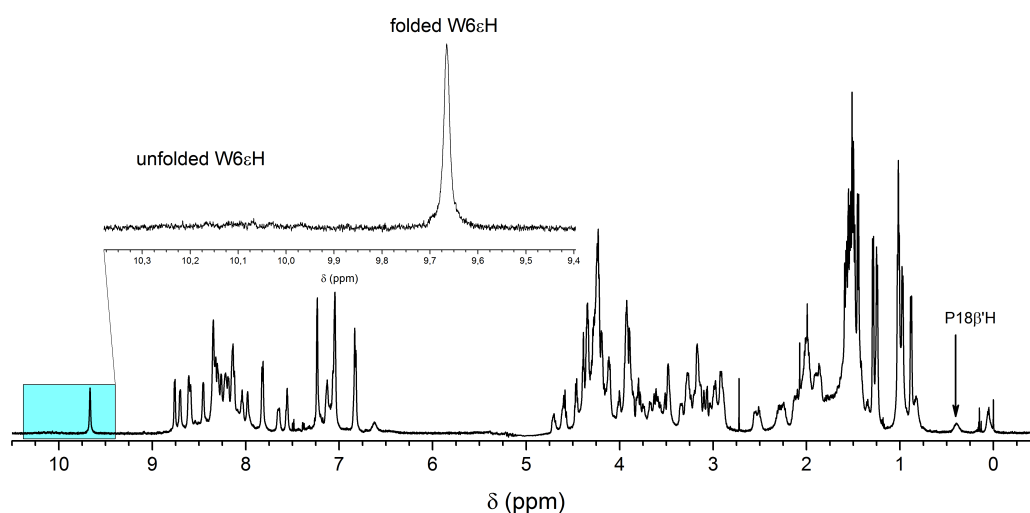
In summary, all experimental data characterize AFP-Tc-7 to be a stably, completely 100% folded protein. Excitingly, all aspects of our chimera approach are validated. The amino acid replacements in the chimera helix are accepted without adverse effects on the Trp-cage's moiety. Not only is the chimera helix maximally folded, so is the helical extension. Thus, the goal of stabilizing the helical structure of the AFP sequence *via* the helix propagation mechanism has been achieved to the maximally possible extent.

## 1.2. AFP-Tc-6 – DTASDAAA-AAYLTWLA-DGGPSSGRPPPS

In the AFP-Tc-6 variant, the folding relevant Tyr and the Trp are inserted in the position X<sup>4</sup> and X<sup>6</sup> of the AFP sequence (Table II.1.5). Due to the helical extension of eight amino acids is the helix 16 residues long overall. The sequence of the AFP is C-terminally terminated with the cage loop motif and yields the 28 amino acids chimera protein.

**Table II.1.5.:** Sequence of AFP-Tc-6 and Tc10b. Replacements in AFP-Tc-6 sequence with Trp-cage residues are marked in bold.

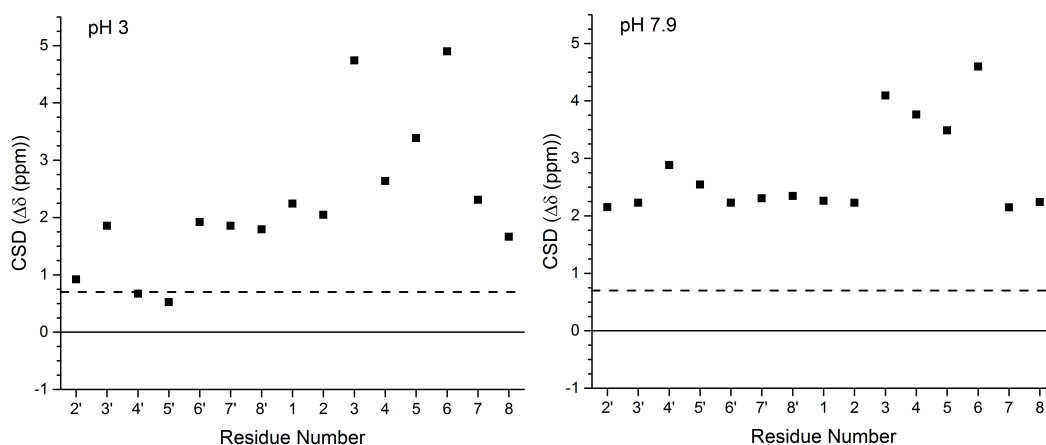
Numbering	1'	1	9
<b>AFP-Tc-6</b>	DTASDAAA-	<b>AAYLTWLA</b>	-DGGPSSGRPPPS
<b>Tc10b</b>		DAYAQWLK-	DGGPSSGRPPPS
<b>AFP motif</b>	DTASDAXX-XAXXTAXX		



**Figure II.1.9.:** <sup>1</sup>H NMR spectra of AFP-Tc-6 at 274 K and pH 3 with enlargement of the indole proton region.

The <sup>1</sup>H NMR spectrum of AFP-Tc-6 at 274 K and pH 3 exhibits the typical separation and shift of the characteristic resonances of the Trp-cage motif (Figure II.1.9). The

characteristic upfield shifted indole proton W6H $\epsilon$  of the Trp side chain is observed at  $\delta = 9.66$  ppm. The ratio between the fast equilibrium signal and the *cis*-Pro unfolded conformations is with 88 : 12 in the expected range. Furthermore, the Pro (18) H $\beta$ ' is visible at  $\delta = 0.40$  ppm. These three observations establish the presence of a very stable fold.



**Figure II.1.10.:**  $^{13}\text{C}\alpha$  CSD plot of the complete helical region from residue 2' to 8 of AFP-Tc-6 pH 3 (left) and pH 7.9 (right) at 274 K.

In order to determine and visualize the structure of the helical region (residue 2' to 8), the chemical shift deviations for the  $^{13}\text{C}\alpha$  of AFP-Tc-6 were analyzed for AFP-Tc-6 at 274 K and at pH 3 and 7.9. At pH 3, all residues display a positive deviation (Figure II.1.10, left). The chimera helix and the helical extension up to the beginning of the capping sequence (residue 6' to 8) continuously exceed the helical limit. Furthermore, the important cage loop residues Tyr and Trp show the expected higher shift deviation in contrast to the other residues. Towards the N-terminus, the CSDs decrease. In particular, Ser at position 4' and the second Asp at position 5' do not exceed the helical limit. This indicates a less stable fold at the first helical turn and suggests some terminal fraying.

At pH 7.9, the  $^{13}\text{C}\alpha$  deviation of all residues increases. The complete region exceeds the helical limit, indicating a significant stabilization of the entire helix at neutral pH. Particularly, the first turn of the helix experiences a significant stabilization.

We quantify the fraction folded for the structural elements as before using the proton H $\alpha$  CSDs for the cage loop and the  $^{13}\text{C}\alpha$  CSDs for the two helical regions. As reference values serve Tc10b for the loop and AFP-Tc-7 for the helix.

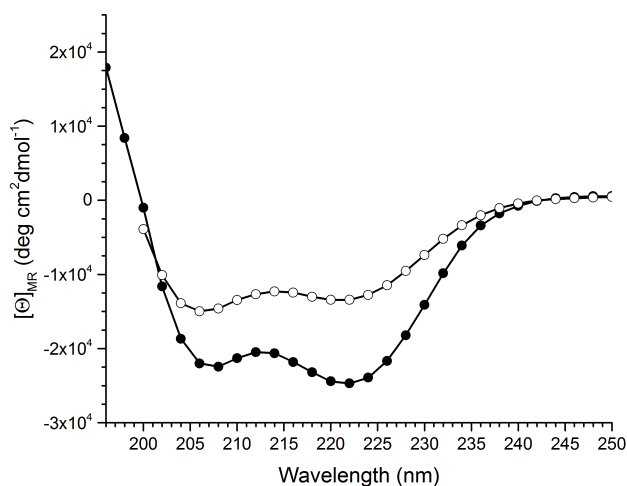
**Table II.1.6.:** Sum of the CSDs ( $^1\text{H}$ ) for the cage loop and its fraction folded and the average chemical deviations ( $\overline{\Delta\delta}$ ,  $^{13}\text{C}$ ) and fraction folded ( $\chi_F$ ) for the complete helix, the chimera helix and the helical extension for the AFP-Tc-6 at 274 K and pH 3 and 7.9.

pH	Cage Loop		Helix	
	$\Delta\delta(\text{ppm})$	$\chi_F^{\text{cage}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{helix}}(\%)$
3	-8.86	92	2.2	80
7.9	-9.49	99	2.8	99

pH	Chimera Helix		Helical Extension	
	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{chim.}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{exten.}}(\%)$
3	3.0	96	1.4	55
7.9	3.1	99	2.4	96

Evaluation of the fractions folded for AFP-Tc-6 shows a stably folded cage loop at both pH values (Table II.1.6). The increase of the fraction folded at neutral pH is attributed to the existing salt bridge in the cage loop. At pH 3, the helix exhibits a smaller fraction folded value compared to the cage loop. This is a result of the less stable helical extension ( $\chi_F^{\text{exten.}} = 55\%$ ). At pH 7.9, the helical extension is significantly stabilized and leads to a completely folded helical region ( $\chi_F^{\text{helix}} = 99\%$ ). These trends in stability and fraction folded at both pH values and different parts in the protein are largely comparable with the chimera AFP-Tc-7.



**Figure II.1.11.:** CD spectra of Tc10b (hollow circles) and AFP-Tc-6 (filled circles) recorded at 274 K and pH 7.9. Plotted is the wavelength against the mean residue molar ellipticity.

CD spectra were recorded at 274 K in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer at pH 7.9. AFP-Tc-6 exhibits a typical helical curve progression with two minima at  $\lambda = 208 \text{ nm}$  and

$\lambda = 222 \text{ nm}$  (Figure II.1.11, filled circle). Compared to Tc10b (hollow circles), the minima values at  $\lambda = 208 \text{ nm}$  and  $\lambda = 222 \text{ nm}$  are significantly smaller, thus indicating a large increase of helicity per residue.

With the measured  $[\theta]_{MR}(obs) = -24,696 \text{ deg cm}^2 \text{ dmol}^{-1}$ , a helical content of 58% is obtained (following the method described by LOU), which is equivalent to 16.4 helical residues in the protein. Considering the expected number of helical residues, including the C-terminal residue of the helix (D9), we obtain a stability of 96% for AFP-Tc-6.

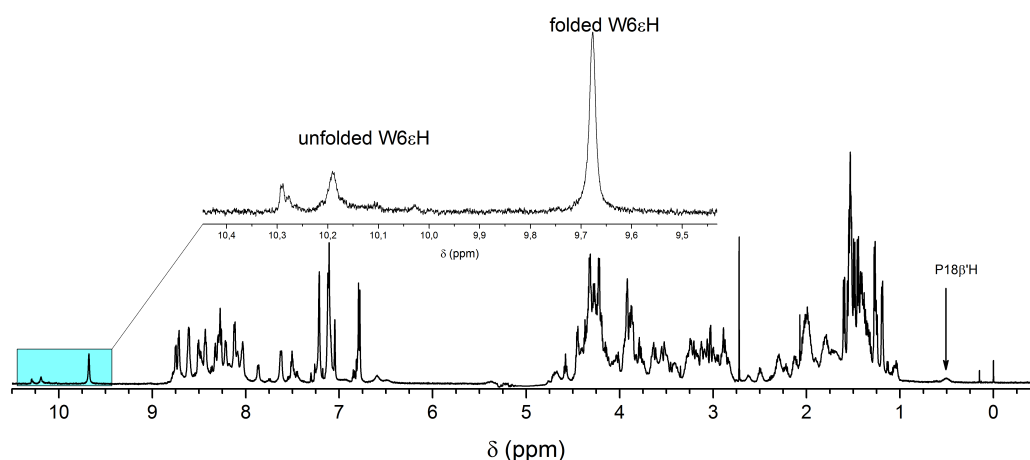
Overall, the structural characterizations show a successful incorporation of the AFP sequence in the Trp-cage. The replacements in the Trp-cage helix and the extension of the helix are well accepted. The chimera protein AFP-Tc-6 is stably folded with a contiguous helix on the entire expected helix length, which proves the success of the chimera approach.

### 1.3. AFP-Tc-5 – DTASDA-AAYAAWTA-DGGPSSGRPPPS

In AFP-Tc-5 are the Tyr and Trp residues located in the position X<sup>3</sup> and X<sup>5</sup> in the AFP motif (Table II.1.7). This results in an N-terminal helical extension of six residues. With the completion of the twelve residue cage loop motif at the C-terminus, AFP-Tc-5 exhibits a total length of 26 amino acids

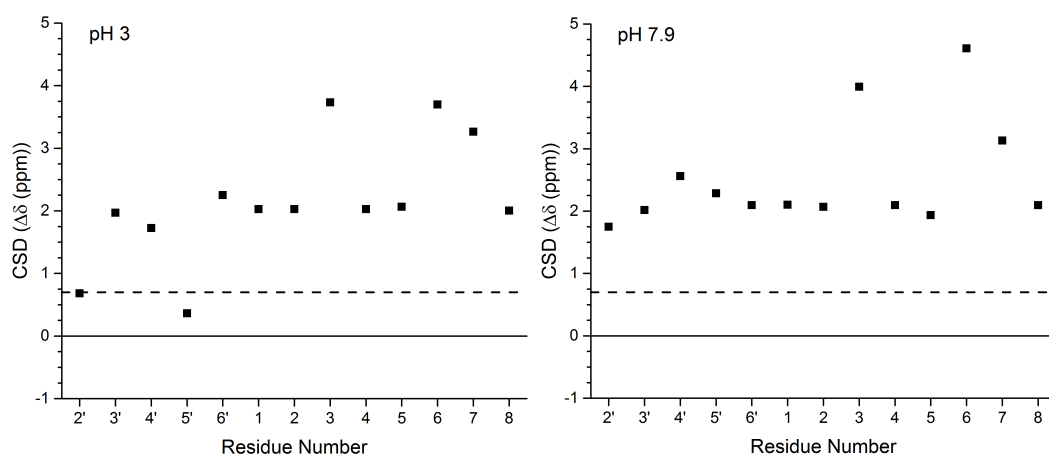
**Table II.1.7.:** Sequence of AFP-Tc-5 and Tc10b. Replacements within the Trp-cage sequence are marked in bold.

Numbering	1'	1	9
<b>AFP-Tc-5</b>	DTASDA-	<b>AAYA</b>	<b>AWTA</b> -DGGPSSGRPPPS
<b>Tc10b</b>		DAYAQWLK-	DGGPSSGRPPPS
<b>AFP motif</b>	DTASDA-	XXXAXXTA	



**Figure II.1.12.:**  $^1\text{H}$  NMR spectra of TcAFP5b at 274 K and pH 3. Magnification of indole proton region.

The  $^1\text{H}$  NMR spectrum of AFP-Tc-5 at 274 K and pH 3 (Figure II.1.12) exhibits the typical characteristics of a folded Trp-cage. The upfield shifted indole proton W6H $\epsilon$  of the Trp side chain is observed at  $\delta = 9.68$  ppm. Furthermore, the characteristic Pro (18) H $\beta'$  shift resonate at  $\delta = 0.51$  ppm. Between 10.0 and 10.4 ppm, the *cis*-Pro unfolded conformations are observable and are more prominent and sharper (10.2 and 10.3 ppm) as for AFP-Tc-7 and AFP-Tc-6. The ratio between the folded and unfolded species is determined to be 67 : 33, which may suggest a largely (but not completely) stable Trp-cage fold.



**Figure II.1.13.:**  $^{13}\text{C}\alpha$  CSD plot of the helical region from residue 2' to 8 for AFP-Tc-5 pH 3 (left) and pH 8 (right) at 274 K.

Analysis of the  $^{13}\text{C}\alpha$  chemical shift deviations of AFP-Tc-5 was performed at 274 K at pH 3 and 7.9, respectively. At pH 3, the deviations for the complete helical region are positive throughout the whole sequence. The helical limit is exceeded from residue 6' through the entire chimera helix until residue 8, implying a continuous helical fold. The N-terminal capping sequence shows decreased helicity, as the CSDs for residue 2' and 5' do not exceed the helical limit. This indicates a destabilization or local disturbance of the secondary structure in these positions near the N-terminus.

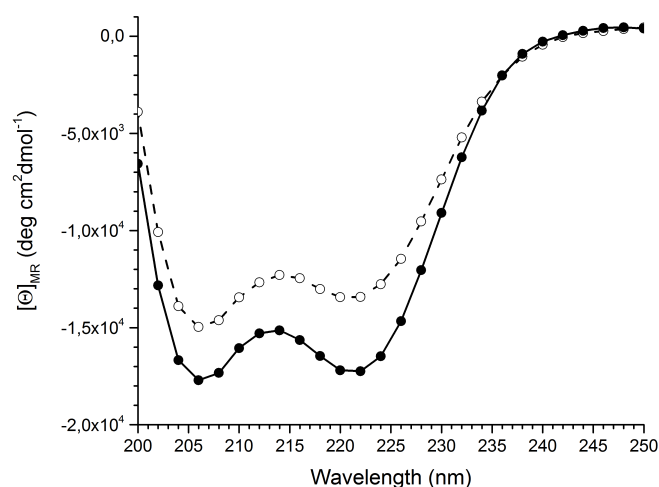
At pH 7.9, the chimera helical region from residue 1 to 8 exhibits a similar distribution and deviation. The N-terminal residues exceed the helical limit and experience a great stabilization effect relative to pH 3.

As for AFP-Tc-7 and AFP-Tc-6, the fractions folded for the different parts in the chimera protein were determined using  $^1\text{H}$  and  $^{13}\text{C}\alpha$  CSDs.

**Table II.1.8.:** Sum of the CSDs ( $^1\text{H}$ ) for the cage loop and its fraction folded and the average chemical deviations ( $\overline{\Delta\delta}$ ,  $^{13}\text{C}$ ) and fraction folded ( $\chi_F$ ) for the complete helix, the chimera helix and the helical extension for the AFP-Tc-5 at 274 K and pH 3 and 7.9.

pH	Cage Loop		Helix	
	$\Delta\delta(\text{ppm})$	$\chi_F^{\text{cage}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{helix}}(\%)$
3	-8.54	89	2.1	77
7.9	-8.27	85	2.5	90
pH	Chimera Helix		Helical Extension	
	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{chim.}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{exten.}}(\%)$
3	2.6	83	1.4	57
7.9	2.8	88	2.1	87

The fraction folded for the helical extension is with  $\chi_F^{\text{exten.}} = 57\%$  at pH 3 and  $\chi_F^{\text{exten.}} = 87\%$  surprisingly low (Table II.1.8). The low stability (and flexibility) of the extension also impacts the stability of the chimera helix, which exhibits a smaller fraction folded compared to the cage loop. Thus, we suggest that a destabilizing effect is present in the N-terminal extension, which influences the first residues in the chimera helix at low pH. This also affects the stability of the cage loop, in particular at neutral pH, as the fraction folded is significantly smaller compared to the prior described variants.



**Figure II.1.14.:** CD spectra of Tc10b (hollow circles) and AFP-Tc-5 (filled circles) recorded at 274 K and pH 7.9. Plotted is the wavelength against the mean residue molar ellipticity.

The CD spectrum (at 274 K and pH 7.9) displays a typical helical curvature with two minima at  $\lambda = 208 \text{ nm}$  and  $\lambda = 222 \text{ nm}$  (Figure II.1.14, filled circles). Compared to Tc10b (hollow circles), both minima are deeper, signifying a larger helical content per residue and consequently of the entire protein.

Following the prior described equations of LOU with the observed mean residue molar ellipticity  $[\theta]_{MR}(obs) = -17,240 \text{ deg cm}^2 \text{ dmol}^{-1}$ , we determined a helical content of 41% for AFP-Tc-5. In consideration of the sequence length of the chimera, 10.7 residues are in a helical conformation. However, we expected 15 residues to be in a helical conformation for AFP-Tc-5, including the Asp at position 9. Comparing the calculated and expected number of helical residues, a protein stability of only 71% is obtained.

It is apparent that the fraction folded of the helix (90%, NMR) and the helical content (71%, CD) significantly differ in their magnitude in this variant. In particular, the fractions folded for the helical extension are unusually and strongly pH dependent. This suggests that a local effect in this region influences the overall stability of the protein, which will be pursued in detail later.

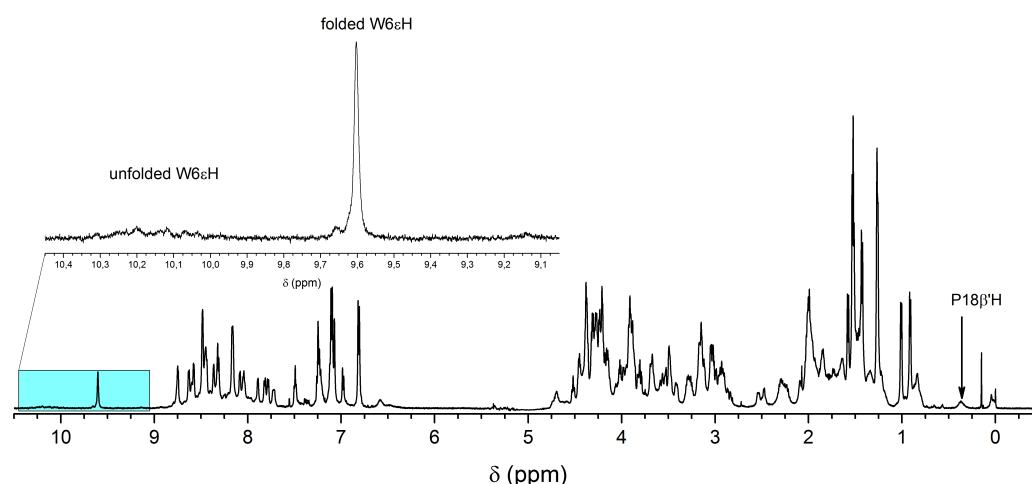


### 1.4. AFP-Tc-4 – DTASD-AAYA~~A~~WLT-DGGPSSGRPPPS

AFP-Tc-4 is with 25 amino acids our shortest design for the junction of the AFP repeat with two Thr and the Trp-cage motif (Table II.1.9). In the AFP motif, the Tyr and Trp residues are inserted at the position X<sup>2</sup> and X<sup>4</sup>, respectively. This leads to an N-terminal extension of five residues. At the C-terminal end of the helix, the cage loop sequence completes the chimera.

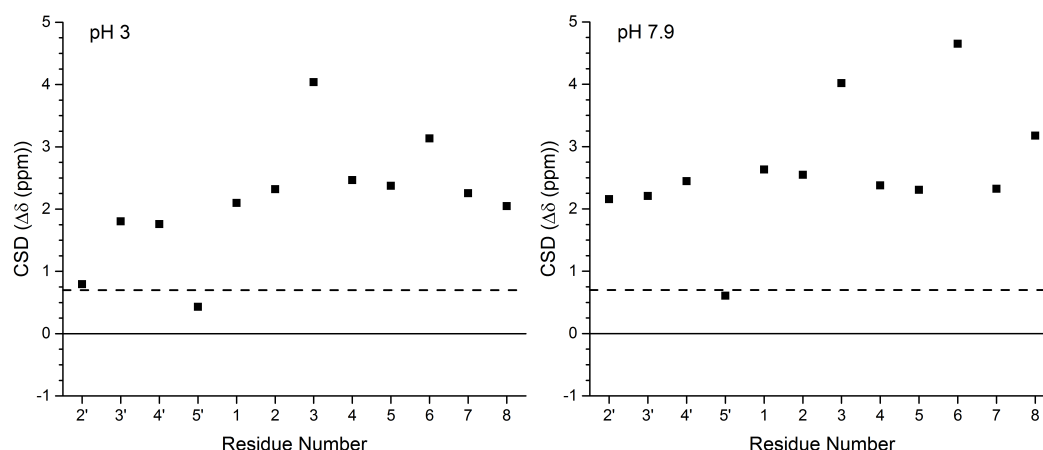
**Table II.1.9.:** Sequence of AFP-Tc-4 and Tc10b. Replacements within the Trp-cage sequence are marked in bold.

Numbering	1'	1	9
<b>AFP-Tc-4</b>	DTASD-	<b>AAYA<del>A</del>WLT</b>	-DGGPSSGRPPPS
<b>Tc10b</b>		DAYAQLK	-DGGPSSGRPPPS
<b>AFP motif</b>	DTASD-	AXXXAXXT	



**Figure II.1.15.:** <sup>1</sup>H-NMR spectra of AFP-Tc-4 at 274 K and pH 3 with enlargement of the indole proton region.

The <sup>1</sup>H NMR at pH 3 spectrum shows all characteristics of a folded Trp-cage, recognizable by the indole resonance W6H $\epsilon$  at  $\delta = 9.60$  ppm (Figure II.1.15). Furthermore, a upfield shifted resonance at  $\delta = 0.37$  ppm is observed. This is assigned to the Pro  $\beta'$  at position 18. As observed for AFP-Tc-5, a significant amount of *cis*-Pro unfolded protein conformations is present. The ratio between the fast equilibrium folding conformation and the *cis*-Pro unfolded conformations is determined to be 73 : 27. These characteristics display in general the Trp-cage fold of the AFP-Tc-4 with indications for minor destabilizations of the fold.



**Figure II.1.16.:**  $^{13}\text{C}_\alpha$  CSD plot of the complete helical region from residue 2' to 8 for AFP-Tc-4 pH 3 (left) and pH 8 (right) at 274 K.

Analysis of the CSD plot of the  $^{13}\text{C}_\alpha$  atoms for AFP-Tc-4 at pH 3 reveals positive deviations for all residues (Figure II.1.16, left). From residue 1-8 in the chimera helix, the  $^{13}\text{C}_\alpha$  CSDs exceed the helical limit. Furthermore, both aromatic residues, Tyr and Trp, exhibit a high chemical shift deviation, induced by the tertiary structure. The N-terminal helical extension also shows a helical tendency, but the residue Asp at position 5' does not exceed the helical limit and interrupts an otherwise contiguous helix.

At pH 7.9, the deviations remain continuously positive. The residues 1-8 of the Trp-cage helix exceed the helical limit, indicating a stable, helical fold. However, the Asp residue at position 5' again shows a significantly lower chemical shift and barely reaches the helical limit. This may indicate that Asp 5' has an unusual torsion angles, which minimally lead to a kink in the helix.

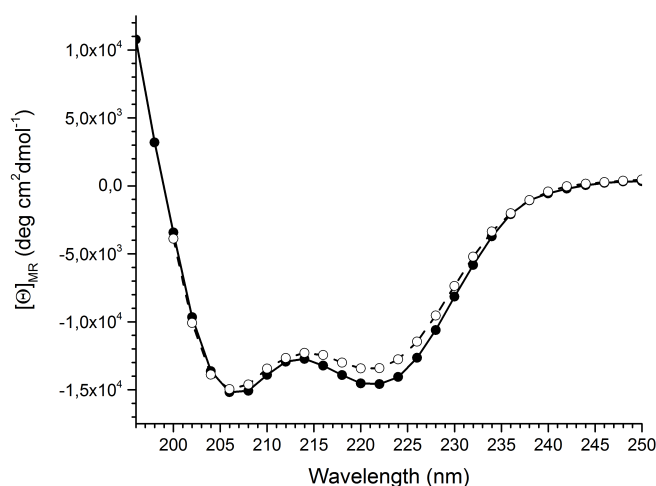
The fraction folded were determined for the complete protein as described previously.

**Table II.1.10.:** Sum of the CSDs ( $^1\text{H}$ ) for the cage loop and its fraction folded and the average chemical deviations ( $\overline{\Delta\delta}$ ,  $^{13}\text{C}$ ) and fraction folded ( $\chi_F$ ) for the complete helix, the chimera helix and the helical extension for the AFP-Tc-4 at 274 K and pH 3 and 7.9.

pH	Cage Loop		Helix	
	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{cage}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{helix}}(\%)$
3	-8.58	91	2.1	76
7.9	-9.24	96	2.6	94
pH	Chimera Helix		Helical Extension	
	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{chim.}}(\%)$	$\overline{\Delta\delta}(\text{ppm})$	$\chi_F^{\text{exten.}}(\%)$
3	2.6	83	1.2	48
7.9	3.0	96	1.9	75

The fraction folded for the different parts in the AFP-Tc-4 chimera reveal a stably folded cage loop that increases its stability at neutral pH. As observed for the other variants,

the helical extension exhibits a significantly lower fraction folded at pH 3, which indicates terminal fraying, which also has an impact on the chimera helix. Therefore, the fraction folded of the chimera helix is lower compared to the cage loop. We suggest that the flexibility of the helical extension also affects the first N-terminal residues of the chimera helix. At pH 7.9, the extension increases in stability, attributed to the capping effect of the first residue at position 1' (Asp). As a consequence thereof, the chimera helix is more stable and as such the complete helix. Nevertheless, the increase in stability of the helical extension is significantly smaller compared to all other variants. This effect is mainly attributed to the putative kink at position 5'.



**Figure II.1.17.:** CD spectra of Tc10b (hollow circles) and AFP-Tc-4 (filled circles) recorded at 274 K and pH 7.9. Plotted is the wavelength against the mean residue molar ellipticity.

CD spectra of AFP-Tc-4 and Tc10b for comparison were recorded at 274 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer (Figure II.1.17). Both proteins exhibit an almost identical curve progression with two minima at  $\lambda = 208 \text{ nm}$  and at  $\lambda = 222 \text{ nm}$ . This shape is characteristic for a helical fold. The marginally deeper minimum at  $\lambda = 222 \text{ nm}$  of the AFP-Tc-4 is unexpected, as the chimera has more residues in a helical conformation by design compared to Tc10b (13 and 8, respectively).

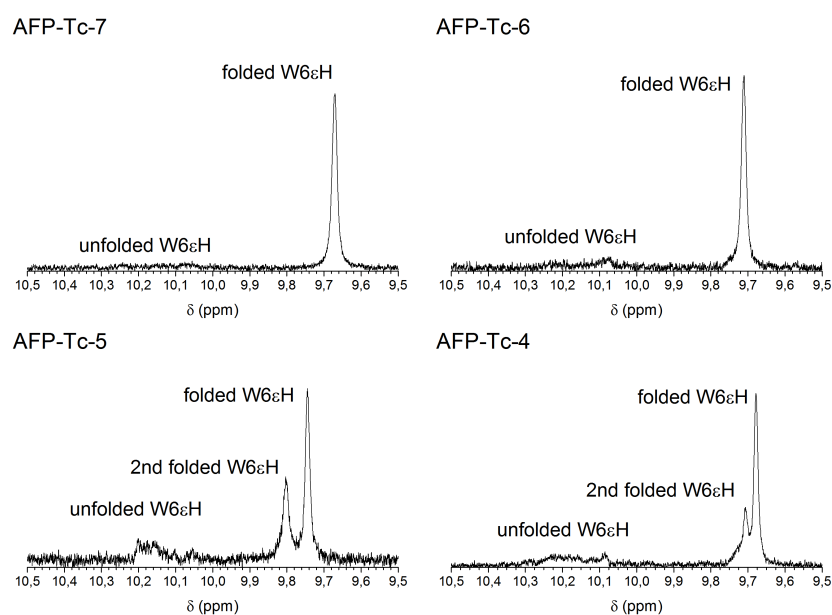
The helical content for this variant is determined to be 35%, or 8.7 residues in a helical conformation. For AFP-Tc-4, we expected 14 residues to be helical (including the D9). Comparing the calculated and expected helical residues, a stability of 62% is determined. This observation is in strong contradiction to the obtained fraction folded of the helix (94%), as determined by NMR spectroscopy. Therefore, unknown effects must be present, reducing the signal and, as such, the helicity.

Overall, the spectroscopic characterization by NMR proves the presence of a mostly

folded Trp-cage for the AFP-Tc-4. However, the data reveals a partial disruption and destabilization of the fold in the region of the helical extension. It results from destabilizing effects at neutral pH, which was also observed for AFP-Tc-5 and is discussed in the following chapter in detail.

### 1.5. Discussion of the Fold at Neutral pH

The four chimera proteins surprisingly exhibit differing trends when comparing the stability at pH 3 and pH 7.9. For AFP-Tc-6 and AFP-Tc-7, a large stabilization effect on the overall fold is observed at neutral pH. In contrast, AFP-Tc4 and AFP-Tc-5 show only a smaller increase or even a decrease of stability, respectively. Furthermore, the determined ratio of the folded fast equilibrium and the unfolded *cis*-Pro conformations is significantly higher for the latter variants. Therefore, the  $^1\text{H}$  NMR spectra of the four chimera proteins at neutral pH are reviewed in more detail.



**Figure II.1.18.:**  $^1\text{H}$  NMR spectra of AFP-Tc-7 (top, left), AFP-Tc-6 (top, right), AFP-Tc-5 (bottom, left) and AFP-Tc-4 (bottom, right) between 9.5 ppm and 10.5 ppm (region of the indole proton) at pH 7.9 and 274 K .

A comparison of the indole proton region of the four chimera proteins show that the stably folded AFP-Tc-6 and AFP-Tc-7 exhibit spectra with similar characteristics (Figure II.1.18). In these spectra, a single, upfield shifted resonance for the folded W6 $\epsilon$ H

and markedly low amounts of *cis*-Pro unfolded conformations are observed. Surprisingly, in the spectra of AFP-Tc4 and AFP-Tc-5, a second upfield shifted resonance of the indole proton is visible. Judged by the chemical shift of this second resonance, it is assigned to a folded protein in an alternative conformation. This second conformation is somewhat less stably folded compared to the main conformation, which was described in the previous chapters. Accompanied with a second, less stably folded conformation, the *cis*-Pro unfolded equilibrium is also affected, indicating a higher amount of the unfolded conformations for these variants compared to AFP-Tc-6 and AFP-Tc-7.

However, the presence of the second folded conformation is surprising. Indications for its origin can be found by evaluating the chemical shifts of the protein. We observed a second set of resonances for several residues that are affected by the ring current of the Trp residue, and foremost, for residues in close proximity to the C-terminus. This includes residues in the helix, i.e. Asp at position 5'. Due to the visibility in the NMR, this second folded conformation is in a slow equilibrium, which is only possible in our proteins with a X-Pro peptide bond in *cis* configuration. Combined with the second set of resonances, we assume that one X-Pro peptide bond at the C-terminal end is in a *cis* configuration. This is unexpected, as we anticipated that any *cis*-X-Pro peptide bond would lead to a complete unfolding of the protein.

In order to estimate the stability, and thus the possibility of a X-Pro peptide bond in *cis* configuration, we modelled AFP-Tc-4 and AFP-Tc-5 and their conformers with X-Pro (position 18 or 19) in *cis* configuration. Important for the stability of the protein is the cage loop and therefore the quality of the shielding of the Trp side chain. A measure for this is the accessible surface area (ASA) of the Trp side chain and the related free folding energy of the protein. The values were obtained using the web-based VADAR program (volume, area, dihedral angle reporter).<sup>94</sup>

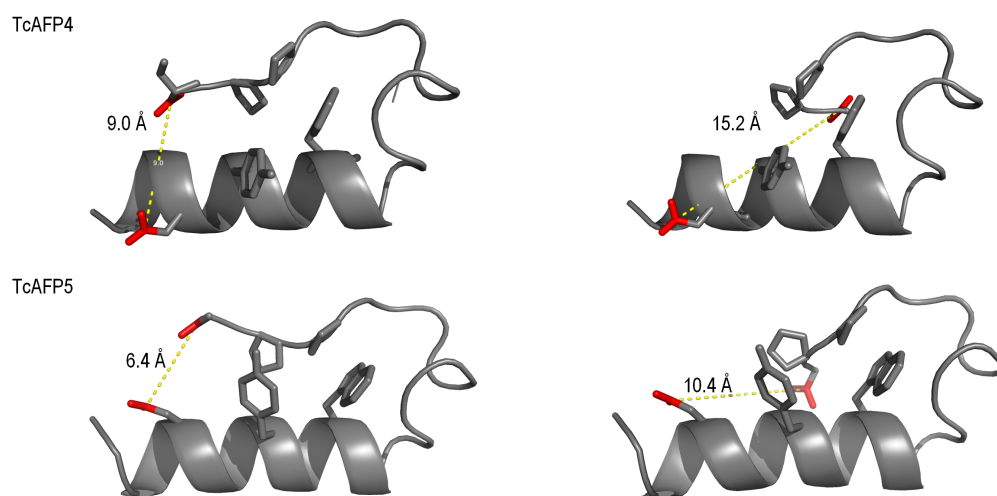
**Table II.1.11.:** Fraction of the ASA of the Trp side chain and the free folding energy of AFP-Tc-4 and AFP-Tc-5 with the X-Pro peptide bond at position 18 or 19 in *trans* or *cis* configuration using VADAR.<sup>94</sup>

	<b>ASA (Trp) (%)</b>	<b>Free folding energy (kJ/mol)</b>
AFP-Tc-4	11	-10.47
AFP-Tc-4-P19 <i>cis</i>	12	-10.43
AFP-Tc-4-P18 <i>cis</i>	9	-10.44
AFP-Tc-5	21	-8.37
AFP-Tc-5-P19 <i>cis</i>	16	-10.18
AFP-Tc-5-P18 <i>cis</i>	18	-8.76

The evaluation of the ASA of AFP-Tc-4 shows that the Trp side chain is similarly inaccessible for the solvent in all conformations. This demonstrates that the Trp residue is buried in all three different models of AFP-Tc-4. Thus, in all models the hydrophobic cluster can be formed, being the driving force of the folding. The free folding energy

for the three models exhibit negative values of similar magnitude. Both criteria suggest that these models of AFP-Tc-4 are able to fold and likely to be present at neutral pH. Evaluating the models for AFP-Tc-5 displays the Trp side chain being similarly buried for all the models. This leads to the conclusion that all modelled structures are able to form a hydrophobic cluster involving the Trp side chain. This proves that AFP-Tc-5 can form such structured conformations, which may be present at neutral pH.

Energetically, the *cis* configuration of the X-Pro peptide bond is unfavourable compared to the *trans* configuration. Thus, a driving force for this isomerization needs to be localized in the proteins. As the second conformation is only visible at neutral pH, we assume that an unfavoured charge-charge interaction is present and scanned the models for the presence of such. At the N-terminus, two negatively charged Asp residues are located at position 1' and 5'. The Asp side chain at position 5' is in close proximity to the equally charged carboxyl group of the C-terminus in AFP-Tc-4 and AFP-Tc-5. Representatively, the models with all *trans* and with X-Pro (19) in its *cis* configuration are depicted.



**Figure II.1.19.:** Models of AFP-Tc-4 (top) and AFP-Tc-5 (bottom) with the secondary structure presented in cartoon style. The side chains of Asp (5'), Tyr (3), Trp (6), Pro (18 and 19) Ser (20) and the free carboxyl group at the C-terminus are sketched in stick style. The carboxyl group of the Asp side chain and the C-terminus are colored in red. The distance between the Asp 5' side chain and the C-terminus is marked with a yellow dotted line. The Pro (19) is in *trans* configuration for the left models, and in *cis* configuration for the right models.

The models of AFP-Tc-4 and AFP-Tc-5 with all peptide bonds in *trans* configuration display the close proximity between the Asp side chain and the C-terminus (Figure II.1.19, left models). The distance between the two carboxyl groups was determined to be 9.0 Å for AFP-Tc-4 and 6.4 Å for AFP-Tc-5, respectively. Changing

the configuration of the X-Pro peptide bond to *cis* (Figure II.1.19, right models), the distance between the Asp at position 5' and the free C-terminus is increased to 15.2 Å for AFP-Tc-4 and 10.4 Å for AFP-Tc-5. As the potential energy between charges is inversely proportional to the distance, both chimeras can circumvent the repulsive charge-charge interaction.

Overall, the modelled structures and the determined ASA and free folding energy prove that alternative folded conformations in the protein are possible. The effect of the charge-charge interaction favours most likely an isomerization of an X-Pro peptide bond to the *cis* configuration. This also affects the remaining X-Pro peptide bonds in the protein. Their *trans* to *cis* equilibrium can be influenced towards the *cis* configuration, which in turn yields a higher amount of unfolded *cis* Pro conformations. Furthermore, this unfavoured charge-charge interaction possibly leads to a disruption of the fold. Indications for this disruption were observed for AFP-Tc-4, whose unexpectedly small CSD for Asp at position 5' could be caused by a kink in the helix at this position. AFP-Tc-5, in contrast, exhibits a continuous helix, but the fraction folded for the cage loop was surprisingly low. We assume that the negative effect of Asp at position 5' for AFP-Tc-5 has a greater impact, as the side chain is oriented and in close proximity to the C-terminus. This might rather lead to a disturbance in the cage loop than in the helix.

**Table II.1.12.:** Percentage of *cis*-Pro unfolded conformations present in the four AFP-Tc chimeras and their helical content (CD) and fraction folded of the helix (NMR)  $\chi_F^{helix}$  at 274 K and pH 7.9.

Chimera	<i>cis</i> -Pro (%)	Helical Content (%)	Fraction Folded $\chi_F^{helix}$ (%)
AFP-Tc-4	25	61	94
AFP-Tc-5	18	71	90
AFP-Tc-6	11	96	99
AFP-Tc-7	8	100	99

These observations and interpretations explain also the contrary observations in the characterizations of the structural properties with different spectroscopic measures (Table II.1.12). The high amount of unfolded *cis*-Pro conformations in AFP-Tc-4 and AFP-Tc-5 has an impact on the determined helical content by CD spectroscopy, as the CD signal is an average of all conformations and the unfolded conformations do not contribute to the signal. Therefore, the fraction folded determined by NMR can be significantly higher compared to the CD signal, as it describes the conformation in the fast folding equilibrium, in which the *cis*-Pro unfolded conformations are not included.

In conclusion, the two chimera proteins AFP-Tc-6 and AFP-Tc-7 are both very stably folded, showing a propagation of the helical fold to the N-terminal extension. This

results in a contiguous helical fold of the incorporated sequence of the biologically active peptide. These results show that our pursued concept to join the sequences of the Trp-cage with those of the AFP is performs extraordinary well. The identified reasons of the unfavoured interactions that lead to the less stable AFP-Tc-4 and AFP-Tc-5 proteins will be incorporated in future designs in order to avoid these complications.

### 1.6. Activity of the AFP-Tc Chimera Proteins

AFPs have the unique ability to bind to different planes of an ice crystal nucleus, i.e. the prism, pyramidal or basal plane. The crystal growth is retarded at the binding surface and, dependent on it, the forming (growing) ice crystal develops a characteristic shape. For AFP Type 1 HPLC6, the AFP variant from which the ice binding repeat used in the four AFP-Tc chimeras is derived, a hexagonal bypramidal crystal shape is characteristic. The activity of the AFP Type 1 HPLC6 is known to be strongly structure and length dependent.<sup>30</sup> The minimal sequence required for activity was determined to be 28 out of 37 amino acids.<sup>30</sup> In contrast, the AFP-Tc-4 to AFP-Tc-7 variants have a total length of the ice binding helix from 13 to 17 residues, respectively. The variants are therefore significantly shorter than the established minimal length. Thus, using this criterion, the variants would be inactive. However, it was the underlying hypothesis of the chimera approach that stabilization of the structure of short AFP sequences could facilitate measurable AFP activity.

A simple assay, particularly in the case of weak AFP activity, is the observation of the ice crystal morphology.

**Table II.1.13.:** Sequences of 15 amino acids reference peptides. Differences between the sequences are marked in bold. The ratio of measured and theoretical helicity at the absorption at  $\lambda = 222\text{ nm}$  in the CD spectra and prediction following LOU at pH 7.9 and 274 K is given and represents the percent of helical content.

Name	Sequence	Helical Content (%)
15er	DTASDAAAAAALTAD-NH <sub>2</sub>	37
15erYW	DTASDAAAYAA <b>WT</b> AD-NH <sub>2</sub>	21

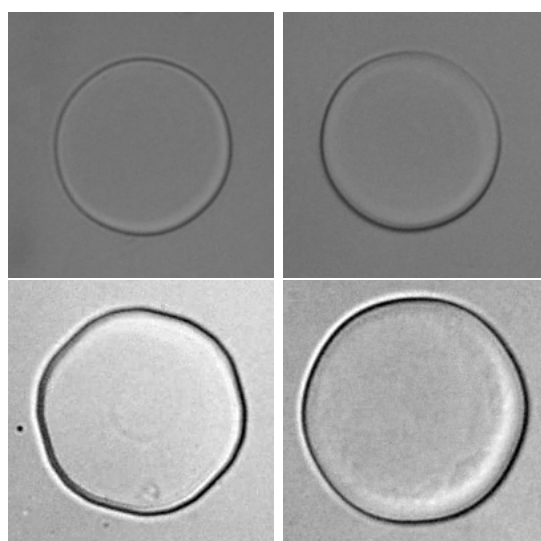
Two peptides with the AFP sequence repeat used in our design of the AFP-Tc chimera were synthesized (Table II.1.13). They serve as reference peptides to assess the effect of the Trp-cage modifications on peptide stability and AFP activity. Therefore, both peptides are C-terminally amidated to avoid a C-terminal charge that is not present in the AFP-Tc. The differences between the peptide sequences is the Tyr and Trp replacement in the AFP repeat, which is similarly in our AFP-Tc variants.

The structural characterization of the peptides by CD spectroscopy shows mostly flexible peptides. However, the 15er peptide comprising the first 15 residues of AFP Type 1 HPLC6 has a helical content of 37%. The Trp-cage based replacements made



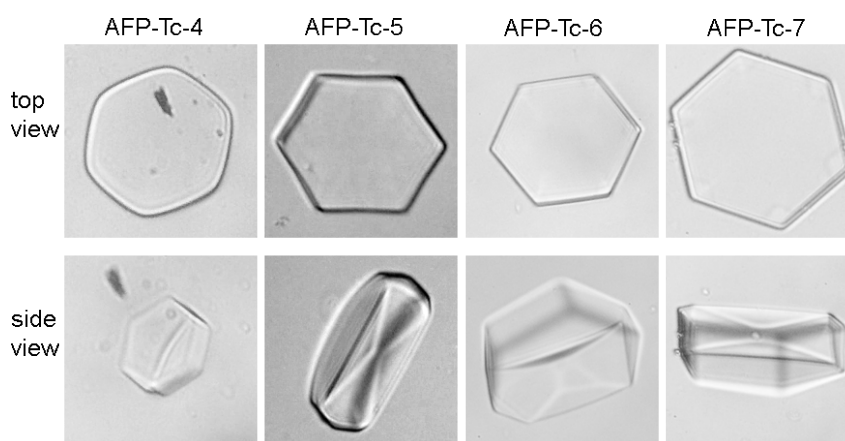
in 15erYW reduce the helical content by 16% in agreement with the lower helical propensity of the aromatic amino acids.

The ice activity assays were performed in 0.1 M  $\text{NH}_4\text{HCO}_3$  at pH 7.9 with a protein concentration of 80  $\text{mg}/\text{mL}$ . A detailed description can be found in the Experimental Section (section 3.9).



**Figure II.1.20.:** Pictures of reference ice crystals: in the presence of 0.1 M  $\text{NH}_4\text{HCO}_3$  buffered water (top, left), Tc10b (top, right), 15er (bottom, left) and 15erYW (bottom right).

Figure II.1.20 shows four pictures, each of an ice crystal grown from a single ice nucleus at temperatures slightly below the freezing point and observed under the microscope. Tc10b (top row, right) shows a perfectly round disk shaped crystal that continuously grows over time. Shape and rate are identical to ice crystal growth in buffered water. This proves that the Trp-cage sequence expectedly does not have any ice binding activity. The AFP peptides 15er and 15erYW (bottom row) lead to ice crystals that are noticeably not perfectly round, first indications of a shaping effect towards a hexagonal form are observable. This is more pronounced for the 15er peptide (bottom, left) that also showed the higher helical content. The slight shaping activity is surprising for a peptide of this length based on the literature results.



**Figure II.1.21.:** Ice crystals during the growth in the presence of AFP-Tc-4 (first column), AFP-Tc-5 (second column), AFP-Tc-6 (third column) and AFP-Tc-7. Top row shows the view perpendicular to the basal plane. Bottom row shows pictures of a tilted side view of the crystals.

In the presence of an AFP-Tc chimera, the morphology of the growing ice crystal changes significantly. Hexagonal shaped disks are clearly observed for all four variants (Figure II.1.21). Occasionally, tilted ice crystals were captured (bottom row Figure II.1.21). The side view clearly highlights the disk shape of the crystals that results from a lower growth on the basal plane (in the  $c$ -direction) compared to the prism face growth (in the  $a$ -axes direction). Nevertheless, the  $a$ -axes growth is reduced as shown by the clear expression of the hexagonal faces. This was further established by the measurements of the growth rates (Masters thesis, K. BAMBERG). Consequently, one might expect that the expressed crystal faces represent the binding surface of the AFP-Tc chimera. The clear prism-like character of the faces may suggest chimera binding to prism planes ( $10\bar{1}0$ ). However, the parent AFP Type 1 HPLC6 is known to bind to the  $20\bar{2}1$  planes, which differ only by their pyramidal tilt towards the  $c$ -axis. The difference between the planes is subtle and could be caused by either the changed length of the AFP sequence or the short crystal dimensions in the  $c$ -direction.

The sharpest edges and thus the strongest shaping effects are observed for AFP-Tc-6 and AFP-Tc-7, whereas for AFP-Tc-5 and in particular for AFP-Tc-4 more rounded edges are observed (top row, Figure II.1.21). Nevertheless, all four variants show ice shaping abilities far more pronounced than the reference peptides, which proves the success of the chimera design. Further, the strength of the shaping ability appears to correlate with the stability of the AFP-Tc proteins determined in the previous section (AFP-Tc-7  $\approx$  AFP-Tc-6 > AFP-Tc-5 > AFP-Tc-4). Moreover, as the fold stability is sufficient to explain the observed activity differences between the variants, other differences in AFP-Tc chimera must play a smaller role, e.g. AFP sequence length, orientation between AFP active side and cage loop.

## **Part III.**

### **Summary and Outlook**



## 1. Linking and Switching

The goal of this dissertation was the design of a stabilizing and switching module for  $\alpha$ -helical peptides. We aimed to create a protein module, which comprises a highly stable helical fold and the possibility to incorporate a switchable molecule. Upon fusion with unstructured, biologically active peptides, this module should be able to induce a stable helical fold in the fused peptide.

The focus of the first part was to introduce a photo responsive azobenzene cross-linker in the Trp-cage miniprotein, which was chosen to be the protein platform for the module concept. Initially, we used the WOOLLEY strategy that employs iodoacetamide as reactive groups and Cys residues as anchor residues. Several problems were encountered pursuing this strategy in our platform. The Cys replacements at the chosen positions in the Trp-cage led to differently folded conformations. During the cross-linking reaction with the reducing agent TCEP, the linker reacted and formed soluble by-products. These by-products were not separable by the available purification methods. Therefore, the cross-linking chemistry was changed to NHS esters, which selectively react with primary amines, such as Lys residues.

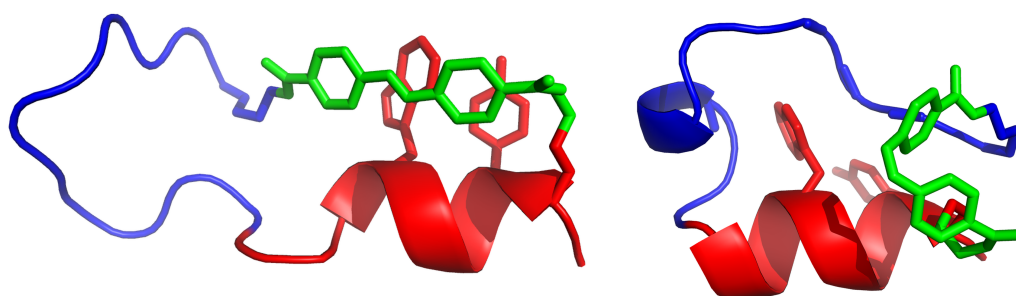
Three amino acid replacements were made in the Trp-cage variant Tc10b and led to the new variant TcKKA. As anchor points to facilitate cross-linking with NHS esters, two Lys residues were introduced at position 2 and 20. The NHS chemistry proves to be exceptionally suitable for cross-linking bifunctional azobenzenes with the TcKKA. During the linking reaction, no signs for side reactions of the linker were observed. The cyclized product could easily be separated by HPLC purification from the starting material and side products, such as multiple linked or uncyclized combinations of linker and Trp-cage. The first employed linker was the flexible LOMANT reagent, which demonstrated these advantages of the cross-linking chemistry. The ensuing cross-linking with the synthesized azobenzene to introduce switchability in the Trp-cage module was successfully performed.

The properties of the generated proteins (linked and unlinked) were characterized with various spectroscopic methods, such as UV/vis-, CD- and NMR-spectroscopy. With the latter, the stability of the secondary (helix) and tertiary structure (cage loop) of the Trp-cage was assessed using chemical shift deviations (CSD) of fold-characteristic proton resonances. In all NMR-spectra of Trp-cage variants, additional, clearly unfolded resonances were surprisingly visible. With different approaches, we could assign these resonances to be Trp-cage conformations with a *cis*-X-Pro peptide bond. We utilized the amount of these unfolded conformations as an additional stability criterion of the Trp-cage fold.

**Table III.1.1.:** Summary of the fraction folded  $\chi_F$  of the helix and the cage loop of Tc10b and TcKKA (unlinked and linked with 4,4'-NHS-azobenzene) at pH 3 and 298 K.

Trp-cage	$\chi_F^{helix}$ (%)	$\chi_F^{cage}$ (%)
<b>Tc10b</b>	84	71
TcKKA	80	67
<i>cis</i> -TcKKA-azo	92	88
<i>trans</i> -TcKKA-azo	82	16

The structural characterization of the unlinked TcKKA variant demonstrated that all replacements and modifications in the parent sequence are exceptionally well accepted and is of similar stability compared to the published parent sequence Tc10b. With the azobenzene cross-linker incorporated, switchability of structure was introduced in the protein.



**Figure III.1.1.:** Possible models of TcKKA-azo with the azobenzene linker in *trans* (left) and *cis* configuration (right) based on the results of the spectroscopic data. The secondary structure is represented in cartoon style and the side chains of Tyr and Trp as well as the azobenzene in stick style. The helix is colored in red, the C-terminal loop in blue and the azobenzene in green.

The spectroscopic results of the cross-linked TcKKA-azo proved that the pursued strategy to stabilize the protein structure in *cis* configuration was successful (Figure III.1.1). The protein is not only correctly folded, but its stability is significantly increased compared to the unlinked Trp-cage. In the *trans* configuration of the linker, the tertiary structure of the protein is disrupted, leading to a flexible cage loop (colored blue). Surprisingly, the resulting absence of tertiary structure interactions did not lead to an unfolding of the helix. Instead, the azobenzene linker provided an alternative stabilization of the helical fold by interactions with the two aromatic residues Tyr and Trp. TcKKA-azo could be irradiated (switched) multiple times between these two conformations and we observed no signs of aggregation in both states of the protein. Further, the excitation wavelength for the *trans* to *cis* isomerization is notably hypsochromic shifted. This is explained by putative interactions between the linker and the aromatic residues in the Trp-cage. A surprising and interesting observation was the further absence of

---

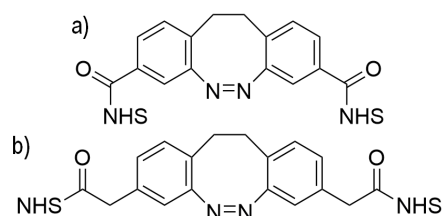
measurable thermal back relaxation. This is advantageous as the protein remains in its folded conformation, which allows a selective switch between the unfolded and folded conformation that remains stable over time.

Overall, we demonstrated that the pursued approach of switching the structure of the Trp-cage is possible, creating a protein module with good and partially surprising photochromic properties. This module can be fused with peptides to remotely induce or destabilize their structure.

Some observations of the azobenzene cross-linked TcKKA require further investigations to better understand the properties of this switching module. Detailed insights into the mechanism of the switching process for TcKKA-azo are of great interest. The determination of the chiral properties will provide an understanding of the switching mechanism and can be assessed by CD measurements in the absorption range of the azobenzene. This experiment will show, if one enantiomer (*P* or *M*) of the linker in its *cis* configuration is preferred and thus, whether the switching mechanism is bi- or uni-directional. An excess of one diastereomeric form would indicate that the protein moiety has an influence on the direction of the photoisomerization. Another well suited method to gain insights into the switching mechanism is ultrafast time resolved spectroscopy. Advantageous for these experiments is that the azobenzene linker as well as the aromatic side chains of Trp and Tyr are observable. The experiments will reveal if the azobenzene linker is influenced by the protein and if cooperative interactions, such as electronic coupling of the linker with the protein residues are present. Additional quantum mechanics/molecular mechanics calculations of the switching and folding process will support the interpretation of the results of CD spectroscopy and ultrafast time resolved spectroscopy.

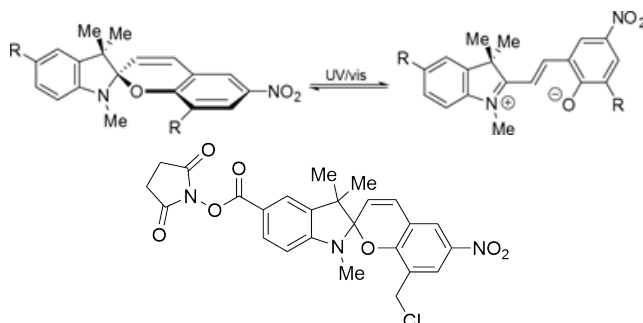
It is possible to improve the switching module and enhance the structural differences between the protein states or to increase the switching efficiency of the photochrome. The structural difference between the folded and unfolded states of TcKKA-azo were smaller than expected due to the surprising presence of residual helical structure in the unfolded state. This was attributed to interactions between the linker and the aromatic residues Tyr and Trp. One possibility to reduce these interactions and achieve unfolding of the helix would be a replacement of the aromatic Tyr residue by an aliphatic residue. This residue should possess some ability to shield the Trp side chain in the folded conformation, but should only weakly interact with the azobenzene linker in the *trans*-state. Alternatively, one can selectively destabilize the helix of the Trp-cage at one or more positions with helix breaking residues. A single replacement that accomplishes both tasks is the replacement of the existing Tyr at position 3 by a Leu or Ile residue.

In our pursued switching approach, we focused mainly on azobenzene as cross-linker. However, other photochromic molecules can be employed and may partially possess superior properties.



**Scheme III.1.1.:** Scheme of NHS functionalized diazocine. a) current molecule, b) work in progress.

An alternative switching moiety is diazocine (Scheme III.1.1, synthesis by W. MOORMAN, ongoing work). Preliminary studies revealed that the linking between the Trp-cage and diazocine can be performed similarly to the cross-linking with azobenzene. However, it was found that the cross-linker in *cis*-configuration was too short for stabilization. Thus, a methylene group will be inserted (diazocine b)) in order to elongate the linker. This project will be continued in collaboration with W. MOORMAN from the group of R. Herges.



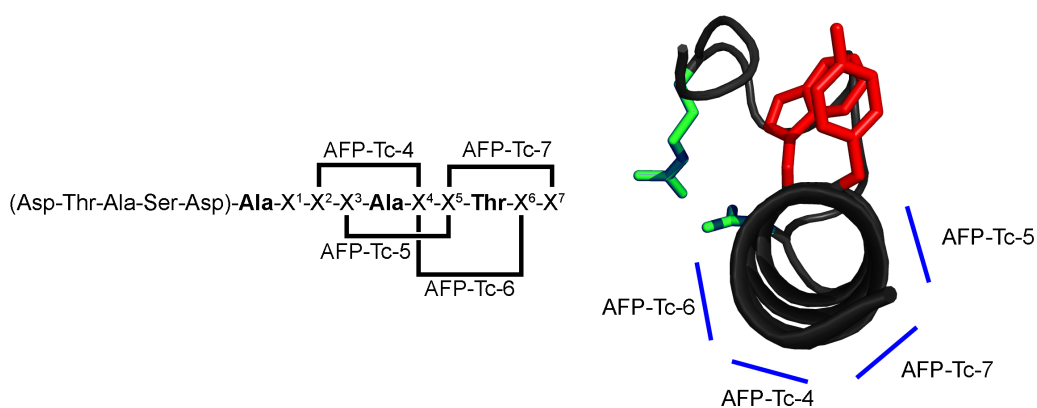
**Scheme III.1.2.:** Top: Scaffold of spiropyran. By irradiation, a ring opening leads to merocyanine. Bottom: Scheme of a bifunctionalized spiropyran with NHS ester and an allylic chloride.

A second applicable scaffold for a cross-linker is a spiropyran (Scheme III.1.2). In the closed spiropyran form, the linker should stabilize the fold, whereas the open merocyanine form destabilizes it. This linker possesses advantages in that the linker geometry is different compared to azobenzene, and that a change in the charge distribution is present between the two forms. Incorporated into the protein, the merocyanine form is expected to reduce interactions with the aromatic residues and thus should lead to a complete unstructured protein state.



## 2. AFP-Cage Chimera Proteins

In the second part of this dissertation, a fusion approach linking the Trp-cage module with a biologically active peptide with  $\alpha$ -helical propensity was presented. We chose to join the sequences of the first eleven amino acid repeat of the AFP Type 1 HPLC6 with the Trp-cage variant Tc10b. The chosen AFP is known to exhibit a hydrophobic surface, which can bind to an ice crystal plane. Therefore, we pursued a chimera strategy, which is distinguished by a sequence overlap of the N-terminal helix of the Trp-cage and the biologically active peptide. The advantages of this strategy are a reduction of the overall peptide length, a simplification of the synthesis, and the availability of the complete helical region of the resulting chimera protein for the ice binding activity. The resulting chimera protein can be divided into three parts, the N-terminal extension, the chimera helix and the cage loop.



**Figure III.2.1.:** Left: Sequence pattern of the used ice binding motif including the N-terminal capping sequence (in brackets). The positions marked in bold are essential, whereas the positions marked as  $X^n$  are arbitrary for the activity and can be replaced. Right: View along the helix axis (from N to C) of the chimera protein with marked orientations (blue lines) of the ice binding surface. The aromatic residues Trp and Tyr (marked in red) and the salt bridge between Asp-Arg (marked in green) are conserved.

In order to show the wide applicability of this concept, four different variants with different orientations of the ice binding surface were designed and synthesized (Figure III.2.1). All four AFP-Tc chimera proteins were structurally characterized by NMR and CD spectroscopy to assess their structure. We observed that all variants exhibit a stable fold and all characteristic properties of the Trp-cage tertiary structure. Most importantly, all AFP-Tc chimera proteins exhibit a helical fold that completely propagates through the N-terminal extension proving the success of our concept.

Minor limitations were observed for the chimera proteins AFP-Tc-4 and AFP-Tc-5 as they exhibit alternatively folded conformations at neutral pH. These conformations are attributed to the presence of a charge-charge interaction of the N-terminal helical extension with the C-terminal cage loop. We found that residues at the junction of helical extension and chimera helix are oriented in the structure towards the charged C-terminus and lead to a destabilizing charge-charge repulsion. In future designs, residues in these positions need to be carefully selected with respect to their charge and size in order to prevent this interaction.

The activity of the chimera proteins was assessed by observing the morphology of growing ice crystals *via* an optical microscope. Astonishingly, all four AFP-Tc chimeras were active, as we observed hexagonal ice crystals in their presence. Differences in their activity were indicated by details of the crystal shape. Whereas in the presence of AFP-Tc-7 and AFP-Tc-6 a perfect hexagonal shape was observed, the hexagonal crystal in the presence of AFP-Tc-4 and AFP-Tc-5 exhibit somewhat rounded edges. The differences in the ice-shaping abilities is fully consistent with the determined fold stability of the variants. This suggests that other design differences, such as orientation of the active side to the Trp-cage loop is not decisive for activity. It also suggests that two third of the helix surface is accessible for protein activity.

Overall, the spectroscopic and activity results proved the hypothesis that a structure stabilization of the conformation of a helical peptide can be used to increase its biological activity and that this can be achieved by fusion with a highly structured protein module.

The ice binding properties of the presented AFP-Tc chimera proteins require further characterization. It is of particular interest to determine the binding planes of the proteins on the ice crystal surface. The hexagonal shape of the ice crystals in the presence of the AFP-Tc may suggest a preferred binding to the prism plane of the crystal. Since the differences between the prism and pyramidal plane are subtle, an unambiguous statement of a preferred binding plane is not possible with the assay performed in this work.

One possibility to assess the preferred binding plane are ice-etching experiments analyzed by fluorescence microscopy. In this experiment, AFPs bind to a single ice crystal attached to a cooling finger. The pattern of the bound AFPs in the resulting crystal will be visualized by excitation of a fluorescence label and will reveal the binding plane and possible orientation of the bound AFPs. This has already been applied to different AFPs fused with green fluorescent protein (GFP).<sup>95,96</sup> Our AFP-Tc are particularly suitable for such experiments, as the Trp side chain possesses intrinsic fluorescence properties depending on its local environment, and thus the fold.<sup>97</sup> Therefore, no

---

additional tagging with a GFP is necessary.

The spectroscopic data of AFP-Tc-4 and AFP-Tc-5 revealed that residues at the junction of helical extension and chimera helix can interfere with the tertiary structure of the C-terminal cage loop leading to a reduced stability and alternatively folded conformations. To avoid such interactions, different strategies are possible. The existing Asp residue at the appropriate position in the helix can be replaced by an Ala or the C-terminus can be altered, i.e. amidated, reducing the charge-charge interactions. Alternatively, the charge-charge interactions between the N-terminal extension and the C-terminus can also be utilized to enhance the fold stability. Within the helical extension, the residue oriented towards the C-terminus will be replaced by a side chain bearing a positive charge, such as Lys. Thus, a salt bridge between this residue and the C-terminus might form, stabilizing the cage fold and as a consequence the helix. This effect was already observed for the Trp-cage TcKKA. These studies are currently pursued by K. BAMBERG

Although the chimera proteins presented herein showed significant binding to the ice surface, the activity levels of AFP Type 1 typical bipyramidal crystal could not yet be reached. One reason might be a different charge distribution at the C-terminal end of the helix. Whereas AFP Type 1 HPLC6 bears a positively charged Arg, the chimera proteins have a negatively charged Asp at this position. This Asp residue is involved in a salt bridge with an Arg residue in the Trp-cage loop. A reversion of the orientation of this salt bridge will alter the charge at the C-terminal end of the helix while preserving the fold-stabilizing salt bridge. This change will place the Arg residue into an identical position as AFP Type 1 HPLC6 and may lead to a different preferred binding plane.

An alternative explanation for the absence of a pyramidal crystal can be slight difference in the twist of the helix due to the Trp-cage tertiary fold. This would result in a different ice binding surface, and thus different binding preferences. The reliable detection of such differences requires structural information at high resolution. Precise protein structures can be obtained by solution NMR with  $^{13}\text{C}$  or  $^{15}\text{N}$  labelled proteins and by X-ray crystallography. Superposition of obtained structures with the crystal structure of wild-type AFP Type 1 HPLC6 will prove whether the structure is identical in the steric properties of the ice binding, helical surface.

The chimera concept possesses the advantage of an independent fold and stability that is conferred to the sequence of the biological peptide. This stabilizing effect can be utilized to investigate the ice binding side of the AFP by selective replacements in the sequence, without causing a concurrent change in helical conformation. Foremost, investigations of length dependency of the ice binding side and determination of the

minimal sequence requirement for activity can now be performed under preservation of the helical fold. An extension of the ice binding sequence is of great interest as this should lead to an increased activity. This project and the combination of AFP-Tc chimeras with GFP label are pursued by J. GRONOW.

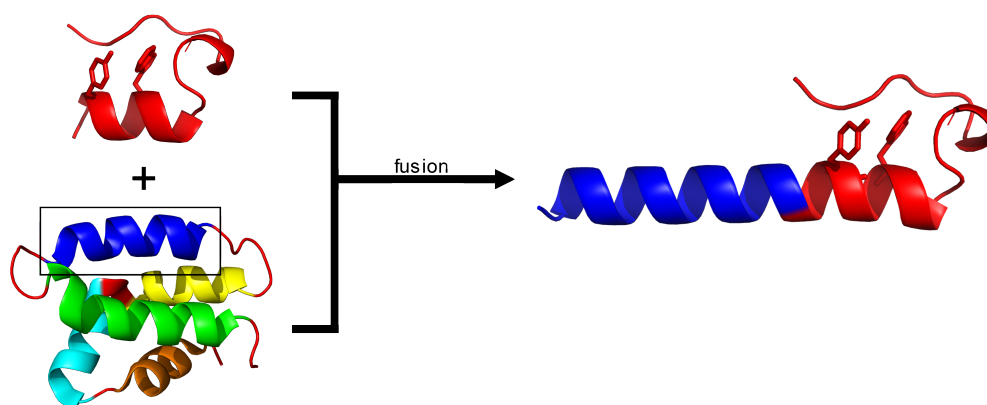
In the first part, the possibility to introduce a switching function in the Trp-cage module was presented and in the second part the strategy to fuse the Trp-cage with the AFP sequence. Merging the switching and chimera concept will be the next step in this project. Transferring and combining the results of the switching approach to the functionalized chimera will give access to a switchable AFP in which the antifreeze activity can be activated or removed. Crucial for a successful combination in the AFP-Tc chimera is the placement and orientation of the necessary anchor residues for the linker with respect to the ice binding surface. The cross-linker should not interfere sterically with the ice binding surface and influence the ice activity. The orientation of the cross-linker in the present approach is comparable to the existing intra-molecular salt bridge, and thus should be accommodated well. Accordingly, the presented switch-cage module appears compatible with all four chimeras.

In preliminary work towards a switchable AFP-Tc, two Lys residues were introduced in the AFP-Tc-5 chimera sequence, resulting in AFP-Tc-5-KK. This chimera protein was fully characterized by NMR spectroscopy at pH 3. In addition, CD spectra were recorded and ice binding assays were performed at pH 7.9 (see Appendix, subsection 1.3.5). The AFP-Tc-5-KK variant will be cross-linked with the NHS functionalized azobenzene and the effect of the switching on the ice activity investigated (K. BAMBERG, Master's thesis, in progress).

A high switching efficiency is a crucial property for effective applications of molecular switches. Ideally, either a pure *on* or pure *off* state of the protein structure and activity should be reached. The approach presented in this work fulfills several of these ideals. Almost full conversion to the *trans* state of the TcKKA-azo can be achieved photochemically and thermally. Advantages in the context of AFP chimera are that this state will be fully inactive and stable over time. By light induced isomerization to the *cis* configuration, the protein will be switched "on" and structures with high spatial and temporal control. As the activity is distinctly present only in this state, it is not essential to achieve full photochemical conversion or that the AFP-Tc chimera need to be fully active AFPs. This is a direct advantage of our design choosing the fully reachable, thermodynamically stable state of the photochrome as the "off"-state of protein activity. Coincidentally, it is very advantageous that the functional "on"-state unexpectedly does not show any thermal back relaxation. Thus, we succeeded to create a true switch between an "off"- and an "on"-functional state, in which both states are stable over time.

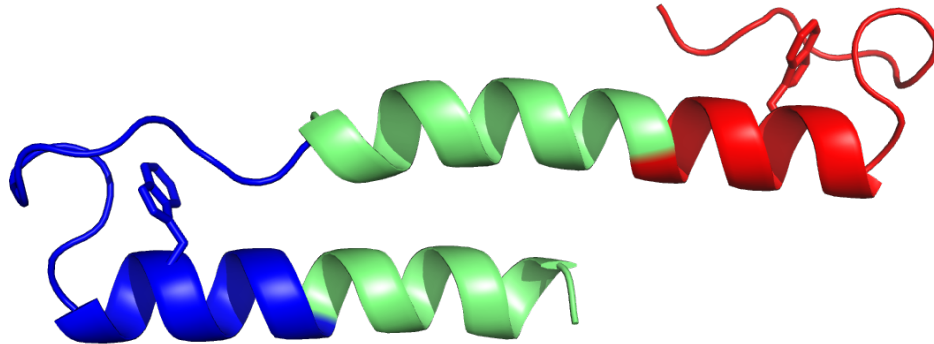
### 3. Other Fusion Partner and Possibilities

Due to the modular approach, the fusion concept can be used not only with AFPs, but transferred to any peptide sequence of interest with helical propensity. In particular, antimicrobial peptides (AMPs) are suitable. Initially, A. SIMPSON and A. GIANESINI worked on a fusion approach of AMP sequences with the Trp-cage.



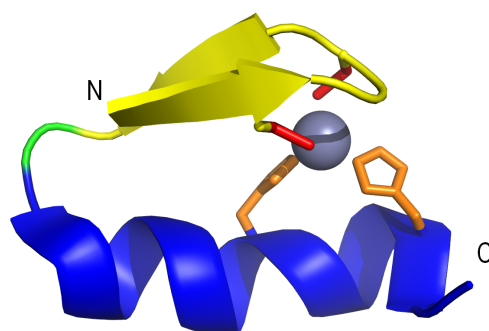
**Figure III.3.1.:** Fusion concept for stabilizing helical AMP sequences, exemplarily shown for the 77 residue long protein Amoebapore A. The colored  $\alpha$ -helical structures are: helix I (green), helix II (blue), helix III (yellow), helix IV (cyan), helix V (orange). (PDB: 1OF9.<sup>98</sup>) The helix II of Amoebapore is discussed to play a key roll in the activity of the AMP. Thus, we extract its sequence and fuse it N-terminally to the Trp-cage module.

In contrast to the chimera approach, pursued in this work, the combination of both sequences was accomplished by using a direct fusion strategy. Therefore, the AMP sequence is simply N-terminally added to the Trp-cage without a sequential overlap. A. SIMPSON worked on a fusion protein between helix II of the *Amoebapore A* protein (Figure III.3.1, blue colored helix) and the Trp-cage TcKKA. Therefore, we decided to extract the sequence and fuse it N-terminally to the Trp-cage TcKKA. This allows the characterization and determination of the role in the function of the solely helix. The work on stabilizing AMP sequences was continued by A. GIANESINI during his Master's thesis. He fused the short AMP *KR-12*, truncated from the human LL-37 cathelicidin<sup>99</sup> and the scorpions' AMP *IsCT*<sup>100</sup> with the Trp-cage TcKKA. Further studies in this project will be conducted by N. PREUSSKE, who will perform studies of the activity and *in vitro* lifetime of the fusion proteins of AMPs and the Trp-cage.



**Figure III.3.2.:** Concept for a serial connection of two Trp-cage modules (blue and red) for anti parallel  $\alpha$ -helical dimers (green).

Another advantage of the Trp-cage module concept is that it can be extended to multiple modules and biological peptides. It then can be used also to pre-orientate intra-molecular folding domains, such as dimers. For this, multiple possibilities are conceivable. In one example, the attachment of a second Trp-cage module at the N-terminus of a Trp-cage fusion protein results in a serial connection (Figure III.3.2). Thereby, the start of the second Trp-cage sequence (colored in blue) leads to a termination of the first helix, due to the Pro residues, which are unfavoured in a helical fold. Simultaneously, these Pro residues will be the transition to the second Trp-cage motif (polyproline-II-helix) and the starting point of the second helix in opposite direction. This leads to a close proximity of the two helical regions with an anti parallel orientation. The longitudinal shift and orientation of both helices can be adjusted by the starting residue of the second Trp-cage.



**Figure III.3.3.:** Cartoon representation of the zinc-finger structure. The loops are colored in green, sheets in yellow, helix in blue, Cys involved for complexation with Zn (grey) is colored in red and His is colored in orange; PDB: 1ZNF.<sup>101</sup>

The modular concept of attaching a switching module to short peptides with structural propensity can be also extended to other short, folded miniproteins. A suitable for a similar module is the 25 amino acid long zinc-finger (Figure III.3.3).<sup>101</sup> The structure of the zinc finger consists of an anti-parallel  $\beta$ -sheet at the N-terminus and a C-terminal  $\alpha$ -helix. Its fold is not mainly induced and stabilized by hydrophobic shielding in contrast to the Trp-cage, but by complexation of a zinc ion. With the zinc-finger as stabilization and folding module, peptide sequences can be fused N-terminally to the helix of the zinc-finger (in contrast to C-terminal fusion with the Trp-cage that was pursued here). The helix of the zinc-finger thereby provides a structural nucleation point and the fold can propagate into the attached sequence, similar to the Trp-cage concept. As for the Trp-cage, two different strategies are possible to fuse the zinc-finger with an  $\alpha$ -helical peptide: a direct, C-terminal attachment, or sequential overlap of the C-terminal helix with the N-terminal part of the peptide.

In this dissertation, we showed that the creation of a protein module to stabilize helical peptide sequences is possible and that switchability is easily incorporated in this module. The module provides structural stabilization of the biological active peptide, which results in a significant increase in activity. This concept has enormous potential, as it can be further developed and tailored for future applications.





**Part IV.**

**Experimental Section**

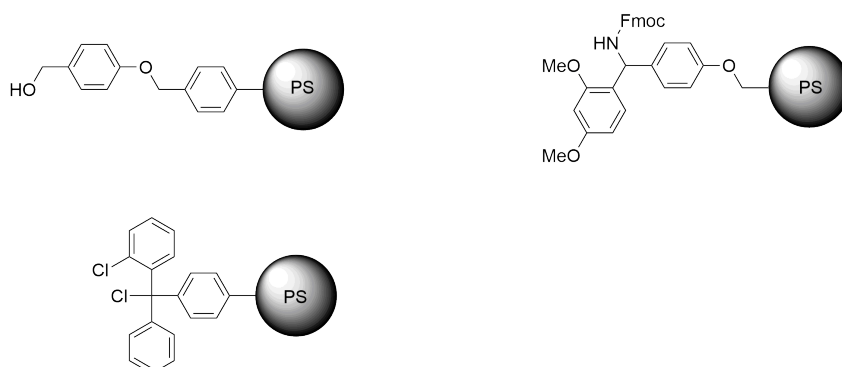


## 1. Solid Phase Peptide Synthesis

The peptides were synthesized using Fmoc (Fluorenylmethyloxycarbonyl) based SPPS. This method was originally invented by R.B. MERRIFIELD<sup>102</sup> and allows the synthesis of sequences up to a length of 50 amino acids. The advantage of the solid phase synthesis is the possibility of fully automation of the synthesis. In this work, a semi-automatic approach was used with the semi-automated peptide synthesizer INITIATOR+ SP WAVE by Biotage. In the following, the general procedure is described, details for each peptide synthesis can be found in the Appendix (chapter 3).

### 1.1. Preparation of the Resin

As solid phase, a polystyrene-based resin was used with different surface functionalizations (Scheme IV.1.1). The Wang-Resin for unproblematic amino acids at the C-terminus, the 2-Chlorotrityl resin for amino acids, which tend to perform racemization during synthesis (i.e. Cys) and the Rink-Amide resin for an amidated C-terminus.



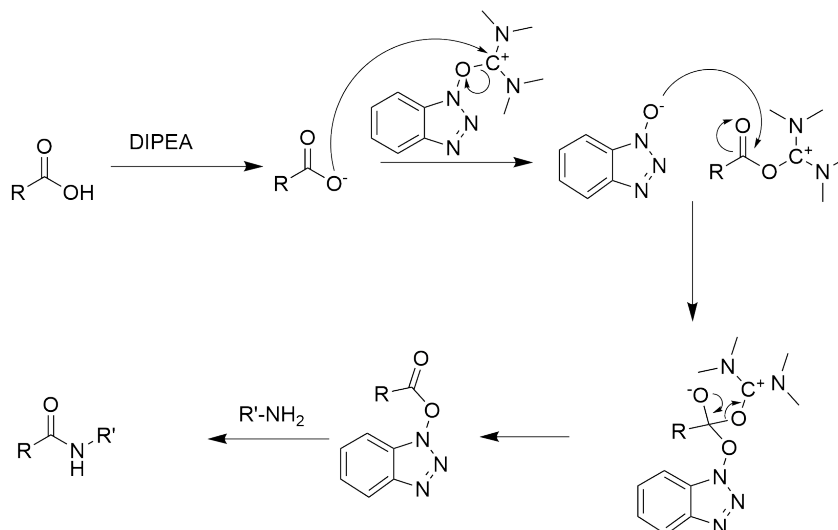
**Scheme IV.1.1.:** Used resins for SPPS. Wang resin (top, left), Rink-Amide resin (top, right) and 2-Chlorotrityl resin (bottom, left).

Prior to the coupling of the amino acids, the resin was swollen in dichloromethane (DCM) and washed with dimethylformamide (DMF) to increase the volume- and surface-accessibility. In case of the Rink-Amide resin, a deprotection step of the Fmoc group with 25% pyridine in DMF is included.

### 1.2. Coupling Cycle

The coupling was performed at 75°C with microwave assisted heating. Each coupling step consists of two steps. First, the coupling of the Fmoc protected

amino acids (5 eq.<sup>a</sup>) was performed with 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) (4.9 eq.<sup>a</sup>) and N,N-diisopropylethylamine (DIPEA) followed by a washing cycle with DMF.



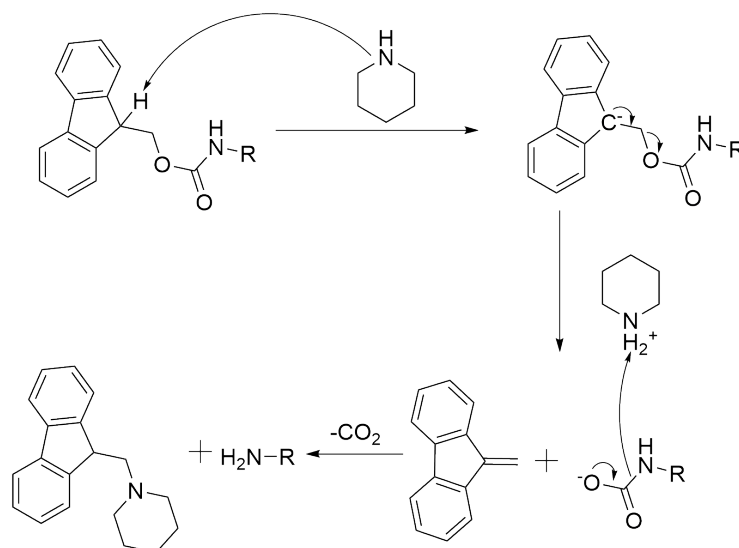
**Scheme IV.1.2.:** Coupling-mechanism with HATU and DIPEA.

The Fmoc-protected amino acid is deprotonated by DIPEA and forms a carboxylate anion. The carboxylate anion attacks HATU and forms the O-acyl(tetramethyl)isouronium salt. The formed anion reacts with the activated carboxylic acid and tetramethyl urea is released. An amine group can then react with the formed, activated ester to the amide peptide bond (Scheme IV.1.2).

Difficult coupling steps with a yield lower than 90% and couplings of unpolar amino acids (i.e. alanine) were performed twice to improve the overall yield and ease up the purification process. In case of several unpolar amino acids in a row, a capping step with acetic anhydride and pyridine was inserted to quench unreacted sides.

<sup>a</sup>Equivalents per binding side on the resin.

### 1.3. Removal of the Fmoc Protecting Group



**Scheme IV.1.3.:** Mechanism of the removal of the Fmoc protecting group using piperidine.

In a second step, the Fmoc group was removed with 25% piperidine in DMF at 75 °C. This step is extended in case of non-quantitative deprotection. Piperidine attacks the acidic fluorenyl proton of the Fmoc protected amino acid. This results in a  $\beta$ -elimination of the fluorenyl ring leading to the dibenzofulvene and elimination of CO<sub>2</sub> and therefore to the free amine. The highly reactive dibenzofulvene is scavenged by piperidine (Scheme IV.1.3).

Functional side chains (i.e. carboxylic acids) are protected orthogonally to avoid side reactions.

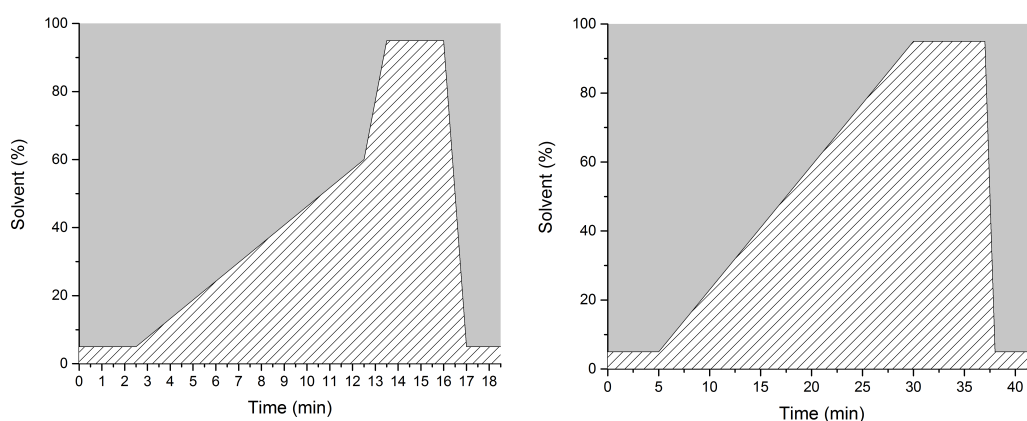
### 1.4. Cleavage and Deprotection of the Peptide

The peptide was cleaved and fully deprotected by treating the resin with trifluoroacetic acid (TFA) and scavengers<sup>b</sup> at ambient temperature. Afterwards, the resin was filtrated and the combined filtrates were treated with diethylether to precipitate and wash the peptide. The crude peptide was dissolved in water and lyophilized. The peptides were purified using high-performance liquid chromatography (HPLC).

<sup>b</sup>Wang & 2-Chlorotrityl resin: 95 v/v% TFA, 2.5 v/v% water and 2.5 v/v% Triisopropylsilane (TIS).  
Rink-Amide resin: 90 v/v% TFA, 5 v/v% 1,3-dimethoxy benzene (DMB), 2.5% water, 2.5 v/v% TIS.

## 1.5. HPLC and ESI-MS

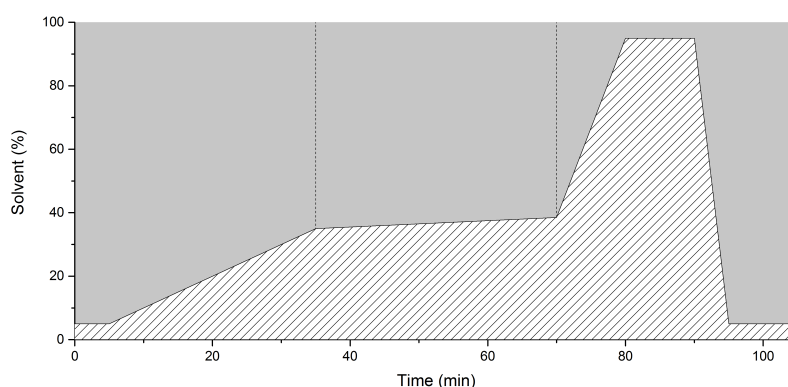
The mobile phase of the HPLC consists of H<sub>2</sub>O and acetonitril (Acn) with 0.01*v/v*% TFA. As a solid phase, poly(styrene-divinylbenzene) (PLRP-S) is used. Detailed description of the columns and the equipment can be found in the chapter "Methods and Machines".



**Figure IV.1.1.:** Solvent profile of the short (left) and long analytical method (right). The grey area represents the percentage of H<sub>2</sub>O and the hatched area the percentage of Acn.

For the analytical approach, an analytical column was used and two different methods were performed. Polar proteins were characterized with a short method (Figure IV.1.1) with a flow rate of 1 mL/min starting with 95% H<sub>2</sub>O and 5% Acn for 2.5 min, followed by a linear gradient from 2.5 to 12.5 min to 60% Acn. To eluate unpolar species, the amount of Acn is set to 95% for 3.5 min. For re-equilibration, the column is washed for 1.5 min with 95% H<sub>2</sub>O and 5% Acn.

In the case of unpolar molecules, a longer analytical method was used. This method with a flow rate of 1 mL/min starts with 95% H<sub>2</sub>O and 5% Acn for 5 min followed by a linear gradient from 5 to 30 min to 95% Acn. The column is washed with 95% Acn for 6 min and for re-equilibrating the column 4 min with 95% H<sub>2</sub>O and 5% Acn (Figure IV.1.1).



**Figure IV.1.2.:** Exemplary profile of a preparative run. The grey area represents the percentage of H<sub>2</sub>O and the hatched area the percentage of Acn. Fractionized samples were collected between 35 and 70 *min*.

The purification was performed on a semi-preparative PLRP-S column (see chapter Used Methods and Machines) and a modified method by MANT was used.<sup>103</sup> In this method, the retention time ( $t_R$ ) of the desired product was determined on the analytical column and the injection time of 2.5 *min* was subtracted. The corresponding percentage of Acn was corrected by 12% for the long and 17% for the short analytical method. For 30 *min*, a linear gradient from 5% to the determined percentage Acn, followed by a linear gradient of 0.1%/min for 35 *min* to eluate the desired product was used. For washing and re-equilibration, the column was flushed with 95% Acn for 10 *min* followed by 5% Acn for additional 10 *min* (Figure IV.1.2). During the 0.1% linear gradient, slices of 0.25 – 0.5 *min* were collected.

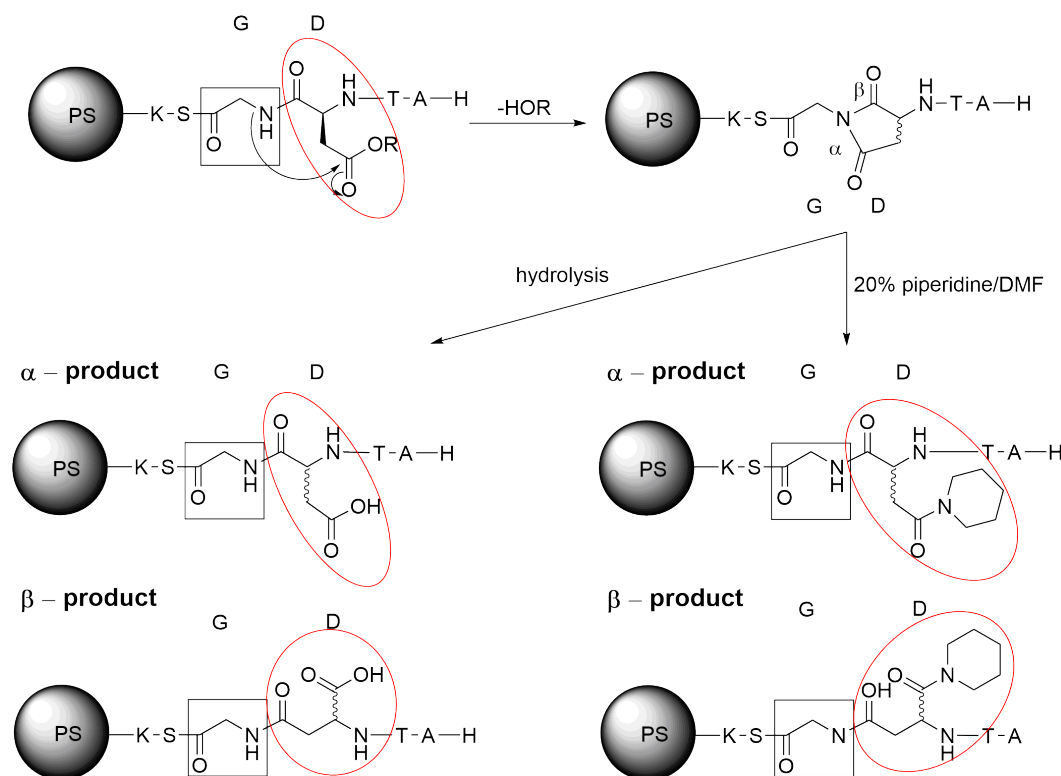
The collected fractions were analyzed by HPLC and MALDI-ToF or HPLC-ESI-MS to verify the purity. Subsequently, the pure fractions were flash frozen with liquid nitrogen and the solvent was removed by lyophilization and a white, fluffy solid was obtained.

**Table IV.1.1.:** Isolated yields of the peptide synthesis.

Peptide	Isolated yield
Tc10b	9mg (4.3 $\mu$ mol, 2.3%)
Tc10b-S20C-21G	4mg (1.9 $\mu$ mol, 2.0%)
Tc10b-A2K-S20C(S-tBu)-21G	8mg (3.5 $\mu$ mol, 4.0%)
Tc10b-A2C(S-tBu)-S20K-21G	9mg (3.9 $\mu$ mol, 4.5%)
Tc10b-S20K	4mg (1.9 $\mu$ mol, 2.0%)
TcKKA	11mg (5.1 $\mu$ mol, 5.5%)
15er	13mg (9.7 $\mu$ mol, 6.5%)
15erYW	9mg (6.0 $\mu$ mol, 4.5%)

## 1.6. Sequence Difficulties

All the Trp-Cage sequences contain the motif Asp-Gly. This motif causes major problems during the Fmoc based SPPS.



**Scheme IV.1.4.:** Mechanism of the base catalyzed aspartimide formation of the Asp residue (red ellipse) with the prior coupled Gly (black square), followed by a ring-opening by hydrolysis or deprotection with piperidine.

During the peptide synthesis, the protected side chain of the Asp residue can be attacked base or acid catalyzed by the amide of the previous amino acid and form a five membered ring (aspartimide). The ring can either be opened by hydrolysis or by nucleophilic attack of piperidine during the Fmoc removal, whereas the first result in a racemic mixture of  $\alpha$ - and  $\beta$ -peptide, the latter yields in a racemic mixture of  $\alpha$ - and  $\beta$ -piperidide (Scheme IV.1.4).<sup>104-106</sup>

The various side products not only significantly decrease the yield, but show similar retention times on the RP-HPLC and it is therefore impossible to purify those peptides. To reduce this problem, Asp can be replaced by Glu. However, a similar side reaction can occur: the protected side chain can be attacked by the amide of the previous amino acid (see aspartimide) or by the own amide. The second ring closure results in a capping of the peptide chain and further growth is prevented. Nevertheless, the yield decreases significantly. Another possibility to prevent this side reaction is the use



of backbone protected amino acids. In case of Glu, the previous coupled amino acid and Glu itself must be backbone protected, while in the case of Asp only the previous amino acid requires a protection. For synthetic reasons and comparability with the literature, Asp was used and the prevention of the named side reactions was archived by using the building block Fmoc-Asp(OtBu)-(Dmb)Gly-OH. As an example for this side reaction, a hexa-peptide with the sequence AT-XY-SK was synthesized. For XY, either the di-amino acid Fmoc-Asp(OtBu)-(Dmb)Gly-OH, Fmoc-Asp(OtBu)-OH and Fmoc-Gly-OH, or Fmoc-Glu(OtBu)-OH and Fmoc-Gly-OH were used. The hexapeptides were synthesized using standard protocol microwave assisted coupling and deprotection on a Wang-resin as described previously. After the coupling of Lys and Ser, the resin was split in three equals and separately, I) Fmoc-Asp(OtBu)-(Dmb)Gly-OH moiety, II) Asp and Gly, III) Glu and Gly were coupled. The three peptides were cleaved from the resin and deprotected as described elsewhere (section 1.4).

**Table IV.1.2.:** ESI-MS of the hexa-peptide ATD-GSK, ATDGSK and ATEGSK. First column indicates detected mass peaks above 400  $m/z$ , second column indicates the area in percent of the peak and third column indicates the assigned molecule. All peaks >1% were assigned and normalized to 100 %. a) Molecule plus sodium, b) incomplete deprotection of tBu-groups (one remaining), c) piperidide, d) aspartimide.

I) ATD-GSK			II) ATDGSK			III) ATEGSK		
Mass ( $m/z$ )	Area (%)	Assignment	Mass ( $m/z$ )	Area (%)	Assignment	Mass ( $m/z$ )	Area (%)	Assignment
600.5	0.8	M <sup>a</sup>	645.7	3.1	M <sup>c</sup>	592.6	9.2	M
578.5	17.8	M	578.5	5.6	M	561.5	1.2	M-S <sup>b</sup>
547.5	2.5	M-S <sup>b</sup>	560.5	4.5	M <sup>d</sup>	535.5	3.2	M-G
513.5	2.6	M-S <sup>a</sup>	558.5	7.8	M-S <sup>c</sup>	527.5	1.7	M-S <sup>a</sup>
507.5	2.7	M-A	547.5	0.8	M-S <sup>b</sup>	521.5	1.3	M-A
498.5	0.9	M-T <sup>a</sup>	520.5	0.9	M-G	519.5	1.4	M-E <sup>b</sup>
491.5	60.0	M-S	519.5	7.9	M-D <sup>b</sup>	505.5	35.2	M-S
473.5	2.5	M-K	513.5	1.1	M-S <sup>a</sup>	485.5	2.2	M-E <sup>a</sup>
420.4	4.5	M-S-A	507.5	1.3	M-A	477.5	1.1	M-A-T <sup>b</sup>
406.4	5.7	M-DG	491.5	19.5	M-S	463.5	19.7	M-E
			489.5	1.4	M-A <sup>d</sup>	448.5	13.0	M-S-G
			477.5	0.8	M-T	434.5	1.6	M-A-S
			473.5	18.1	M-K / M-S <sup>d</sup>	432.5	2.8	M-E-S <sup>b</sup>
			463.5	16.8	M-D	420.4	1.6	M-A-T
			434.5	1.4	M-S-G	406.4	4.7	M-E-G
			432.5	2.1	M-S-D <sup>b</sup>			
			420.4	2.2	M-S-A			
			406.4	2.6	M-D-G			
			402.4	2.3	M-S-A <sup>d</sup>			

The resulting crude peptides were characterized by ESI-MS (Table IV.1.2) and the masses above 400  $m/z$  were assigned. First, it is observed that the coupling of Ser on Lys is incomplete with a yield below 40%. All other coupling steps are in the expected range

of above 90%. However, these results coincide well with studies of difficult couplings in SPPS using *N,N'*-dicyclohexylcarbodiimide (DCC)/Benzotriazol-1-ol (HOBt) as coupling reagents of J.W. VAN NISPEN.<sup>107</sup> Secondly, a minor part shows an incomplete deprotection of tBu groups.

Comparing the crude hexapeptides with separate coupling Gly and Asp and the building block Asp(OtBu)-(Dmb)Gly shows the problems which occur due to the aspartimide formation. With the building block, the correct sequence was obtained in a yield of  $\approx 18\%$ , while without the building block the yield decreased to a third to 6%. Further, the side products (piperidide and aspartimide) were identified in a combined yield of  $\approx 8\%$ . A replacement of Asp with Glu suppresses the named amide side reactions, with a yield of  $\approx 9\%$ . This is a result of the insufficient coupling steps of Glu on Gly.

## 2. Procedure of Cross-Linking the Trp-Cage

### 2.1. Cross-Linking with Activated Halides on Cys

For the cross-linking of the switchable organic molecule (azobenzene, diazobenzene and spiropyranes), reaction conditions typical for marking biomolecules were chosen. Iodoacetamide groups are the common counterpart for a nucleophilic replacement on the thiol of a Cys residue and generally react at a pH of 7.9 (0.1 M  $\text{NH}_4\text{CO}_3$  buffer) in the presence of the reducing reagent TCEP to prevent the formation of disulfide bridges during synthesis. DMSO can be added as co-solvent for the linker if necessary. 1 eq. of the protein was dissolved in the buffer and stirred with TCEP for 1h. The cross-linker (1.1 eq.) was solved in DMSO and added drop-wise to the protein and the reaction mixture was stirred for several hours. Monitoring by HPLC and MALDI-MS determined the end point of the reaction, when no conversion was observed. The solvent mixture was removed by lyophilization. The remaining solid was dissolved in water and filtered. Purification was carried out by HPLC as described elsewhere (section 1.5).

### 2.2. Cross-Linking with NHS Ester on Lys

The second methodology that was used is the reaction of primary amines with NHS ester. Reactions with NHS ester were carried out in a mixture of buffered water at pH 7.4 (20 mM PBS buffer and 1M NaCl) and DMSO.

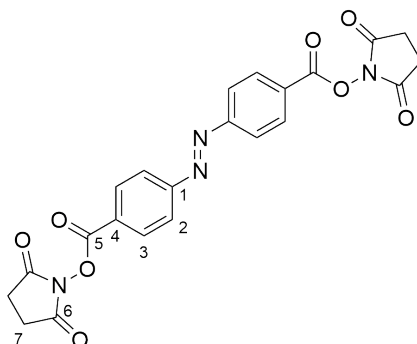
1 eq. of the protein was dissolved in the buffer. The linker was either dissolved in buffer or in DMSO and drop-wise added to the protein solution. The reaction was monitored by HPLC-MS. When no further conversion was observed, the reaction mixture was diluted with water and filtered. The solvent was removed by lyophilization and the solid redissolved in water. After filtration, the crude product was purified by HPLC as described elsewhere (section 1.5).

Isolated yield (TcKKA-Lomant): 0.5mg (0.2  $\mu\text{mol}$ , 23.1%)

Isolated yield (TcKKA-azo): 1mg (0.4  $\mu\text{mol}$ , 45.1%)

### 2.3. Linker Synthesis

#### 2.3.1. 4,4'-Bis(2,5-dioxopyrrolodin-1-yl)azobenzene



A synthesis by K. SAKURAI was adapted and modified as follows:<sup>108</sup> 4,4'-Dibenzoic acid azobenzene (0.50 g, 1.9 mmol) was dissolved in DMSO (50 mL). 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) (0.57 g, 3.7 mmol) and N-hydroxysuccinimide (0.43 g 3.7 mmol) were dissolved in 10 mL DMSO and added dropwise over the course of 10 min to the solution. The reaction mixture was stirred at 75 °C for 20 h. The solvent was evaporated and the crude product was purified by column chromatography (eluent: chloroform). 0.20 g (0.5 mmol, 27%) of a red solid were obtained.

TLC (chloroform, SiOH<sub>2</sub>):  $R_f = 0.45$

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta = 8.34\text{--}8.30$  (m, 4H, H-3), 8.09–8.06 (m, 4H, H-2), 2.94 (s, 8H, H-10) ppm.

<sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>):  $\delta = 169.0$  (C-6), 161.2 (C-5), 155.7 (C-1), 131.8 (C-4), 127.6 (C-3), 123.4 (C-2), 25.7 (C-7) ppm.

HRMS (EI sector):  $m/z(\%) = [M]^+$  calcd for [C<sub>22</sub>H<sub>16</sub>N<sub>4</sub>O<sub>8</sub>]<sup>+</sup> 464.0968, found 464.0963 (10%);  $[M - C_4H_4NO_3]^+$  calcd for [C<sub>18</sub>H<sub>12</sub>N<sub>3</sub>O<sub>5</sub>]<sup>+</sup> 350.0777, found 350.0771 (100%),.

IR (atr):  $\tilde{\nu} = 3100$  (vw), 2923 (w), 1770 (s), 1733 (vs), 1600 (s) 1200 (vs)  $cm^{-1}$ .

Melting point T = 281 °C .

## 3. Used Methods and Instrumentation

### 3.1. HPLC

The purification and analysis of the proteins were carried out by high performance liquid chromatography (HPLC) using a reversed phase column. The HPLC device is a LACHROM ELITE by VWR-HITACHI with a L2400 UV-detector. Additionally, a fraction auto-collector FOXY R1 by TELEDYNE ISCO and an autosampler 717-PLUS by WATERS complete the HPLC set-up. In this work, different columns were used, which vary in size and column material. Two columns have an immobile phase of poly(styrene-divinylbenzene) (PLRP-S) particles, with a particle size of  $8\ \mu\text{m}$  and pore diameter of  $30\ \text{nm}$ . The dimensions of the analytical column are  $150\ \text{mm} \times 4.6\ \text{mm}$  and for the semi-preparative  $325\ \text{mm} \times 30\ \text{mm}$  by AGILENT. The second pair of columns are C-18 NUCLEOSIL column material by WATERS with a particle size of  $7\ \mu\text{m}$  and  $100\ \text{nm}$  pore diameter was used. The analytical column has the dimensions  $250\ \text{mm} \times 4.6\ \text{mm}$ , whereas the semi-preparative column has dimensions of  $250\ \text{mm} \times 32\ \text{mm}$ . For quantitative detection of the molecules, an UV detector at  $\lambda = 200\ \text{nm}$  or  $\lambda = 365\ \text{nm}$  detector was used.

### 3.2. Mass Analysis

For determination of the molecule mass, either a directly connected electron spray ionization mass spectrometer EXPRESSIONL CMS by ADVION (ESI-MS) with a hexapole as separator or manually a matrix-assisted laser desorption ionization (MALDI) with a time of flight (ToF) mass separator BIFLEX III by BRUKER was used. High resolution mass was determined using ACCU-TOF 4GGCV by JEOL with electron ionization (EI) of  $70\ \text{eV}$  and a ToF mass separator.

### 3.3. Lyophilizer

Lyophilization of the samples were performed on the lyophilizer ALPHA 2-4 by Christ equipped with a RZ6 vacuum pump by VACUUBRAND and a typical vacuum of  $0.01\ \text{mbar}$ .

### 3.4. UV/vis Spectroscopy

UV/vis spectra were recorded on a LAMBDA 14 two channel spectrometer by PerkinElmer. It is equipped with a thermostat by Büchi. All measurements were performed at  $25\ ^\circ\text{C} \pm 0.1\ ^\circ\text{C}$  and carried out in a quartz cuvettes with a path length of

10 mm by HELMA ANALYTICS. Measurements were performed under the same conditions and with the same stock solution as for the CD measurements to determine their concentration. The extinction coefficient for the Trp-cages were taken from NEIDIGH<sup>39</sup> to be  $\epsilon_{278\text{ nm}} = 6760 \frac{\text{cm}^2}{\text{mmol}}$ . In addition for following the tuning of the excitation wavelength of TcKKA-azo, an UV-2401PC spectrophotometer by Shimadzu was used.

### 3.5. IR Spectroscopy

Infrared (IR) spectra were measured on a PARAGON 1000 FT-IR spectrometer by PERKIN ELMER. The spectrometer is equipped with an A-531-G GOLDEN-GATE attenuated total reflection (ATR) unit.

### 3.6. Melting Points

The melting point of the linker was determined on an IA6304 capillary melting point apparatus by ELECTROTHERMAL. The measured melting point was not further corrected.

### 3.7. CD Spectroscopy

CD spectra were recorded on a Jasco J-720 spectropolarimeter. The temperature was adjusted with a thermostat WKL 26 by Haake and a PTC-423S Peltier element by Jasco. Spectra of TcKKA and its linked variants were recorded at 298 K and pH 3. TcAFP4-7 were recorded at 274 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer.

All measurements were carried out in quartz cuvettes with a path length of 1.0 mm by HELMA ANALYTICS. The lower limit of the spectrum was tested manually, hence it should not exceed 700 V as the data become unreliable. Spectra were recorded with a data pitch of 2 nm and a scanning speed of 5 nm/min. Accumulation of three repeats and background subtraction result in the full spectra.

The resulting signal is converted to mean residue ellipticity  $[\theta]_{MR}$  (Equation IV.3.1), in which  $\theta$  is the measured CD signal,  $c$  the concentration,  $l$  the path length of the cuvette and  $n$  the number of residues.

$$[\theta]_{MR} = \frac{100 * \theta}{C * l * n} \quad (\text{IV.3.1})$$

### 3.8. NMR Spectroscopy

The NMR data were recorded on Fourier transform NMR (FT-NMR) spectrometers AV III 600 with a frequency of 600 MHz equipped with a triple ( $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ ) resonance

cryo probe head and on a DRX500 500 MHz spectrometer with a triple resonance probe head ( $^1\text{H}$ ,  $^{13}\text{C}$  and  $^n\text{X}$ , with X=any nucleus and n=the corresponding nucleon number) by BRUKER. Both spectrometers are equipped with a variable temperature unit. The 600 MHz spectrometer has a temperature range from 274.2 K to 328 K. For higher temperature measurements up to 358 K, the 500 MHz spectrometer was used. All samples, if not mentioned elsewhere, were measured in 90/10 ratio  $\text{H}_2\text{O}/\text{D}_2\text{O}$  at pH 3, 7 and 7.9. The pH was adjusted with  $\text{D}_2\text{O}$  solutions containing 0.01-1 N NaOD or DCl. Measurements at pH 7.9 were carried out in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The pH was measured with a two point calibration (7.0, 4.0 and 9.8) on ACCUMET MODEL 10 by FISHER SCIENTIFIC. As internal standard for proton referencing, 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) or 3-trimethylsilyl-propionic acid (TSP) were used as reference in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  and tetramethylsilane in organic solvents.  $^{13}\text{C}$  and  $^{15}\text{N}$  were referenced indirectly according to the IUPAC recommendations for the frequency ratio to  $^1\text{H}$ .<sup>109</sup>

### 3.8.1. Protein NMR

All proteins were characterized by 1D  $^1\text{H}$  NMR,  $^1\text{H},^1\text{H}$  correlation spectroscopy (COSY),  $^1\text{H},^1\text{H}$  total correlation spectroscopy (TOCSY) and  $^1\text{H},^1\text{H}$  nuclear Overhauser enhanced spectroscopy (NOESY). For the  $^1\text{H},^1\text{H}$  TOCSY, a DIPSI2 (decoupling in the presence of scalar interactions) for isotropic mixing of 60 ms was used.<sup>110</sup> The mixing time in the  $^1\text{H},^1\text{H}$  NOESY was set to 200 ms. The tr-NOESY was recorded as pseudo 2D experiment. For saturation, a low power Gauss-shaped pulse with a length of 75 ms was used. The total saturation time was set to 4 s. All  $^1\text{H}$  and  $^1\text{H},^1\text{H}$  correlation spectra were recorded using an excitation sculpting with gradient pulses (esgp) for water suppression.<sup>111</sup>  $^1\text{H},^{13}\text{C}/^{15}\text{N}$  spectra were recorded as hetero single quantum coherence (HSQC) with a flip-back pulse for water in the  $^1\text{H},^{15}\text{N}$  HSQC and presaturation of water in the  $^1\text{H},^{13}\text{C}$ -HSQC.

Assignment of the resonances in the  $^1\text{H}$  spectra followed the guide of K. WÜTHRICH.<sup>112</sup> Saturation transfer difference spectra (tr-NOESY) were recorded using a Gaussian shaped excitation pulse.<sup>113</sup> Best transfer and selectivity results were obtained with a saturation time of 4 s and a pulse power of 63.84 – dBW.

### 3.8.2. Reference Shifts for the Trp-Cage Fold

The published  $^1\text{H}$  NMR shifts were used and divided according to the folding pattern they represent.<sup>43</sup> The  $\text{H}\alpha$  shifts from residue 2 till 8 (Table IV.3.1) serve as reference for the N-terminal helix. The cage fold is represented by the upfield shifted protons due to the ring current of Trp: the  $\text{H}\alpha$  of Leu (7),  $\text{H}\alpha'$  of Gly (11),  $\alpha$  of Pro (18),  $\beta'$  of Pro (18) and  $\delta,\delta'$  of Pro (19). In addition,  $\alpha$  and  $\beta$  of Pro (12) and  $\alpha$  of Arg (16)

represent the downfield shifted cage shifts (Table IV.3.2). The shifts are referenced to the random coil shifts by D. WISHART<sup>77</sup> leading to the CSD and summed up resulting in  $\Delta\delta$ . Thus, the CSDs of Tc10b represents a 99.5% folded Trp-cage and by setting observed chemical shift deviations of new variants in ratio, the stability for new variants at any condition can be determined.

**Table IV.3.1.:** Sum of CSDs for selected protons in the helical region representing the helix fold of Tc10b at 280 K at pH 7.<sup>43</sup>

Helix	$\Delta\delta$ (ppm).							$\Sigma$
	A2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$	
pH 7	-0.043	-0.522	-0.200	-0.442	-0.423	-0.953	-0.386	<b>-2.969</b>

**Table IV.3.2.:** Sum of CSDs for selected protons in the cage structure representing the cage fold of Tc10b at 280 K at pH 7.<sup>43</sup>

Cage upfield	$\Delta\delta$ (ppm)						$\Sigma$
	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$	
pH 7	-0.953	-3.483	-2.329	-1.697	-0.664	-0.449	<b>-9.575</b>

### 3.9. Ice Activity Assay

Ice activities were measured using a nanolitre osmomter by CLIFTON. It is equipped with a PELTIER element cooling station, which is cooled to 10 °C by a MINISTAT 230 by Huber with glycol as thermal fluid. The cooling station is mounted on an ECLIPSE NI microscope with a D7000 camera by NIKON. Pictures were recorded with 50x magnification with the CAMERA CONTROL PRO 2 software by NIKON.

All measurements were carried out at pH 7.9 in 0.1 M NH<sub>4</sub>HCO<sub>3</sub> buffer. The sample holder's bores are filled with a high viscosity immersion oil (type B) and serves with 20  $\mu$ m as internal reference for the crystal size. With the help of a glass capillary and a HAMILTON syringe a nanoliter sized drop was placed in the bore without touching the border. The sample was deep-frozen to -40 °C. The frozen sample was heated up with caution until one single crystal remained. For observation of the ice crystal growth, the temperature was adjusted below the freezing point.<sup>26</sup>

### 3.10. Photo-Irradiation Equipment

Irradiation experiments were performed with LED light sources with an optical power of 1050 mW (NICHIA NC4U033A;  $\lambda = 365$  nm) and 900 mW (LUXEON LXML-PR01;  $\lambda = 450$  nm).

In addition, a FLOUROMAX-4 spectrofluorometer manufactured by HORIBA JOBIN



YVON was used for tuning of the wavelength for irradiation between  $\lambda = 365\text{ nm}$  and  $\lambda = 340\text{ nm}$ . It is equipped with a 150 W ozone-free xenon arc-lamp.

### 3.11. Molecular Modelling and MD Simulations

Ongoing from the first structure of the NMR ensembles of Tc5b by NEIDIGH<sup>39</sup> (PDB entry: 1L2Y) all models presented in this work were prepared. With the Swiss-PDB-viewer SPDBV<sup>114</sup> the models were extended.

All structures were prepared in Schrödingers Maestro<sup>115</sup> and the structures minimized. Molecular dynamic simulations were performed with Schrödingers MacroModel.<sup>116</sup> Prior to the simulations, an energy minimization was performed with maximum 2500 iterations following the POLAK-RIBIER conjugate gradient (PRCG) method<sup>117</sup> and a derivative convergence. The MD simulation follows stochastic dynamics with a time step of 1.5 *fs* at variable temperatures. For side chain conformations, a temperature of 200 K was chosen. Simulations of TcKKA linked with LOMANT and azobenzene MD were performed at 350 K. The MD simulations runs for a total simulation time of 20,000 ps with a 10 ps equilibration time. For structure sampling, 100 structures per MD simulation were saved and minimized as described earlier. All simulations were performed using the force field OPTIMIZED POTENTIALS FOR LIQUID SIMULATIONS (OPLS 2005).<sup>118</sup> In the simulations, azobenzene was calculated with constraints for the *cis* conformer. Thus, the angle of C-N-N (for both rings) were set to 121.9°, and the dihedral angle C-N-N-C to 6.5° and C-C-N-N to 53.3°, respectively.<sup>119</sup>

The energy of the modelled peptides was calculated after a geometry optimization. As potential, the OPLS 2005 force field was used with extended cut-off for the non-bonded interactions.

The accessible surface area and free folding energy of the models were determined using the web-based VADAR program.<sup>94</sup> The VADAR program is a compilation of different algorithms to evaluate protein structures based on the coordinate data. For determination of the accessible surface area, the SHRAKE definition for the VAN DER WAALS radii was used.

Pictures of the proteins and protein models were generated using PyMol.<sup>120</sup>



**Part V.**

**Appendix**



## 1. Spectra and NMR Shifts of the Trp-cage Variants

In the following, the HPLC-profiles, mass spectra (MALDI or ESI), NMR, CD and UV/vis spectra of all here discussed substances are presented if not already displayed in the respective chapter. Furthermore, the proton, carbon and nitrogen chemical shifts of all peptides and proteins in different conditions (pH and temperature) are listed.

**Table V.1.1.:** Nomenclature of the AFP-Tc chimera proteins used in this dissertation and in the measurements.

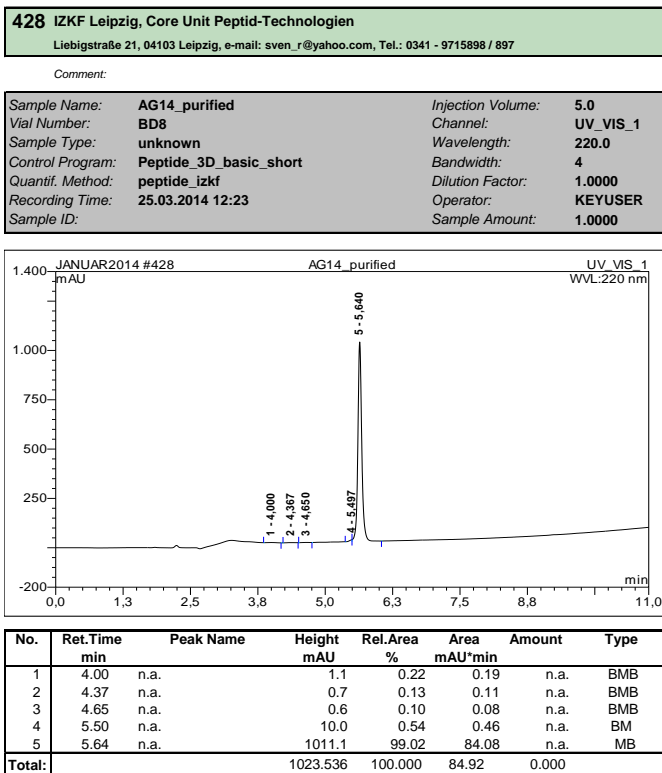
<b>Work</b>	<b>Measurement</b>
AFP-Tc-4	TcAFP4
AFP-Tc-5	TcAFP5
AFP-Tc-6	TcAFP7
AFP-Tc-7	TcAFP6

The nomenclature of the here presented AFP-Tc chimera proteins differs from the used nomenclature of the measurements (Table V.1.1). In this work, the chimera proteins are named according the sequence and numbered by the position of the Trp residue in the AFP sequence. For the measurements, the proteins were named "TcAFPX", whereas "X" is an ascending integer value, depending on the time of design and synthesis.

## 1.1. Tc5b and its Variants

### 1.1.1. Tc5b

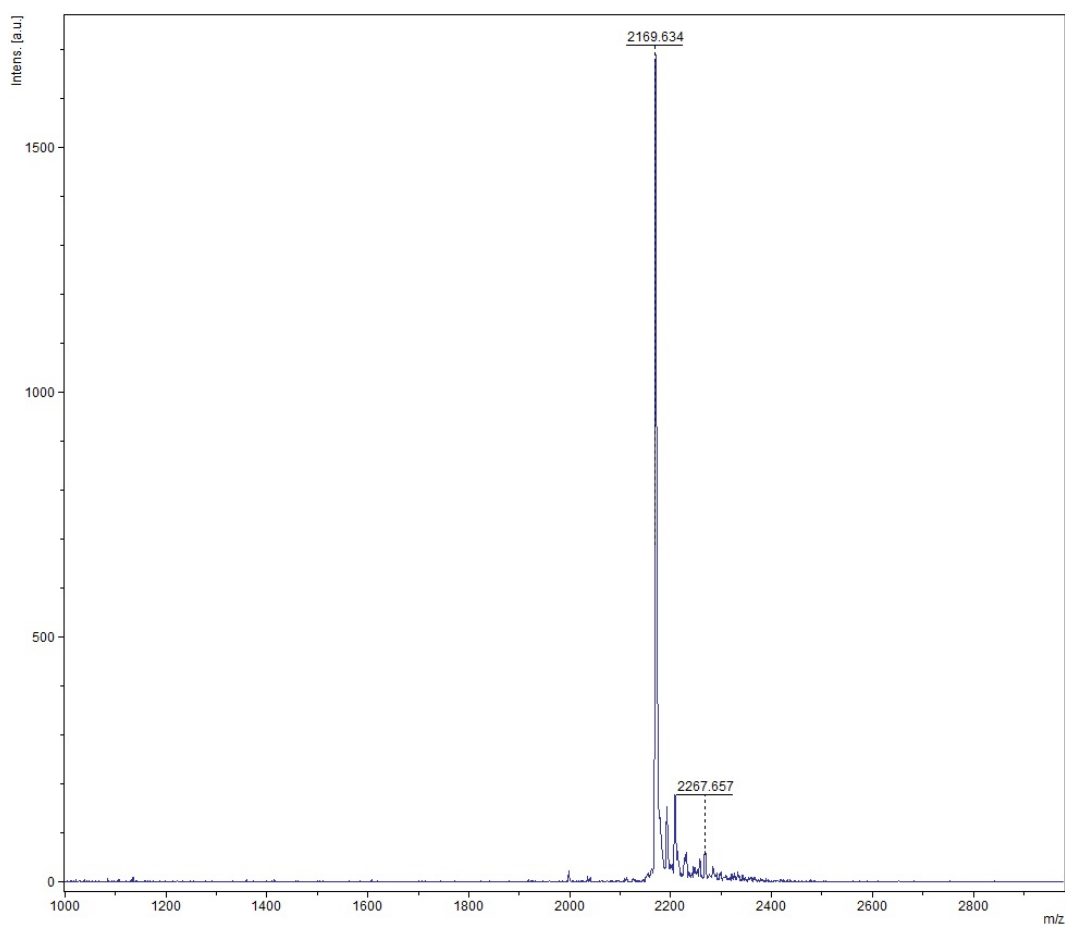
Operator:KEYUSER Timebase:U-3000RS Sequence:JANUAR2014

Page 1-1  
27.02.2017 11:58

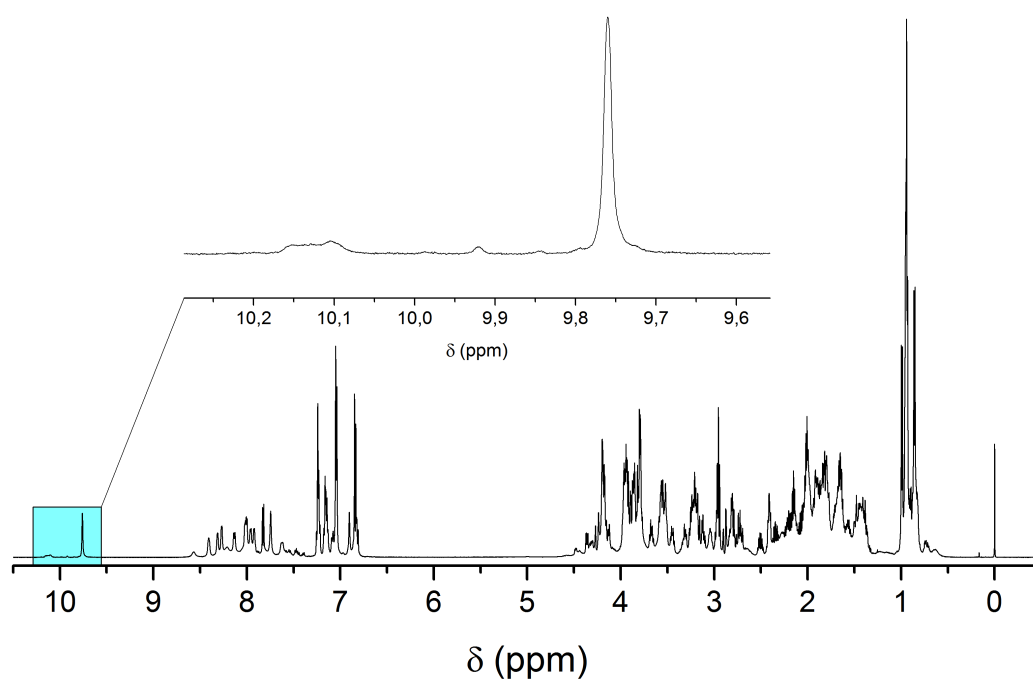
AG14\_purified.XLS/Integration

Chromeleon (c) Dionex 1996-2006  
Version 6.80 SR15 Build 4656 (243203)

**Figure V.1.1.:** HPLC chromatogram of purified Trp-cage Tc5b. Synthesis performed by S. ROTHEMUND.



**Figure V.1.2.:** Mass spectrum of the purified Trp-cage Tc5b. Synthesis performed by S. ROTHEMUND.  $m/z$ :  $[M + H]^+$  calcd for  $[Tc5b + H]^+$  2170.4, found 2169.6.



**Figure V.1.3.:**  $^1\text{H-NMR}$  spectrum of Tc5b at 298 K and pH 7 with enlargement of the indole proton region.



**Table V.1.2.:** Chemical shifts of Tc5b at 298 K and pH 3.

Residue		chemical shift $\delta$ (ppm)								
		NH	N	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asn			4.353	3.083, 3.018			7.457, 6.873		
2	Leu	8.815	123.32	4.286	1.504	1.682	0.943, 0.893			
3	Tyr	8.419	122.74	4.300	3.055		7.047	6.820		
4	Ile	8.033	118.26	3.869	1.316	1.875, 1.526, 0.833	0.891			
5	Gln	7.958	121.44	4.040	2.061	2.356	7.482, 6.914			
6	Trp	8.026	120.54	4.413	3.468, 3.215		7.344	9.845, 7.149	7.225, 7.129	7.336
7	Leu	8.207	121.44	3.7415	1.7177	1.5218	0.8954, 0.8393			
8	Lys	8.232	121.30	4.064	1.838	1.421, 1.507	1.650	2.954	7.533	
9	Asp	8.020	121.80	4.689	2.990, 2.856					
10	Gly	7.872	107.54	3.703, 3.384						
11	Gly	8.126	110.73	3.384, 2.454						
12	Pro			4.494	2.392, 1.990	2.055	3.653, 3.343			
13	Ser	8.067	114.38	4.454	3.904					
14	Ser	8.177	117.10	4.326	3.919, 3.742					
15	Gly	8.122	110.04	4.125, 3.898						
16	Arg	8.058	120.64	4.773	1.877, 1.772	1.683	3.232	7.294		
17	Pro			4.658	2.279, 1.818	1.996	3.820, 3.611			
18	Pro			3.946	1.410, 1.117	1.711	3.946			
19	Pro			4.393	2.251, 1.923	1.965	3.414, 3.252			
20	Ser	8.103	118.76	4.327	3.892, 3.827					

**Table V.1.3.:** Chemical shifts of Tc5b at 298 K and pH 8.

	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asn	exch.	3.834	2.801					
2	Leu	exch.	4.184	1.484	1.809	0.997, 0.942			
3	Tyr	8.572	4.192	3.171, 3.126		7.047	6.846		
4	Ile	8.002	3.798	1.966	1.683, 1.388, 0.954	0.952			
5	Gln	7.926	3.952	2.112	2.414	7.749, 6.905			
6	Trp	8.019	4.317	3.563, 3.211		7.051	9.761, 7.236	7.248, 7.147	7.163
7	Leu	8.273	3.528	1.920	1.631	0.952, 0.863			
8	Lys	8.315	3.969	1.916	1.572, 1.454	1.650	2.955	exch.	
9	Asp	7.961	4.568	2.884, 2.719					
10	Gly	7.627	4.141, 3.537						
11	Gly	8.410	3.225, 1.228						
12	Pro		4.600	2.511, 2.063	2.155	3.799, 3.451			
13	Ser	exch.	4.468	3.927					
14	Ser	exch.	4.201	3.887, 3.564					
15	Gly	7.925	4.248, 3.852						
16	Arg	8.132	4.996	1.870, 1.795	1.683	3.320, 3.246	exch.		
17	Pro		4.742	2.345, 1.814	2.013	3.864, 3.676			
18	Pro		2.826	1.424, 0.638	1.795, 1.717	3.553			
19	Pro		4.364	2.223, 1.997	1.878	3.222, 3.044			
20	Ser	7.830	4.185	3.795					

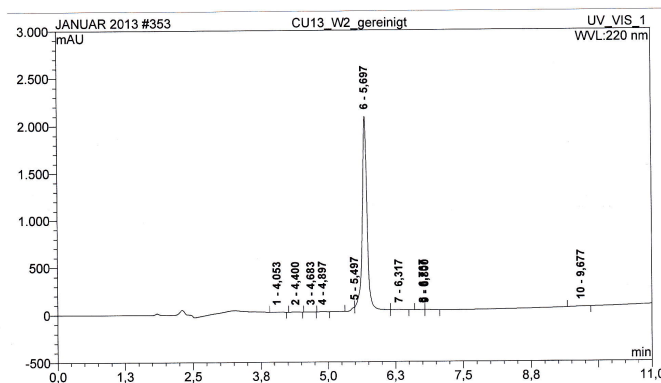
1.1.2. Synthesis Details of Tc5b-L2C-S20C

Operator:Administrator Timebase:U-3000RS Sequence:JANUAR 2013 Page 1-1  
22.10.2013 11:06 AM

**353 IZKF Leipzig, Core Unit Peptid-Technologien**  
Liebigstraße 21, 04103 Leipzig, e-mail: sven\_r@yahoo.com, Tel.: 0341 - 9715898 / 897

Comment:

Sample Name:	CU13_W2_gereinigt	Injection Volume:	30,0
Vial Number:	RE1	Channel:	UV_VIS_1
Sample Type:	unknown	Wavelength:	220.0
Control Program:	Peptide_3D_basic_short	Bandwidth:	4
Quantif. Method:	peptide_izkf	Dilution Factor:	1,0000
Recording Time:	22.10.2013 10:38	Operator:	Administrator
Sample ID:		Sample Amount:	1,0000



No.	Ret.Time min	Peak Name	Height mAU	Rel.Area %	Area mAU*min	Amount	Type
1	4,05	n.a.	0,9	0,07	0,15	n.a.	BMB
2	4,40	n.a.	0,6	0,04	0,08	n.a.	BMB
3	4,68	n.a.	0,4	0,03	0,06	n.a.	BMB
4	4,90	n.a.	2,2	0,09	0,20	n.a.	BMB
5	5,50	n.a.	42,2	1,09	2,41	n.a.	BM
6	5,70	n.a.	2059,2	97,79	216,31	n.a.	M
7	6,32	n.a.	4,8	0,46	1,03	n.a.	MB
8	6,77	n.a.	1,2	0,06	0,13	n.a.	BM
9	6,80	n.a.	1,3	0,11	0,24	n.a.	MB
10	9,68	n.a.	3,5	0,27	0,59	n.a.	BMB
<b>Total:</b>			2116,300	100,000	221,20	0,000	

IZKF/Integration

Chromeleon (c) Dionex 1996-2006  
Version 6.80 SR7 Build 2528 (146369)

Figure V.1.4.: HPLC chromatogram of purified Trp-cage Tc5b-L2C-S20C. Synthesis performed by S. ROTHEMUND.

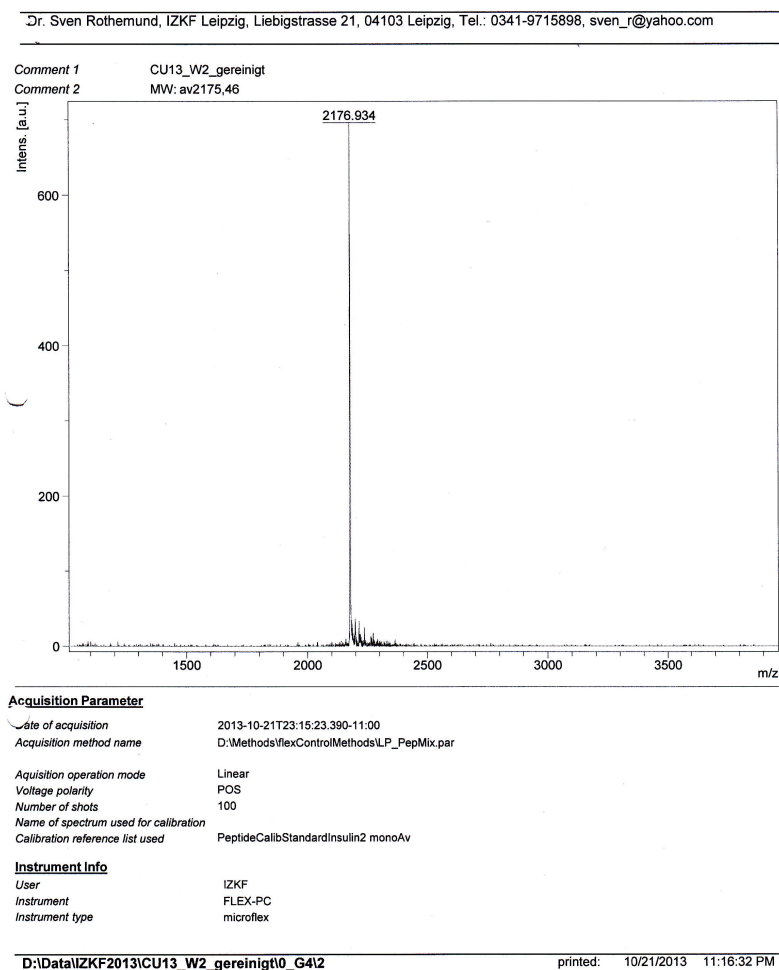
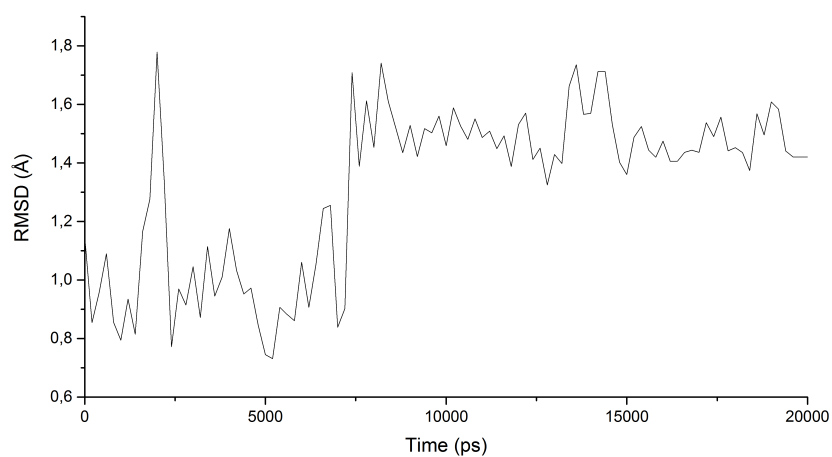
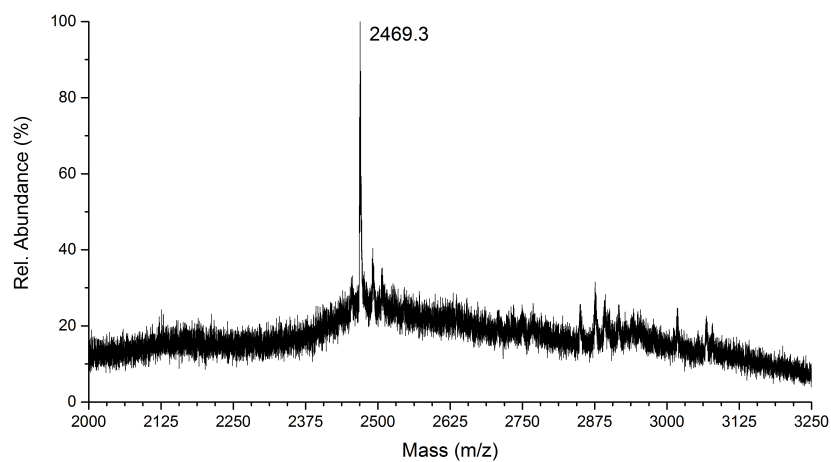


Figure V.1.5.:  $m/z$ :  $[M + H]^+$  calcd for  $[Tc5b - L2C - S20C + H]^+$  2176.5, found 2176.9.

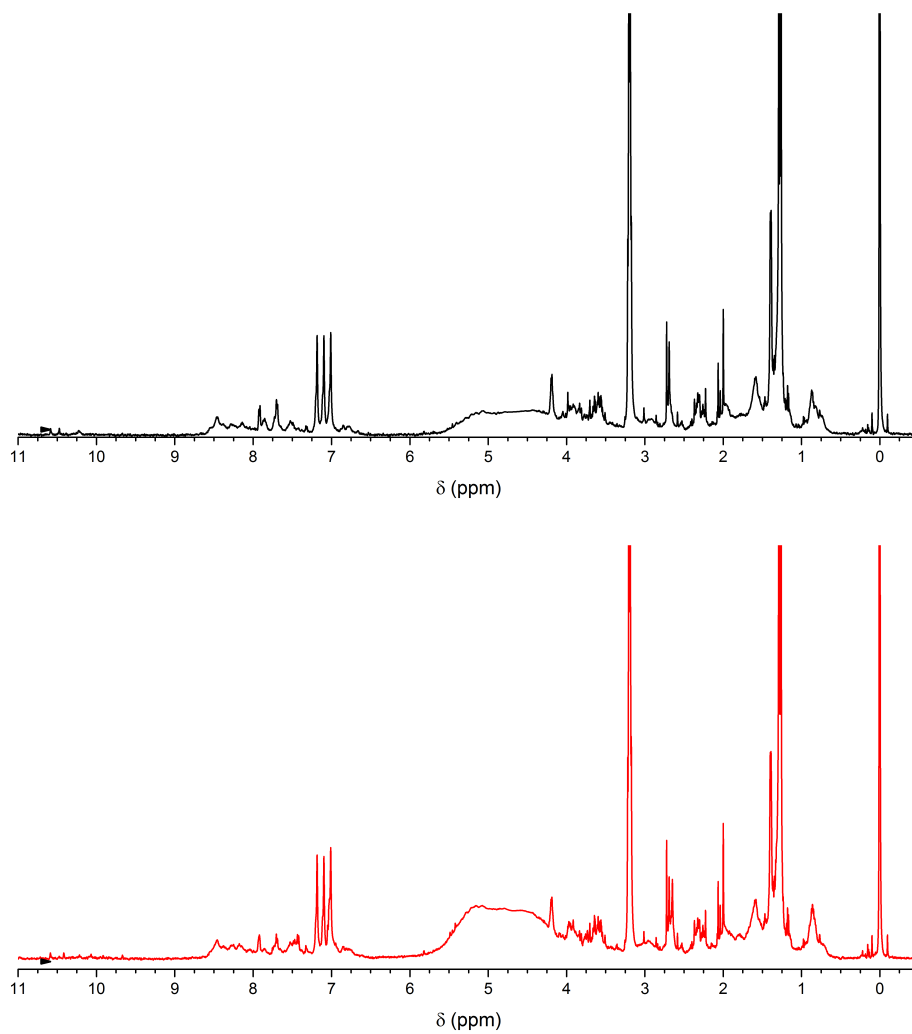


**Figure V.1.6.:** root-mean-square deviation of atomic positions (RMSD) of the  $C\alpha$  atoms during the MD simulation of Tc5b-L2C-S20C. The simulation was performed with a total duration of 20,000 ps and every 200 ps a structure was saved and the RMSD with respect to the first model determined.

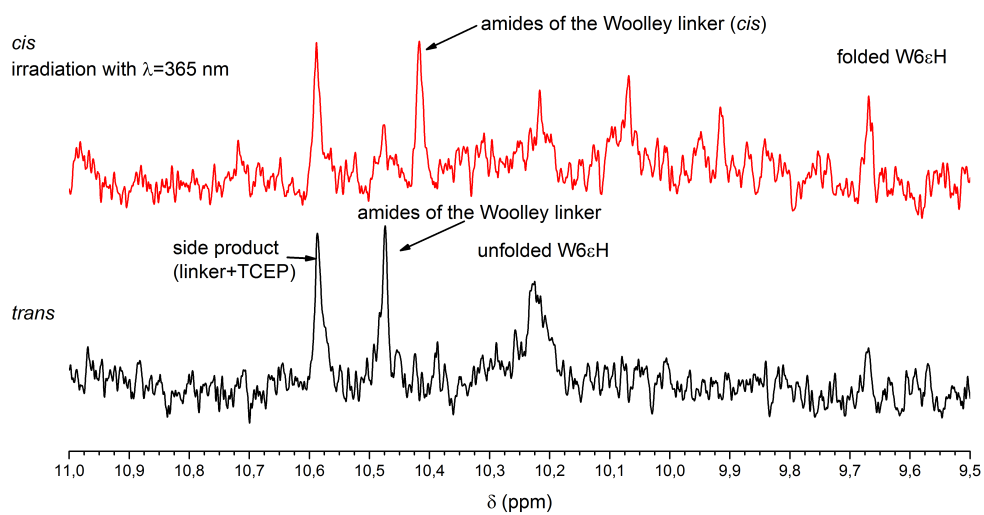
### 1.1.3. MALDI-ToF Spectra of Purified Cross-Linked Tc5b-L2C-S20C



**Figure V.1.7.:** MALDI-ToF spectrum of the fraction collected with a  $t_R = 17.8$  min, containing cross-linked Tc5b-L2C-S20C with the WOOLLEY linker.  $m/z$ :  $[M + H]^+$  calcd for  $[Tc5b - L2C - S20C - Woolley - linker + H]^+$  2468.1, found 2469.3.

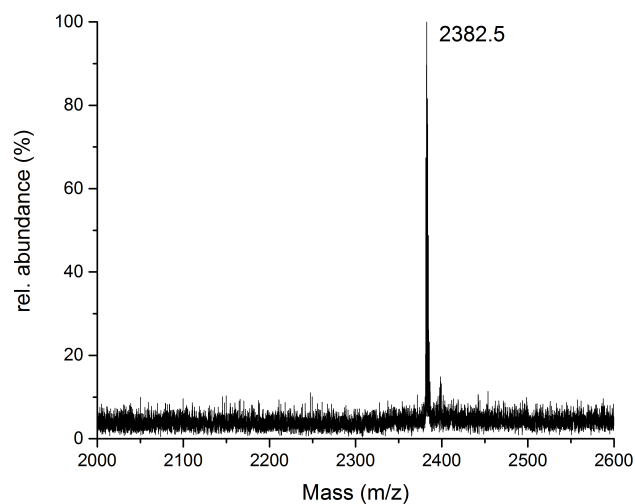


**Figure V.1.8.:**  $^1\text{H}$  NMR spectra between of Tc5b-L2C-S20C cross-linked with the WOOLLEY-linker (top) in *trans* (black) and *cis* (red) at 298 K and pH 3.

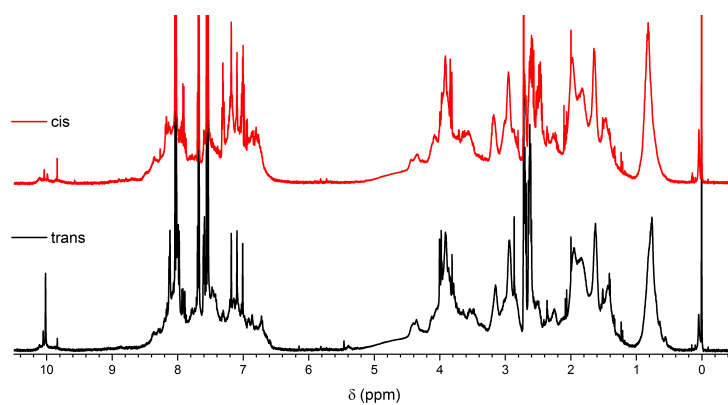


**Figure V.1.9.:** Enlarged  $^1\text{H}$  NMR of the indole and amide region of the purified cross-linked Tc5b-L2C-S20C with the WOOLLEY linker in *trans* (bottom/black) and *cis* (top/red) at 298 K and pH 3.





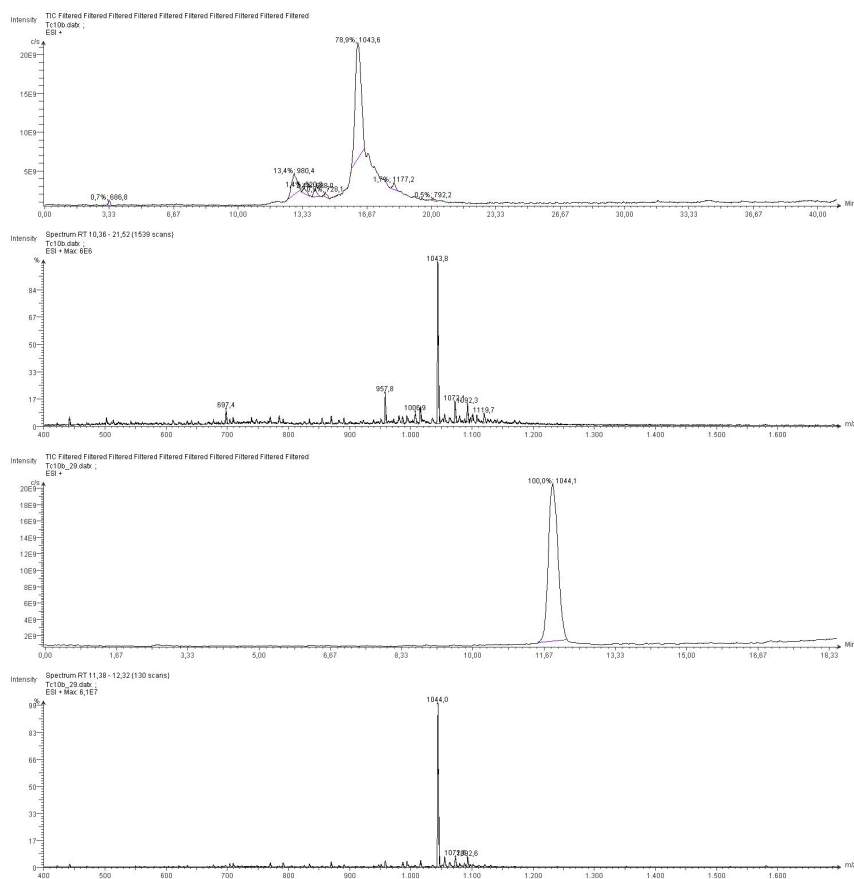
**Figure V.1.10.:** MALDI-ToF spectrum of cross-linked Tc5b-L2C-S20C with 4,4'-dibromomethyl azobenzene linker.  $m/z$ :  $[M + H]^+$  calcd for  $[Tc5b - L2C - S20C - (4,4' - dibromomethylazobenzene) + H]^+$  2382.7, found 2382.5



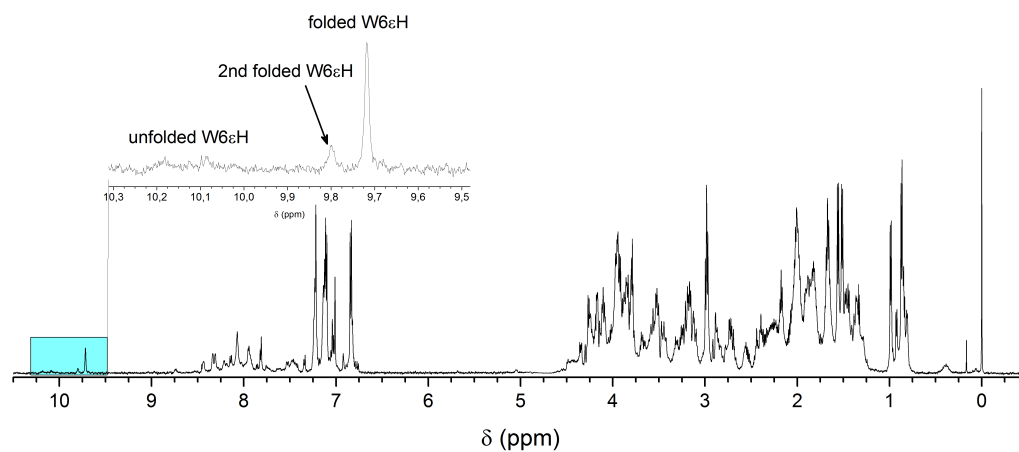
**Figure V.1.11.:**  $^1H$  NMR spectra of Tc5b-L2C-S20C cross-linked with 4,4'-dibromomethyl azobenzene (top) in *trans* (black) and *cis* (red) at 298 K and pH 3.

## 1.2. Tc10b and its Variants

### 1.2.1. Tc10b



**Figure V.1.12.:** Total ion chromatogram of crude Tc10b (top) and purified (bottom).  $m/z$ :  $[M + 2H]^{2+}$  calcd for  $[Tc5b - L2C - S20C + 2H]^{2+}$  1044.1, found 1044.1  
 . Isolated yield: 9mg (4.3  $\mu$ mol, 2.3%).



**Figure V.1.13.:**  $^1\text{H}$ -NMR spectrum of Tc10b at 298 K and pH 7 with enlargement of the indole proton region.

**Table V.1.4.:** Sum of CSDs for selected protons in the helical region representing the helix fold of Tc10b at 280 K at pH 7.<sup>43</sup>

Helix	$\Delta\delta$ (ppm).							$\Sigma$
	A2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	Q5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$	
pH 7	-0.043	-0.522	-0.200	-0.442	-0.423	-0.953	-0.386	<b>-2.969</b>

**Table V.1.5.:** Sum of CSDs for selected protons in the cage structure representing the cage fold of Tc10b at 280 K at pH 7.<sup>43</sup>

Cage	$\Delta\delta$ (ppm)						$\Sigma$
	upfield						
	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$	
pH 7	-0.953	-3.483	-2.329	-1.697	-0.664	-0.449	<b>-9.575</b>
downfield							
	P12 $\alpha$	P12 $\beta$	R16 $\alpha$				
pH 7	0.229	0.251	0.436				<b>0.916</b>

**Table V.1.6.:**  $^1\text{H}$  chemical shifts of Tc10b at 298 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.332	3.194, 3.053					
2	Ala	8.936	4.276	1.471					
3	Tyr	8.595	4.116	3.087		7.077	6.813		
4	Ala	8.170	4.101	1.512					
5	Gln	7.935	4.003	2.143	2.411				
6	Trp	8.068	4.292	3.511, 3.175		7.283	9.731, 7.094	7.221, 7.113	7.095
7	Leu	8.289	3.514	1.748, 1.448	1.539	0.926, 0.858			
8	Lys	7.977	4.041	1.880	1.453	1.659	2.972		
9	Asp	7.854	4.680	3.050, 2.884					
10	Gly	7.668	4.111, 3.596						
11	Gly	8.105	3.116, 1.722						
12	Pro		4.528	2.451, 2.012	2.082	3.685, 3.280			
13	Ser	7.885	4.443	3.909					
14	Ser	8.146	4.276	3.919, 3.682					
15	Gly	8.021	4.187, 3.864						
16	Arg	8.033	4.843	1.895, 1.795	1.686	3.250			
17	Pro		4.652	2.257, 1.794	1.992	3.837, 3.625			
18	Pro		3.064	1.464, 0.887	1.712	3.064, 3.070			
19	Pro		4.385	2.223, 1.941	1.855	3.263, 3.066			
20	Ser	8.184	4.375	3.916, 3.822					

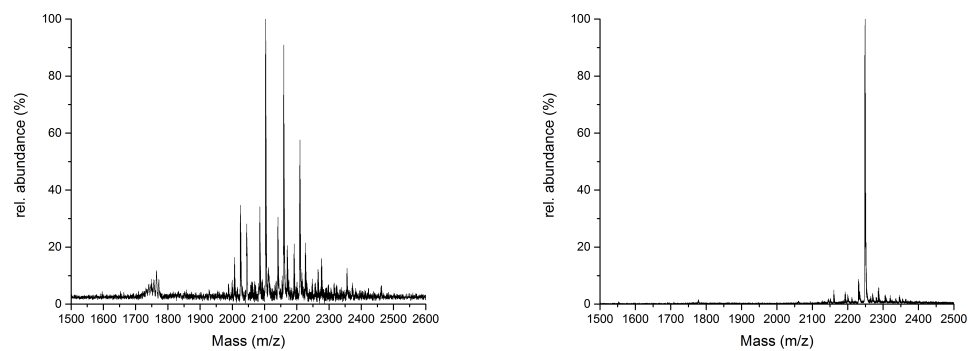
**Table V.1.7.:**  $^1\text{H}$ -chemical shifts of Tc10b at 298 K and pH 8

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.078	2.900	2.869				
2	Ala	exch.	4.270	1.511					
3	Tyr	8.738	4.090	3.197, 3.110		7.100	6.804		
4	Ala	8.306	4.113	1.561					
5	Gln	8.053	3.928	2.228, 2.156	2.409	7.810	6.921		
6	Trp	8.074	4.264	3.170, 3.558		7.010	9.717, 7.036	7.037, 7.221	7.222
7	Leu	8.330	3.436	1.876, 1.384	1.636	0.980, 0.867			
8	Lys	8.068	3.957	1.959, 1.914	1.670	1.510	2.982		
9	Asp	7.948	4.546	2.900, 2.711					
10	Gly	7.528	4.156, 3.490						
11	Gly	8.433	3.145, 0.898						
12	Pro		4.621	2.533, 2.065	2.171	3.820, 3.450			
13	Ser	7.747	4.213	3.781					
14	Ser	8.211	4.166	3.864, 3.516					
15	Gly	7.937	4.280, 3.816						
16	Arg	8.136	5.034	1.909, 1.847	1.676	3.250, 3.316	exch.		
17	Pro		4.765	2.346, 1.804	2.009	3.870, 3.680			
18	Pro		2.582	1.357, 0.405	1.761, 1.671	3.52			
19	Pro		4.348	2.215, 1.982	1.839, 1.852	3.160, 2.968			
20	Ser	7.818	4.178	3.787					

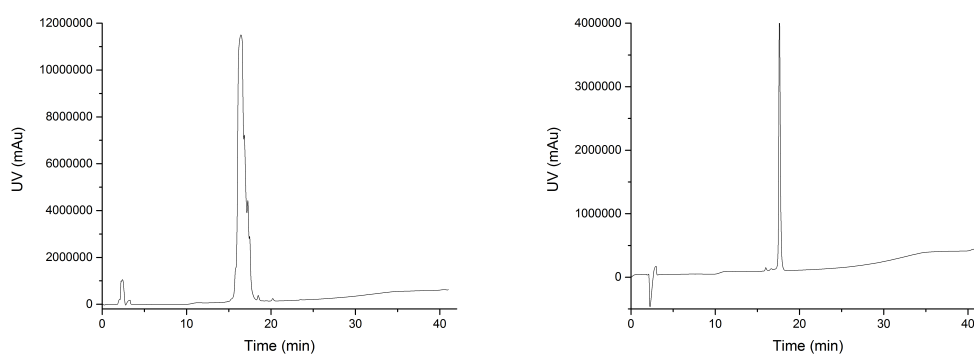
**Table V.1.8.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of Tc10b at 298 K and pH 3.

Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
	NH	other	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$	
1 Asp	n.d.		52.70	38.68						
2 Ala	124.01		18.40	18.40						
3 Tyr	121.63		62.06	38.28		133.01	118.39			
4 Ala	121.44		18.40	18.40						
5 Gln	118.00	112.88	58.16	28.53	33.79					
6 Trp	122.10	131.05	53.29	28.14		127.36	122.68	124.24	114.49, 120.93	
7 Leu	120.11		57.58	42.18	26.19	25.41, 23.66				
8 Lys	121.20	n.d.	58.35	31.61	25.02	29.12	42.18			
9 Asp	117.29		n.d.	38.28						
10 Gly	106.82		44.46							
11 Gly	111.51		43.74							
12 Pro	n.d.		64.79	31.85	27.36	50.36				
13 Ser	113.66		58.94	63.62						
14 Ser	117.02		59.91	64.79						
15 Gly	109.87		45.25							
16 Arg	120.26	124.99	n.d.	31.22	26.78	43.54				
17 Pro	n.d.		n.d.	30.68	27.17	50.56				
18 Pro	n.d.		60.30	29.70	n.d.	n.d.				
19 Pro	n.d.		62.64	n.d.	n.d.	49.78				
20 Ser	117.18		58.35	64.01						

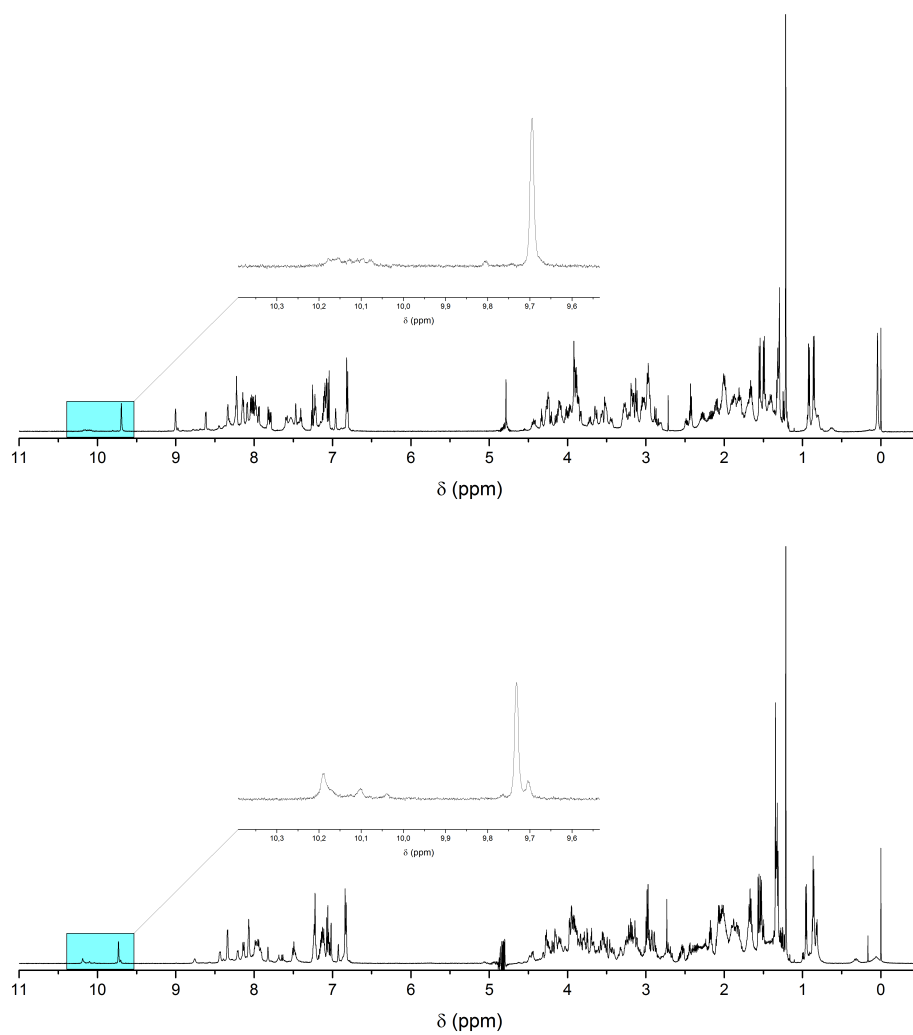
**1.2.2. Tc10b-S20C-21G**



**Figure V.1.14.:** MALDI-ToF spectrum of the crude Tc10b-S20C-21G (left) and purified (right).  
 $m/z$ :  $[M + H]^+$  calcd for  $[Tc10b - S20C - 21G + H]^+$  2160.3, found 2159.7.



**Figure V.1.15.:** HPLC chromatogram of the crude Tc10b-S20C-21G (left) and purified (right).  
Isolated yield 4mg (1.9  $\mu$ mol, 2.0%).



**Figure V.1.16.:**  $^1\text{H}$  NMR of Tc10b-S20C-21G at pH 3 and 298 K (top) and pH 7 (bottom).



**Table V.1.9.:**  $^1\text{H}$  chemical shifts of Tc10b-S20C-21G at 298 K and pH 3.

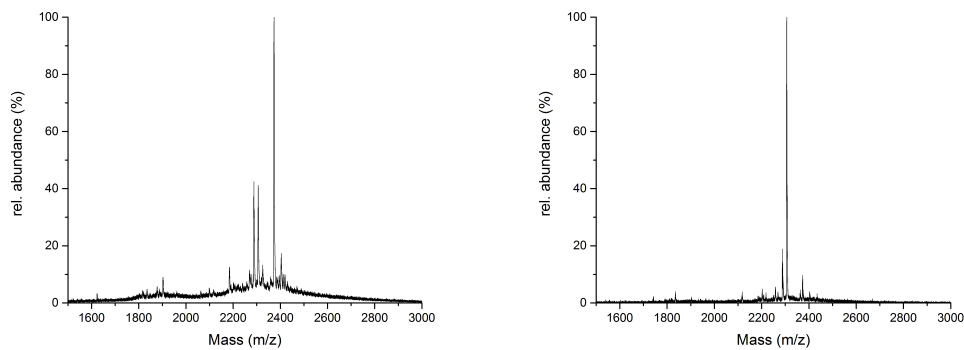
Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.335	3.185, 3.030					
2	Ala	9.004	4.272	1.494					
3	Tyr	8.616	4.109	3.121		7.052	6.813		
4	Ala	8.224	4.090	1.549					
5	Gln	7.939	3.982	2.171	2.431				
6	Trp	8.088	4.252	3.549, 3.172		7.076	9.697, 7.087	7.261, 7.115	7.223
7	Leu	8.338	3.442	1.783, 1.421	1.550	0.918, 0.857			
8	Lys	7.989	4.019	1.895	1.491	1.657	2.976	7.536	
9	Asp	7.818	4.663	3.050, 2.863					
10	Gyl	7.590	4.131, 3.545						
11	Gly	8.152	3.033, 1.374						
12	Pro		4.547	2.482, 2.025	2.105	3.714, 3.284			
13	Ser	7.791	4.448	3.917					
14	Ser	8.140	4.245	3.909, 3.639					
15	Gly	7.974	4.228, 3.846						
16	Arg	8.040	4.892	1.905, 1.818	1.702	3.265	7.405		
17	Pro		4.677	2.276, 1.806	2.000	3.862, 3.645			
18	Pro		2.817	1.409, 0.625	1.808, 1.685	3.507			
19	Pro		4.422	2.103, 2.014	1.871	3.184, 2.961			
20	Cys	8.013	4.598	3.147, 2.970					
21	Gly	8.226	3.888						

**Table V.1.10.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of Tc10b-S20C-21G at 298 K and pH 3.

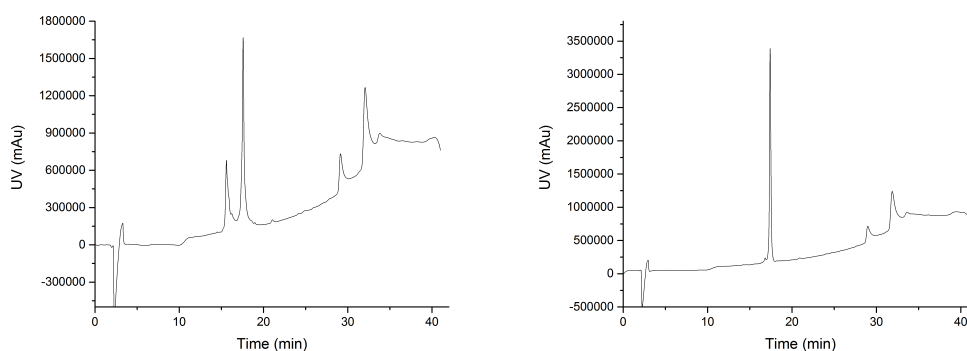
Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
	NH	other	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$	
1 Asp	n.d.		n.d.	n.d.						
2 Ala	123.84		n.d.	18.16						
3 Tyr	121.84		n.d.	38.14		132.97	118.15			
4 Ala	121.34		n.d.	18.16						
5 Gln	117.84	113.34	n.d.	n.d.	33.51					
6 Trp	122.25	131.31	n.d.	n.d.		127.36				
7 Leu	119.72		57.55	n.d.		25.41, 23.52	120.78	124.20	114.21, 122.54	
8 Lys	121.44		n.d.	n.d.	n.d.	n.d.	n.d.			
9 Asp	117.34		n.d.	n.d.						
10 Gyl	106.43		44.08							
11 Gly	112.22		n.d.							
12 Pro	n.d.		n.d.	n.d.	n.d.	n.d.				
13 Ser	113.34		n.d.	n.d.						
14 Ser	117.06		n.d.	n.d.						
15 Gly	109.78		45.02							
16 Arg	120.00	117.00	n.d.	n.d.	n.d.	n.d.				
17 Pro	n.d.		n.d.	n.d.						
18 Pro	n.d.		n.d.	29.38	n.d.	n.d.				
19 Pro	n.d.		n.d.	n.d.	n.d.	49.96				
20 Cys	118.00		n.d.	n.d.						
21 Gly	112.90		n.d.							

**Table V.1.11.:** <sup>1</sup>H-chemical shifts of Tc10b-S20C-21G at 298 K and pH 7.

	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.066	2.901, 2.874					
2	Ala	exch.	4.266	1.518					
3	Tyr	8.748	4.111	3.201, 3.131		7.055	6.817		
4	Ala	8.327	4.081	1.551					
5	Gln	8.046	3.906	2.232, 2.155	2.412	7.813, 6.912			
6	Trp	8.065	4.243	3.565, 3.164		7.006	9.720, 7.030	7.134, 7.131	7.213
7	Leu	8.326	3.406	1.864, 1.353	1.533	0.940, 0.896			
8	Lys	8.052	3.935	1.960, 1.896	1.653	1.492	2.973		
9	Asp	7.933	4.545	2.897, 2.702					
10	Gyl	7.49	4.155, 3.467						
11	Gly	8.422	3.111, 0.800						
12	Pro		4.612	2.530, 2.066	2.650	3.815, 3.440			
13	Ser	7.954	4.611	3.171, 2.936					
14	Ser	8.198	4.146	3.869, 3.491					
15	Gly	7.911	4.281, 3.805						
16	Arg	8.121	5.046	1.895, 1.829	1.829, 1.666	3.312, 3.235	exch.		
17	Pro		4.759	2.346, 1.801	2.008	3.871			
18	Pro		2.524	1.332, 0.309	1.787, 1.691	3.531			
19	Pro		4.436	2.063	1.865	3.137, 2.905			
20	Cys	8.005	4.247	1.808, 1.683					
21	Gly	7.976	3.747, 3.667						

**1.2.3. Tc10b-A2K-S20C(S-tBu)-21G**

**Figure V.1.17.:** MALDI-ToF spectrum of the crude Tc10b-A2K-S20C(S-tBu)-21G (left) and purified (right).  $m/z$ :  $[M + H]^+$  calcd for  $[Tc10b - A2K - K8A - S20C - 21G + H]^+$  2305.6, found 2306.3.



**Figure V.1.18.:** HPLC-chromatogram of the crude Tc10b-A2K-S20C(S-tBu)-21G (left) and purified (right). Isolated yield 8mg (3.5  $\mu$ mol, 4.0%).

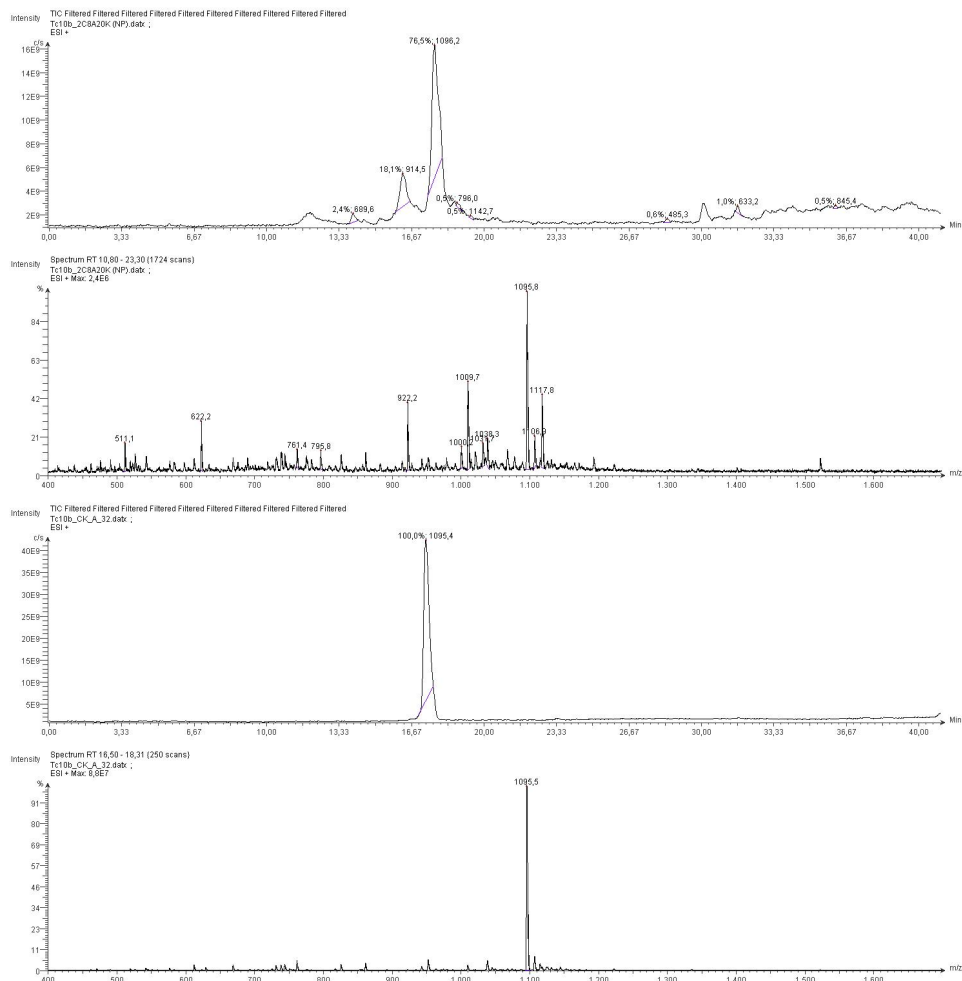
**Table V.1.12.:** <sup>1</sup>H-chemical shifts of Tc10b-A2K-S20C-21G at 298 K and pH 3.

Residue			chemical shift $\delta$ (ppm)						
1	Asp	exch.	4.336	3.128, 2.991					
2	Lys	8.888	4.219	1.786	1.412	1.663	2.975	7.537	
3	Tyr	8.448	4.241	3.120, 3.059		6.811	7.080		
4	Ala	8.213	4.100	1.453					
5	Gln	7.953	4.033	2.102	2.379				
6	Trp	8.008	4.371	3.478, 3.205		7.135	9.823, 7.328	7.327, 7.123	7.227
7	Leu	8.205	3.658	1.470	1.690	0.897, 0.846			
8	Lys	7.989	4.069	1.858	1.450	1.657	2.967		
9	Asp	7.950	4.686	3.003, 2.865					
10	Gyl	7.786	4.077, 3.664						
11	Gyl	8.088	3.272, 2.205						
12	Pro		4.503	2.413, 2.004	2.054	3.661, 3.313			
13	Ser	7.999	4.447	3.910					
14	Ser	8.159	4.312	3.92, 3.729					
15	Gly	8.091	4.140, 3.885						
16	Arg	8.038	4.823	1.887, 1.772	1.682	3.229	7.307		
17	Pro		4.650	2.271, 1.822	1.998	3.833, 3.616			
18	Pro		2.910	1.591, 1.203	1.660	4.208			
19	Pro		4.427	2.133, 1.986	1.909	3.318, 3.144			
20	Cys	8.079	4.627	3.162, 3.002		1.247 (S- tBu)			
21	Gly	8.252	3.908						

**Table V.1.13.:** <sup>1</sup>H-chemical shifts of Tc10b-A2K-S20C-21G at 298 K and pH 7.

Residue		chemical shift $\delta$ (ppm)							
1	Asp	exch.	n.d.						
2	Lys	8.296	3.9561	1.917, 1.501	1.657	2.962	exch.		
3	Tyr	8.704	4.173	3.222, 3.104		7.123	6.822		
4	Ala	8.358	4.08	1.522					
5	Gln	8.011	3.931	2.213, 2.209	2.420	7.792, 6.906			
6	Trp	7.958	4.283	3.541, 3.179		7.038	9.720, 7.220	7.487, 7.148	7.636
7	Leu	8.296	3.453	1.842, 1.875	1.585	0.940, 0.849			
8	Lys	8.042	3.958	1.938, 1.886	1.658	1.439	2.969		
9	Asp	7.929	4.544	2.878, 2.699					
10	Gyl	7.535	4.144, 2.504						
11	Gyl	8.357	3.135, 1.058						
12	Pro		4.588	2.510, 2.054	2.147	3.738, 3.400			
13	Ser	7.982	4.639	3.177, 2.937					
14	Ser	8.183	4.184	3.877, 3.542					
15	Gly	8.007	3.763, 3.687						
16	Arg	8.100	4.988	1.863, 1.773	1.67	3.31, 3.228	exch		
17	Pro		4.745	2.336, 1.817	2.004	3.787, 3.673			
18	Pro		2.675	1.389, 0.550	1.832, 1.728	3.545			
19	Pro		4.447	2.288, 2.023	1.865	3.116, 2.917			
20	Cys	8.181	4.18	3.689, 3.544		1.244 (S- tBu)			
21	Gly	7.970	3.755						

## 1.2.4. Tc10b-A2C(S-tBu)-S20K



**Figure V.1.19.:** Total ion chromatogram of crude Tc10b-A2C(S-tBu)-S20K (top) and purified (bottom).  $m/z$ :  $[M + 2H]^{2+}$  calcd for  $[Tc10b - A2C - K8A - S20K + 2H]^{2+}$  1096.2, found 1095.5. Isolated yield 9mg (3.9  $\mu$ mol, 4.5%).

**Table V.1.14.:**  $^1\text{H}$ -chemical shifts of Tc10b-A2C(S-tBu)-S20K at 298 K and pH 3.

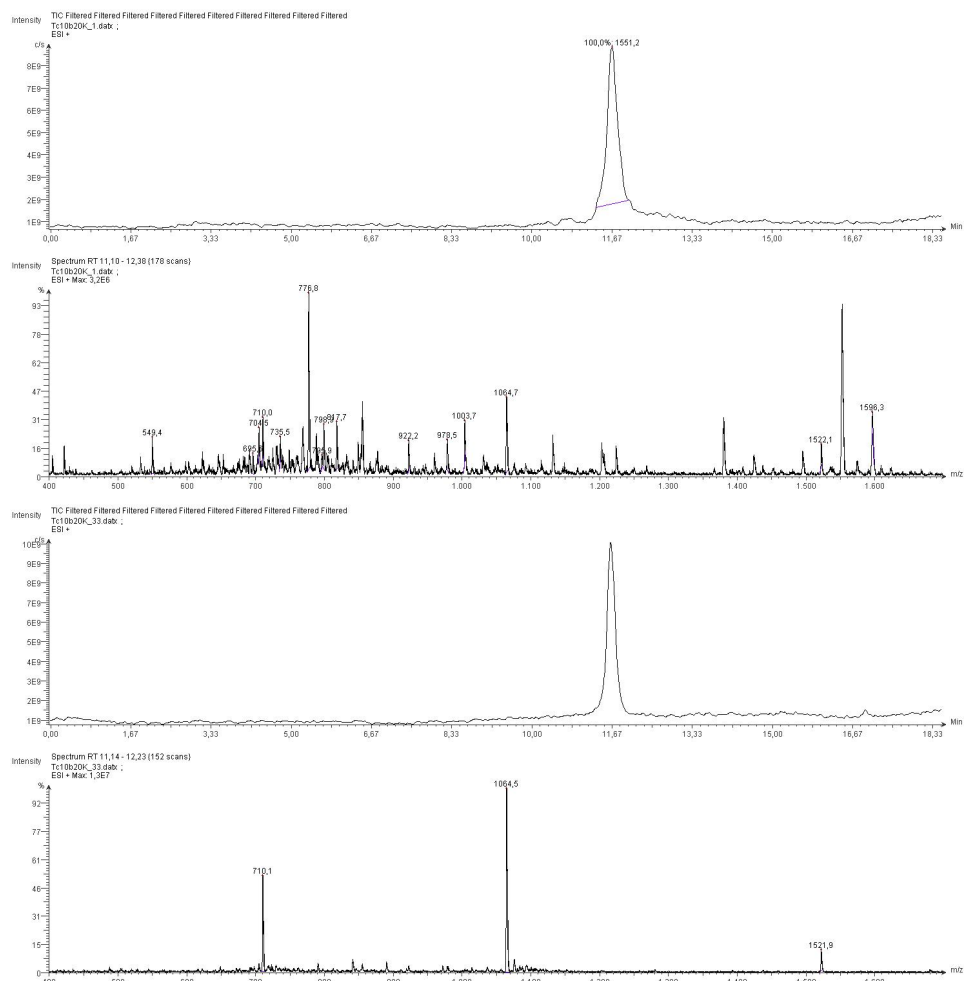
Residue		chemical shift $\delta$ (ppm)							
1	D	exch.	4.347	3.054, 2.937					
2	C	8.879	4.632	3.163, 3.115		1.320 (S- tBu)			
3	Y	8.619	4.191	3.106, 3.061		7.089	6.812		
4	A	8.320	4.095	1.473					
5	Q	7.904	4.084	2.156	2.393				
6	W	8.081	4.3347	3.452		7.116	9.841, 7.313	7.219, 7.224	7.123
7	L	8.290	3.604	1.660	1.473	0.895, 0.853			
8	A	7.975	4.120	1.442					
9	D	7.846	4.703	3.015, 2.861					
10	G	7.763	4.085, 3.650						
11	G	8.075	3.205, 2.119						
12	P		4.488	2.408, 1.9955	2.4044	3.649, 3.275			
13	S	7.975	4.437	3.899					
14	S	8.14	4.294	3.911, 3.728					
15	G	8.074	4.136, 3.874						
16	R	8.04	4.819	1.869, 1.783	1.679	3.223	7.391		
17	P		4.658	2.278, 1.808	2.001	3.8325, 3.617			
18	P		3.413	1.514, 1.187	1.753, 1.846	3.519, 3.579			
19	P		4.348	2.205	1.856	3.353, 3.134			
20	K	8.239	4.176	1.774, 1.710	1.405	1.642	2.953	7.506	



**Table V.1.15.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of Tc10b-A2C(S-tBu)-S20K at 298 K and pH 3.

Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
	NH	other	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$	
1 D			n.d.	n.d.						
2 C	118.59		n.d.	42.71						
3 Y	124.15		n.d.	n.d.		133.12	118.35			
4 A	121.63		54.23	18.43						
5 Q	118.73	112.64	57.85	28.83	34.23					
6 W	121.87	130.66	62.59	28.17		127.32	114.43	122.34	124.23, 120.91	
7 L	120.52		57.32	42.26	24.73	25.13, 23.58				
8 A	122.92		53.78	17.98						
9 D	116.39		n.d.	38.34						
10 G	107.22		44.82							
11 G	109.94		44.26							
12 P	n.d.		n.d.	31.83	27.30	50.38				
13 S	114.04		59.13	63.49						
14 S	117.06		59.50	64.40						
15 G	110.88		45.42							
16 R	120.55	114.98	n.d.		26.90	n.d.				
17 P	n.d.		n.d.	30.48	27.30	50.45				
18 P	n.d.		n.d.	30.15						
19 P	n.d.		52.93	n.d.	n.d.	50.23				
20 K	121.60		56.19	33.14	24.87	n.d.	39.32			

## 1.2.5. Tc10b-S20K

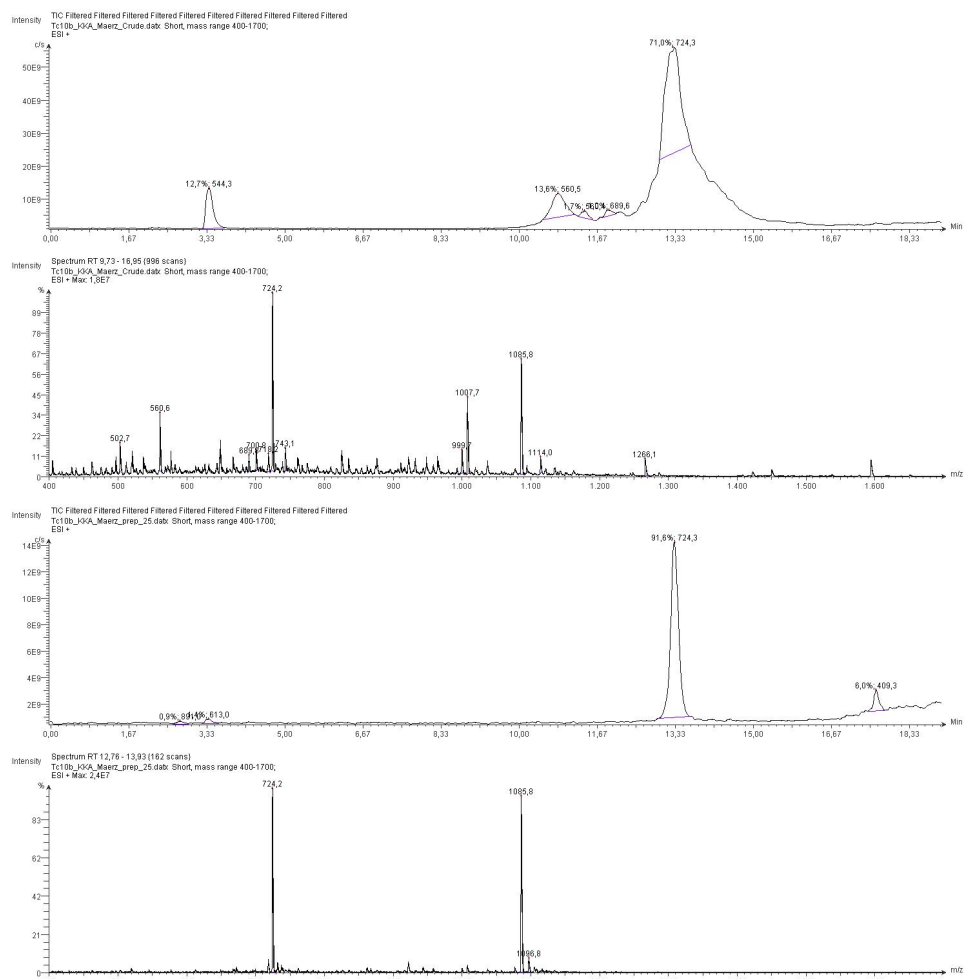


**Figure V.1.20.:** Total ion chromatogram of crude Tc10b-S20K (top) and purified (bottom).  
 $m/z$ :  $[M + 2H]^{2+}$  calcd for  $[Tc10b - S20K + 2H]^{2+}$  1064.7, found 1065.5.  
 Isolated yield 4mg (1.9  $\mu\text{mol}$ , 2.0%).

**Table V.1.16.:** <sup>1</sup>H-chemical shifts of Tc10b-20K at 298 K and pH 3.

	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.328	3.251, 3.050					
2	Ala	9.070	4.253	1.500					
3	Tyr	8.838	4.039	3.108		7.098	6.813		
4	Ala	8.268	4.115	1.582					
5	Gln	8.029	3.971	2.206, 2.142	2.435				
6	Trp	8.153	4.229	3.562, 3.141		7.057	9.684, 7.114	7.121, 7.235	7.228
7	Leu	8.411	3.396	1.815, 1.421	1.577	0.964, 0.878			
8	Lys	8.033	4.012	1.941, 1.882	1.488	1.669	2.976		
9	Asp	7.847	4.647	3.059, 2.872					
10	Gly	7.593	4.155, 3.514						
11	Gly	8.216	2.998, 1.156						
12	Pro		4.570	2.496, 2.042	4.573	3.738, 3.303			
13	Ser	7.826	4.453	3.924					
14	Ser	8.191	4.228	3.909, 3.619					
15	Gly	8.017	4.252, 3.848						
16	Arg	8.109	4.920	1.891, 1.845	1.740, 1.682	3.275	7.512		
17	Pro		4.670	2.259, 1.781	2.004	3.873, 3.666			
18	Pro		2.661	1.334, 0.486	1.708, 1.748	3.502			
19	Pro		4.338	2.204, 1.913	1.820	3.194, 2.950			
20	Lys	8.269	4.146	1.811, 1.679	1.387	1.655	2.947	7.205	

## 1.2.6. TcKKA



**Figure V.1.21.:** Total ion chromatogram of crude TcKKA (top) and purified (bottom).  $m/z$ :  $[M + 2H]^{2+}$  calcd for  $[Tc10b - Ac - A2K - K8A - S20K + 2H]^{2+}$  1085.7, found 1085.8. Isolated yield 11mg (5.1  $\mu$ mol, 5.5%).

**Table V.1.17.:** <sup>1</sup>H-chemical shifts of TcKKA at 298 K and pH 3.

	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.4062	4.6525	2.8836					
2	Lys	8.5063	4.1568	1.8234, 1.7703	1.4479	1.6741	2.9905	7.5739	
3	Tyr	8.5036	4.1962	3.1441, 3.0641		7.0864	6.814		
4	Ala	8.1768	4.0992	1.4946					
5	Gln	8.0786	4.0712	2.1098	2.4371, 2.3452	7.4330, 6.8807			
6	Trp	7.9501	4.3055	3.4808, 3.2028		7.110	9.7563, 7.1685	7.2974, 7.2239	7.1196
7	Leu	8.3419	3.5222	1.7222, 1.4507	1.5465	0.9214, 0.8587			
8	Ala	8.0356	4.1108	1.4472					
9	Asp	7.7653	4.7132	3.0162, 2.8491					
10	Gly	7.6995	4.1098, 3.6065						
11	Gly	8.0797	3.1181, 1.7940						
12	Pro		4.5144	2.4444, 2.0160	2.0724	3.6848, 3.2715			
13	Ser	7.8952	4.4367	3.907					
14	Ser	8.1334	4.2807	3.9167, 3.6947					
15	Gly	8.0231	4.1833, 3.8635						
16	Arg	8.0443	4.8384	1.9033, 1.7979	1.6902	3.2473			
17	Pro		4.6732	2.2747, 1.8088	2.0034	3.8490, 3.6430	7.3373		
18	Pro		3.1086	1.4530, 0.9392	1.8107, 1.0794	3.5211			
19	Pro		4.3303	2.1838, 1.8908	1.8591	3.2609, 3.0614			
20	Lys	7.9882	4.1376	1.7755, 1.6829	1.3739	1.6219	2.9412	7.4731	

**Table V.1.18.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of TcKKA at 298 K and pH 3.

Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
	NH	other	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$	
1 Asp	125.81		n.d.	39.33						
2 Lys	121.56		58.38	n.d.	n.d.	n.d.	42.10			
3 Tyr	122.67		61.34	38.04		133.06	118.38			
4 Ala	121.91		54.63	18.10						
5 Gln	118.68	112.48	58.29	28.54	33.98					
6 Trp	121.94	130.99	n.d.	n.d.		127.29	120.90	114.39, 124.10	122.38	
7 Leu	120.56		57.51	42.59	26.36	25.33, 23.56				
8 Ala	122.99		54.39	18.10						
9 Asp	116.40		n.d.	38.52						
10 Gly	106.88		44.72							
11 Gly	111.27		43.92							
12 Pro	n.d.		n.d.	31.76	n.d.	n.d.				
13 Ser	113.73		n.d.	63.14						
14 Ser	117.04		59.39	65.12						
15 Gly	109.86		45.21							
16 Arg	120.38	114.95	n.d.	n.d.	n.d.	50.43				
17 Pro	n.d.		n.d.	30.85	n.d.	n.d.				
18 Pro	n.d.		n.d.	29.74	n.d.	n.d.				
19 Pro	n.d.		62.32	31.59	n.d.	n.d.				
20 Lys	124.58		57.19	n.d.	n.d.	n.d.	42.10			

**Table V.1.19.:** <sup>1</sup>H-chemical shifts of TcKKA at 278 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.571	4.659	2.930, 2.929					
2	Lys	8.665	4.153	1.883, 1.809	1.586, 1.482	1.697	3.009	7.674	
3	Tyr	8.767	4.113	3.114, 3.101		7.099	6.816		
4	Ala	8.304	4.106	1.548					
5	Gln	8.166	4.060	2.144	2.481, 2.370		7.371, 7.018		
6	Trp	7.993	4.241	3.531, 3.169		7.075	9.693, 7.243	7.236, 7.040	7.127
7	Leu	8.510	3.394	1.777, 1.443	1.583	0.961, 0.887			
8	Ala	8.114	4.107	1.468					
9	Asp	7.722	4.709	3.040, 2.876					
10	Gly	7.649	4.150, 3.530						
11	Gly	8.151	2.965, 1.252						
12	Pro		4.558	2.486, 2.052	2.111, 2.116	3.272			
13	Ser	7.845	4.448	3.918					
14	Ser	8.181	4.239	3.913, 3.644					
15	Gly	8.032	4.241, 3.848						
16	Arg	8.120	4.909	1.901, 1.842	1.728, 1.686	3.296, 3.247	7.497		
17	Pro		4.693	2.274, 1.793	2.005	3.881, 3.664			
18	Pro		2.698	1.351, 0.557	1.745, 1.662	3.521, 3.507			
19	Pro		4.318	2.199, 1.890	1.831	3.188, 2.958			
20	Lys	8.167	4.117	1.781, 1.655	1.374	1.621	2.931	7.557	

**Table V.1.20.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of TcKKA at 278 K and pH 3.

Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
1	Asp	125.68		n.d.	39.01					
2	Lys	121.01		58.85	30.64	26.07		41.86		
3	Tyr	123.29		62.26	37.88		118.14	133.02		
4	Ala	121.06		n.d.	17.75					
5	Gln	118.44	112.79	58.25	n.d.	33.82				
6	Trp	121.85	130.91	59.92	42.99		127.32	123.83, 113.80	122.71	120.69
7	Leu	119.70		57.41	42.05		23.40, 25.32			
8	Ala	122.52		54.35	17.59					
9	Asp	116.01		n.d.	n.d.					
10	Gly	106.11		n.d.						
11	Gly	111.74		n.d.						
12	Pro	n.d.		31.75	n.d.	27.36	50.54			
13	Ser	112.91		n.d.	62.80					
14	Ser	116.76		n.d.	n.d.					
15	Gly	109.48		n.d.						
16	Arg	119.69	114.28	n.d.	n.d.	28.92	43.18			
17	Pro	n.d.		n.d.	30.69	26.90	50.04			
18	Pro	n.d.		n.d.	29.25	26.53	n.d.			
19	Pro	n.d.		62.19	31.56	26.87	49.74			
20	Lys	124.41		56.53	n.d.	21.50	28.95	41.87		



**Table V.1.21.:** Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. Shifts marked in bold couldn't be defined due to broadening or overlap with H<sub>2</sub>O.

Temperature / K	Proton	$\delta$ (ppm)	$\Delta\delta$ ppm	Proton	$\delta$ (ppm)	$\Delta\delta$ (ppm)
274.2						
	K2 $\alpha$	4.162	-0.158	L7 $\alpha$	3.367	-0.973
	Y3 $\alpha$	4.096	-0.454	G11 $\alpha'$	0.983	-3.147
	A4 $\alpha$	4.113	-0.207	P18 $\alpha$	2.528	-2.202
	Q5 $\alpha$	4.039	-0.311	P18 $\beta'$	0.398	-1.512
	W6 $\alpha$	4.236	-0.425	P19 $\delta$	2.932	-0.668
	L7 $\alpha$	3.367	-0.973	P19 $\delta'$	3.154	-0.446
	A8 $\alpha$	4.089	-0.232			
	$\Sigma$		-2.760	$\Sigma$		-8.947
	$\chi_U$		0.078	$\chi_U$		0.070
278						
	K2 $\alpha$	4.153	-0.167	L7 $\alpha$	3.377	-0.963
	Y3 $\alpha$	4.101	-0.449	G11 $\alpha'$	1.030	-3.101
	A4 $\alpha$	4.110	-0.211	P18 $\alpha$	2.556	-2.175
	Q5 $\alpha$	4.048	-0.302	P18 $\beta'$	0.433	-1.477
	W6 $\alpha$	4.249	-0.411	P19 $\delta$	2.935	-0.665
	L7 $\alpha$	3.377	-0.963	P19 $\delta'$	3.140	-0.460
	A8 $\alpha$	4.083	-0.237			
	$\Sigma$		-2.739	$\Sigma$		-8.840
	$\chi_U$		0.085	$\chi_U$		0.081
283						
	K2 $\alpha$	4.153	-0.167	L7 $\alpha$	3.385	-0.955
	Y3 $\alpha$	4.113	-0.437	G11 $\alpha'$	1.105	-3.025
	A4 $\alpha$	4.113	-0.207	P18 $\alpha$	2.639	-2.091
	Q5 $\alpha$	4.040	-0.310	P18 $\beta'$	0.494	-1.417
	W6 $\alpha$	4.247	-0.413	P19 $\delta$	3.175	-0.425
	L7 $\alpha$	3.385	-0.955	P19 $\delta'$	2.948	-0.652
	A8 $\alpha$	4.080	-0.240			
	$\Sigma$		-2.729	$\Sigma$		-8.564
	$\chi_U$		0.088	$\chi_U$		0.110
288						
	K2 $\alpha$	4.154	-0.166	L7 $\alpha$	3.412	-0.928
	Y3 $\alpha$	4.138	-0.412	G11 $\alpha'$	1.195	-2.935
	A4 $\alpha$	4.105	-0.215	P18 $\alpha$	2.698	-2.032
	Q5 $\alpha$	4.039	-0.311	P18 $\beta'$	0.562	-1.349
	W6 $\alpha$	4.253	-0.407	P19 $\delta$	2.959	-0.642
	L7 $\alpha$	3.412	-0.928	P19 $\delta'$	3.157	-0.443
	A8 $\alpha$	4.080	-0.240			
	$\Sigma$		-2.679	$\Sigma$		-8.327
	$\chi_U$		0.105	$\chi_U$		0.135

**Table V.1.22.:** Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. Shifts marked in bold couldn't be defined due to broadening or overlap with H<sub>2</sub>O.

Temperature / K	Proton	$\delta$ (ppm)	$\Delta\delta$ ppm	Proton	$\delta$ (ppm)	$\Delta\delta$ (ppm)
293						
	K2 $\alpha$	4.161	-0.159	L7 $\alpha$	3.439	-0.901
	Y3 $\alpha$	4.153	-0.397	G11 $\alpha'$	1.351	-2.779
	A4 $\alpha$	4.099	-0.221	P18 $\alpha$	2.796	-1.935
	Q5 $\alpha$	4.045	-0.305	P18 $\beta'$	0.666	-1.244
	W6 $\alpha$	4.270	-0.390	P19 $\delta$	2.980	-0.620
	L7 $\alpha$	3.439	-0.901	P19 $\delta'$	3.171	-0.429
	A8 $\alpha$	4.084	-0.237			
	$\Sigma$		-2.609	$\Sigma$		-7.908
	$\chi_U$		0.129	$\chi_U$		0.178
298						
	K2 $\alpha$	4.159	-0.161	L7 $\alpha$	3.473	-0.868
	Y3 $\alpha$	4.168	-0.382	G11 $\alpha'$	1.489	-2.641
	A4 $\alpha$	4.104	-0.217	P18 $\alpha$	2.903	-1.827
	Q5 $\alpha$	4.054	-0.296	P18 $\beta'$	0.757	-1.153
	W6 $\alpha$	4.292	-0.368	P19 $\delta$	3.003	-0.597
	L7 $\alpha$	3.473	-0.868	P19 $\delta'$	3.193	-0.407
	A8 $\alpha$	4.084	-0.236			
	$\Sigma$		-2.527	$\Sigma$		-7.492
	$\chi_U$		0.156	$\chi_U$		0.221
300						
	K2 $\alpha$	4.166	-0.154	L7 $\alpha$	3.512	-0.828
	Y3 $\alpha$	4.178	-0.372	G11 $\alpha'$	1.714	-2.416
	A4 $\alpha$	4.100	-0.220	P18 $\alpha$	3.050	-1.680
	Q5 $\alpha$	4.057	-0.293	P18 $\beta'$	0.892	-1.018
	W6 $\alpha$	4.299	-0.361	P19 $\delta$	3.047	-0.553
	L7 $\alpha$	3.512	-0.828	P19 $\delta'$	3.244	-0.356
	A8 $\alpha$	4.106	-0.214			
	$\Sigma$		-2.441	$\Sigma$		-6.851
	$\chi_U$		0.185	$\chi_U$		0.288
303						
	K2 $\alpha$	4.150	-0.170	L7 $\alpha$	3.513	-0.827
	Y3 $\alpha$	4.192	-0.358	G11 $\alpha'$	1.657	-2.473
	A4 $\alpha$	4.103	-0.217	P18 $\alpha$	3.030	-1.700
	Q5 $\alpha$	4.052	-0.298	P18 $\beta'$	0.875	-1.035
	W6 $\alpha$	4.309	-0.351	P19 $\delta$	3.040	-0.560
	L7 $\alpha$	3.513	-0.827	P19 $\delta'$	3.223	-0.377
	A8 $\alpha$	4.097	-0.223			
	$\Sigma$		-2.444	$\Sigma$		-6.973
	$\chi_U$		0.184	$\chi_U$		0.275

**Table V.1.23.:** Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. Shifts marked in bold couldn't be defined due to broadening or overlap with H<sub>2</sub>O.

Temperature / K	Proton	$\delta$ (ppm)	$\Delta\delta$ ppm	Proton	$\delta$ (ppm)	$\Delta\delta$ (ppm)
308						
	K2 $\alpha$	4.156	-0.164	L7 $\alpha$	3.594	-0.746
	Y3 $\alpha$	4.234	-0.316	G11 $\alpha'$	2.047	-2.083
	A4 $\alpha$	4.094	-0.226	P18 $\alpha$	3.297	-1.433
	Q5 $\alpha$	4.063	-0.287	P18 $\beta'$	1.123	-0.787
	W6 $\alpha$	4.336	-0.324	P19 $\delta$	3.119	-0.481
	L7 $\alpha$	3.594	-0.746	P19 $\delta'$	3.297	-0.303
	A8 $\alpha$	4.116	-0.204			
	$\Sigma$		-2.267	$\Sigma$		-5.834
	$\chi_U$		0.243	$\chi_U$		0.394
313						
	K2 $\alpha$	4.155	-0.165	L7 $\alpha$	3.622	-0.718
	Y3 $\alpha$	4.258	-0.293	G11 $\alpha'$	2.074	-2.056
	A4 $\alpha$	4.097	-0.223	P18 $\alpha$	3.325	-1.405
	Q5 $\alpha$	4.078	-0.272	P18 $\beta'$	1.142	-0.768
	W6 $\alpha$	4.360	-0.300	P19 $\delta$	3.128	-0.472
	L7 $\alpha$	3.622	-0.718	P19 $\delta'$	3.298	-0.302
	A8 $\alpha$	4.122	-0.198			
	$\Sigma$		-2.169	$\Sigma$		-5.721
	$\chi_U$		0.276	$\chi_U$		0.405
318						
	K2 $\alpha$	4.153	-0.167	L7 $\alpha$	3.723	-0.618
	Y3 $\alpha$	4.310	-0.240	G11 $\alpha'$	2.515	-1.615
	A4 $\alpha$	4.087	-0.234	P18 $\alpha$	3.237	-1.493
	Q5 $\alpha$	4.087	-0.264	P18 $\beta'$	1.422	-0.488
	W6 $\alpha$	4.417	-0.243	P19 $\delta$	3.237	-0.363
	L7 $\alpha$	3.723	-0.618	P19 $\delta'$	3.393	-0.207
	A8 $\alpha$	4.120	-0.200			
	$\Sigma$		-1.965	$\Sigma$		-4.784
	$\chi_U$		0.344	$\chi_U$		0.503
323						
	K2 $\alpha$	4.154	-0.166	L7 $\alpha$	3.754	-0.586
	Y3 $\alpha$	4.320	-0.230	G11 $\alpha'$	2.560	-1.570
	A4 $\alpha$	4.106	-0.214	P18 $\alpha$	3.762	-0.968
	Q5 $\alpha$	4.099	-0.251	P18 $\beta'$	1.506	-0.404
	W6 $\alpha$	4.430	-0.230	P19 $\delta$	3.235	-0.365
	L7 $\alpha$	3.754	-0.586	P19 $\delta'$	3.405	-0.195
	A8 $\alpha$	4.127	-0.193			
	$\Sigma$		-1.870	$\Sigma$		-4.087
	$\chi_U$		0.376	$\chi_U$		0.575

**Table V.1.24.:** Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. Shifts marked in bold couldn't be defined due to broadening or overlap with H<sub>2</sub>O.

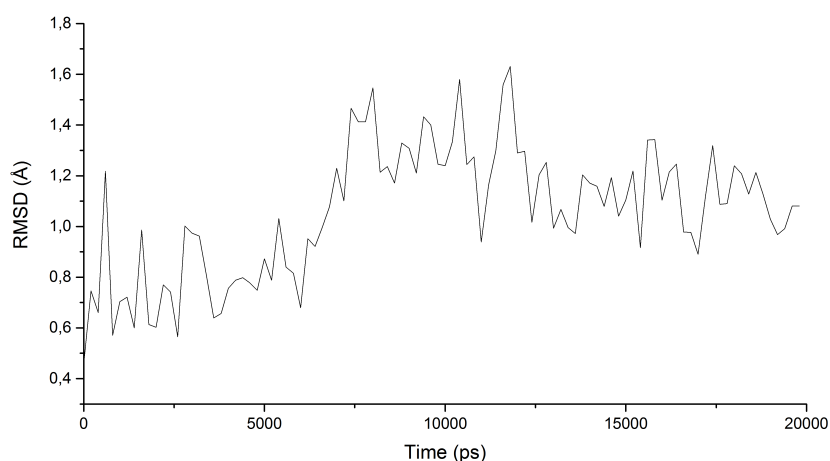
Temperature / K	Proton	$\delta$ (ppm)	$\Delta\delta$ ppm	Proton	$\delta$ (ppm)	$\Delta\delta$ (ppm)
328						
	K2 $\alpha$	4.154	-0.166	L7 $\alpha$	3.826	-0.514
	Y3 $\alpha$	4.354	-0.196	G11 $\alpha'$	2.809	-1.321
	A4 $\alpha$	4.115	-0.205	P18 $\alpha$	3.828	-0.902
	Q5 $\alpha$	4.113	-0.237	P18 $\beta'$	1.607	-0.303
	W6 $\alpha$	4.637	-0.023	P19 $\delta$	3.301	-0.299
	L7 $\alpha$	3.826	-0.514	P19 $\delta'$	3.452	-0.148
	A8 $\alpha$	4.144	-0.176			
	$\Sigma$		-1.516	$\Sigma$		-3.488
	$\chi U$		0.494	$\chi U$		0.638
338						
	K2 $\alpha$	4.157	-0.163	L7 $\alpha$	3.985	-0.356
	Y3 $\alpha$	<b>4.550</b>	0.000	G11 $\alpha'$	3.328	-0.802
	A4 $\alpha$	4.123	-0.197	P18 $\alpha$	<b>4.730</b>	0.000
	Q5 $\alpha$	4.147	-0.203	P18 $\beta'$	<b>1.910</b>	0.000
	W6 $\alpha$	4.648	-0.013	P19 $\delta$	3.339	-0.261
	L7 $\alpha$	3.985	-0.356	P19 $\delta'$	3.727	0.127
	A8 $\alpha$	4.170	-0.150			
	$\Sigma$		-1.082	$\Sigma$		-1.291
	$\chi U$		0.560	$\chi U$		0.768

**Table V.1.25.:** <sup>1</sup>H-chemical shifts of TcKKA at 274 K and pH 7.9.

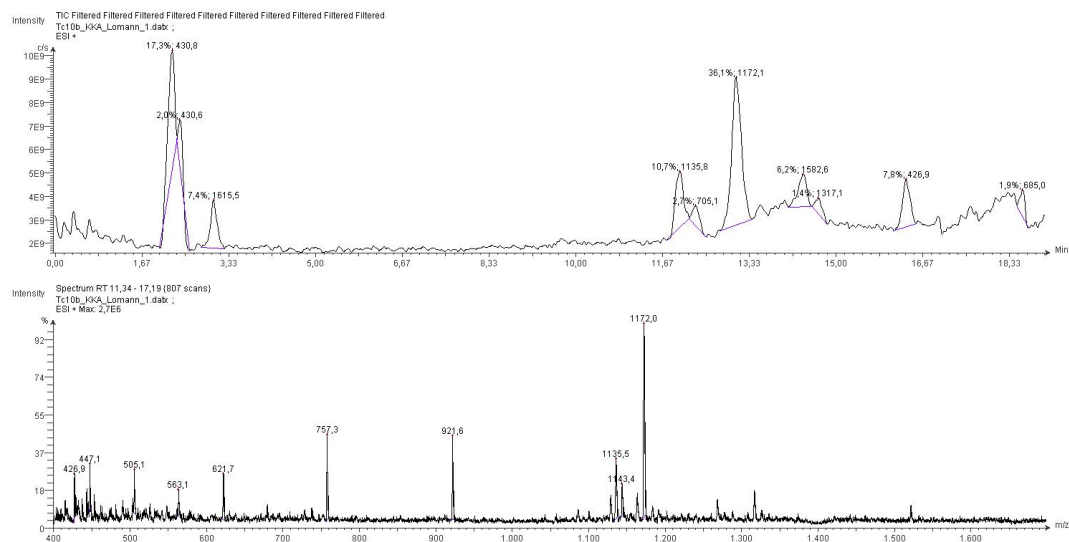
Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.500	4.589	2.781					
2	Lys	8.126	4.083	1.787	1.419	1.690	2.996	exch.	
3	Tyr	exch.	4.087	3.201, 3.102		7.109	6.822		
4	Ala	8.445	4.119	1.559					
5	Gln	8.171	4.983	2.402	2.2041, 2.1126				
6	Trp	7.970	4.244	3.524, 3.1543		6.998	9.740, 7.225	7.196, 7.182	7.139
7	Leu	8.474	3.378	1.871, 1.375	1.627	0.993, 0.877			
8	Ala	8.218	4.016	1.480					
9	Asp	7.877	4.553	2.916, 2.724					
10	Gly	7.582	4.177, 3.454						
11	Gly	8.453	3.119, 0.674						
12	Pro		4.644	2.533, 2.076	2.169	3.847, 3.476			
13	Ser	exch.	4.482	3.938					
14	Ser	8.261	4.137	3.838, 3.488					
15	Gly	8.003	4.288, 3.800						
16	Arg	8.189	5.091	1.921, 1.811	1.662	3.24, 3.205	exch.		
17	Pro		4.775	2.354, 1.792	2.010	3.879, 3.688			
18	Pro		2.420	1.282, 0.248	1.738, 1.663	3.534			
19	Pro		4.310	2.168, 1.937	1.855, 1.780	3.147, 2.925			
20	Lys	8.033	4.056	1.742	1.335	1.626	2.916	exch.	

**Table V.1.26.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of TcKKA at 274 K and pH 7.9.

Residue	chemical shift $\delta$ (ppm)									
	$^{15}\text{N}$		$^{13}\text{C}$							
	NH	other	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$	
1 Asp	127.23		54.70	41.10	n.d.	n.d.	n.d.			
2 Lys	126.70		57.65	n.d.	18.31					
3 Tyr	n.d.		62.68	n.d.		118.12	132.89			
4 Ala	121.78		55.38	17.82						
5 Gln	119.20		58.02	33.53	31.50					
6 Trp	122.01	131.74	62.15	27.83		127.46	124.01	122.33	114.21	
7 Leu	119.02		57.68	n.d.	n.d.	25.78, 23.00				
8 Ala	124.56		54.94	17.42						
9 Asp	117.82		54.58	41.96						
10 Gly	105.98		43.94							
11 Gly	114.45		42.91							
12 Pro	n.d.		65.11	32.02	n.d.	50.97				
13 Ser	115.80		59.34	63.24						
14 Ser	117.59		60.27	65.22						
15 Gly	109.79		45.60							
16 Arg	119.40		n.d.	n.d.	n.d.	n.d.				
17 Pro	n.d.		n.d.	30.60	n.d.	50.50				
18 Pro	n.d.		60.03	28.95	n.d.	49.57				
19 Pro	n.d.		62.29	31.72	n.d.	49.77				
20 Lys	126.50		57.63	33.61	19.04	n.d.	n.d.			

**Figure V.1.22.:** RMSD of the  $\text{C}\alpha$  atoms during the MD simulation of TcKKA. The simulation was performed with a total duration of 20,000 ps and every 200 ps a structure was saved and the RMSD with respect to the first model determined.

## 1.2.7. TcKKA-Lomant

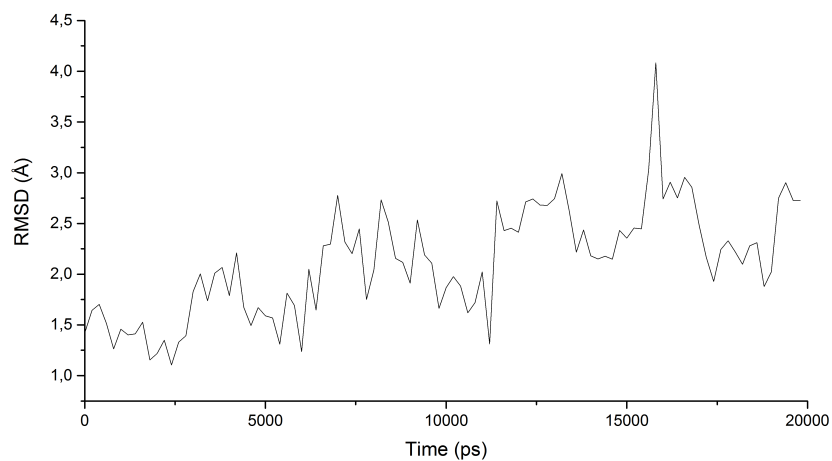


**Figure V.1.23.:** Total ion chromatogram of crude TcKKA linked with LOHMANN'S REAGENT.  $m/z$ :  $[M+2H]^{2+}$  calcd for  $[TcKKA-Lomant+2H]^{2+}$  1172.8, found 1172.0. Isolated yield 0.5mg (0.2  $\mu$ mol, 23.1%).

**Table V.1.27.:** <sup>1</sup>H-chemical shifts of TcKKA-Lomant at 298 K and pH 3.

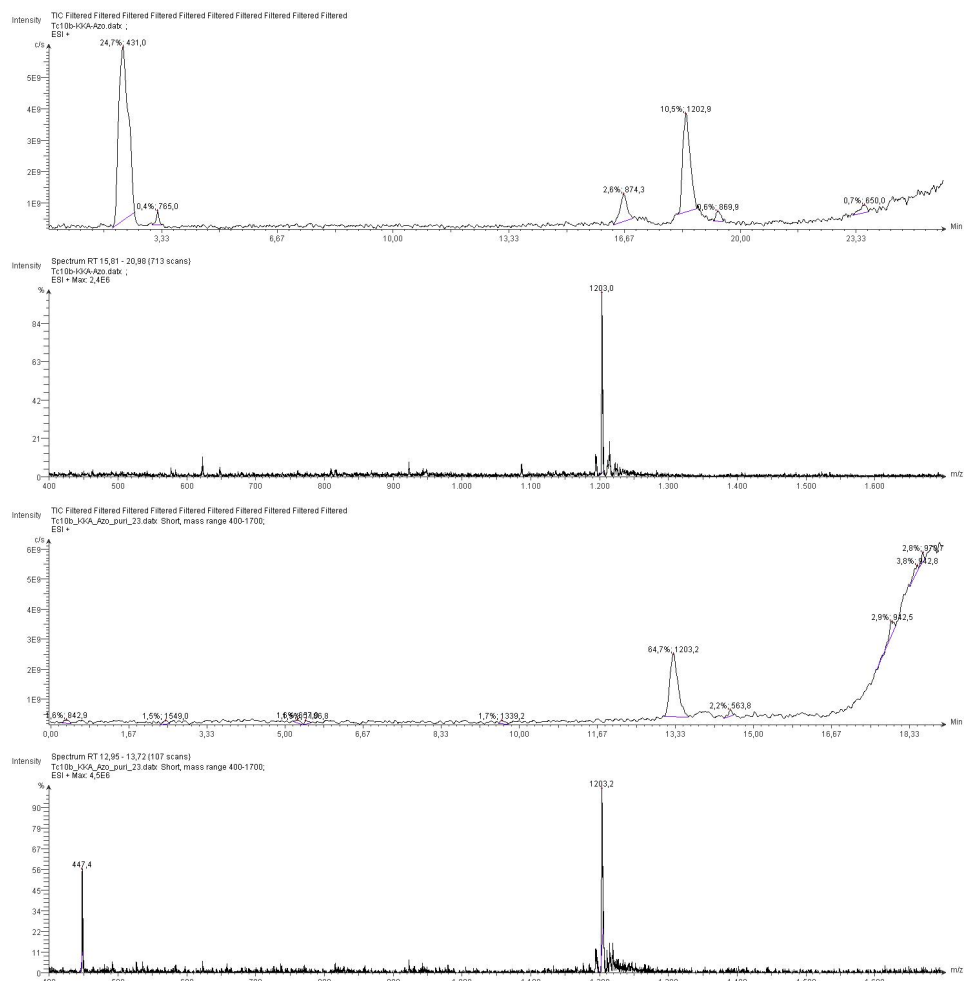
	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.351	4.669	2.872					
2	Lys	8.481	4.102	1.828	1.405	1.570, 1.537	3.217	8.147	
3	Tyr	8.578	4.141	3.182, 3.067		7.085	6.832		
4	Ala	8.224	4.090	1.567					
5	Gln	8.158	4.034	2.153	2.451, 2.386	7.480, 6.893			
6	Trp	7.953	4.235	3.531, 3.186		7.049	9.664, 7.131	7.218, 7.119	7.061
7	Leu	8.463	3.379	1.806, 1.410	1.610	0.938, 0.862			
8	Ala	8.113	4.076	1.472					
9	Asp	7.682	4.673	3.002, 2.812					
10	Gly	7.565	4.143, 3.502						
11	Gly	8.201	2.977, 1.067						
12	Pro		4.721	2.297, 1.801	1.991	3.860, 3.650			
13	Ser	7.720	4.453	3.923					
14	Ser	8.134	4.202	3.895, 3.591					
15	Gly	7.933	4.250, 3.826						
16	Arg	8.074	4.929	1.891	3.267	1.751, 1.671	3.275	7.499	
17	Pro		4.567	2.505, 2.049	2.125	3.752, 3.300			
18	Pro		2.679	1.387, 0.629	1.775	3.513			
19	Pro		4.348	2.156, 1.905	1.782	3.049, 2.871			
20	Lys	7.591	4.115	1.752, 1.670	1.292	1.483	3.176	8.061	
	Lohmant linker	2.758	2.950						





**Figure V.1.24.:** RMSD of the  $C\alpha$  atoms during the MD simulation of TcKKA-Lomant. The simulation was performed with a total duration of 20,000 ps and every 200 ps a structure was saved and the RMSD with respect to the first model determined.

## 1.2.8. TcKKA-Azobenzene



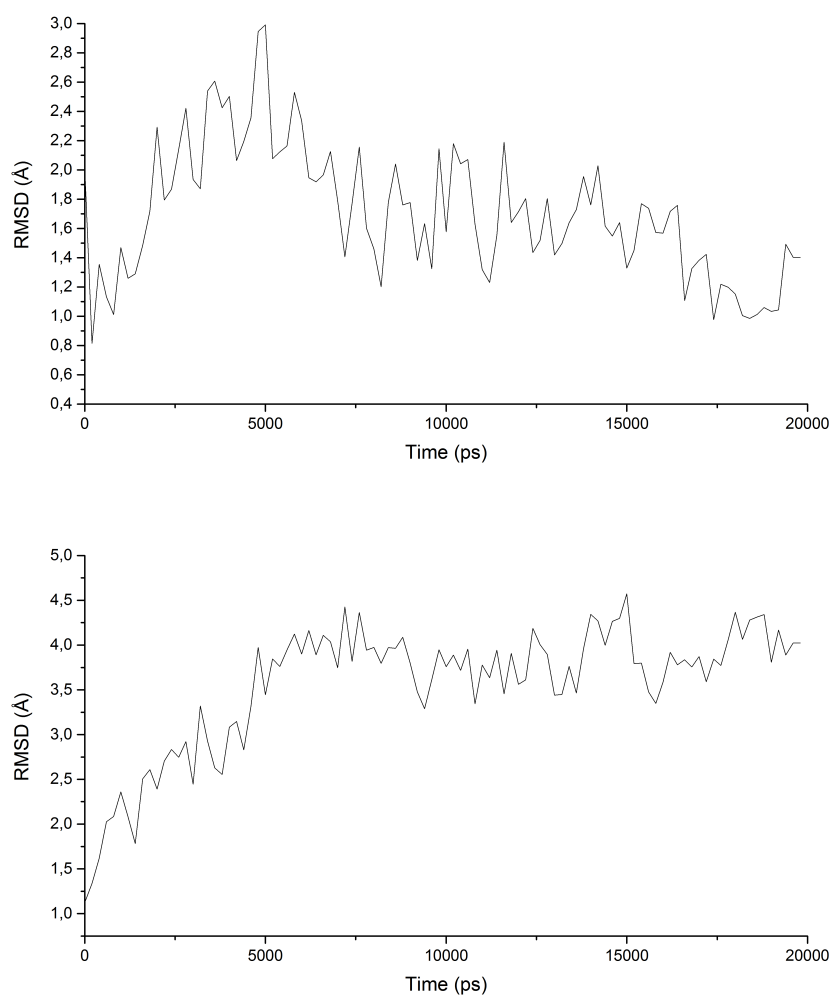
**Figure V.1.25.:** Total ion chromatogram of crude TcKKA linked with azobenzene (top) and purified (bottom).  $m/z$ :  $[M + 2H]^{2+}$  calcd for  $[TcKKA - azo + 2H]^{2+}$  1202.8, found 1203.2. Isolated yield 1mg (0.4  $\mu$ mol, 45.1%).

**Table V.1.28.:** <sup>1</sup>H-chemical shifts of TcKKA-Azo in *trans* at 298 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.239	4.576	2.726					
2	Lys	8.419	4.123	1.895, 1.756	1.359	1.672, 1.401	3.525, 3.421	8.655	
3	Tyr	8.198	4.172	3.003, 2.890		6.823	6.633		
4	Ala	7.987	3.926	1.411					
5	Gln	8.113	4.023	2.063	2.353, 2.287	n.d.			
6	Trp	8.054	4.374	3.351, 3.200		7.210	10.118, 7.245	7.254, 6.969	7.031
7	Leu	8.089	3.706	1.571	1.366	0.758, 0.725			
8	Ala	7.796	4.083	1.389					
9	Asp	7.989	4.554	2.669					
10	Gly	7.913	3.860, 3.759						
11	Gly	7.997	3.623, 3.359						
12	Pro		4.409	2.262, 1.987	1.987	3.519, 3.394			
13	Ser	8.379	4.446	3.910, 3.871					
14	Ser	8.123	4.371	3.904, 3.854					
15	Gly	8.296	3.970, 3.863						
16	Arg	7.908	4.494	1.673	1.578, 1.834	2.948	7.000		
17	Pro		4.600	2.217, 1.759	1.885	3.689, 3.427			
18	Pro		4.639	2.294, 1.910	1.999	3.726, 3.545			
19	Pro		4.386	2.125, 1.870	1.806	3.638, 3.478			
20	Lys	7.754	4.199	1.893, 1.727	1.386	1.730, 1.622	3.449	8.321	
Azobenzene			7.992 (02, 2,2')	7.869 (02, 3,3')	7.367 (20, 6,6')	7.652 (20, 7,7')			

**Table V.1.29.:**  $^1\text{H}$ -chemical shifts of TcKKA-Azo in *cis* at 298 K and pH 3.

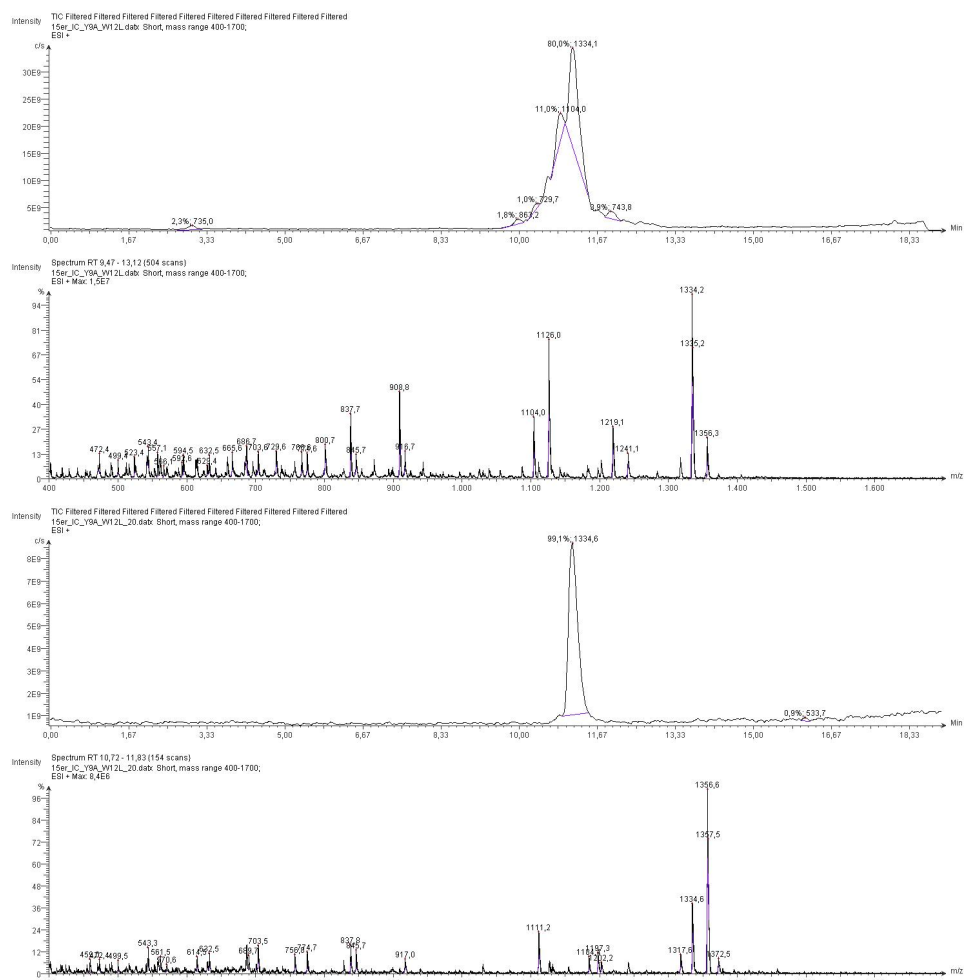
	Residue	chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	8.281	4.605	2.776					
2	Lys	8.535	4.116	1.847	1.470, 1.400	1.656, 1.609	3.382	8.506	
3	Tyr	8.523	4.055	3.121, 2.984		6.945	6.752		
4	Ala	8.256	4.071	1.559					
5	Gln	8.180	4.006	2.169	2.422				
6	Trp	7.986	4.283	3.528, 3.226		7.112	9.789, 7.217	7.225, 7.073	7.123
7	Leu	8.395	3.394	1.816, 1.392	1.595	0.927, 0.922			
8	Ala	8.141	4.187	1.479					
9	Asp	7.800	4.591	2.866, 2.703					
10	Gly	7.557	4.140, 3.484						
11	Gly	8.302	3.054, 1.051						
12	Pro		4.712	1.956, 1.693	1.879	3.867, 3.609			
13	Ser	7.724	4.464	3.921					
14	Ser	8.139	4.040	3.879, 3.565					
15	Gly	7.951	4.239, 3.817						
16	Arg	8.085	4.979	1.874	1.762, 1.666	3.198	7.254		
17	Pro		4.578	2.142	2.053	3.779, 3.360			
18	Pro		2.855	1.300, 0.507	1.625, 1.519	3.449			
19	Pro		4.112	1.976	n.d.	3.126, 2.904			
20	Lys	7.667	4.064	1.695, 1.608	1.264	1.537	3.382, 3.259	8.279	
	Azobenzene	7.584 (02, 2,2')	6.822 (02, 3,3')		6.929 (20, 6,6')	7.638 (20, 7,7')			



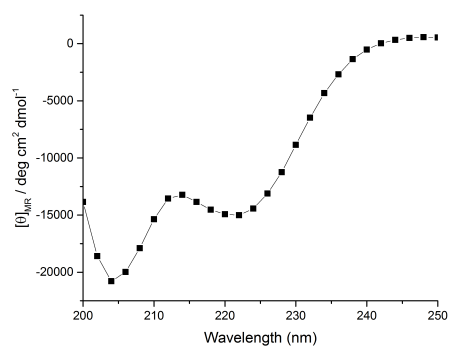
**Figure V.1.26.:** RMSD of the  $C\alpha$  atoms during the MD simulation of TcKKA-azo in *cis* configuration of the *M* (top) enantiomer and *P* enantiomer (bottom). The simulation was performed with a total duration of 20,000 ps and every 200 ps a structure was saved and the RMSD with respect to the first model determined.

### 1.3. AFP-Tc Chimera Proteins

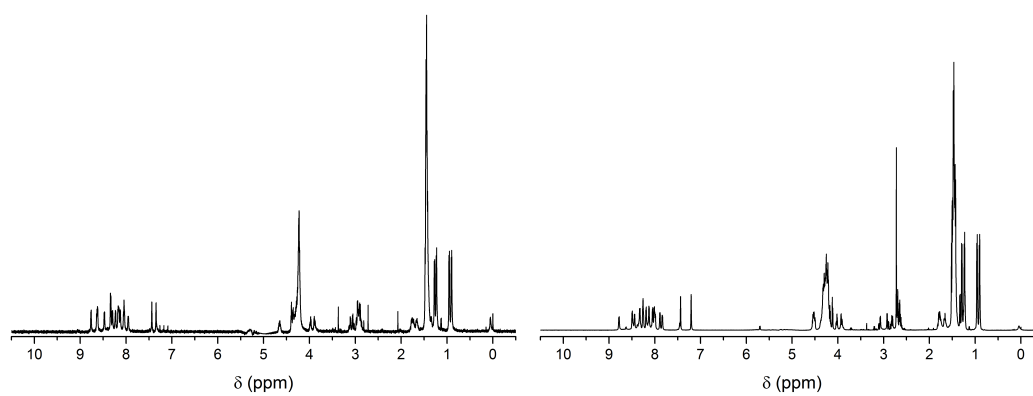
#### 1.3.1. 15er Reference Peptide



**Figure V.1.27.:** Total ion chromatogram of crude 15er reference peptide (top) and purified (bottom).  $m/z$ :  $[M + 1H]^+$  calcd for  $[15er + 1H]^+$  1334.4, found 1334.6. Isolated yield 13mg (9.7  $\mu\text{mol}$ , 6.5%).



**Figure V.1.28.:** CD spectrum of 15er reference peptide at 274 K and pH 3.



**Figure V.1.29.:** <sup>1</sup>H-NMR spectra of 15er at 274 K and pH 3 (left) or pH 7.9 (0.1 M NH<sub>4</sub>HCO<sub>3</sub> buffer) (right).

**Table V.1.30.:**  $^1\text{H}$  chemical shifts of 15er reference peptide at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.387	3.120, 3.051					
2	Thr	8.764	4.227	4.227	1.264				
3	Ala	8.619	4.294	1.437					
4	Ser	8.638	4.355	3.983, 3.880					
5	Asp	8.473	4.643	2.931					
6	Ala	8.337	4.244	1.456					
7	Ala	8.297	4.238	1.449					
8	Ala	8.176	4.233	1.469					
9	Ala	8.126	4.223	1.443					
10	Ala	8.159	4.245	1.448					
11	Ala	8.039	4.244	1.448					
12	Leu	7.954	4.326	1.659	1.745	0.955, 0.901			
13	Thr	8.051	4.228	4.228	1.231				
14	Ala	8.231	4.273	1.427					
15	Asp	8.335	4.663	2.918					

**Table V.1.31.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of 15er at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
1	Asp	n.d.		52.57	38.95				
2	Thr	115.79		63.43	n.d.	21.74			
3	Ala	126.07		53.78	18.37				
4	Ser	116.30		59.86					
5	Asp	123.16		52.58					
6	Ala	124.65		53.78	18.37				
7	Ala	122.69		53.78	18.37				
8	Ala	122.83		53.78	18.37				
9	Ala	122.22		53.78	18.37				
10	Ala	122.01		53.78	18.37				
11	Ala	121.98		53.78	18.37				
12	Leu	119.78		56.42	n.d.		24.87, 23.24		
13	Thr	114.54		63.48	n.d.	21.56			
14	Ala	125.33		53.78	18.37				
15	Asp	118.33		52.69	38.11				



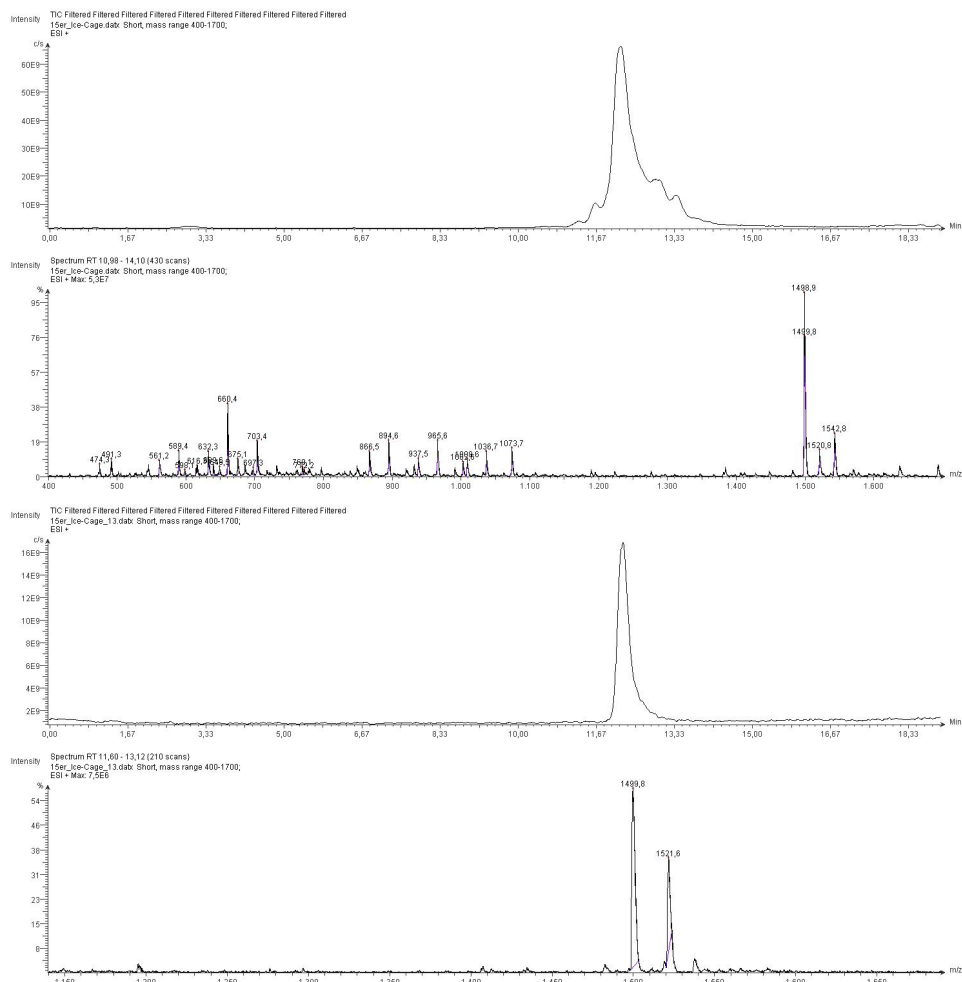
**Table V.1.32.:**  $^1\text{H}$ -chemical shifts of 15er reference peptide at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	5.706	4.37	2.629					
2	Thr	8.159	4.281	1.263					
3	Ala	8.447	4.277	1.458					
4	Ser	8.784	4.325	4.010, 3.916					
5	Asp	8.496	4.515	2.820, 2.684					
6	Ala	8.262	4.257	1.491					
7	Ala	8.192	4.231	1.479					
8	Ala	8.128	4.226	1.502					
9	Ala	8.016	4.218	1.474					
10	Ala	8.049	4.219	1.458					
11	Ala	7.891	4.265	1.469					
12	Leu	7.839	4.338	1.795	1.649	0.947, 0.913			
13	Thr	7.992	4.307	4.307	1.237				
14	Ala	8.254	4.279	1.436					
15	Asp	8.332	4.541	2.648					

**Table V.1.33.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of 15er at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)							
1	Asp	n.d.		56.55	41.18	21.91			
2	Thr	n.d.		65.06	69.21				
3	Ala	125.06		53.20	18.29				
4	Ser	117.53		60.91	62.24				
5	Asp	125.12		56.49	39.87				
6	Ala	124.39		53.26	18.29				
7	Ala	122.51		54.35	18.29				
8	Ala	122.36		54.35	18.29				
9	Ala	121.49		53.89	18.29				
10	Ala	121.49		53.89	18.29				
11	Ala	121.22		53.20	18.29				
12	Leu	119.43		55.99	42.16	26.87	24.93, 23.27		
13	Thr	113.75		62.29	69.80	21.62			
14	Ala	125.88		54.59	18.29				
15	Asp	119.84		54.43	41.25				

## 1.3.2. 15erYW Reference Peptide



**Figure V.1.30.:** Total ion chromatogram of crude 15er reference peptide (top) and purified (bottom).  $m/z$ :  $[M + H]^+$  calcd for  $[15erYW + 1H]^+$  1499.5, found 1499.8. Isolated yield 9mg (6.0  $\mu\text{mol}$ , 4.5%).

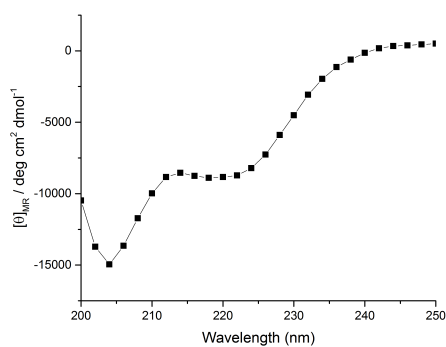


Figure V.1.31.: CD spectrum of 15erYW reference peptide at 274 K and pH 3.

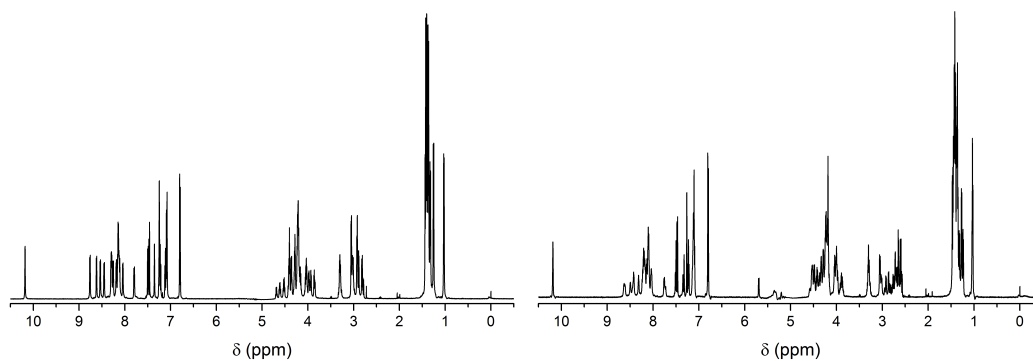


Figure V.1.32.:  $^1\text{H}$ -NMR spectra of 15erYW at 274 K and pH 3 (left) or pH 7.9 (0.1 M  $\text{NH}_4\text{HCO}_3$  buffer) (right).

**Table V.1.34.:**  $^1\text{H}$ -chemical shifts of 15erYW reference peptide at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	exch.	4.404	3.011					
2	Thr	8.762	4.287	4.219	1.255				
3	Ala	8.618	4.289	1.426					
4	Ser	8.535	4.369	3.936, 3.853					
5	Asp	8.451	4.695	2.922					
6	Ala	8.288	4.236	1.426					
7	Ala	8.302	4.227	1.394					
8	Ala	8.155	4.2133	1.364					
9	Tyr	8.116	4.415	3.011		7.079	6.798		
10	Ala	8.143	4.175	1.3244					
11	Ala	8.133	4.207	1.397					
12	Trp	8.186	4.523	3.288		7.249	10.183, 7.467	7.232, 7.113	7.483
13	Thr	7.796	3.978	3.978	1.018				
14	Ala	8.045	4.042	1.369					
15	Asp	8.256	4.609	2.890, 2.800					

**Table V.1.35.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of 15erYW at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)								
		$^{15}\text{N}$		$^{13}\text{C}$						
1	Asp	n.d.		52.46	38.30					
2	Thr	115.85		63.07	69.47	21.74				
3	Ala	126.64		53.37	18.62					
4	Ser	115.68		59.11	63.16					
5	Asp	122.15		53.62	38.08					
6	Ala	124.32		53.31	18.62					
7	Ala	122.62		53.31	18.62					
8	Ala	125.17		53.31	18.62					
9	Tyr	119.63		58.80	38.55		118.12	133.10		
10	Ala	122.68		53.01	18.82					
11	Ala	122.68		53.31	18.62					
12	Trp	120.45	129.71	58.62	29.03		127.14	114.48	124.58, 121.93	120.77
13	Thr	116.26		62.10	69.79	21.20				
14	Ala	125.35		53.13	18.62					
15	Asp	118.08		52.83	38.11					

**Table V.1.36.:** <sup>1</sup>H-chemical shifts of 15erYW reference peptide at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Asp	5.695	4.356	2.640, 2.588					
2	Thr	8.139	4.289	4.289	1.241				
3	Ala	8.496	4.286	1.449					
4	Ser	8.3635	4.331	3.985, 3.885					
5	Asp	8.422	4.535	2.724					
6	Ala	8.319	4.219	1.393					
7	Ala	8.611	4.333	1.422					
8	Ala	8.088	4.225	1.412					
9	Tyr	8.137	4.212	3.046		7.107	6.801		
10	Ala	8.111	4.178	1.365					
11	Ala	8.046	4.235	1.415					
12	Trp	8.224	4.546	8.288		7.266	10.187, 7.476	7.231, 7.121	7.489
13	Thr	7.759	4.006	4.006	1.029				
14	Ala	8.207	4.237	1.463					
15	Asp	8.202	4.484	2.651					

**Table V.1.37.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of 15erYW at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)								
1	Asp	n.d.		56.53	41.34					
2	Thr	n.d.		61.91	69.39	21.56				
3	Ala	n.d.		53.94	18.58					
4	Ser	n.d.		59.07	62.73					
5	Asp	n.d.		55.61	40.48					
6	Ala	n.d.		53.48	18.58					
7	Ala	n.d.		53.02	18.58					
8	Ala	n.d.		53.02	18.58					
9	Tyr	n.d.		58.91	38.42		118.23	133.18		
10	Ala	n.d.		53.12	18.58					
11	Ala	n.d.		53.02	18.58					
12	Trp	n.d.	n.d.	58.56	29.13		127.17	114.37	124.65, 121.97	120.62
13	Thr	n.d.		61.76	70.07	21.33				
14	Ala	n.d.		53.02	18.58					
15	Asp	n.d.		54.29	41.17					

## 1.3.3. AFP-Tc-4

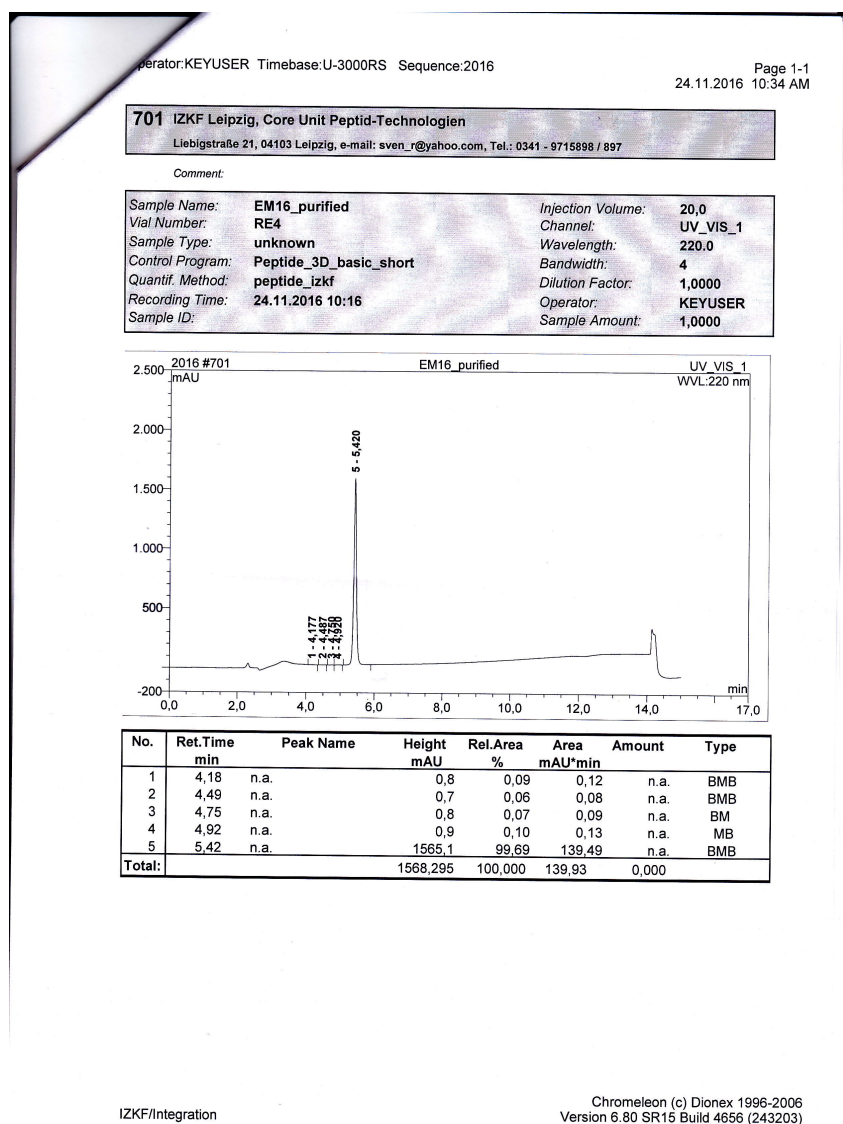


Figure V.1.33.: HPLC chromatogram of purified AFP-Tc-4, synthesized by S. Rothmund.



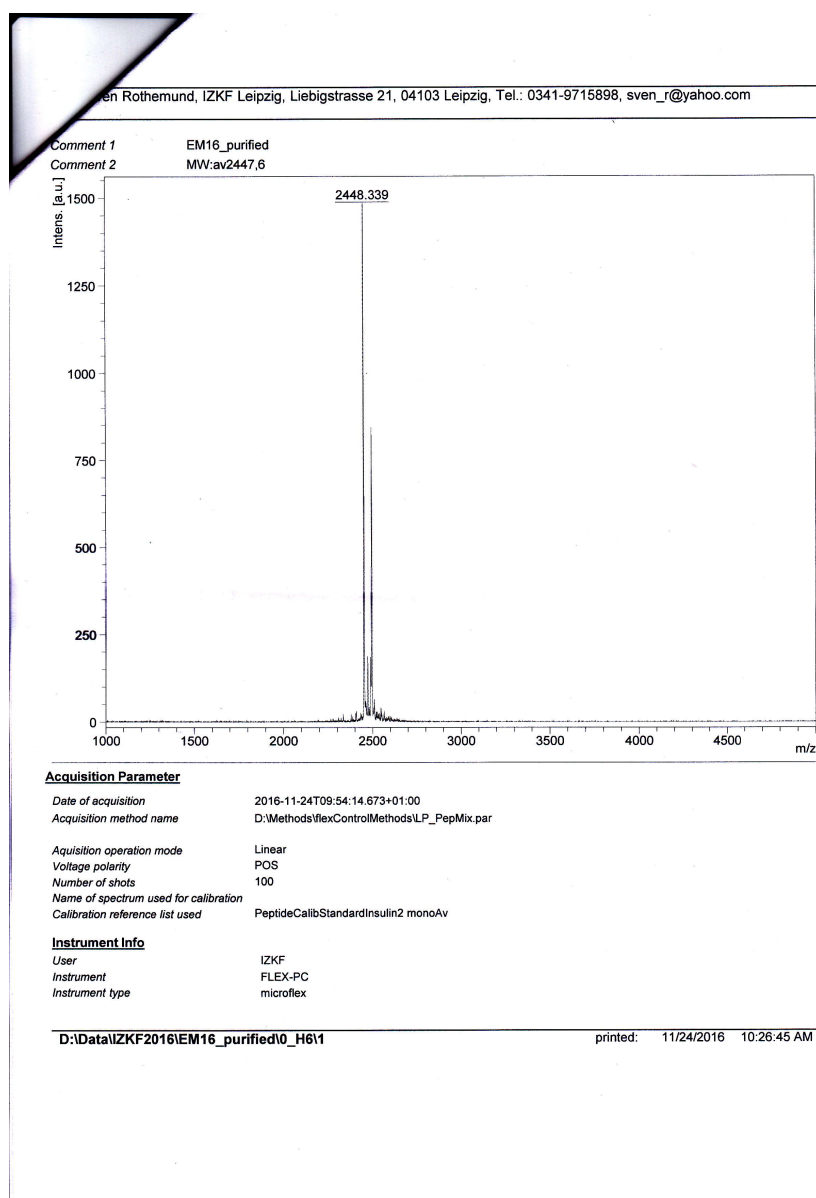
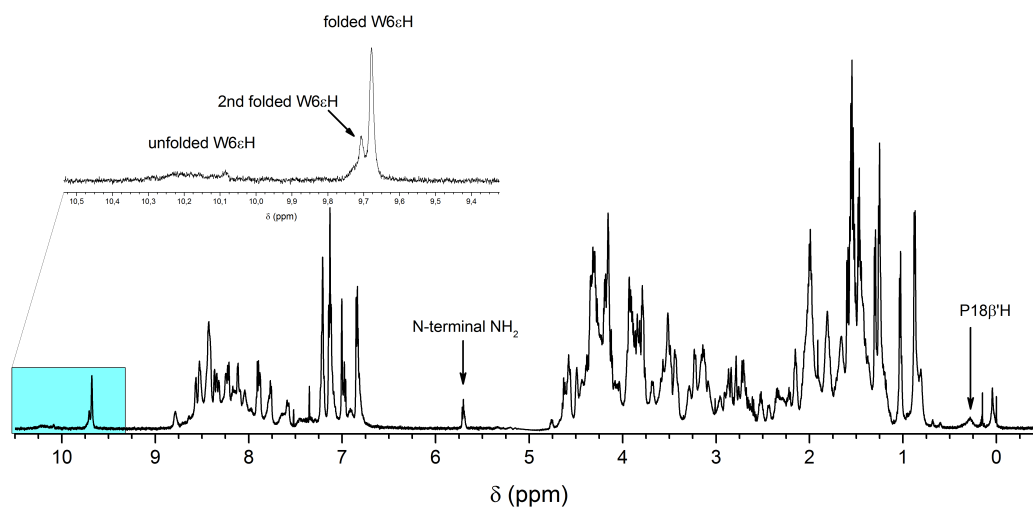


Figure V.1.34.:  $m/z$ :  $[M + H]^+$  calcd for  $[TcAFP4]^+$  2447.5, found 2448.4.



**Figure V.1.35.:**  $^1\text{H}$  NMR spectrum of AFP-Tc-4 at 274 K and pH 7.9 with enlargement of the indole proton region.

**Table V.1.38.:** <sup>1</sup>H-chemical shifts of TcAFP4b at 274K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.052	3.162					
2'	Thr	8.749	4.208	4.208	1.069				
3'	Ala	8.577	4.266	1.427					
4'	Ser	8.630	4.370	3.990, 3.893					
5'	Asp	8.439	4.682	3.036, 2.975					
1	Ala	8.486	4.279	1.519					
2	Ala	8.600	4.274	1.428					
3	Tyr	8.457	4.047	3.154		7.099	6.814		
4	Ala	8.366	4.168	1.578					
5	Ala	8.087	4.146	1.518					
6	Trp	7.893	4.206	3.557, 3.156		7.073	9.605, 7.244	7.229, 7.114	6.985
7	Leu	8.484	3.413	1.932, 1.482	1.647	1.009, 0.914			
8	Thr	8.324	4.016	4.313	1.262				
9	Asp	7.782	4.735	3.038, 2.927					
10	Gly	7.721	4.176, 3.519						
11	Gly	8.163	2.847, 1.337						
12	Pro		4.527	2.471, 2.001	2.080	3.660, 3.107			
13	Ser	7.817	4.455	3.919					
14	Ser	8.168	4.234	3.929, 3.687					
15	Gly	8.041	4.253, 3.844						
16	Arg	8.162	4.916	1.909, 1.859	1.725, 1.681	3.285	7.491		
17	Pro		4.713	2.291, 1.781	2.006	3.883, 3.677			
18	Pro		2.534	1.349, 0.370	1.727, 1.634	3.49			
19	Pro		4.379	2.235, 1.972	1.854	3.156, 2.919			
20	Ser	8.315	4.315	3.898, 3.805					



**Table V.1.40.:**  $^1\text{H}$ -chemical shifts of TcAFP4b at 298K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.100	3.126					
2'	Thr	8.630	4.298	4.218	1.239				
3'	Ala	8.419	4.292	1.420					
4'	Ser	8.378	4.391	3.959, 3.882					
5'	Asp	8.307	4.724	2.996					
1	Ala	8.385	4.249	1.492					
2	Ala	8.343	4.260	1.483					
3	Tyr	8.250	4.116	3.115		7.073	6.811		
4	Ala	8.234	4.127	1.511					
5	Ala	8.017	4.149	1.495					
6	Trp	7.841	4.252	3.522, 3.191		7.094	9.675, 7.278	7.290, 7.221	7.106
7	Leu	8.347	3.518	1.843, 1.490	1.589	0.956, 0.883			
8	Thr	8.197	4.049	4.289	1.257				
9	Asp	7.814	4.733	3.018, 2.897					
10	Gly	7.717	4.116, 3.590						
11	Gly	8.085	3.010, 1.761						
12	Pro		4.492	2.440, 1.998	2.058	3.649, 3.198			
13	Ser	7.841	4.447	3.917					
14	Ser	8.126	4.336	3.896, 3.814					
15	Gly	8.009	4.198, 3.856						
16	Arg	8.059	4.854	1.919, 1.806	1.701	3.251	7.339		
17	Pro		4.672	2.288, 1.800	2.005	3.856, 3.644			
18	Pro		2.924	1.426, 0.742	1.776, 1.676	3.501			
19	Pro		4.379	2.224, 1.946	1.855	3.216, 2.998			
20	Ser	8.118	4.272	3.926, 3.709					

**Table V.1.41.:** <sup>1</sup>H-chemical shifts of TcAFP4b at 274K and pH 7.9. Residues marked in italic belong to the second folded conformation with Pro (19) in *cis*.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.708	4.371	2.657, 2.595					
2'	Thr	exch.	4.120	4.198	1.301				
3'	Ala	exch.	4.282	1.476					
4'	Ser	8.779	4.347	4.055, 3.921					
5'	Asp	8.518	4.568	2.888, 2.769					
5'	<i>Asp</i>	8.441	4.603	2.785					
1	Ala	8.410	4.173	1.570					
2	Ala	8.346	4.331	1.562					
3	Tyr	8.565	4.087	3.227		7.141	6.846		
4	Ala	8.369	4.197	1.586					
5	Ala	8.117	4.181	1.546					
6	Trp	7.882	4.237	3.566, 3.140		7.007	9.689, 7.219	7.210, 7.135	6.970
7	Leu	8.433	3.426	2.012, 1.402	1.683	1.033, 0.877			
8	Thr	8.531	3.889	4.323	1.262				
9	Asp	7.784	4.578	2.847, 2.700					
9	<i>Asp</i>	7.764	4.579	2.856, 2.698					
10	Gly	7.580	4.158, 3.456						
10	<i>Gly</i>	7.609	4.154, 3.475						
11	Gly	8.413	3.108, 0.821						
12	Pro		4.623	2.518, 2.070	2.154	3.805, 3.435			
13	Ser	8.161	4.228	3.850					
14	Ser	8.247	4.156	3.858, 3.506					
15	Gly	8.044	4.276, 3.803						
16	Arg	8.219	5.049	1.917	1.815, 1.649	3.290, 3.219	7.650		

**Table V.1.42.:** Continued  $^1\text{H}$ -chemical shifts of TcAFP4b at 274K and pH 7.9. Residues marked in *italic* belong to the second folded conformation with Pro (19) in *cis*.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
17	Pro		4.760	2.346, 1.792	1.984	3.868, 3.685			
18	Pro		2.441	1.368, 0.284	1.723, 1.663	3.521			
19	Pro		4.312	2.222, 2.045	1.970, 1.824	3.140, 2.962			
20	Ser	7.904	4.151	3.783					

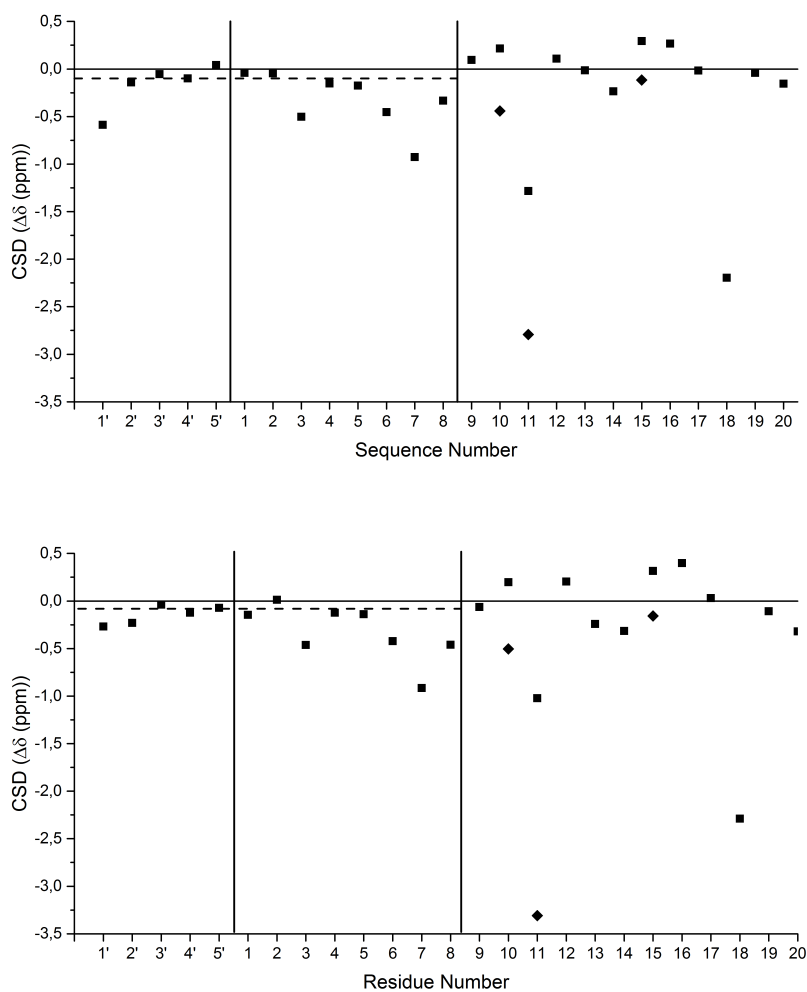
**Table V.1.43.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-4 at 274 K and pH 8

Residue		chemical shift $\delta$ (ppm)									
		$^{15}\text{N}$			$^{13}\text{C}$						
1'	Asp	n.d.		n.d.	n.d.						
2'	Thr	n.d.		65.26	69.13	21.82					
3'	Ala	n.d.		54.71	17.68						
4'	Ser	117.81		60.75	62.48						
5'	Asp	125.15		54.71	n.d.						
1	Ala	n.d.		55.13	17.68						
2	Ala	122.63		55.05	17.68						
3	Tyr	123.24		62.62	n.d.		118.27	133.04			
4	Ala	120.99		54.88	17.68						
5	Ala	120.99		54.80	17.68						
6	Trp	121.14	131.63	62.45	27.65		127.58	123.84	114.17, 122.74	120.64	
7	Leu	119.43		58.02	41.65		25.67, 23.12				
8	Thr	119.11		66.28	68.64	21.87					
9	Asp	121.13		44.24	n.d.						
10	Gly	105.90		43.26							
11	Gly	114.22		n.d.							
12	Pro	n.d.		n.d.	31.82	27.36	50.86				
13	Ser	117.59		n.d.	65.13						
14	Ser	n.d.		n.d.	64.71						
15	Gly	109.94		45.33							
16	Arg	119.82	n.d.	n.d.	n.d.	n.d.	n.d.				
17	Pro	n.d.		n.d.	30.51	27.13	50.56				
18	Pro	n.d.		60.18	28.81	26.52	49.58				
19	Pro	n.d.		62.71	31.51	27.05	49.87				
20	Ser	120.93		n.d.	64.73						

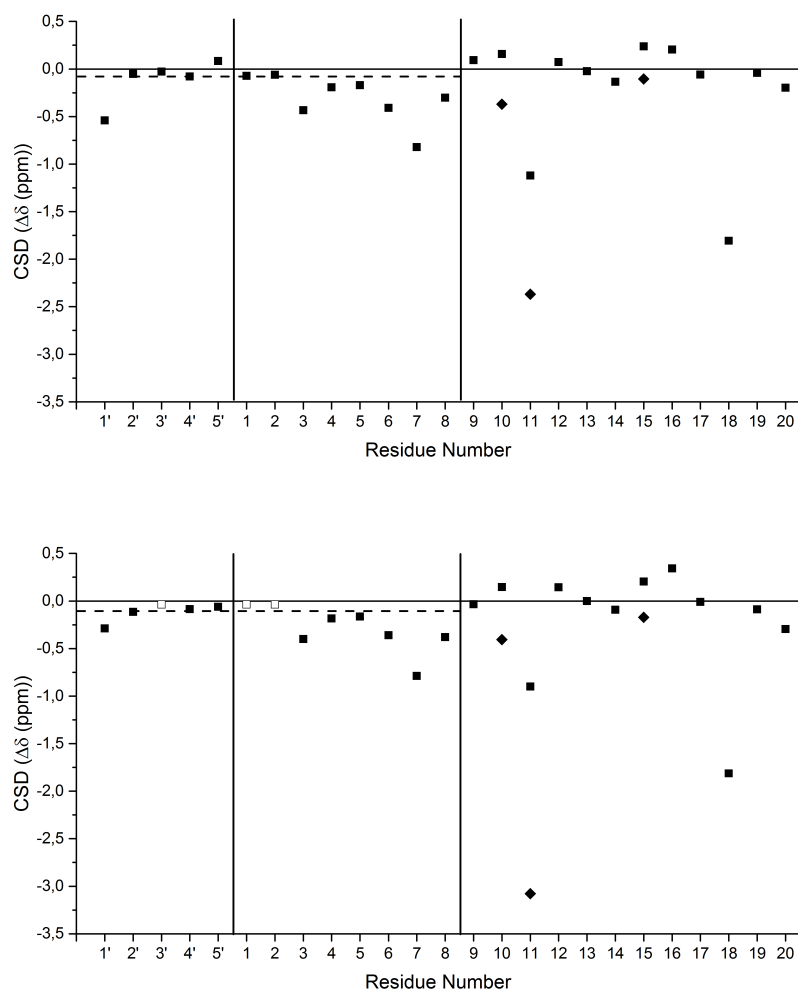
**Table V.1.44.:**  $^1\text{H}$ -chemical shifts of TcAFP4b at 298K and pH 7.9. Shifts marked with an asterisk are averaged.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.604	4.352	2.618					
2'	Thr	exch.	4.236	4.236	1.266				
3'	Ala	exch.	4.283*	1.480*					
4'	Ser	exch.	4.384	3.998, 3.908					
5'	Asp	exch.	4.578	2.782, 2.703					
1	Ala	exch.	4.283*	1.480*					
2	Ala	exch.	4.283*	1.480*					
3	Tyr	8.396	4.149	3.164		7.096	6.831		
4	Ala	8.241	4.135	1.518					
5	Ala	8.012	4.157	1.502					
6	Trp	7.842	4.300	3.517, 3.179		7.049	9.744, 7.271	7.253, 7.221	7.118
7	Leu	8.307	3.551	1.883, 1.440	1.619	0.967, 0.853			
8	Thr	8.317	3.970	4.293	1.237				
9	Asp	7.784	4.606	2.795, 7.662					
10	Gly	7.652	4.106, 3.555						
11	Gly	8.283	3.229, 1.052						
12	Pro		4.564	2.461, 2.021	2.111	3.751, 3.390			
13	Ser	exch.	4.468	3.914					
14	Ser	exch.	4.378	3.983, 3.878					
15	Gly	7.808	4.164, 3.788						
16	Arg	8.121	4.992	1.844	1.777, 1.650	3.239			
17	Pro		4.721	2.320, 1.810	1.999	3.846, 3.652			
18	Pro		2.917	1.466, 0.738	1.787, 1.703	3.538			
19	Pro		4.332	2.208, 1.999	1.931, 1.849	3.208, 3.060			
20	Ser	7.769	4.175	3.794					





**Figure V.1.36.:**  $H\alpha$  CSD plot of AFP-Tc-4 at pH 3 (top) and pH 7.9 (bottom) at 274 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. The dashed line at  $y = -0.1$  ppm represents the helical limit. Solid line between residue 5' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.



**Figure V.1.37.:**  $H\alpha$  CSD plot of AFP-Tc-4 at pH 3 (top) and pH 7.9 (bottom) at 298 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. Averaged  $H\alpha$  shifts for certain Ala are represented as non-filled squares. The dashed line at  $y=-0.1$  ppm represents the helical limit. Solid line between residue 5' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.

**Table V.1.45.:** CSDs for selected protons in the cage structure representing the cage fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9.

Cage	$\Delta\delta$ (ppm).						$\Sigma$
	upfield						
	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$	
pH 3, 274 K	-0.927	-2.794	-2.197	-1.540	-0.444	-0.681	<b>-8.582</b>
pH 3, 298 K	-0.822	-2.369	-1.806	-1.168	-0.384	-0.602	<b>-7.152</b>
pH 7.9, 274 K	-0.914	-3.310	-2.289	-1.626	-0.460	-0.638	<b>-9.237</b>
pH 7.9, 298 K	-0.789	-3.078	-1.813	-1.172	-0.392	-0.540	<b>-7.784</b>
downfield							
	P12 $\alpha$	P12 $\beta$	R16 $\alpha$				
pH 3, 274 K	0.107	0.181	0.266				<b>0.554</b>
pH 3, 298 K	0.072	0.150	0.204				<b>0.426</b>
pH 7.9, 274 K	0.203	0.228	0.398				<b>0.829</b>
pH 7.9, 298 K	0.144	0.171	0.342				<b>0.657</b>

**Table V.1.46.:** CSDs for selected protons in the helical region representing the helix fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9.

Helix	$\Delta\delta$ (ppm).							$\Sigma$
	A2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	A5 $\alpha$	W6 $\alpha$	L7 $\alpha$	T8 $\alpha$	
pH 3, 274 K	-0.046	-0.503	-0.152	-0.175	-0.454	-0.927	-0.334	<b>-2.591</b>
pH 3, 298 K	-0.060	-0.434	-0.193	-0.171	-0.408	-0.822	-0.302	<b>-2.390</b>
pH 7.9, 274 K	0.011	-0.463	-0.123	-0.139	-0.424	-0.914	-0.461	<b>-2.512</b>
pH 7.9, 298 K	-0.037	-0.401	-0.185	-0.163	-0.360	-0.789	-0.380	<b>-2.315</b>

**Table V.1.47.:** CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9.

Helix	$\Delta\delta$ (ppm).					$\Sigma$
	T2' $\alpha$	A3' $\alpha$	S'4 $\alpha$	D5' $\alpha$	A1 $\alpha$	
pH 3, 274 K	-0.143	-0.054	-0.100	0.042	-0.041	<b>-0.295</b>
pH 3, 298 K	-0.052	-0.028	-0.079	0.084	-0.071	<b>-0.146</b>
pH 7.9, 274 K	-0.230	-0.039	-0.123	-0.072	-0.147	<b>-0.610</b>
pH 7.9, 298 K	-0.114	-0.037	-0.086	-0.062	-0.037	<b>-0.336</b>

**Table V.1.48.:** CSDs for the  $^{13}\text{C}\alpha$  atoms in the helix representing the helix fold of AFP-Tc-4 at 274 K and at pH 3 and 7.9. The  $\overline{\Delta\delta}$  and the  $\chi_F$  for the helix, chimera and helical extension are calculated. The  $\chi_F$  was determined using the CSD of AFP-Tc-7 at pH 7.9 and 274 K as reference.

<b>TcAFP4</b>		
	<b>pH 3.</b>	<b>pH 8</b>
	$\delta$ (ppm)	$\delta$ (ppm)
<b>Helical Extension</b>	0.79	2.16
	1.80	2.21
	1.76	2.45
	0.43	0.61
<b>Chimera Helix</b>	2.10	2.63
	2.32	2.55
	4.04	4.02
	2.47	2.38
	2.37	2.30
	3.14	4.65
	2.26	2.32
	2.05	3.18
<b>Helix</b> $\overline{\Delta\delta}$ (ppm)	2.13	2.62
$\chi_F^{helix}$ (%)	76	94
<b>Chimera Helix</b> $\overline{\Delta\delta}$ (ppm)	2.59	3.00
$\chi_F^{chimera}$ (%)	83	96
<b>Helical Extension</b> $\overline{\Delta\delta}$ (ppm)	1.20	1.85
$\chi_F^{exten.}$ (%)	48	75

## AFP-Tc-4 VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

```

(The expected values represent those numbers which would be expected for highly refined Xray and NMR protein structures. See the help pages for more information on expected values.)

Statistic	Observed	Expected
# Helix	13 ( 52%)	-
# Beta	0 ( 0%)	-
# Coil	12 ( 48%)	-
# Turn	4 ( 16%)	-

## HYDROGEN BONDS (hbonds)

Statistic	Observed	Expected
Mean hbond distance	2.3 sd=0.4	2.2 sd=0.4
Mean hbond energy	-1.4 sd=1.0	-2.0 sd=0.8
# res with hbonds	20 ( 80%)	18 ( 75%)

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## DIHEDRAL ANGLES

Statistic	Observed	Expected
Mean Helix Phi	-67.8 sd=10.7	-65.3 sd=11.9
Mean Helix Psi	-34.5 sd=13.2	-39.4 sd=25.5
# res with Gauche+ Chi	7 ( 53%)	7 ( 55%)
# res with Gauche- Chi	2 ( 15%)	2 ( 20%)
# res with Trans Chi	4 ( 30%)	3 ( 25%)
Mean Chi Gauche+	-56.2 sd=11.6	-66.7 sd=15.0
Mean Chi Gauche-	66.0 sd=23.2	64.1 sd=15.7
Mean Chi Trans	169.7 sd=14.1	168.6 sd=16.8
Std. dev of chi pooled	14.14	15.70
Mean Omega ( omega >90)	178.3 sd=7.1	180.0 sd=5.8
# res with  omega <90	1 ( 4%)	-

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	1971.7 Angs**2	1927.5 Angs**2
ASA of backbone	258.5 Angs**2	-
ASA of sidechains	1713.2 Angs**2	-
ASA of C	1341.4 Angs**2	-
ASA of N	51.1 Angs**2	-
ASA of N+	40.4 Angs**2	-
ASA of O	428.9 Angs**2	-
ASA of O-	110.0 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1333.7 Angs**2	1202.8 Angs**2
Exposed polar ASA	355.8 Angs**2	394.3 Angs**2
Exposed charged ASA	282.2 Angs**2	374.6 Angs**2
Side exposed nonpolar ASA	1330.4 Angs**2	-
Side exposed polar ASA	150.6 Angs**2	-
Side exposed charged ASA	232.2 Angs**2	-
Fraction nonpolar ASA	0.68	0.61 sd=0.03
Fraction polar ASA	0.18	0.20 sd=0.05
Fraction charged ASA	0.14	0.19 sd=0.05
Mean residue ASA	78.9 sd=38.2	-
Mean frac ASA	0.5 sd=0.2	-
% side ASA hydrophobic	41.40	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2348.1 Angs**2	-
Extended polar ASA	1060.8 Angs**2	-
Extended charged ASA	358.4 Angs**2	-
Extended side nonpolar ASA	2328.6 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2537.0 Angs**3	2472.1 Angs**3
Mean residue volume	101.5 sd=39.5	125.0 sd=40.0
Mean frac volume	0.9 sd=0.1	1.0 sd=0.1
Molecular weight	2447.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
6:151-176.

```

*****
* STEREO/PACKING QUALITY INDEX *
*****

( 3 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...: * *
SEQUENCE: DTASDAAYAA WLTDGGPSSG RPPPS 25
TORSION.: 3333333333 3333333333 33333
OMEGA...: 0333333333 3333333333 33320
VDW.....: 3333333333 3333333333 33333
-----
TOTAL...: 6999999999 9999999999 99986

```

```

*****
* 3D PROFILE QUALITY INDEX *
*****

( 9 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:
SEQUENCE: DTASDAAYAA WLTDGGPSSG RPPPS 25
ENV.....: 7665666666 6677776677 76777

```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in phipsi core	25 (100%)	23 ( 90%)
# res in phipsi allowed	0 ( 0%)	2 ( 7%)
# res in phipsi generous	0 ( 0%)	0 ( 1%)
# res in phipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	22 ( 88%)	24 ( 96%)
# res in omega allowed	1 ( 4%)	1 ( 3%)
# res in omega generous	0 ( 0%)	0 ( 0%)
# res in omega outside	2 ( 8%)	0 ( 1%)
# packing defects	0	1
Free energy of folding	-10.47	-8.73
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243

```
*****
*       SIDE CHAIN INFO PANEL       *
*****
```

RES.	RES.	SCND	BTURN	SIDE	FRAC.	SFE	CHI1	ENV.	DISULF	S-S
NUM.	NAME	STRUC	BTURN	SURF (ASA)	SURF (ASA)	SFE	ANGLE	CLASS	BOND	DIST
-----										
Chain A										
5	ASP	C		130.4	1.10	-1.1	-53.8	E0C		
6	THR	H		113.0	1.00	-0.4	-39.5	E0H		
7	ALA	H		83.5	0.98	-0.0	360.0	E0H		
8	SER	H		41.6	0.44	-0.7	-67.9	P1H		
9	ASP	H		71.2	0.60	-0.7	-47.4	P2H		
10	ALA	H		57.8	0.67	-0.4	360.0	E0H		
11	ALA	H		28.5	0.33	-0.9	360.0	P1H		
12	TYR	H		112.6	0.55	-1.5	-179.7	P2H		
13	ALA	H		51.7	0.60	-0.5	360.0	E0H		
14	ALA	H		58.3	0.68	-0.4	360.0	E0H		
15	TRP	H		25.0	0.11	-2.9	176.5	B1H		
16	LEU	H		78.6	0.46	-1.5	-72.5	P1H		
17	THR	H		97.5	0.86	-0.1	49.6	E0H		
18	ASP	H		56.8	0.48	-1.3	-148.8	P1H		
19	GLY	C		32.6	0.62	-0.3	360.0	E0C		
20	GLY	C		0.8	0.02	-0.8	360.0	P1C		
21	PRO	C	III	94.3	0.71	-0.6	26.2	E0C		
22	SER	C	III	76.6	0.82	-0.3	-51.8	E0C		
23	SER	C	III	10.8	0.12	-0.8	-173.6	P1C		
24	GLY	C	III	41.9	0.80	-0.2	360.0	E0C		
25	ARG	C		131.2	0.64	-2.3	82.5	P2C		
26	PRO	C		108.5	0.82	-0.4	28.0	E0C		
27	PRO	C		65.7	0.50	-1.1	23.2	P2C		
28	PRO	C		57.5	0.43	-1.2	29.7	P1C		
29	SER	C		86.8	0.92	-0.3	-60.6	E0C		

```
*****
*   END VADAR   *
*****
```



## AFP-Tc-4-P18cis VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

```

(The expected values represent those numbers which would be expected for highly refined Xray and NMR protein structures. See the help pages for more information on expected values.)

Statistic	Observed	Expected
# Helix	13 ( 52%)	-
# Beta	0 ( 0%)	-
# Coil	12 ( 48%)	-
# Turn	4 ( 16%)	-

## HYDROGEN BONDS (hbonds)

Statistic	Observed	Expected
Mean hbond distance	2.2 sd=0.4	2.2 sd=0.4
Mean hbond energy	-1.8 sd=1.1	-2.0 sd=0.8
# res with hbonds	20 ( 80%)	18 ( 75%)

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## DIHEDRAL ANGLES

Statistic	Observed	Expected
Mean Helix Phi	-65.1 sd=9.8	-65.3 sd=11.9
Mean Helix Psi	-36.9 sd=12.9	-39.4 sd=25.5
# res with Gauche+ Chi	7 ( 53%)	7 ( 55%)
# res with Gauche- Chi	2 ( 15%)	2 ( 20%)
# res with Trans Chi	4 ( 30%)	3 ( 25%)
Mean Chi Gauche+	-50.4 sd=15.0	-66.7 sd=15.0
Mean Chi Gauche-	66.6 sd=22.4	64.1 sd=15.7
Mean Chi Trans	170.9 sd=15.3	168.6 sd=16.8
Std. dev of chi pooled	16.20	15.70
Mean Omega ( omega >90)	178.2 sd=7.2	180.0 sd=5.8
# res with  omega <90	2 ( 8%)	-

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	1999.9 Angs**2	1927.5 Angs**2
ASA of backbone	241.0 Angs**2	-
ASA of sidechains	1758.8 Angs**2	-
ASA of C	1347.7 Angs**2	-
ASA of N	48.5 Angs**2	-
ASA of N+	43.6 Angs**2	-
ASA of O	420.4 Angs**2	-
ASA of O-	139.7 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1339.9 Angs**2	1219.9 Angs**2
Exposed polar ASA	339.6 Angs**2	400.0 Angs**2
Exposed charged ASA	320.4 Angs**2	380.0 Angs**2
Side exposed nonpolar ASA	1377.0 Angs**2	-
Side exposed polar ASA	139.1 Angs**2	-
Side exposed charged ASA	242.8 Angs**2	-
Fraction nonpolar ASA	0.67	0.61 sd=0.03
Fraction polar ASA	0.17	0.20 sd=0.05
Fraction charged ASA	0.16	0.19 sd=0.05
Mean residue ASA	80.0 sd=37.1	-
Mean frac ASA	0.5 sd=0.2	-
% side ASA hydrophobic	39.83	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2309.6 Angs**2	-
Extended polar ASA	1060.8 Angs**2	-
Extended charged ASA	396.9 Angs**2	-
Extended side nonpolar ASA	2328.6 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2491.6 Angs**3	2472.1 Angs**3
Mean residue volume	99.7 sd=36.9	125.0 sd=40.0
Mean frac volume	0.9 sd=0.0	1.0 sd=0.1
Molecular weight	2447.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
6:151-176.

```

*****
* STEREO/PACKING QUALITY INDEX *
*****

( 3 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:          * *
SEQUENCE:  DTASDAAYAA WLTDGGPSSG RPPPS      25
TORSION.:  3333333333 3333333333 33333
OMEGA...:  1333333333 3333333233 30220
VDW.....:  3333333333 3333333333 33333
-----
TOTAL...:  7999999999 9999999899 96886

```

```

*****
* 3D PROFILE QUALITY INDEX *
*****

( 9 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:
SEQUENCE:  DTASDAAYAA WLTDGGPSSG RPPPS      25
ENV.....:  6655656666 6677777677 76777

```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in phipsi core	25 (100%)	23 ( 90%)
# res in phipsi allowed	0 ( 0%)	2 ( 7%)
# res in phipsi generous	0 ( 0%)	0 ( 1%)
# res in phipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	19 ( 76%)	24 ( 96%)
# res in omega allowed	3 ( 12%)	1 ( 3%)
# res in omega generous	1 ( 4%)	0 ( 0%)
# res in omega outside	2 ( 8%)	0 ( 1%)
# packing defects	0	1
Free energy of folding	-10.44	-8.73
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243

\*\*\*\*\*  
 \* SIDE CHAIN INFO PANEL \*  
 \*\*\*\*\*

RES.	RES.	SCND	BTURN	SIDE	FRAC.	SFE	CHI1	ENV.	DISULF	S-S
NUM.	NAME	STRUC	BTURN	SURF (ASA)	SURF (ASA)	SFE	ANGLE	CLASS	BOND	DIST
-----										
Chain A										
5	ASP	C		130.2	1.10	-1.2	-62.9	E0C		
6	THR	H		112.6	0.99	-0.4	-36.7	E0H		
7	ALA	H		85.9	1.00	0.0	360.0	E0H		
8	SER	H		73.8	0.79	-0.5	-55.2	E0H		
9	ASP	H		73.5	0.62	-0.8	-51.2	P2H		
10	ALA	H		58.6	0.68	-0.4	360.0	E0H		
11	ALA	H		58.5	0.68	-0.4	360.0	E0H		
12	TYR	H		120.4	0.59	-1.4	-178.5	P2H		
13	ALA	H		52.5	0.61	-0.5	360.0	E0H		
14	ALA	H		58.7	0.69	-0.4	360.0	E0H		
15	TRP	H		19.8	0.09	-3.0	179.7	B1H		
16	LEU	H		77.9	0.46	-1.5	-69.6	P1H		
17	THR	H		99.1	0.87	-0.1	50.8	E0H		
18	ASP	H		58.6	0.50	-1.4	-148.1	P1H		
19	GLY	C		31.8	0.61	-0.3	360.0	E0C		
20	GLY	C	III	0.8	0.01	-0.8	360.0	P1C		
21	PRO	C	III	35.7	0.27	-1.6	25.5	P1C		
22	SER	C	III	74.6	0.79	-0.3	-51.2	E0C		
23	SER	C	III	10.2	0.11	-0.8	177.4	P1C		
24	GLY	C		41.4	0.79	-0.2	360.0	E0C		
25	ARG	C		133.1	0.65	-2.3	82.5	P2C		
26	PRO	C		91.5	0.69	-0.7	28.2	P2C		
27	PRO	C		84.3	0.64	-0.8	31.7	P2C		
28	PRO	C		77.2	0.58	-0.9	34.6	P2C		
29	SER	C		98.1	1.04	-0.3	-25.8	E0C		

\*\*\*\*\*  
 \* END VADAR \*  
 \*\*\*\*\*

## AFP-Tc-4-P19cis VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

```

(The expected values represent those numbers which would be expected for highly refined Xray and NMR protein structures. See the help pages for more information on expected values.)

Statistic	Observed	Expected
# Helix	13 ( 52%)	-
# Beta	0 ( 0%)	-
# Coil	12 ( 48%)	-
# Turn	4 ( 16%)	-

## HYDROGEN BONDS (hbonds)

Statistic	Observed	Expected
Mean hbond distance	2.2 sd=0.4	2.2 sd=0.4
Mean hbond energy	-2.0 sd=1.2	-2.0 sd=0.8
# res with hbonds	20 ( 80%)	18 ( 75%)

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## DIHEDRAL ANGLES

Statistic	Observed	Expected
Mean Helix Phi	-60.5 sd=10.5	-65.3 sd=11.9
Mean Helix Psi	-40.9 sd=21.0	-39.4 sd=25.5
# res with Gauche+ Chi	7 ( 53%)	7 ( 55%)
# res with Gauche- Chi	2 ( 15%)	2 ( 20%)
# res with Trans Chi	4 ( 30%)	3 ( 25%)
Mean Chi Gauche+	-55.8 sd=10.5	-66.7 sd=15.0
Mean Chi Gauche-	43.7 sd=3.8	64.1 sd=15.7
Mean Chi Trans	169.8 sd=11.8	168.6 sd=16.8
Std. dev of chi pooled	9.84	15.70
Mean Omega ( omega >90)	176.4 sd=5.5	180.0 sd=5.8
# res with  omega <90	2 ( 8%)	-

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	1905.9 Angs**2	1927.5 Angs**2
ASA of backbone	211.2 Angs**2	-
ASA of sidechains	1694.6 Angs**2	-
ASA of C	1310.5 Angs**2	-
ASA of N	41.5 Angs**2	-
ASA of N+	46.3 Angs**2	-
ASA of O	396.9 Angs**2	-
ASA of O-	110.5 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1304.9 Angs**2	1162.6 Angs**2
Exposed polar ASA	308.4 Angs**2	381.2 Angs**2
Exposed charged ASA	292.5 Angs**2	362.1 Angs**2
Side exposed nonpolar ASA	1330.7 Angs**2	-
Side exposed polar ASA	124.9 Angs**2	-
Side exposed charged ASA	239.0 Angs**2	-
Fraction nonpolar ASA	0.68	0.61 sd=0.03
Fraction polar ASA	0.16	0.20 sd=0.05
Fraction charged ASA	0.15	0.19 sd=0.05
Mean residue ASA	76.2 sd=36.3	-
Mean frac ASA	0.5 sd=0.2	-
% side ASA hydrophobic	42.03	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2309.6 Angs**2	-
Extended polar ASA	1060.8 Angs**2	-
Extended charged ASA	396.9 Angs**2	-
Extended side nonpolar ASA	2328.6 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2504.5 Angs**3	2472.1 Angs**3
Mean residue volume	100.2 sd=37.9	125.0 sd=40.0
Mean frac volume	0.9 sd=0.1	1.0 sd=0.1
Molecular weight	2447.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
6:151-176.

```

*****
* STEREO/PACKING QUALITY INDEX *
*****

( 3 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:          * *
SEQUENCE:  DTASDAAYAA WLTDGGPSSG RPPPS      25
TORSION.:  3333333333 3332333333 33333
OMEGA...:  2333333333 3333333333 33020
VDW.....:  3333333333 3333333233 33333
-----
TOTAL...:  8999999999 99989999899 99686

```

```

*****
* 3D PROFILE QUALITY INDEX *
*****

( 9 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:
SEQUENCE:  DTASDAAYAA WLTDGGPSSG RPPPS      25
ENV.....:  7665666666 6676766677 76777

```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in pipsi core	24 ( 96%)	23 ( 90%)
# res in pipsi allowed	1 ( 4%)	2 ( 7%)
# res in pipsi generous	0 ( 0%)	0 ( 1%)
# res in pipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	21 ( 84%)	24 ( 96%)
# res in omega allowed	2 ( 8%)	1 ( 3%)
# res in omega generous	0 ( 0%)	0 ( 0%)
# res in omega outside	2 ( 8%)	0 ( 1%)
# packing defects	0	1
Free energy of folding	-10.43	-8.73
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243

```
*****
*       SIDE CHAIN INFO PANEL       *
*****
```

RES.	RES.	SCND	BTURN	SIDE	FRAC.	SFE	CHI1	ENV.	DISULF	S-S
NUM.	NAME	STRUC	BTURN	SURF (ASA)	SURF (ASA)	SFE	ANGLE	CLASS	BOND	DIST
-----										
Chain A										
5	ASP	C		122.5	1.04	-1.3	-71.0	E0C		
6	THR	H		115.5	1.02	-0.3	-39.3	E0H		
7	ALA	H		84.1	0.98	-0.0	360.0	E0H		
8	SER	H		70.3	0.75	-0.4	-48.1	E0H		
9	ASP	H		77.4	0.66	-0.9	-57.3	P2H		
10	ALA	H		60.0	0.70	-0.4	360.0	E0H		
11	ALA	H		34.3	0.40	-0.8	360.0	P1H		
12	TYR	H		112.4	0.55	-1.5	-178.0	P2H		
13	ALA	H		53.0	0.62	-0.5	360.0	E0H		
14	ALA	H		46.9	0.55	-0.6	360.0	E0H		
15	TRP	H		26.4	0.12	-2.9	-177.2	B1H		
16	LEU	H		80.2	0.47	-1.4	-62.8	P1H		
17	THR	H		97.7	0.86	-0.1	46.4	E0H		
18	ASP	H		59.7	0.51	-1.4	-152.7	P1H		
19	GLY	C		35.6	0.68	-0.3	360.0	E0C		
20	GLY	C	III	1.9	0.04	-0.8	360.0	P1C		
21	PRO	C	III	97.1	0.73	-0.6	27.4	E0C		
22	SER	C	III	79.3	0.84	-0.3	-51.4	E0C		
23	SER	C	III	10.9	0.12	-0.8	-171.2	P1C		
24	GLY	C		43.2	0.82	-0.1	360.0	E0C		
25	ARG	C		125.0	0.61	-2.3	41.0	P2C		
26	PRO	C		106.3	0.80	-0.4	29.7	E0C		
27	PRO	C		64.0	0.48	-1.1	22.1	P2C		
28	PRO	C		59.9	0.45	-1.2	36.7	P2C		
29	SER	C		30.9	0.33	-0.9	-60.6	P1C		

```
*****
*   END VADAR   *
*****
```



1.3.4. AFP-Tc-5

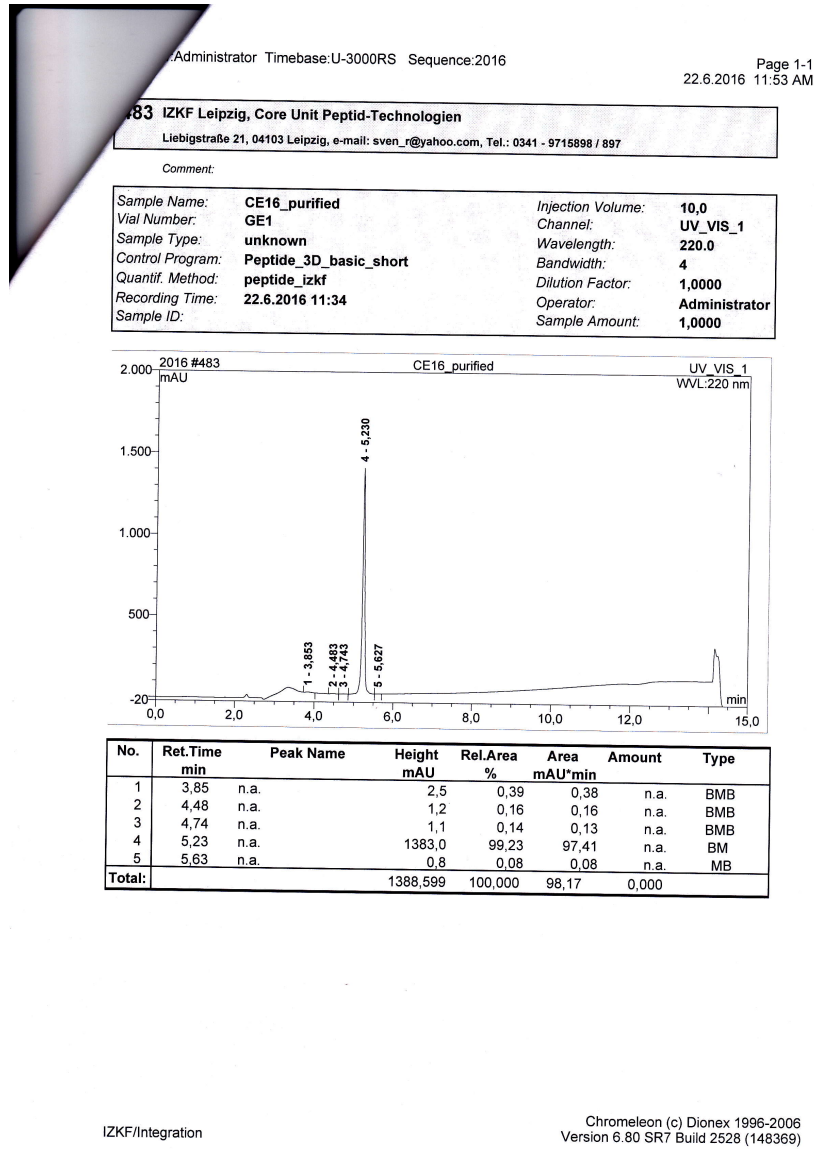


Figure V.1.38.: HPLC chromatogram of purified AFP-Tc-5, synthesized by S. Rothemund

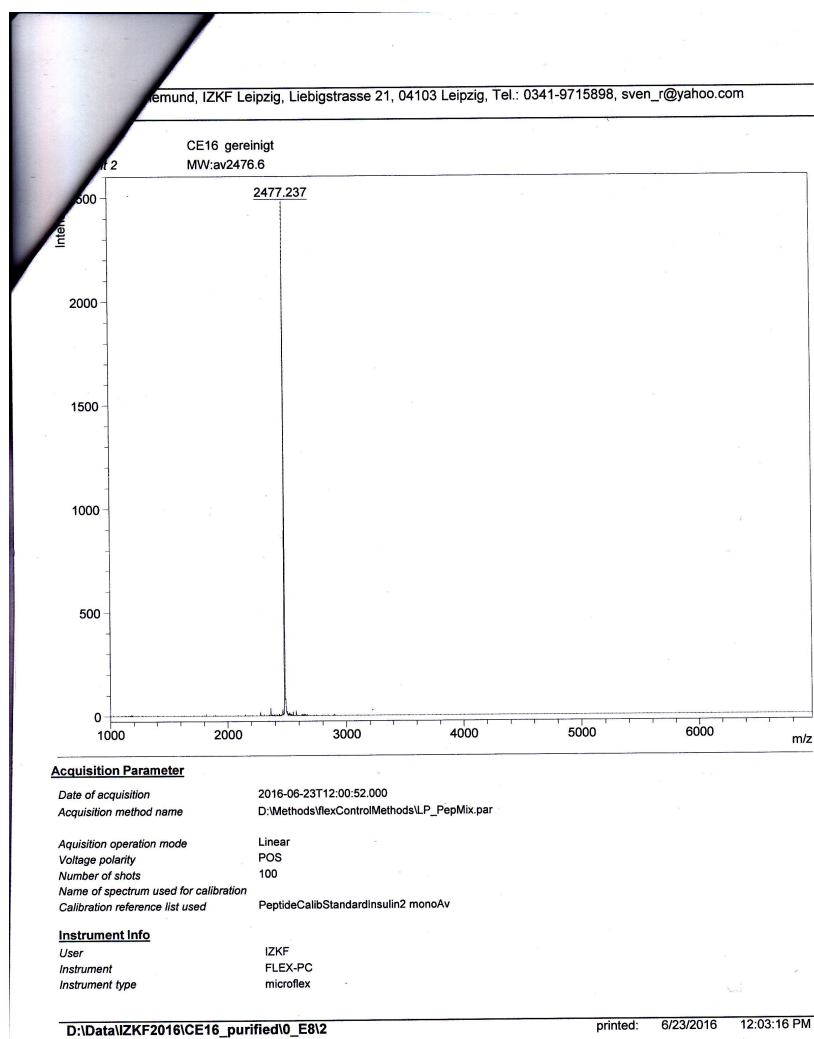
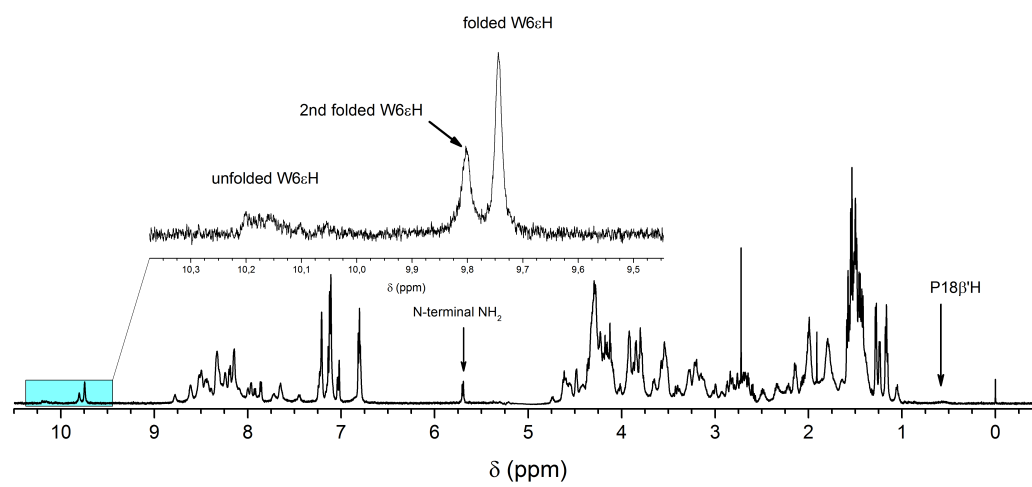


Figure V.1.39.:  $m/z$ :  $[M + H]^+$  calcd for  $[TcAFP5]^+$  2476.5, found 2477.2.



**Figure V.1.40.:**  $^1\text{H}$  NMR spectrum of AFP-Tc-5 at 274 K and pH 7.9 with enlargement of the indole proton region.

**Table V.1.49.:**  $^1\text{H}$ -chemical shifts of TcAFP5b at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.384	3.003					
2'	Thr	8.749	4.204	4.204	1.273				
3'	Ala	8.607	4.283	1.445					
4'	Ser	8.717	4.351	4.020					
5'	Asp	8.426	4.680	2.959, 2.886					
6'	Ala	8.326	4.312	1.555					
1	Ala	8.506	4.312	1.599					
2	Ala	8.295	4.268	1.547					
3	Tyr	8.618	4.072	3.204, 3.110		7.118	6.786		
4	Ala	8.437	4.305	1.533					
5	Ala	8.119	4.224	1.518					
6	Trp	8.029	4.249	3.450, 3.201		7.045	9.678, 7.209	7.218, 7.118	7.048
7	Thr	8.713	3.242	4.290	1.210				
8	Ala	8.277	4.151	1.474					
9	Asp	7.620	4.752	3.036, 2.868					
10	Gly	7.623	4.136, 3.535						
11	Gly	8.087	3.082, 1.120						
12	Pro		4.570	2.488, 2.034	2.128	3.743, 3.403			
13	Ser	7.866	4.449	3.925					
14	Ser	8.212	4.223	3.902, 3.627					
15	Gly	8.042	4.232, 3.843						
16	Arg	8.118	4.900	1.895, 1.832	1.735, 1.590	3.264	7.505		
17	Pro		4.715	2.300, 1.785	1.998	3.857, 3.652			
18	Pro		2.615	1.353, 0.505	1.758, 1.698	3.502			
19	Pro		4.327	2.216, 1.958	1.795	3.071, 2.833			
20	Ser	8.267	4.318	3.885, 3.784					

**Table V.1.50.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-5 at 278 K and pH 3.

Residue		chemical shift $\delta$ (ppm)									
		$^{15}\text{N}$			$^{13}\text{C}$						
1'	Asp	n.d.		52.51	n.d.						
2'	Thr	115.51		63.78	69.49	21.74					
3'	Ala	125.68		54.47	17.87						
4'	Ser	116.57		60.03	62.49						
5'	Asp	123.51		54.46	n.d.						
6'	Ala	122.78		54.75	17.87						
1	Ala	123.22		54.53	17.87						
2	Ala	124.25		54.53	17.87						
3	Tyr	122.34		62.33	n.d.		118.55	133.01			
4	Ala	122.61		54.53	17.87						
5	Ala	122.22		54.56	17.87						
6	Trp	120.76	131.31	61.50	27.99		127.29	124.08	114.05, 121.07	122.42	
7	Thr	116.28		66.36	68.88	21.45					
8	Ala	125.62		54.38	17.87						
9	Asp	115.95		n.d.	n.d.						
10	Gly	106.37		44.41							
11	Gly	112.38		43.95							
12	Pro	n.d.		64.61	31.79	27.36	50.65				
13	Ser	113.61		59.44	63.03						
14	Ser	117.30		59.92	64.70						
15	Gly	109.95		45.25							
16	Arg	120.20	114.78	n.d.							
17	Pro	n.d.		62.48	30.59	27.10	50.57				
18	Pro	n.d.		60.24	29.29	26.59	n.d.				
19	Pro	n.d.		63.96	31.70	27.01	49.78				
20	Ser	117.33		58.15	64.01						

**Table V.1.51.:**  $^1\text{H}$ -chemical shifts of TcAFP5b at 298K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1	Ala	8.288	4.258	1.468					
2	Ala	8.118	4.254	1.446					
3	Tyr	8.281	4.247	3.146, 3.054		7.098	6.796		
4	Ala	8.250	4.247	1.482					
5	Ala	8.009	4.205	1.459					
6	Trp	7.992	4.389	3.393, 3.245		7.120	9.818, 7.297	7.310, 7.216	7.119
7	Thr	8.312	3.580	4.197	1.143				
8	Ala	8.112	4.147	1.435					
9	Asp	7.821	4.739	2.995, 2.854					
10	Gly	1.790	4.072, 3.680						
11	Gly	8.027	3.379, 2.228						
12	Pro		4.689	2.281, 1.817	1.994	3.818, 3.605			
13	Ser	8.020	4.441	3.906					
14	Ser	8.168	4.300	3.916, 3.726					
15	Gly	8.093	4.122, 3.880						
16	Arg	8.038	4.756	1.883	1.769, 1.672	3.229	7.288		
17	Pro		4.494	2.400, 2.018	2.075	3.677, 3.405			
18	Pro		3.409	1.547, 1.220	1.824	3.6009, 3.509			
19	Pro		4.357	2.221, 1.934	1.846	3.297, 3.106			
20	Ser	8.132	4.368	3.896, 3.819					

**Table V.1.52.:** <sup>1</sup>H-chemical shifts of TcAFP5b at 274 K and pH 7.9. Residues marked in italic belong to the second folded conformation with Pro (19) in *cis*.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.699	4.362	2.643, 2.600					
2'	Thr	exch.	4.126	4.183	1.828				
3'	Ala	exch.	4.2863	1.4567					
4'	Ser	8.779	4.335	4.031, 3.936					
5'	Asp	8.524	4.550	2.813, 2.747					
5'	<i>Asp</i>	8.444	4.533	2.721					
6'	Ala	8.332	4.302	1.559					
1	Ala	8.605	4.321	1.418					
2	Ala	8.327	4.337	1.576					
3	Tyr	8.489	4.125	3.197, 3.146		7.126	6.809		
3	<i>Tyr</i>	8.464	4.135	3.152		7.104	6.793		
4	Ala	8.501	4.313	1.588					
5	Ala	8.145	4.233	1.522					
6	Trp	7.969	4.299	3.429, 3.211		7.024	9.748, 7.220	7.213, 7.176	7.115
6	<i>Trp</i>	8.000	4.324	3.399, 3.214		7.043	9.806, n.d.	n.d.	n.d.
7	Thr	8.616	3.291	4.262	1.174				
8	Ala	8.343	4.096	1.501					
9	Asp	7.656	4.625	2.849, 2.682					
9	<i>Asp</i>	7.732	4.612	2.823, 2.675					
10	Gly	7.642	4.132, 3.516						
10	<i>Gly</i>	7.712	4.109, 3.516						
11	Gly	8.243	3.255, 1.472						
12	Pro		4.739	2.339, 1.793	1.990	3.846, 3.659			
13	Ser	exch.	4.176	3.848, 3.537					
14	Ser	8.149	4.271	3.837					
15	Gly	8.097	4.2256, 3.823						
16	Arg	8.197	4.980	1.856, 1.791	1.646	3.285			

**Table V.1.53.:** Continued  $^1\text{H}$ -chemical shifts of TcAFP5b at 274 K and pH 7.9. Residues marked in *italic* belong to the second folded conformation with Pro (19) in *cis*.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
17	Pro		4.614	2.493, 2.069	2.150	3.792, 3.538			
18	Pro		2.676	1.416, 0.557	1.734	3.571			
19	Pro		4.297	2.217, 2.001	1.915, 1.807	3.125, 2.933			
20	Ser	7.862	4.135	3.794					

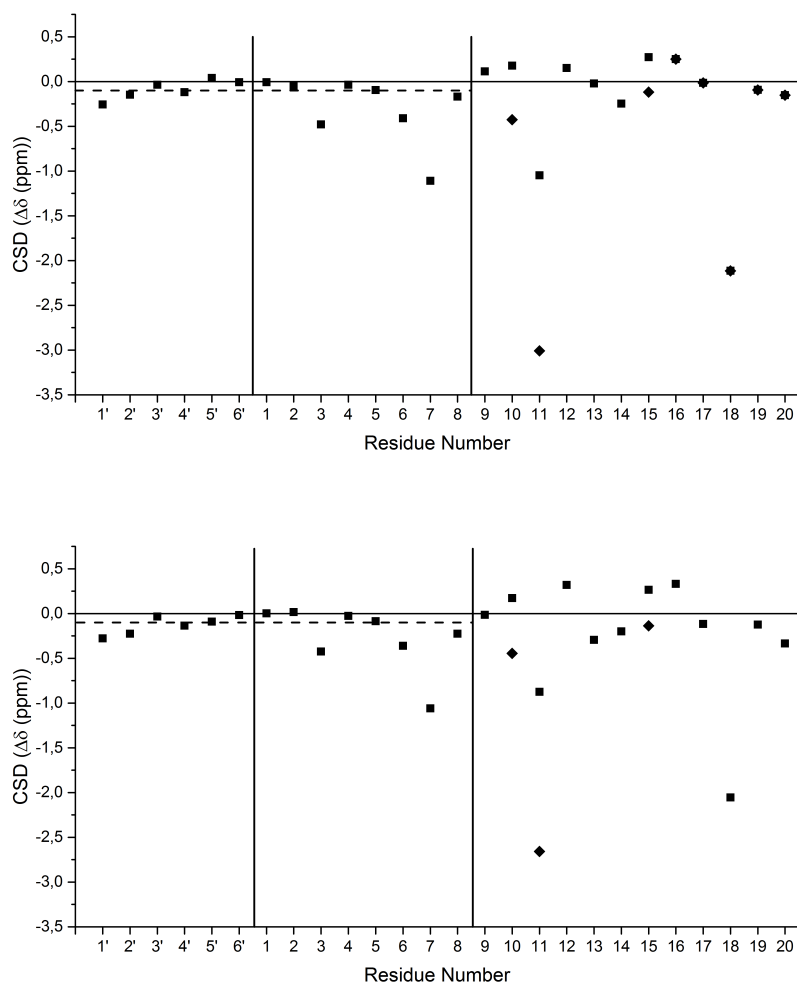
**Table V.1.54.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-5 at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)									
		$^{15}\text{N}$		$^{13}\text{C}$							
1'	Asp	n.d.		n.d.	n.d.						
2'	Thr	n.d.		64.849	69.3477	21.568					
3'	Ala	n.d.		54.5155	17.8529						
4'	Ser	n.d.		60.862	62.2272						
5'	Asp	n.d.		56.3869	n.d.						
6'	Ala	n.d.		54.5968	17.8529						
1	Ala	n.d.		53.9357	17.8529						
2	Ala	n.d.		54.4341	17.8529						
3	Tyr	n.d.		62.2453	n.d.		118.241	9133.0221			
4	Ala	n.d.		54.5968	17.8529						
5	Ala	n.d.		54.4341	17.8529						
6	Trp	n.d.	n.d.	61.3386	28.2757		127.427	3123.8367	114.15, 120.77	122.4171	
7	Thr	n.d.		66.2322	69.2445	21.3616					
8	Ala	n.d.		54.5968	17.8529						
9	Asp	n.d.		n.d.							
10	Gly	n.d.		44.0647							
11	Gly	n.d.		n.d.							
12	Pro	n.d.		61.1492	30.546	27.0373	50.4628				
13	Ser	n.d.		n.d.	64.7039						
14	Ser	n.d.		n.d.	n.d.						
15	Gly	n.d.		44.9934	n.d.						
16	Arg	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.				
17	Pro	n.d.		64.9494	31.7844	27.3469	50.6692				
18	Pro	n.d.		n.d.	29.3077	26.5214	49.9468				
19	Pro	n.d.		62.5866	31.578	26.7278	49.5341				
20	Ser	n.d.		n.d.	64.7039						

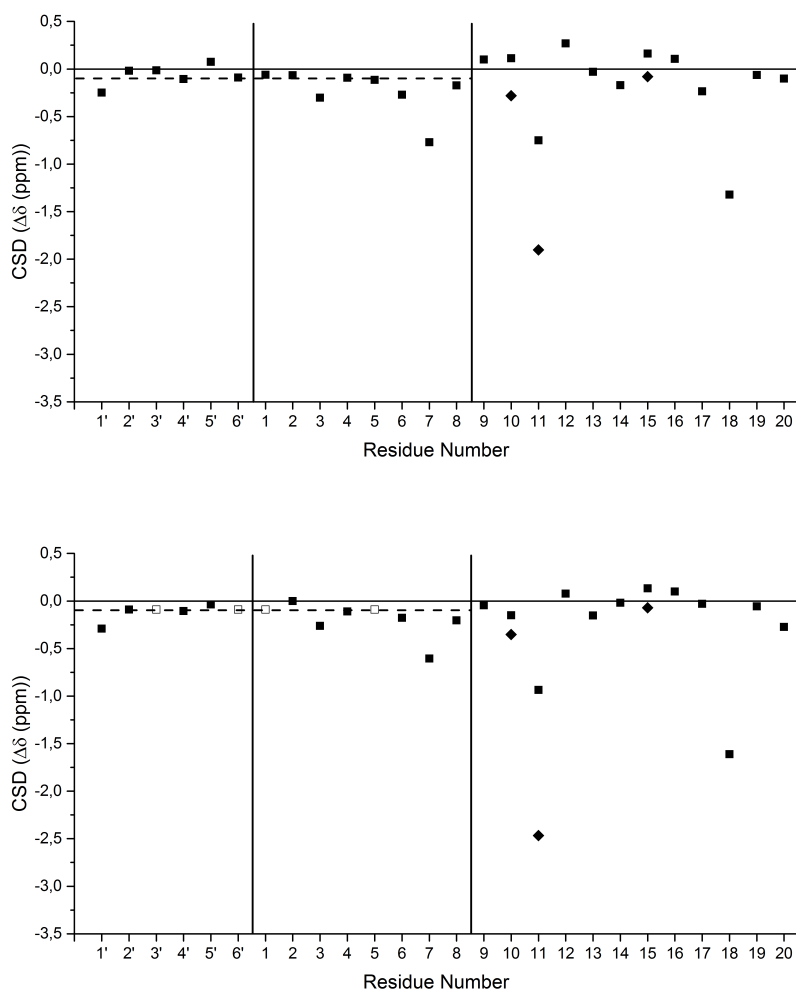


**Table V.1.55.:**  $^1\text{H}$ -chemical shifts of TcAFP5b at 298 K and pH 7.9. Shifts marked with an asterisk are averaged.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.599	4.350	2.614					
2'	Thr	exch.	4.260	4.220	1.249				
3'	Ala	exch.	4.231*	1.443*					
4'	Ser	exch.	4.364	3.932, 3.860					
5'	Asp	exch.	4.603	2.737, 2.661					
6'	Ala	exch.	4.231*	1.443*					
1	Ala	exch.	4.231*	1.443*					
2	Ala	exch.	4.318	1.425					
3	Tyr	8.141	4.289	3.104, 3.047		7.100	6.805		
4	Ala	exch.	4.231*	1.443*					
5	Ala	exch.	4.231*	1.443*					
6	Trp	7.971	4.482	3.354, 3.257		7.134	9.915, 7.356	7.344, 7.214	7.123
7	Thr	exch.	3.745	4.182	1.113				
8	Ala	exch.	4.117	1.427					
9	Asp	7.871	4.595	2.749, 2.644					
10	Gly	exch.	3.81, 3.608						
11	Gly	exch.	3.196, 1.662						
12	Pro		4.497	2.367, 2.006	2.063	3.678, 3.714			
13	Ser	exch.	4.318	3.887, 3.714					
14	Ser	exch.	4.451	3.899					
15	Gly	exch.	4.092, 3.888						
16	Arg	exch.	4.748	1.797, 1.759	1.656	3.216, 3.186			
17	Pro		4.699	2.308, 1.838	1.995	3.797, 3.607			
18	Pro		3.118	1.653, 1.480	1.874	3.675, 3.544			
19	Pro		4.364	2.239, 1.976	1.907	3.422, 3.275			
20	Ser	7.834	4.198	3.816					



**Figure V.1.41.:**  $H\alpha$  CSD plot of AFP-Tc-5 at pH 3 (top) and pH 7.9 (bottom) at 274 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. The dashed line at  $y = -0.1$  ppm represents the helical limit. Solid line between residue 6' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.



**Figure V.1.42.:**  $H_{\alpha}$  CSD plot of AFP-Tc-5 at pH 3 (top) and pH 7.9 (bottom) at 298 K.  $H_{\alpha}$  as black squares and  $H_{\alpha}'$  as black diamonds. Averaged  $H_{\alpha}$  shifts for certain Ala are represented as non-filled squares. The dashed line at  $y=-0.1$  ppm represents the helical limit. Solid line between residue 6' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.

**Table V.1.56.:** CSDs for selected protons in the cage structure representing the cage fold of TcAFP5b at 298 K and 274 K and at pH 3 and 7.9.

Cage	$\Delta\delta$ (ppm).						$\Sigma$
	upfield						
	T7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$	
pH 3, 274 K	-1.109	-3.010	-2.115	-1.405	-0.529	-0.767	<b>-8.935</b>
pH 3, 298 K	-0.771	-1.902	-1.321	-0.363	-0.303	-0.494	<b>-5.155</b>
pH 7.9, 274 K	-1.059	-2.658	-2.055	-1.353	-0.475	-0.667	<b>-8.267</b>
pH 7.9, 298 K	-0.606	-2.4679	-1.612	-0.430	-0.178	-0.325	<b>-5.618</b>
downfield							
	P12 $\alpha$	P12 $\beta$	R16 $\alpha$				
pH 3, 274 K	0.150	-0.073	0.250				<b>0.598</b>
pH 3, 298 K	0.269	-0.198	0.106				<b>0.365</b>
pH 7.9, 274 K	0.319	0.045	0.330				<b>0.698</b>
pH 7.9, 298 K	0.077	0.077	0.098				<b>0.252</b>

**Table V.1.57.:** CSDs for selected protons in the helical region representing the helix fold of AFP-Tc-5 at 298 K and 274 K and at pH 3 and 7.9.

Helix	$\Delta\delta$ (ppm)							$\Sigma$
	A2 $\alpha$	Y3 $\alpha$	A4 $\alpha$	A5 $\alpha$	W6 $\alpha$	T7 $\alpha$	A8 $\alpha$	
pH 3, 274 K	-0.052	-0.478	-0.351	-0.100	-0.411	-1.109	-0.169	<b>-2.349</b>
pH 3, 298 K	-0.066	-0.303	-0.093	-0.115	-0.271	-0.771	-0.173	<b>-1.792</b>
pH 7.9, 274 K	0.017	-0.425	-0.027	-0.087	-0.361	-1.056	-0.224	<b>-2.166</b>
pH 7.9, 298 K	-0.002	-0.261	-0.109*	-0.059	-0.178	-0.606	-0.203	<b>-1.448</b>

**Table V.1.58.:** CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP5b at 298 K and 274 K and at pH 3 and 7.9.

Helix	$\Delta\delta$ (ppm)					$\Sigma$
	T2' $\alpha$	A3' $\alpha$	S'4 $\alpha$	D5' $\alpha$	A6' $\alpha$	
pH 3, 274 K	-0.146	-0.037	-0.120	0.040	-0.008	<b>-0.271</b>
pH 3, 298 K	-0.020	-0.015	-0.106	0.074	0.090	<b>-0.156</b>
pH 7.9, 274 K	-0.225	-0.034	-0.135	-0.090	-0.018	<b>-0.501</b>
pH 7.9, 298 K	-0.130	-0.089	-0.106	-0.037	-0.089	<b>-0.452</b>

**Table V.1.59.:** CSDs for the  $^{13}\text{C}\alpha$  atoms in the helix representing the helix fold of AFP-Tc-5 at 274 K and at pH 3 and 7.9. The  $\overline{\Delta\delta}$  and the  $\chi_F$  for the helix, chimera and helical extension are calculated. The  $\chi_F$  was determined using the CSD of AFP-Tc-7 at pH 7.9 and 274 K as reference.

TcAFP5		
	pH 3	pH 8
	$\delta$ (ppm)	$\delta$ (ppm)
<b>Helical Extension</b>	0.68	1.75
	1.97	2.02
	1.73	2.56
	0.6	2.29
	2.25	2.10
<b>Chimera Helix</b>	2.03	2.10
	2.03	2.07
	3.73	3.99
	2.03	2.10
	2.06	1.93
	3.70	4.61
	3.26	3.13
	2.01	2.10
<b>Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.14	2.52
$\chi_F^{helix}$ (%)	77	90
<b>Chimera Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.61	2.75
$\chi_F^{chimera}$ (%)	83	88
<b>Helical Extension <math>\overline{\Delta\delta}</math> (ppm)</b>	1.40	2.14
$\chi_F^{exten.}$ (%)	57	87

## AFP-Tc-5 VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

```

(The expected values represent those numbers which would be expected for highly refined Xray and NMR protein structures. See the help pages for more information on expected values.)

Statistic	Observed	Expected
# Helix	14 ( 53%)	-
# Beta	0 ( 0%)	-
# Coil	12 ( 46%)	-
# Turn	8 ( 30%)	-

## HYDROGEN BONDS (hbonds)

Statistic	Observed	Expected
Mean hbond distance	2.2 sd=0.3	2.2 sd=0.4
Mean hbond energy	-1.7 sd=1.0	-2.0 sd=0.8
# res with hbonds	20 ( 76%)	19 ( 75%)

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## DIHEDRAL ANGLES

Statistic	Observed	Expected
Mean Helix Phi	-66.1 sd=9.6	-65.3 sd=11.9
Mean Helix Psi	-35.2 sd=12.5	-39.4 sd=25.5
# res with Gauche+ Chi	8 ( 66%)	6 ( 55%)
# res with Gauche- Chi	1 ( 8%)	2 ( 20%)
# res with Trans Chi	3 ( 25%)	3 ( 25%)
Mean Chi Gauche+	-57.6 sd=19.6	-66.7 sd=15.0
Mean Chi Gauche-	50.5 sd=-nan	64.1 sd=15.7
Mean Chi Trans	165.8 sd=13.7	168.6 sd=16.8
Std. dev of chi pooled	-nan	15.70
Mean Omega ( omega >90)	177.3 sd=8.6	180.0 sd=5.8
# res with  omega <90	1 ( 3%)	-

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	2076.5 Angs**2	1943.8 Angs**2
ASA of backbone	281.6 Angs**2	-
ASA of sidechains	1794.8 Angs**2	-
ASA of C	1448.0 Angs**2	-
ASA of N	55.0 Angs**2	-
ASA of N+	38.2 Angs**2	-
ASA of O	432.7 Angs**2	-
ASA of O-	102.6 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1437.7 Angs**2	1266.6 Angs**2
Exposed polar ASA	363.1 Angs**2	415.3 Angs**2
Exposed charged ASA	275.7 Angs**2	394.5 Angs**2
Side exposed nonpolar ASA	1414.6 Angs**2	-
Side exposed polar ASA	151.7 Angs**2	-
Side exposed charged ASA	228.6 Angs**2	-
Fraction nonpolar ASA	0.69	0.61 sd=0.03
Fraction polar ASA	0.17	0.20 sd=0.05
Fraction charged ASA	0.13	0.19 sd=0.05
Mean residue ASA	79.9 sd=37.7	-
Mean frac ASA	0.6 sd=0.3	-
% side ASA hydrophobic	43.14	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
 J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2350.6 Angs**2	-
Extended polar ASA	1097.2 Angs**2	-
Extended charged ASA	358.4 Angs**2	-
Extended side nonpolar ASA	2330.2 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2535.0 Angs**3	2501.4 Angs**3
Mean residue volume	97.5 sd=38.2	125.0 sd=40.0
Mean frac volume	0.9 sd=0.1	1.0 sd=0.1
Molecular weight	2476.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
 6:151-176.

```

*****
* STEREO/PACKING QUALITY INDEX *
*****

( 3 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...: *                               **
SEQUENCE: DTASDAAAYA AWTADGGPSS GRPPPS    26
TORSION.: 3233333333 3333333333 333333
OMEGA...: 0333333333 3333333333 333300
VDW.....: 3333333333 3333333333 233333
-----
TOTAL...: 6899999999 9999999999 899966

```

```

*****
* 3D PROFILE QUALITY INDEX *
*****

( 9 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:
SEQUENCE: DTASDAAAYA AWTADGGPSS GRPPPS    26
ENV.....: 6666676667 7777777777 777777

```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in phipsi core	25 ( 96%)	23 ( 90%)
# res in phipsi allowed	1 ( 3%)	2 ( 7%)
# res in phipsi generous	0 ( 0%)	0 ( 1%)
# res in phipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	23 ( 88%)	25 ( 96%)
# res in omega allowed	0 ( 0%)	1 ( 3%)
# res in omega generous	0 ( 0%)	0 ( 0%)
# res in omega outside	3 ( 11%)	0 ( 1%)
# packing defects	1	1
Free energy of folding	-8.37	-9.72
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243



```
*****
*   SIDE CHAIN INFO PANEL   *
*****
```

RES. NUM.	RES. NAME	SCND STRUC	BTURN BTURN	SIDE SURF(ASA)	FRAC. SURF(ASA)	SFE SFE	CHI1 ANGLE	ENV. CLASS	DISULF BOND	S-S DIST
-----										
Chain 1										
1	ASP	C		131.9	1.12	-1.1	-57.2	E0C		
2	THR	H		111.8	0.99	-0.4	-34.1	E0H		
3	ALA	H		84.0	0.98	-0.0	360.0	E0H		
4	SER	H		72.7	0.77	-0.5	-62.5	E0H		
5	ASP	H		62.5	0.53	-0.6	-37.0	P1H		
6	ALA	H		55.8	0.65	-0.5	360.0	E0H		
7	ALA	H		57.1	0.67	-0.5	360.0	E0H		
8	ALA	H		39.4	0.46	-0.7	360.0	P1H		
9	TYR	H		99.7	0.49	-1.5	-95.8	P1H		
10	ALA	H		58.5	0.68	-0.4	360.0	E0H		
11	ALA	H		59.5	0.70	-0.4	360.0	E0H		
12	TRP	H		48.7	0.21	-2.6	173.8	B1H		
13	THR	H	I	80.5	0.71	-0.4	-56.8	E0H		
14	ALA	H	I	67.5	0.79	-0.3	360.0	E0H		
15	ASP	H	I	61.5	0.52	-1.3	-150.0	P2H		
16	GLY	C	I	29.6	0.56	-0.4	360.0	E0C		
17	GLY	C	III	3.5	0.07	-0.8	360.0	P1C		
18	PRO	C	III	102.0	0.77	-0.5	-23.6	E0C		
19	SER	C	III	76.6	0.81	-0.4	-47.9	E0C		
20	SER	C	III	7.3	0.08	-0.9	-173.6	P1C		
21	GLY	C		42.3	0.81	-0.2	360.0	E0C		
22	ARG	C		137.2	0.67	-2.0	50.5	P2C		
23	PRO	C		108.2	0.82	-0.4	28.1	E0C		
24	PRO	C		46.7	0.35	-1.4	25.8	P1C		
25	PRO	C		47.0	0.35	-1.4	28.2	P1C		
26	SER	C		103.5	1.10	-0.1	-69.2	E0C		

```
*****
*   END VADAR   *
*****
```

AFP-Tc-5-P18*cis* VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

(The expected values represent those numbers which would be expected
for highly refined Xray and NMR protein structures.
See the help pages for more information on expected values.)

|-----|-----|-----|
| Statistic | Observed | Expected |
|-----|-----|-----|
| # Helix   | 14 ( 53%) | -        |
| # Beta    | 0 ( 0%)   | -        |
| # Coil    | 12 ( 46%) | -        |
|-----|-----|-----|
| # Turn    | 4 ( 15%)  | -        |
|-----|-----|-----|

HYDROGEN BONDS (hbonds)

|-----|-----|-----|
| Statistic | Observed | Expected |
|-----|-----|-----|
| Mean hbond distance | 2.1 sd=0.3 | 2.2 sd=0.4 |
| Mean hbond energy   | -2.0 sd=1.1 | -2.0 sd=0.8 |
| # res with hbonds   | 20 ( 76%) | 19 ( 75%) |
|-----|-----|-----|

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and
Thornton JM. Proteins. 1992 Apr;12(4):345-364.

DIHEDRAL ANGLES

|-----|-----|-----|
| Statistic | Observed | Expected | | |
|---|---|---|---|---|
| Mean Helix Phi | -56.2 sd=16.3 | -65.3 sd=11.9 |
| Mean Helix Psi | -43.1 sd=11.7 | -39.4 sd=25.5 |
| # res with Gauche+ Chi | 8 ( 66%) | 6 ( 55%) |
| # res with Gauche- Chi | 1 ( 8%) | 2 ( 20%) |
| # res with Trans Chi | 3 ( 25%) | 3 ( 25%) |
| Mean Chi Gauche+ | -54.4 sd=14.8 | -66.7 sd=15.0 |
| Mean Chi Gauche- | 64.5 sd=-nan | 64.1 sd=15.7 |
| Mean Chi Trans | 162.9 sd=21.8 | 168.6 sd=16.8 |
| Std. dev of chi pooled | -nan | 15.70 |
| Mean Omega (|omega|>90) | 176.3 sd=6.8 | 180.0 sd=5.8 |
| # res with |omega|<90 | 2 ( 7%) | - |
|-----|-----|-----|

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and
Thornton JM. Proteins. 1992 Apr;12(4):345-364.

```

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	2101.4 Angs**2	1943.8 Angs**2
ASA of backbone	236.8 Angs**2	-
ASA of sidechains	1864.6 Angs**2	-
ASA of C	1448.8 Angs**2	-
ASA of N	44.4 Angs**2	-
ASA of N+	41.0 Angs**2	-
ASA of O	425.3 Angs**2	-
ASA of O-	141.9 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1439.9 Angs**2	1281.8 Angs**2
Exposed polar ASA	338.6 Angs**2	420.3 Angs**2
Exposed charged ASA	322.9 Angs**2	399.3 Angs**2
Side exposed nonpolar ASA	1475.4 Angs**2	-
Side exposed polar ASA	138.6 Angs**2	-
Side exposed charged ASA	250.6 Angs**2	-
Fraction nonpolar ASA	0.69	0.61 sd=0.03
Fraction polar ASA	0.16	0.20 sd=0.05
Fraction charged ASA	0.15	0.19 sd=0.05
Mean residue ASA	80.8 sd=35.5	-
Mean frac ASA	0.6 sd=0.2	-
% side ASA hydrophobic	43.14	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
 J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2312.1 Angs**2	-
Extended polar ASA	1097.2 Angs**2	-
Extended charged ASA	396.9 Angs**2	-
Extended side nonpolar ASA	2330.2 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2495.8 Angs**3	2501.4 Angs**3
Mean residue volume	96.0 sd=36.9	125.0 sd=40.0
Mean frac volume	0.9 sd=0.1	1.0 sd=0.1
Molecular weight	2476.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
 6:151-176.

```

*****
* STEREO/PACKING QUALITY INDEX *
*****

( 3 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:          * *
SEQUENCE:  DTASDAAAYA AWTADGGPSS GRPPPS      26
TORSION.:  3233333333 3333333333 333333
OMEGA...:  1333333333 3333333323 330210
VDW.....:  3333333333 3332333333 333333
-----
TOTAL...:  7899999999 9998999989 996876

```

```

*****
* 3D PROFILE QUALITY INDEX *
*****

( 9 = best )
( 0 = worst )

( * = indicates possible problem )

```

```

PRBLM...:
SEQUENCE:  DTASDAAAYA AWTADGGPSS GRPPPS      26
ENV.....:  5666676667 7777777777 777777

```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in phipsi core	25 ( 96%)	23 ( 90%)
# res in phipsi allowed	1 ( 3%)	2 ( 7%)
# res in phipsi generous	0 ( 0%)	0 ( 1%)
# res in phipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	20 ( 76%)	25 ( 96%)
# res in omega allowed	2 ( 7%)	1 ( 3%)
# res in omega generous	2 ( 7%)	0 ( 0%)
# res in omega outside	2 ( 7%)	0 ( 1%)
# packing defects	1	1
Free energy of folding	-8.76	-9.72
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243

\*\*\*\*\*  
 \* SIDE CHAIN INFO PANEL \*  
 \*\*\*\*\*

RES. NUM.	RES. NAME	SCND STRUC	BTURN BTURN	SIDE SURF (ASA)	FRAC. SURF (ASA)	SFE SFE	CHI1 ANGLE	ENV. CLASS	DISULF BOND	S-S DIST
-----										
Chain 1										
1	ASP	C		131.4	1.11	-1.2	-64.0	E0C		
2	THR	H		118.8	1.05	-0.3	-42.4	E0H		
3	ALA	H		83.6	0.98	-0.0	360.0	E0H		
4	SER	H		67.9	0.72	-0.4	-37.1	E0H		
5	ASP	H		68.1	0.58	-1.0	-55.0	P1H		
6	ALA	H		62.9	0.74	-0.4	360.0	E0H		
7	ALA	H		62.9	0.73	-0.4	360.0	E0H		
8	ALA	H		58.6	0.68	-0.4	360.0	E0H		
9	TYR	H		147.0	0.72	-1.1	-84.6	P2H		
10	ALA	H		60.7	0.71	-0.4	360.0	E0H		
11	ALA	H		61.1	0.71	-0.4	360.0	E0H		
12	TRP	H		39.8	0.18	-2.7	176.5	B1H		
13	THR	H		81.9	0.72	-0.4	-53.9	E0H		
14	ALA	H		67.4	0.79	-0.3	360.0	E0H		
15	ASP	H		59.9	0.51	-1.3	-137.8	P2H		
16	GLY	C		32.2	0.61	-0.3	360.0	E0C		
17	GLY	C	III	3.3	0.06	-0.8	360.0	P1C		
18	PRO	C	III	37.9	0.29	-1.5	36.4	P1C		
19	SER	C	III	79.1	0.84	-0.2	-46.1	E0C		
20	SER	C	III	13.2	0.14	-0.8	-174.5	P1C		
21	GLY	C		40.9	0.78	-0.2	360.0	E0C		
22	ARG	C		123.5	0.60	-2.5	64.5	P2C		
23	PRO	C		102.8	0.77	-0.5	28.7	E0C		
24	PRO	C		69.6	0.52	-1.0	28.7	P1C		
25	PRO	C		97.1	0.73	-0.6	28.9	E0C		
26	SER	C		93.0	0.99	-0.3	-51.8	E0C		

\*\*\*\*\*  
 \* END VADAR \*  
 \*\*\*\*\*

## AFP-Tc-5-P19cis VADAR Statistics

```

*****
*           VADAR STATS           *
*****
** Using atomic radii from Shrake **

```

(The expected values represent those numbers which would be expected for highly refined Xray and NMR protein structures. See the help pages for more information on expected values.)

Statistic	Observed	Expected
# Helix	14 ( 53%)	-
# Beta	0 ( 0%)	-
# Coil	12 ( 46%)	-
# Turn	4 ( 15%)	-

## HYDROGEN BONDS (hbonds)

Statistic	Observed	Expected
Mean hbond distance	2.3 sd=0.4	2.2 sd=0.4
Mean hbond energy	-1.4 sd=1.0	-2.0 sd=0.8
# res with hbonds	22 ( 84%)	19 ( 75%)

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## DIHEDRAL ANGLES

Statistic	Observed	Expected
Mean Helix Phi	-66.8 sd=9.7	-65.3 sd=11.9
Mean Helix Psi	-35.4 sd=9.5	-39.4 sd=25.5
# res with Gauche+ Chi	7 ( 58%)	6 ( 55%)
# res with Gauche- Chi	2 ( 16%)	2 ( 20%)
# res with Trans Chi	3 ( 25%)	3 ( 25%)
Mean Chi Gauche+	-57.0 sd=15.0	-66.7 sd=15.0
Mean Chi Gauche-	63.9 sd=9.7	64.1 sd=15.7
Mean Chi Trans	165.3 sd=12.6	168.6 sd=16.8
Std. dev of chi pooled	13.51	15.70
Mean Omega ( omega >90)	177.6 sd=6.8	180.0 sd=5.8
# res with  omega <90	2 ( 7%)	-

Expected values obtained from Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364.

## ACCESSIBLE SURFACE AREA (ASA)

Statistic	Observed	Expected
Total ASA	1971.1 Angs**2	1943.8 Angs**2
ASA of backbone	285.5 Angs**2	-
ASA of sidechains	1685.6 Angs**2	-
ASA of C	1350.1 Angs**2	-
ASA of N	22.2 Angs**2	-
ASA of N+	40.6 Angs**2	-
ASA of O	445.6 Angs**2	-
ASA of O-	112.6 Angs**2	-
ASA of S	0.0 Angs**2	-
Exposed nonpolar ASA	1342.2 Angs**2	1202.4 Angs**2
Exposed polar ASA	373.1 Angs**2	394.2 Angs**2
Exposed charged ASA	255.8 Angs**2	374.5 Angs**2
Side exposed nonpolar ASA	1338.0 Angs**2	-
Side exposed polar ASA	140.2 Angs**2	-
Side exposed charged ASA	207.4 Angs**2	-
Fraction nonpolar ASA	0.68	0.61 sd=0.03
Fraction polar ASA	0.19	0.20 sd=0.05
Fraction charged ASA	0.13	0.19 sd=0.05
Mean residue ASA	75.8 sd=33.2	-
Mean frac ASA	0.5 sd=0.2	-
% side ASA hydrophobic	44.53	-

Expected values obtained from Miller S, Janin J, Lesk AM and Chothia C.  
 J Mol Biol. 1987 Aug 5;196(3):641-656..

## ACCESSIBLE SURFACE AREA FOR EXTENDED CHAIN

Statistic	Observed	Expected
Extended nonpolar ASA	2350.6 Angs**2	-
Extended polar ASA	1097.2 Angs**2	-
Extended charged ASA	358.4 Angs**2	-
Extended side nonpolar ASA	2330.2 Angs**2	-
Extended side polar ASA	231.0 Angs**2	-
Extended side charged ASA	321.4 Angs**2	-

## VOLUME

Statistic	Observed	Expected
Total volume (packing)	2577.6 Angs**3	2501.4 Angs**3
Mean residue volume	99.1 sd=39.5	125.0 sd=40.0
Mean frac volume	0.9 sd=0.1	1.0 sd=0.1
Molecular weight	2476.62	-

Expected values obtained from Richards FM. Annu Rev Biophys Bioeng. 1977;  
 6:151-176.

```
*****
* STEREO/PACKING QUALITY INDEX *
*****
```

```
( 3 = best )
( 0 = worst )
( * = indicates possible problem )
```

```
PRBLM...: * * *
SEQUENCE: DTASDAAAYA AWTADGGPSS GRPPPS 26
TORSION.: 333333333 333333333 333332
OMEGA...: 033333333 333333333 333030
VDW.....: 333333333 333233333 233333
-----
TOTAL...: 699999999 999899999 899695
```

```
*****
* 3D PROFILE QUALITY INDEX *
*****
```

```
( 9 = best )
( 0 = worst )
( * = indicates possible problem )
```

```
PRBLM...:
SEQUENCE: DTASDAAAYA AWTADGGPSS GRPPPS 26
ENV.....: 6666676666 7666677667 776666
```

Statistic	Observed	Expected
Resolution	-	-
R Value	-	-
# res in phipsi core	25 ( 96%)	23 ( 90%)
# res in phipsi allowed	1 ( 3%)	2 ( 7%)
# res in phipsi generous	0 ( 0%)	0 ( 1%)
# res in phipsi outside	0 ( 0%)	0 ( 0%)
# res in omega core	23 ( 88%)	25 ( 96%)
# res in omega allowed	0 ( 0%)	1 ( 3%)
# res in omega generous	0 ( 0%)	0 ( 0%)
# res in omega outside	3 ( 11%)	0 ( 1%)
# packing defects	2	1
Free energy of folding	-10.18	-9.72
# res 95% buried	1	0
# buried charges	0	0

Expected values obtained from 1. Morris AL, MacArthur MW, Hutchinson EG and Thornton JM. Proteins. 1992 Apr;12(4):345-364. 2. Chiche L., Gregoret LM, Cohen FE and Kollman PA. Proc Natl Acad Sci U S A. 1990 Apr;87(8):3240-3243



```
*****
*   SIDE CHAIN INFO PANEL   *
*****
```

RES. NUM.	RES. NAME	SCND STRUC	BTURN	SIDE SURF (ASA)	FRAC. SURF (ASA)	SFE	CHI1 ANGLE	ENV. CLASS	DISULF BOND	S-S DIST
-----										
Chain 1										
1	ASP	C		132.4	1.12	-1.1	-56.7	E0C		
2	THR	H		117.2	1.03	-0.3	-39.9	E0H		
3	ALA	H		85.1	0.99	-0.0	360.0	E0H		
4	SER	H		74.1	0.79	-0.4	-59.9	E0H		
5	ASP	H		63.2	0.54	-0.8	-48.9	P1H		
6	ALA	H		55.4	0.65	-0.5	360.0	E0H		
7	ALA	H		59.1	0.69	-0.4	360.0	E0H		
8	ALA	H		26.9	0.31	-0.9	360.0	P1H		
9	TYR	H		124.5	0.61	-1.4	-87.3	P2H		
10	ALA	H		57.7	0.67	-0.4	360.0	E0H		
11	ALA	H		51.4	0.60	-0.5	360.0	E0H		
12	TRP	H		35.8	0.16	-2.8	173.5	B1H		
13	THR	H		79.6	0.70	-0.5	-57.0	E0H		
14	ALA	H		66.6	0.78	-0.3	360.0	E0H		
15	ASP	H		55.2	0.47	-1.2	-150.8	P2H		
16	GLY	C		29.2	0.56	-0.4	360.0	E0C		
17	GLY	C		2.8	0.05	-0.8	360.0	P1C		
18	PRO	C	I	99.3	0.75	-0.5	-22.3	E0C		
19	SER	C	I	75.5	0.80	-0.4	-49.2	E0C		
20	SER	C	I	14.8	0.16	-0.8	-171.5	P1C		
21	GLY	C	I	40.8	0.78	-0.2	360.0	E0C		
22	ARG	C		89.8	0.44	-2.4	70.8	P1C		
23	PRO	C		111.2	0.84	-0.3	28.8	E0C		
24	PRO	C		49.5	0.37	-1.3	27.5	P1C		
25	PRO	C		52.6	0.40	-1.3	33.7	P2C		
26	SER	C		35.9	0.38	-0.4	57.0	P2C		

```
*****
*   END VADAR   *
*****
```

## 1.3.5. AFP-Tc-5-KK

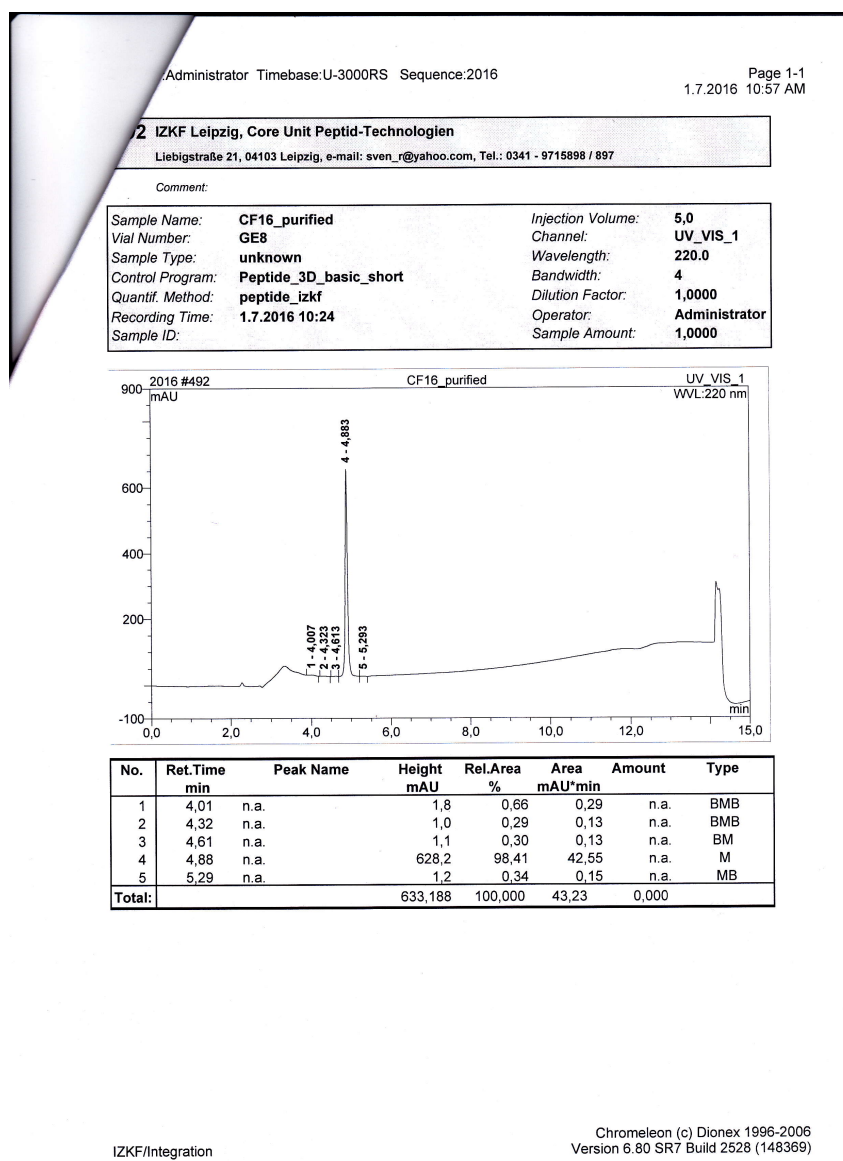


Figure V.1.43.: HPLC chromatogram of purified TcAFP5-KK, synthesized by S. Rothmund

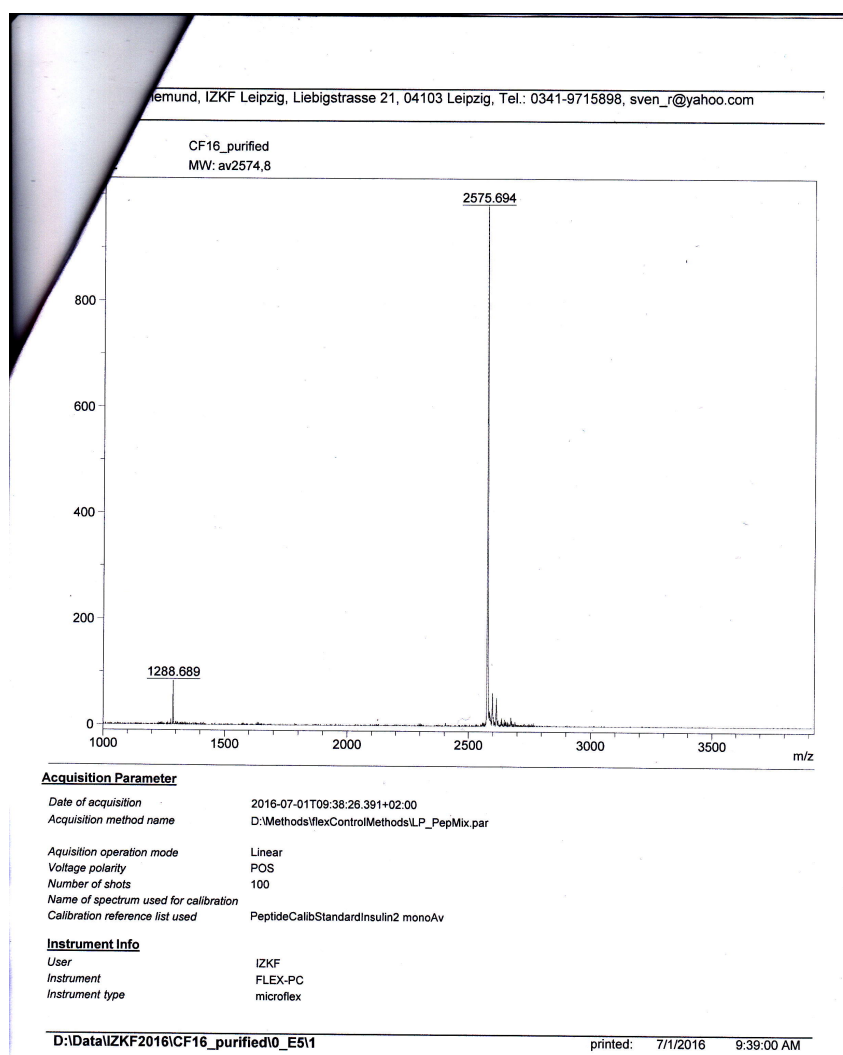
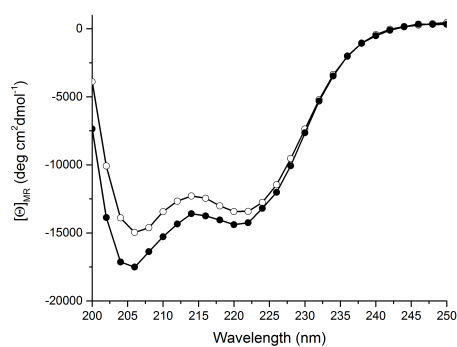
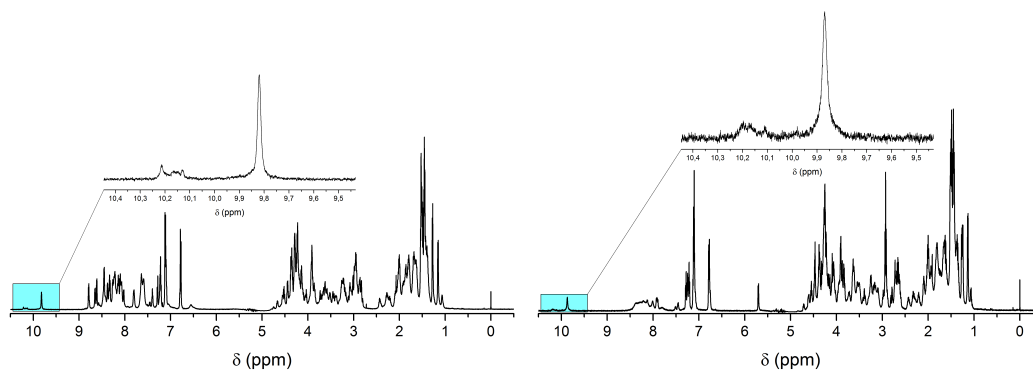


Figure V.1.44.:  $m/z$ :  $[M + H]^+$  calcd for  $[TcAFP5 - KK]^+$  2574.7, found 2575.7.



**Figure V.1.45.:** CD spectra of Tc10b (hollow circles) and AFP-Tc-5-KK (filled circles) at 274 K and pH 3.



**Figure V.1.46.:** <sup>1</sup>H-NMR spectra of AFP-Tc-5-KK at 274 K and pH 3 (left) or pH 7.9 (0.1 M NH<sub>4</sub>HCO<sub>3</sub> buffer) (right). Magnification of indole proton region.

**Table V.1.60.:**  $^1\text{H}$ -chemical shifts of AFP-Tc-5-KK at 274K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	8.425	4.603	2.864					
2'	Thr	8.792	4.240	4.240	1.28				
3'	Ala	8.613	4.285	1.449					
4'	Ser	8.65	4.356	4.071, 3.947					
5'	Asp	8.452	4.56	2.871					
6'	Ala	8.38	4.267	1.520					
1	Ala	8.266	4.285	1.520					
2	Lys	8.329	4.1334	1.79	1.672	1.388	2.959	7.637	
3	Tyr	8.342	4.213	3.191, 3.076		6.781	7.090		
4	Ala	8.455	4.276	1.511					
5	Ala	8.104	4.231	1.485					
6	Trp	8.024	4.373	3.431, 3.209		7.123	9.819, 7.262	7.211, 7.106	7.244
7	Thr	8.475	3.449	3.956	1.601				
8	Ala	8.212	4.142	1.449					
9	Asp	7.806	4.756	3.005, 2.862					
10	Gly	7.803	4.089, 3.636						
11	Gly	8.070	3.245, 1.947						
12	Pro		4.508	2.427, 2.022	2.082	3.681, 3.377			
13	Ser	8.086	4.45	3.899					
14	Ser	8.245	4.293	3.920, 3.727					
15	Gly	8.159	4.154, 3.885						
16	Arg	8.137	4.800	1.865, 1.785	1.674	3.236	7.398		
17	Pro		4.675	2.275, 1.808	1.999	3.851, 3.629			
18	Pro		2.958	1.401, 1.629	1.778, 1.690	4.120			
19	Pro		4.376	2.215, 1.920	1.863	3.430, 3.236			
20	Lys	8.329	4.1334	1.790	1.672	1.388	2.949	7.590	

## 1.3.6. AFP-Tc-7

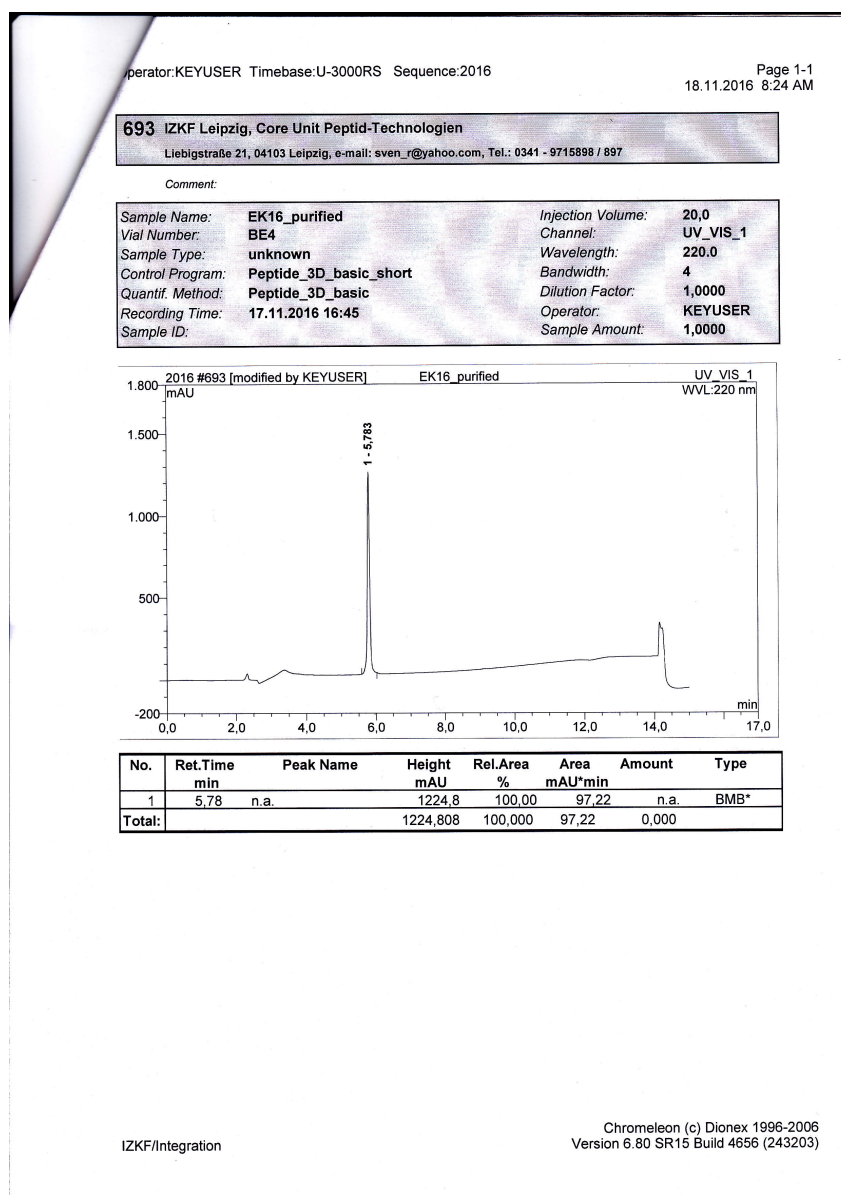


Figure V.1.47.: HPLC chromatogram of purified AFP-Tc-7, synthesized by S. Rothmund

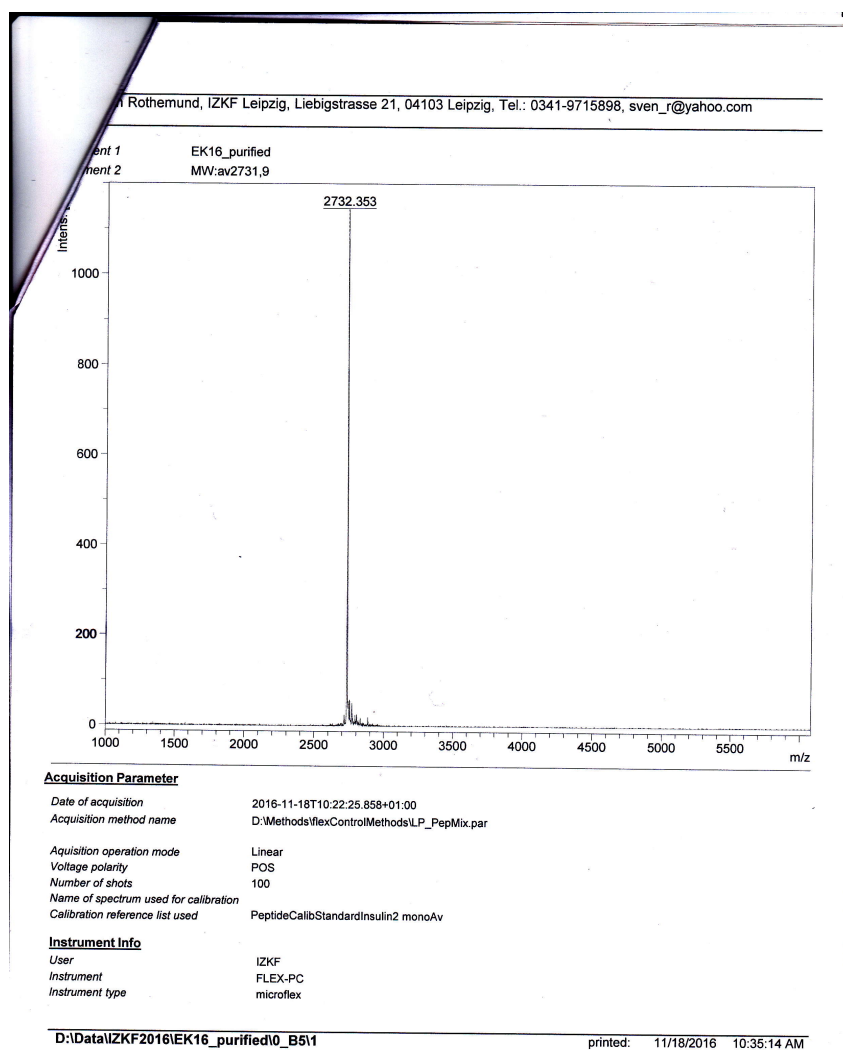


Figure V.1.48.:  $m/z$ :  $[M + H]^+$  calcd for  $[TcAFP6]^+$  2731.9, found 2732.4.

**Table V.1.61.:**  $^1\text{H}$ -chemical shifts of TcAFP6b at 274K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.388	3.194, 3.056					
2'	Thr	8.757	4.200	4.230	1.287				
3'	Ala	8.627	4.277	1.446					
4'	Ser	8.713	4.343	4.002, 3.904					
5'	Asp	8.449	4.607	3.003, 2.902					
6'	Ala	8.387	4.262	1.514					
7'	Ala	8.335	4.262	1.514					
8'	Ala	5.181	4.223	1.526					
9'	Ala	8.236	4.249	1.564					
1	Ala	8.205	4.336	1.551					
2	Ala	8.270	4.325	1.570					
3	Tyr	8.457	4.144	3.297, 3.166		7.128	6.846		
4	Thr	8.710	3.094	4.350	1.339				
5	Ala	8.127	4.23	1.520					
6	Trp	8.001	4.243	3.577, 3.190		7.078	9.636, 7.237	7.232, 7.123	7.017
7	Leu	8.568	3.336	1.903, 1.407	1.615	0.993, 0.877			
8	Ala	8.188	4.126	1.502					
9	Asp	7.566	4.780	3.056	2.870				
10	Gly	7.626	4.158, 3.507						
11	Gly	8.102	2.930, 1.032						
12	Pro		4.591	2.510, 2.051	2.124	3.753, 3.277			
13	Ser	7.826	4.454	3.934					
14	Ser	8.181	4.283	3.917, 3.636					
15	Gly	8.035	4.259, 3.840						
16	Arg	8.148	4.933	1.921, 1.849	1.735, 1.685	3.284	7.513		
17	Pro		4.715	2.298, 1.789	1.999	3.875, 3.673			
18	Pro		2.525	1.352, 0.441	1.727, 1.644	3.494			
19	Pro		4.340	2.233, 2.021	1.918, 1.857	3.125, 2.959			
20	Ser	8.306	4.362	3.910, 3.795					



**Table V.1.62.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-7 at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		$^{15}\text{N}$		$^{13}\text{C}$					
1'	Asp	n.d.		52.51	n.d.				
2'	Thr	115.67		64.20	69.36	21.86			
3'	Ala	125.58		54.69	18.26				
4'	Ser	116.79		60.43	n.d.				
5'	Asp	123.94		55.07	n.d.				
6'	Ala	124.61		54.79	17.86				
7'	Ala	122.62		54.79	17.86				
8'	Ala	122.95		54.75	17.86				
9'	Ala	121.62		54.93	17.86				
1	Ala	122.85		54.92	17.86				
2	Ala	121.95		54.76	17.86				
3	Tyr	123.26		62.91	n.d.		118.29	132.97	
4	Thr	117.01		67.23	68.65	21.89			
5	Ala	124.44		54.79	17.86				
6	Trp	121.54	131.62	61.47	n.d.		127.52	114.31	120.41
									123.87, 122.91
7	Leu	120.39		57.79	41.94		25.93, 23.30		
8	Ala	121.71		54.41	17.86				
9	Asp	116.03		n.d.	n.d.				
10	Gly	106.30		44.48					
11	Gly	112.33		44.06					
12	Pro	n.d.		65.36	31.81	27.21	50.87		
13	Ser	113.30		n.d.	n.d.				
14	Ser	117.26		n.d.	n.d.				
15	Gly	109.96		45.19					
16	Arg	120.13	114.77	n.d.	n.d.	n.d.	n.d.		
17	Pro	n.d.		60.96	30.73	27.01	50.67		
18	Pro	n.d.		59.99	29.00	26.61	49.76		
19	Pro	n.d.		62.58	31.54	26.91	49.96		
20	Ser	116.96		57.98					

**Table V.1.63.:**  $^1\text{H}$ -chemical shifts of TcAFP6b at 298 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp		4.166	3.252, 3.166					
2'	Thr	8.646	4.295	4.239	1.261				
3'	Ala	8.454	4.298	1.423					
4'	Ser	8.425	4.375	3.959, 3.888					
5'	Asp	8.319	4.666	2.937					
6'	Ala	8.292	4.231	1.481					
7'	Ala	8.233	4.224	1.481					
8'	Ala	8.051	4.250	1.484					
9'	Ala	8.112	4.319	1.529					
1	Ala	8.183	4.289	1.533					
2	Ala	8.166	4.250	1.536					
3	Tyr	8.365	4.180	3.249, 3.168		7.098	6.837		
4	Thr	8.515	3.910	4.323	1.331				
5	Ala	8.020	4.205	1.509					
6	Trp	7.937	4.245	3.555, 3.195		7.086	9.647, 7.2582	7.226, 7.112	7.094
7	Leu	8.453	3.381	1.850, 1.421	1.606	0.953, 0.858			
8	Ala	8.136	4.114	1.484					
9	Asp	7.580	4.779	3.038, 2.867					
10	Gly	7.613	4.124, 3.543						
11	Gly	8.044	2.972, 1.328						
12	Pro		4.549	2.483, 2.042	2.107	3.717, 3.247			
13	Ser	7.788	4.451	3.919					
14	Ser	8.109	4.362	3.897, 3.800					
15	Gly	7.972	4.228, 3.849						
16	Arg	8.058	4.875	1.931, 1.821	1.689		3.270	7.367	
17	Pro		4.697	2.279, 1.801	1.997	3.858, 3.649			
18	Pro		2.766	1.404, 0.661	1.760, 1.653	3.491			
19	Pro		4.350	2.227, 1.961	1.869	3.145, 2.985			
20	Ser	8.118	4.251	3.929, 3.667					

**Table V.1.64.:** <sup>1</sup>H-chemical shifts of TcAFP6b at 274 K and pH 7.9.

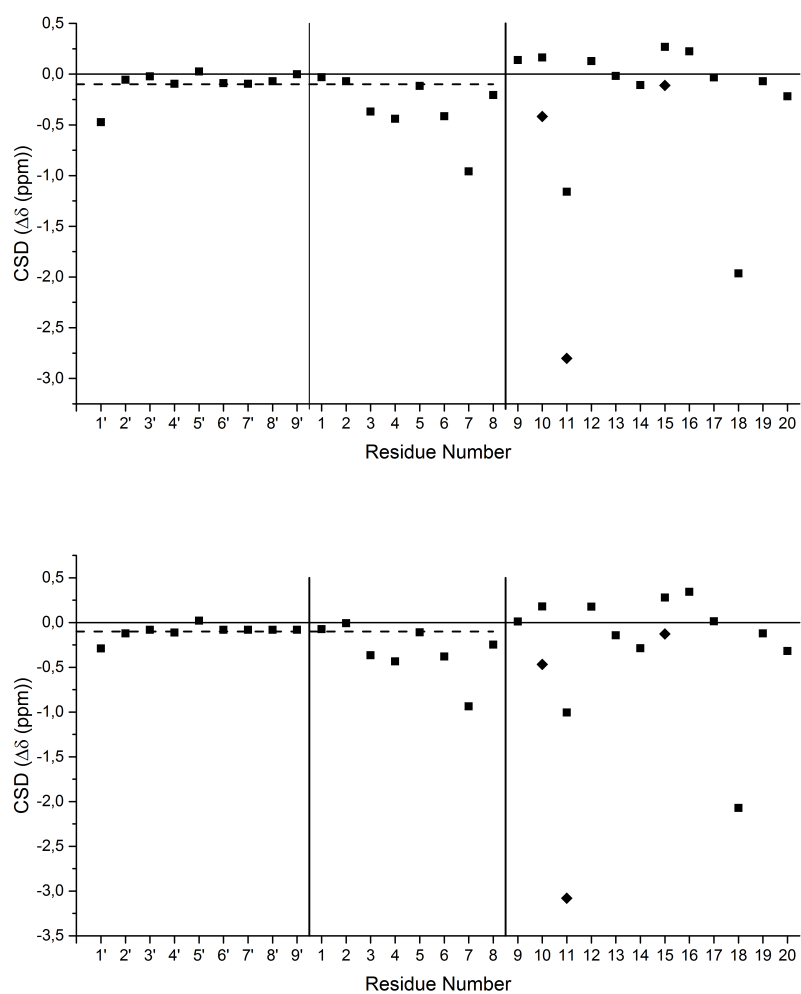
Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.710	4.372	2.659, 2.593					
2'	Thr	exch.	4.090	4.186	1.303				
3'	Ala	8.427	4.278	1.477					
4'	Ser	8.843	4.318	4.046, 3.936					
5'	Asp	8.546	4.507	2.855, 2.713					
6'	Ala	8.311	4.292	1.543					
7'	Ala	8.253	4.274	1.524					
8'	Ala	8.241	4.303	1.552					
9'	Ala	8.179	4.279	1.560					
1	Ala	8.224	4.250	1.567					
2	Ala	8.278	4.343	1.584					
3	Tyr	8.490	4.151	3.306, 3.190		7.151	6.872		
4	Thr	8.724	3.910	4.360	1.340				
5	Ala	8.161	4.235	1.536					
6	Trp	7.793	4.267	3.546, 3.181		7.019	9.675, 7.213	7.227, 7.151	7.001
7	Leu	8.535	3.339	1.949, 1.366	1.655	1.019, 0.868			
8	Ala	8.262	4.081	1.508					
9	Asp	7.543	4.656	2.891, 2.671					
10	Gly	7.581	4.159, 3.455						
11	Gly	8.302	3.098, 0.679						
12	Pro		4.658	2.536, 2.0827	2.173	3.847, 3.461			
13	Ser	8.249	4.153	3.858, 3.497					
14	Ser	7.865	4.135	3.782					
15	Gly	8.033	4.292, 3.800						
16	Arg	8.233	5.067	1.921, 1.829	1.665	3.297, 3.247	7.677		
17	Pro		4.771	2.350, 1.790	1.992	3.868, 3.685			
18	Pro		2.384	1.365, 0.296	1.719, 1.684	3.517			
19	Pro		4.300	2.220, 2.079	1.993, 1.828	3.104, 2.962			
20	Ser	7.759	4.502	3.934					

**Table V.1.65.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-7 at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)								
		$^{15}\text{N}$			$^{13}\text{C}$					
1'	Asp	n.d.		n.d.	n.d.					
2'	Thr	n.d.		65.38	69.04	21.94				
3'	Ala	124.80		54.84	18.15					
4'	Ser	117.97		61.28	62.17					
5'	Asp	125.71		56.80	n.d.					
6'	Ala	124.53		54.89	17.84					
7'	Ala	122.86		54.84	17.84					
8'	Ala	123.33		54.89	17.84					
9'	Ala	122.75		54.84	17.84					
1	Ala	121.57		55.00	17.84					
2	Ala	122.10		55.00	17.84					
3	Tyr	123.51		62.98	38.17		118.37	133.04		
4	Thr	117.09		67.13	68.86	21.94				
5	Ala	124.65		54.95	17.84					
6	Trp	121.43	131.89	62.16	27.72		127.60	114.31	120.47	123.63, 122.96
7	Leu	119.79		57.79	41.92		26.14, 23.05			
8	Ala	123.80		54.73	17.84					
9	Asp	117.71		54.11	n.d.					
10	Gly	105.78		43.98						
11	Gly	113.98		43.08						
12	Pro			64.94	31.71	27.35	51.12			
13	Ser	117.59		n.d.	65.05					
14	Ser	120.72		n.d.	64.79					
15	Gly	109.91		45.31						
16	Arg	119.76	n.d.	n.d.	n.d.	n.d.	n.d.			
17	Pro	n.d.		61.28	30.35	27.17	50.50			
18	Pro	n.d.		60.08	28.78	n.d.	49.47			
19	Pro	n.d.		62.71	31.57	26.89	49.63			
20	Ser	113.10		n.d.	63.17					

**Table V.1.66.:** <sup>1</sup>H-chemical shifts of TcAFP6b at 298 K and pH 7.9. Shifts marked with an asterisk are averaged.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.603	4.351	2.799, 2.623					
2'	Thr	exch.	4.229	4.229	1.270				
3'	Ala	exch.	4.239*	1.511*					
4'	Ser	exch.	4.359	3.983, 3.896					
5'	Asp	exch.	4.660	2.850, 2.653					
6'	Ala	exch.	4.239*	1.511*					
7'	Ala	exch.	4.239*	1.511*					
8'	Ala	exch.	4.239*	1.511*					
9'	Ala	exch.	4.239*	1.511*					
1	Ala	8.133	4.247	1.521					
2	Ala	8.110	4.313	1.535					
3	Tyr	8.389	4.185	3.260, 3.180		7.125	6.861		
4	Thr	8.535	3.916	4.329	1.338				
5	Ala	8.056	4.211	1.513					
6	Trp	7.918	4.281	3.536, 3.198		7.037	9.686, 7.215	7.227, 7.119	7.084
7	Leu	8.424	3.403	1.880, 1.382	1.628	0.971, 0.846			
8	Ala	8.185	4.074	1.496					
9	Asp	7.533	4.653	2.847, 2.650					
10	Gly	7.596	4.141, 3.494						
11	Gly	8.225	3.126, 1.051						
12	Pro		4.598	2.510, 1.053	2.150	3.798, 3.418			
13	Ser	exch.	4.327	3.926					
14	Ser	exch.	4.183	3.879, 3.550					
15	Gly	7.923	4.240, 3.833						
16	Arg	8.148	4.994	1.861, 1.798	1.664	3.269	exch.		
17	Pro		4.744	2.340, 1.809	1.998	3.852, 3.677			
18	Pro		2.657	1.403, 0.559	1.761, 1.685	3.522			
19	Pro		4.300	2.209, 2.022	1.939, 1.840	3.129, 2.996			
20	Ser	7.726	4.153	3.782					



**Figure V.1.49.:**  $H\alpha$  CSD plot of AFP-Tc-7 at pH 3 (top) and pH 7.9 (bottom) at 298 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. The dashed line at  $y = -0.1$  ppm represents the helical limit. Solid line between residue 9' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.

**Table V.1.67.:** CSDs for selected protons in the cage structure representing the cage fold of TcAFP6b at 298 K and 274 K and at pH 3 and 7.9.

<b>Cage</b>	$\Delta\delta$ (ppm)						$\Sigma$	
<b>upfield</b>								
pH	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$		
3	-0.959	-2.802	-1.965	-1.249	-0.455	-0.615		<b>-8.045</b>
7.9	-0.937	-3.079	-2.073	-1.351	-0.471	-0.604		<b>-8.515</b>
<b>downfield</b>								
	P12 $\alpha$	P12 $\beta$	R16 $\alpha$					
3	0.129	0.193	0.225					<b>0.547</b>
7.9	0.178	0.220	0.344					<b>0.742</b>
<b>cage-helix</b> multicolumn8c								
	A2 $\alpha$	Y3 $\alpha$	T4 $\alpha$	A5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$	
3	-0.070	-0.370	-0.440	-0.115	-0.415	-0.959	-0.206	<b>-2.575</b>
7.9	-0.007	-0.365	-0.434	-0.109	-0.379	-0.937	-0.246	<b>-2.476</b>

**Table V.1.68.:** CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP6b at 274 K and at pH 3 and 7.9.

<b>Helix</b>	$\Delta\delta$ (ppm).					$\Sigma$
	T2' $\alpha$	A3' $\alpha$	S'4 $\alpha$	D5' $\alpha$	A6' $\alpha$	
pH 3, 298 K	-0.055	-0.022	-0.095	0.026	-0.089	<b>-0.235</b>
pH 7.9, 298 K	-0.121	-0.081	-0.111	0.020	-0.081	<b>-0.374</b>

**Table V.1.69.:** CSDs for the  $^{13}\text{C}\alpha$  atoms in the helix representing the helix fold of AFP-Tc-7 at 274 K and at pH 3 and 7.9. The  $\overline{\Delta\delta}$  and the  $\chi_F$  for the helix, chimera and helical extension are calculated. The  $\chi_F$  was determined using the CSD of AFP-Tc-7 at pH 7.9 and 274 K as reference (marked with an asterisk).

<b>TcAFP6</b>		
	<b>pH 3</b>	<b>pH 8</b>
	$\delta$ (ppm)	$\delta$ (ppm)
<b>Helical Extension</b>	1.10	2.28
	2.19	2.34
	2.13	2.98
	0.97	2.70
	2.29	2.39
	2.29	2.34
	2.25	2.39
	2.43	2.34
<b>Chimera Helix</b>	2.42	2.50
	2.26	2.50
	4.31	4.38
	4.13	4.03
	2.29	2.90
	3.67	4.36
	2.09	2.09
	1.91	2.23
<b>Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.42	2.80*
$\chi_F^{helix}$ (%)	87	99.5
<b>Chimera Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.89	3.12*
$\chi_F^{chimera}$ (%)	92	99.5
<b>Helical Extension <math>\overline{\Delta\delta}</math> (ppm)</b>	1.96	2.47*
$\chi_F^{exten.}$ (%)	79	99.5



1.3.7. AFP-Tc-6

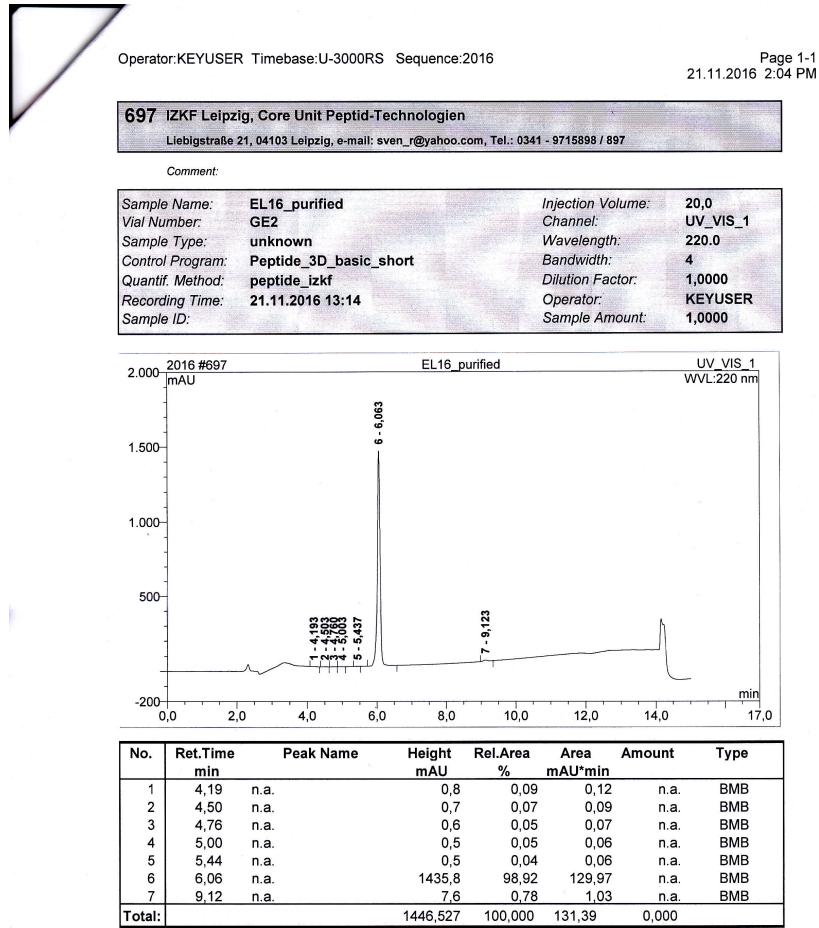


Figure V.1.50.: HPLC chromatogram of purified AFP-Tc-6, synthesized by S. Rothmund

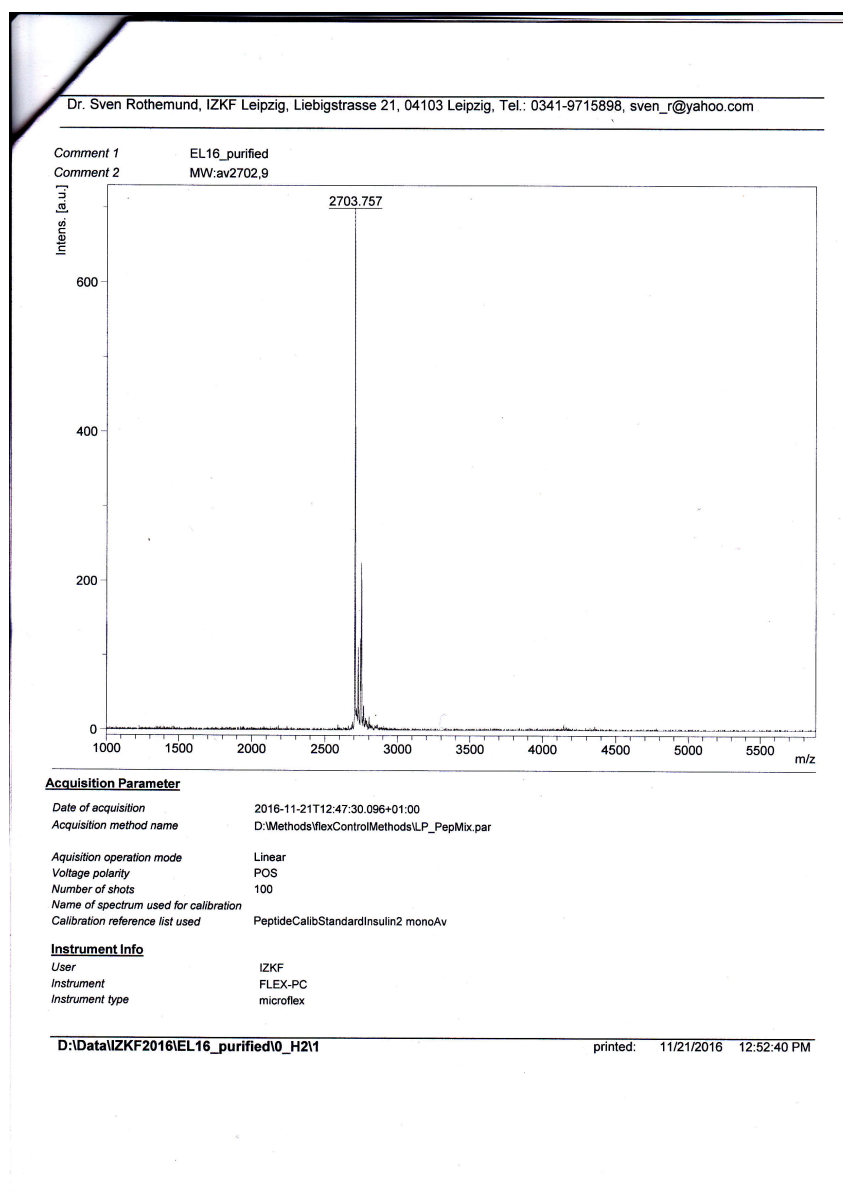


Figure V.1.51.:  $m/z$ :  $[M + H]^+$  calcd for  $[TcAFP7]^+$  2702.9, found 2703.8.

**Table V.1.70.:** <sup>1</sup>H-chemical shifts of AFP-Tc-6 at 274 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.385	3.189, 3.048					
2'	Thr	8.760	4.192	4.226	1.290				
3'	Ala	8.609	4.276	1.447					
4'	Ser	8.702	4.346	4.010, 3.907					
5'	Asp	8.454	4.605	2.995, 2.904					
6'	Ala	8.348	4.239	1.508					
7'	Ala	8.263	4.276	1.486					
8'	Ala	8.150	4.305	1.584					
1	Ala	8.325	4.246	1.562					
2	Ala	8.141	4.236	1.551					
3	Tyr	8.343	4.128	3.229, 3.176		7.063	6.827		
4	Leu	8.227	4.113	2.071, 1.640	1.999	1.018 (both)			
5	Thr	8.347	3.930	4.245	1.247				
6	Trp	7.979	4.178	3.593, 3.141		7.045	9.663, 7.237	7.235, 7.129	7.042
7	Leu	8.590	3.341	1.872, 1.408	1.618	0.973, 0.882			
8	Ala	8.294	4.112	1.513					
9	Asp	7.824	4.7035	3.111, 2.900					
10	Gly	7.649	4.170, 3.501						
11	Gly	8.2119	2.9068, 1.0298						
12	Pro		4.584	2.504, 2.052	2.122	3.746, 3.269			
13	Ser	8.185	4.222	3.903, 3.617					
14	Ser	7.821	4.463	3.931					
15	Gly	8.036	4.256, 3.834						
16	Arg	8.130	4.950	1.902, 1.862	1.747, 1.673	3.278	7.558		
17	Pro		4.719	2.300, 1.792	1.998	3.877, 3.670			
18	Pro		2.547	1.340, 0.396	1.711, 1.606	3.472			
19	Pro		4.351	2.241, 1.999	1.912, 1.860	3.162, 2.977			
20	Ser	8.310	4.353	3.903, 3.796					



**Table V.1.72.:** <sup>1</sup>H-chemical shifts of AFP-Tc-6 at 298 K and pH 3.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	exch.	4.391	3.015					
2'	Thr	8.651	4.298	4.252	1.252				
3'	Ala	8.448	4.301	1.441					
4'	Ser	8.421	4.369	3.975, 3.884					
5'	Asp	8.317	4.658	2.928					
6'	Ala	8.256	4.222	1.467					
7'	Ala	8.034	4.288	1.522					
8'	Ala	8.191	4.094	1.499					
1	Ala	8.105	4.223	1.521					
2	Ala	8.263	4.223	1.535					
3	Tyr	8.222	4.131	3.201		7.025	6.817		
4	Leu	8.220	4.112	2.013, 1.621	1.949	1.000			
5	Thr	8.244	3.937	4.233	1.237				
6	Trp	7.911	4.180	3.573, 3.163		7.060	9.658, 7.251	7.215, 7.119	7.103
7	Leu	8.512	3.361	1.800, 1.439	1.595	0.934, 0.863			
8	Ala	8.175	4.237	1.452					
9	Asp	7.781	4.711	3.080, 2.867					
10	Gly	7.616	4.150, 3.535						
11	Gly	8.137	2.921, 1.282						
12	Pro		4.547	2.489, 2.031	2.099	3.705, 3.216			
13	Ser	7.769	4.445	3.922					
14	Ser	8.109	4.354	3.891, 3.800					
15	Gly	7.970	4.225, 3.831						
16	Arg	8.041	4.8929	1.9268, 1.8282	1.6992	3.2695	7.3926		
17	Pro		4.694	2.280, 1.800	2.000	3.857, 3.651			
18	Pro		2.753	1.364, 0.587	1.724, 1.608	3.467			
19	Pro		4.338	2.221, 1.949	1.865	3.155, 2.980			
20	Ser	8.117	4.251	3.922, 3.656					

**Table V.1.73.:**  $^1\text{H}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.709	4.376		2.661, 2.599				
2'	Thr	exch.	4.112	4.194	1.302				
3'	Ala	8.323	4.268	1.552					
4'	Ser	8.805	4.329	4.045, 3.937					
5'	Asp	8.529	4.524	2.866, 2.733					
6'	Ala	8.308	4.270	1.534					
7'	Ala	8.213	4.283	1.509					
8'	Ala	8.180	4.296	1.600					
1	Ala	8.098	4.254	1.552					
2	Ala	8.323	4.268	1.548					
3	Tyr	8.394	4.133	3.241		7.117	6.870		
4	Leu	8.157	4.137	2.067	1.985	1.017			
5	Thr	8.349	3.966	4.249	1.250				
6	Trp	7.965	4.210	3.567, 3.139		6.984	9.715, 7.233	7.198, 7.162	7.030
7	Leu	8.530	3.375	1.922, 1.364	1.655	1.010, 0.880			
8	Ala	8.322	4.064	1.559					
9	Asp	7.943	4.570	2.866, 2.700					
10	Gly	7.607	4.169, 3.452						
11	Gly	8.454	3.109, 0.676						
12	Pro		4.658	2.540, 2.083	2.183	3.846, 3.492			
13	Ser	8.264	4.126	3.844, 3.467					
14	Ser	7.880	4.129	3.775					
15	Gly	8.029	4.301, 3.800						
16	Arg	8.223	5.096	1.927, 8.175	1.652	3.292, 3.212	7.757		
17	Pro		4.781	2.366, 1.786	1.997	3.875, 3.683			
18	Pro		2.403	1.346, 0.254	1.710, 1.646	3.523			
19	Pro		4.308	2.228, 2.069	1.982, 1.833	3.142, 2.971			
20	Ser	7.753	4.501	3.934					

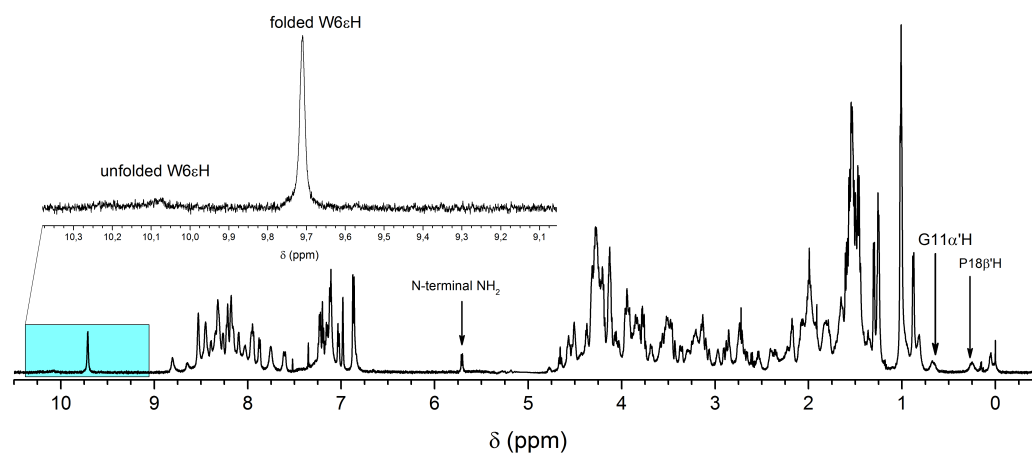
**Table V.1.74.:**  $^{15}\text{N}$  and  $^{13}\text{C}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 7.9.

Residue		chemical shift $\delta$ (ppm)									
		$^{15}\text{N}$		$^{13}\text{C}$							
1'	Asp	n.d.		n.d.	n.d.						
2'	Thr	n.d.		65.25	69.06	21.93					
3'	Ala	124.91		54.73	17.83						
4'	Ser	117.73		61.18	n.d.						
5'	Asp	125.38		56.65	n.d.						
6'	Ala	124.53		54.73	17.83						
7'	Ala	122.95		54.81	17.83						
8'	Ala	122.83		54.85	17.83						
1	Ala	121.46		54.76	17.83						
2	Ala	125.06		54.73	17.83						
3	Tyr	122.75		62.69	n.d.		118.39	132.85			
4	Leu	119.05		59.46	41.21	n.d.	23.21				
5	Thr	118.73		66.59	68.73	22.02					
6	Trp	123.68	131.83	62.61	27.63		127.54	114.28	123.85, 122.92	120.61	
7	Leu	119.02		57.85	41.89	n.d.	25.70, 22.97				
8	Ala	121.10		54.85	17.83						
9	Asp	118.17		54.54	n.d.						
10	Gly	105.95		43.91							
11	Gly	114.63		40.99							
12	Pro	n.d.		n.d.	31.86	27.38	50.92				
13	Ser	117.65		n.d.	64.79						
14	Ser	120.75		n.d.	64.53						
15	Gly	109.91		45.37							
16	Arg	119.58	n.d.	n.d.	n.d.	n.d.	n.d.				
17	Pro	n.d.		n.d.	30.59	27.14	50.75				
18	Pro	n.d.		60.11	28.55	n.d.	49.28				
19	Pro	n.d.		n.d.	31.62	27.08	49.63				
20	Ser	113.10		59.07	62.90						

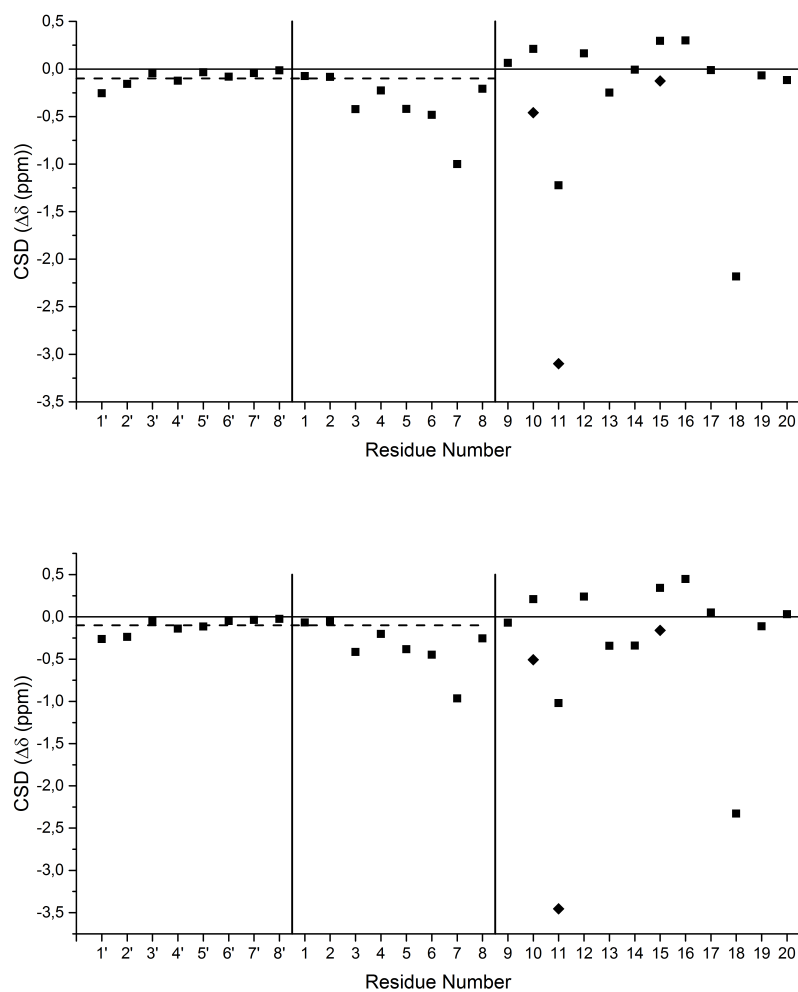
**Table V.1.75.:**  $^1\text{H}$ -chemical shifts of AFP-Tc-6 at 298 K and pH 7.9. Shifts marked with an asterix are averaged.

Residue		chemical shift $\delta$ (ppm)							
		NH	$\alpha$	$\beta$	$\gamma$	$\delta$	$\epsilon$	$\zeta$	$\eta$
1'	Asp	5.605	4.359	2.635, 2.589					
2'	Thr	exch.	4.241	4.241	1.268				
3'	Ala	exch.	4.227*	1.508*					
4'	Ser	exch.	4.370	3.967, 3.905					
5'	Asp	exch.	4.591	2.812, 2.691					
6'	Ala	exch.	4.227*	1.508*					
7'	Ala	exch.	4.227*	1.508*					
8'	Ala	exch.	4.227*	1.508*					
1	Ala	exch.	4.227*	1.508*					
2	Ala	exch.	4.227*	1.508*					
3	Tyr	8.270	4.165	3.208		7.084	6.859		
4	Leu	8.156	4.117	2.012, 1.616	1.943	0.996			
5	Thr	8.239	3.974	4.252	1.271				
6	Trp	7.893	4.243	3.554, 3.169	7.011	9.713, 4.234	7.214, 7.130	7.117	
7	Leu	8.452	3.410	1.852, 1.400	1.622	0.956, 0.857			
8	Ala	8.185	4.073	1.531					
9	Asp	7.853	4.592	2.829, 2.676					
10	Gly	7.606	4.136, 3.500						
11	Gly	8.330	3.133, 1.038						
12	Pro		4.589	2.509, 2.062	2.151	3.788, 3.407			
13	Ser	exch.	4.245	3.982					
14	Ser	exch.	4.245	3.982, 3.866					
15	Gly	7.922	4.225, 3.831						
16	Arg	8.138	5.004	1.880, 1.826	1.647	3.285, 3.231	exch.		
17	Pro		4.741	2.333, 1.989	1.799	3.849, 3.661			
18	Pro		2.666	1.383, 0.511	1.746, 1.645	3.492			
19	Pro		4.302	2.203, 2.010	1.926, 1.857	3.156, 3.012			
20	Ser	7.742	4.157	3.771					

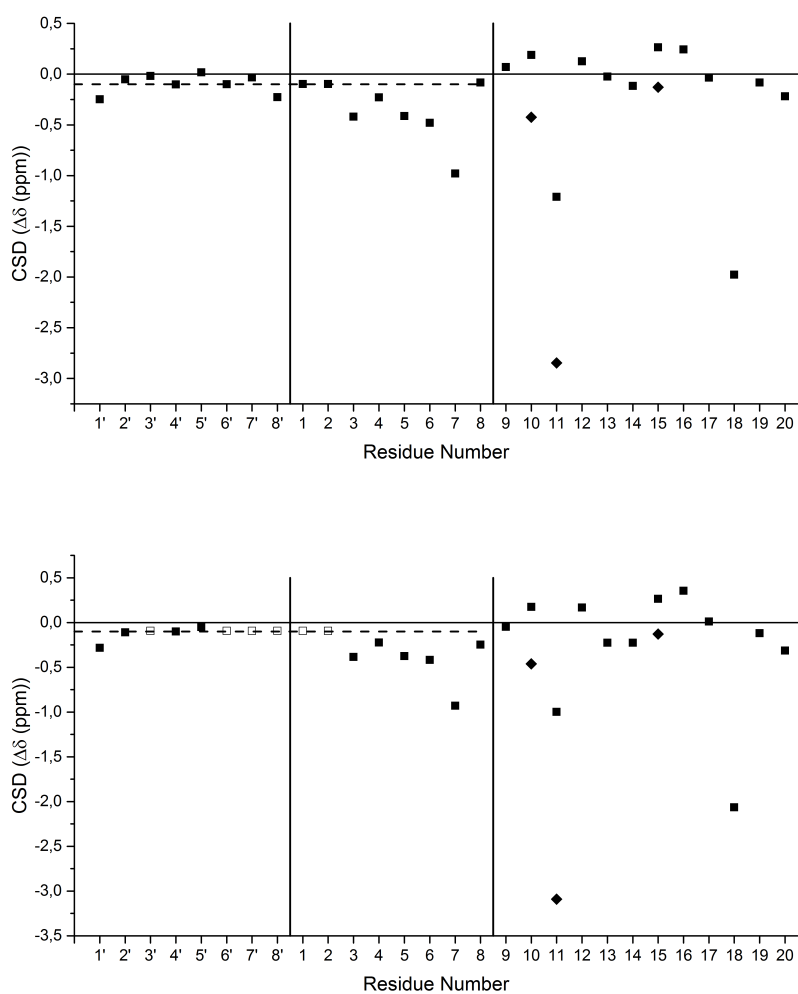




**Figure V.1.52.:**  $^1\text{H-NMR}$  spectrum of AFP-Tc-6 at 274 K and pH 7.9 with enlargement of the indole proton region.



**Figure V.1.53.:**  $H\alpha$  CSD plot of AFP-Tc-6 at pH 3 (top) and pH 7.9 (bottom) at 274 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. The dashed line at  $y = -0.1$  ppm represents the helical limit. Solid line between residue 8' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.



**Figure V.1.54.:**  $H\alpha$  CSD plot of AFP-Tc-7 at pH 3 (top) and pH 7.9 (bottom) at 298 K.  $H\alpha$  as black squares and  $H\alpha'$  as black diamonds. Averaged  $H\alpha$  shifts for certain Ala are represented as non-filled squares. The dashed line at  $y=-0.1$  ppm represents the helical limit. Solid line between residue 8' and 1 represents the junction of the N-terminal extension and the solid line between residue 8 and 9 the conjunction of the helix to the C-terminal cage fold.

**Table V.1.76.:** CSDs for selected protons in the cage structure representing the cage fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9.

Cage	$\Delta\delta$ (ppm)						$\Sigma$
	upfield						
	L7 $\alpha$	G11 $\alpha'$	P18 $\alpha$	P18 $\beta'$	P19 $\delta$	P19 $\delta'$	
pH 3, 274 K	-1.000	-3.100	-2.183	-1.514	-0.438	-0.623	<b>-8.857</b>
pH 3, 298 K	-0.980	-2.848	-1.977	-1.323	-0.445	-0.621	<b>-8.193</b>
pH 7.9, 274 K	-0.966	-3.454	-2.328	-1.656	-0.458	-0.629	<b>-9.490</b>
pH 7.9, 298 K	-0.930	-3.092	-2.064	-1.399	-0.444	-0.589	<b>-8.517</b>
downfield							
	P12 $\alpha$	P12 $\beta$	R16 $\alpha$				
pH 3, 274 K	0.164	0.214	0.300				<b>0.678</b>
pH 3, 298 K	0.127	0.199	0.243				<b>0.568</b>
pH 7.9, 274 K	0.238	0.250	0.446				<b>0.934</b>
pH 7.9, 298 K	0.169	0.250	0.354				<b>0.773</b>

**Table V.1.77.:** CSDs for selected protons in the helical region representing the helix fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9.

Helix	$\Delta\delta$ (ppm).							$\Sigma$
	A2 $\alpha$	Y3 $\alpha$	L4 $\alpha$	T5 $\alpha$	W6 $\alpha$	L7 $\alpha$	A8 $\alpha$	
pH 3, 274 K	-0.084	-0.422	-0.227	-0.420	-0.482	-1.000	-0.208	<b>-2.844</b>
pH 3, 298 K	-0.097	-0.420	-0.228	-0.413	-0.480	-0.980	-0.083	<b>-2.701</b>
pH 7.9, 274 K	-0.052	-0.417	-0.203	-0.384	-0.450	-0.966	-0.256	<b>-2.726</b>
pH 7.9, 298 K	-0.093	-0.385	-0.223	-0.376	-0.417	-0.930	-0.247	<b>-2.671</b>

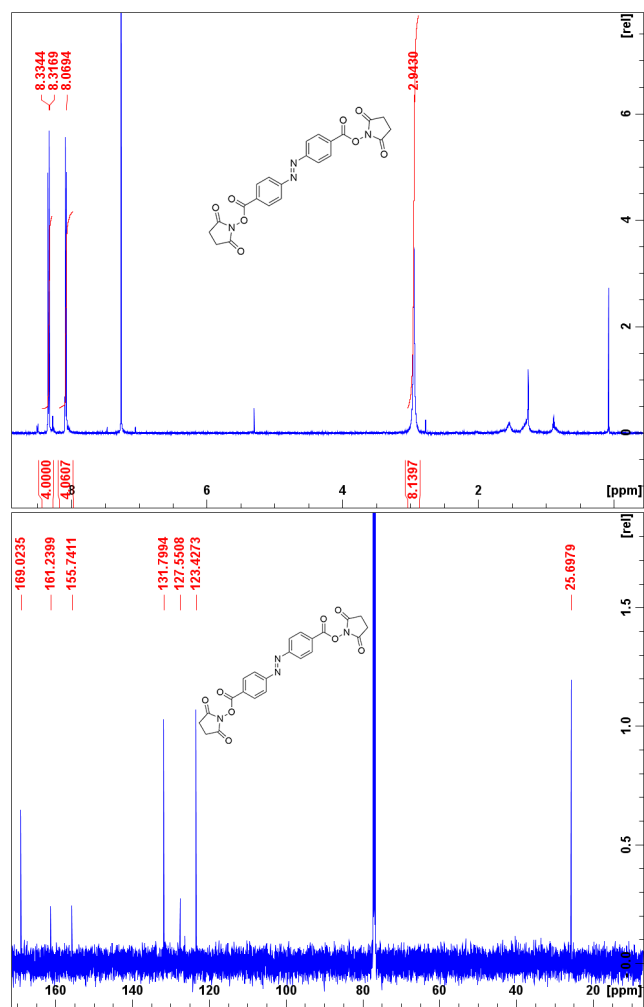
**Table V.1.78.:** CSDs for selected protons in the helical extension representing the N-terminal helix fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9. Shifts marked with an asterisk are averaged.

Helix	$\Delta\delta$ (ppm).					$\Sigma$
	T2' $\alpha$	A3' $\alpha$	S'4 $\alpha$	D5' $\alpha$	A6' $\alpha$	
pH 3, 274 K	-0.158	-0.044	-0.124	-0.035	-0.081	<b>-0.441</b>
pH 3, 298 K	-0.052	-0.019	-0.101	0.018	-0.098	<b>-0.252</b>
pH 7.9, 274 K	-0.238	-0.052	-0.141	-0.116	-0.050	<b>-0.597</b>
pH 7.9, 298 K	-0.109	-0.093*	-0.100	-0.049	-0.093*	<b>-0.445</b>

**Table V.1.79.:** CSDs for the  $^{13}\text{C}\alpha$  atoms in the helix representing the helix fold of AFP-Tc-6 at 274 K and at pH 3 and 7.9. The  $\overline{\Delta\delta}$  and the  $\chi_F$  for the helix, chimera and helical extension are calculated. The  $\chi_F$  was determined using the CSD of AFP-Tc-7 at pH 7.9 and 274 K as reference.

<b>TcAFP7</b>		
	<b>pH 3</b>	<b>pH 8</b>
	$\delta$ (ppm)	$\delta$ (ppm)
<b>Helical Extension</b>	0.92	2.15
	1.86	2.23
	0.67	2.88
	0.52	2.55
	1.92	2.23
	1.86	2.31
	1.79	2.35
<b>Chimera Helix</b>	2.24	2.26
	2.05	2.23
	4.74	4.09
	2.64	3.76
	3.39	3.49
	4.90	4.60
	2.31	2.15
	1.66	2.24
<b>Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.23	2.77
$\chi_F^{helix}$ (%)	80	99
<b>Chimera Helix <math>\overline{\Delta\delta}</math> (ppm)</b>	2.99	3.10
$\chi_F^{chimera}$ (%)	96	99
<b>Helical Extension <math>\overline{\Delta\delta}</math> (ppm)</b>	1.36	2.38
$\chi_F^{exten.}$ (%)	55	96

## 1.4. 4,4'-Bis(2,5-Dioxopyrrolidin-1-yl)Azobenzene

Figure V.1.55.:  $^1\text{H}$  and  $^{13}\text{C}$  NMR of 4,4'-bis(2,5-dioxopyrrolidin-1-yl)azobenzene

## 2. UV/vis-Spectra and Concentrations and Helical Content (CD)

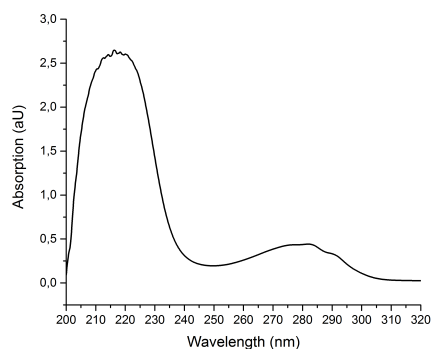
For determination of the concentration for the CD samples, UV/vis spectra were used. For peptides with the aromatic amino acids Trp and Tyr in the Trp-cage fold, an extinction coefficient of  $\epsilon_{278nm} = 6770 \text{ cm}^2 \text{ mmol}^{-1}$  was used. Extinction coefficients for peptides with aromatic residues but not in a Trp-cage fold were calculated using ExPASy's ProtParam.<sup>103</sup> For peptides without aromatic residues, the extinction coefficient of the backbone at 214 nm was calculated using a method by KUIPERS<sup>121</sup> Table V.2.1.

**Table V.2.1.:** Extinction coefficient, corresponding absorption and resulting concentrations of the solutions for CD spectroscopy.

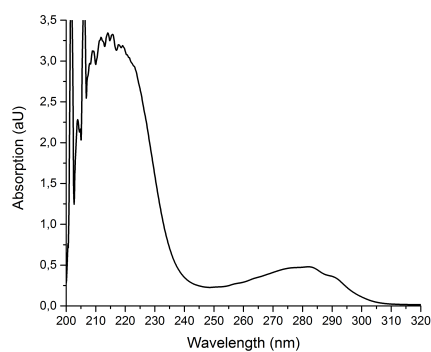
Sequence	Tc10b	TcKKA	AFP-Tc-4	AFP-Tc-5	AFP-Tc-5-KK
$\epsilon$	6770	6770	6770	6770	6770
278 nm	0.435	0.471	0.470	4.356	0.316
214 nm					
Conc. mol/m <sup>3</sup>	0.064	0.070	0.070	0.643	0.047
Sequence	AFP-Tc-7	AFP-Tc-6	15er	15erYW	
$\epsilon$	6770	6770	39320	6990	
278 nm	0.316	0.253	0.344	1.554	
214 nm				0.611	
Conc. mol/m <sup>3</sup>	0.037	0.051	0.040	0.090	

**Table V.2.2.:** Observed and theoretical values of the mean residue ellipticity at 222 nm of AFP-Tc-7 to AFP-Tc-4 and the resulting helical content.

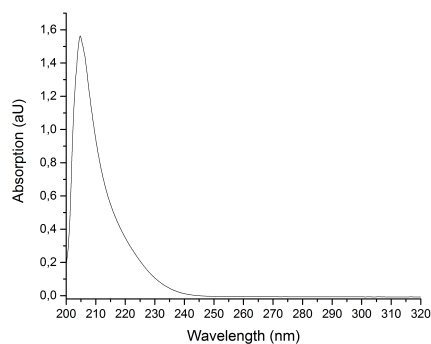
<b>Chimera</b>	AFP-Tc-7	AFP-Tc-6	AFP-Tc-5	AFP-Tc-4
$[\theta]_{MR}(obs) \text{ (deg cm}^2 \text{ dmol}^{-1}\text{)}$	-26,054	-24,696	-17,240	-14,582
$[\theta]_{MR}(ref) \text{ (deg cm}^2 \text{ dmol}^{-1}\text{)}$	-42,241	-42,188	-42,067	-42,000
<b>Helical Content (%)</b>	99	96	71	61



**Figure V.2.1.:** UV/vis spectrum of Tc10b at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.

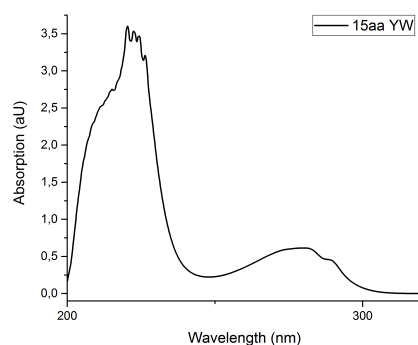


**Figure V.2.2.:** UV/vis spectrum of TcKKA at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.

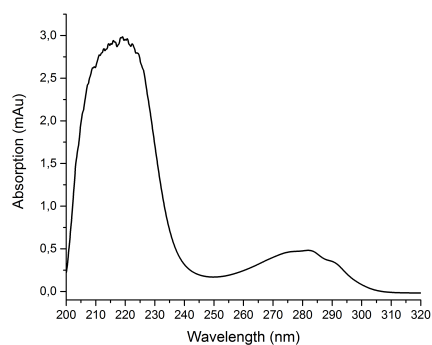


**Figure V.2.3.:** UV/vis spectrum of 15erYW reference peptide at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.

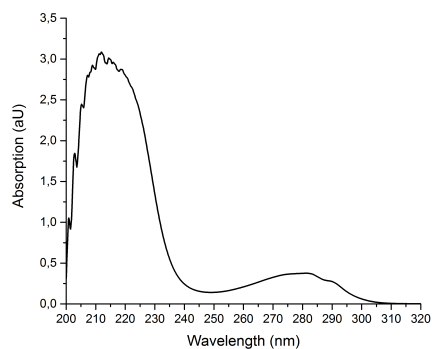




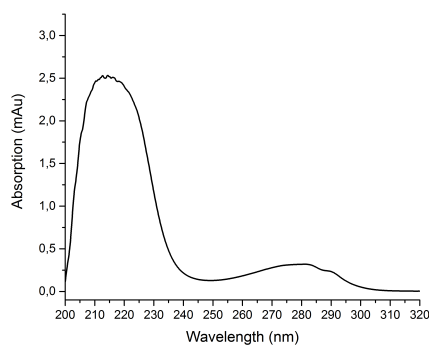
**Figure V.2.4.:** UV/vis spectrum of 15erYW reference peptide at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.



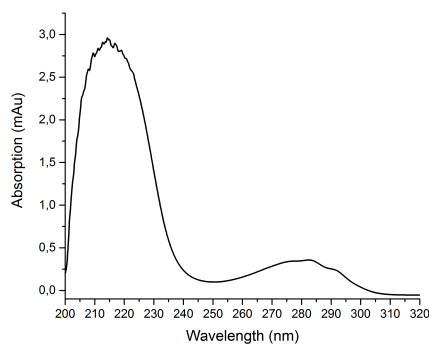
**Figure V.2.5.:** UV/vis spectrum of AFP-Tc-4 at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.



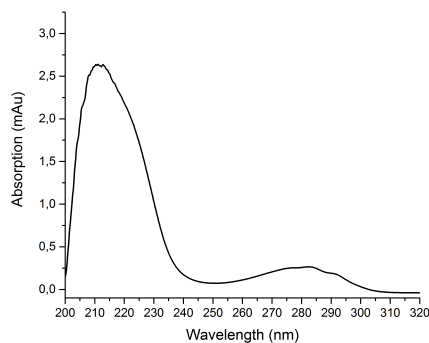
**Figure V.2.6.:** UV/vis spectrum of AFP-Tc-5 at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.



**Figure V.2.7.:** UV/vis spectrum of AFP-Tc-5-KK at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.



**Figure V.2.8.:** UV/vis spectrum of AFP-Tc-6 at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.



**Figure V.2.9.:** UV/vis spectrum of AFP-Tc-7 at 298 K and pH 7.9 in 0.1 M  $\text{NH}_4\text{HCO}_3$  buffer. The spectrum was measured to 600 nm, displayed only to 320 nm.

### 3. Syntheses

#### 3.1. Tc10b

Table V.3.1.: Tc10b

Sequence N to C	DAYAQWLKDGGPSSGRPPPS			
MW peptide (g/mol)	2086.25			
Peptide amount (g) / (mmol)	0.5	0.240		
Resin Type	Wang			
Resin capacity (mmol/g)	0.85			
Resin amount (g)	0.282			
Capping: Pyridine (mL)	1.133			
Aceticanhydride (mL)	1.133			
	C to N	Double	Single	Annotation
4.9 eq. HATU (g)		0.893	0.446	
10 eq. DIPEA (mL)		0.835	0.417	
5 eq. amino acids (g)	S	0.919	0.459	single
	P	0.809	0.404	single
	P	0.809	0.404	single
	P	0.809	0.404	single
	R	1.555	0.777	single/ext.
	G	0.713	0.356	double
	S	0.919	0.459	single
	S	0.919	0.459	single
	P	0.809	0.404	single
	G	0.713	0.356	double/ext.
	DG	1.483	0.741	single
	K	1.123	0.561	single
	L	0.847	0.423	double
	W	1.262	0.631	single
	Q	1.464	0.732	single
	A	0.746	0.373	double
	Y	1.101	0.551	double
	A	0.746	0.373	double
	D	0.986	0.493	single

Table V.3.2.: Tc10b-S20K

Sequence N to C	DAYAQWLKDGGPSSGRPPPK			
MW peptide (g/mol)	2127.06			
Peptide amount (g) / (mmol)	0.1	0.047		
Resin Type	Wang			
Resin capacity (mmol/g)	0.85			
Resin amount (g)	0.055			
Capping: Pyridine (mL)	0.222			
Aceticanhydride (mL)				
	<b>C to N</b>	<b>Double</b>	<b>Single</b>	<b>Annotation</b>
4.9 eq. HATU (g)		0.175	0.088	
10 eq. DIPEA (mL)		0.164	0.082	
5 eq. amino acids (g)	K	0.220	0.110	single
	P	0.159	0.079	single
	P	0.159	0.079	single
	P	0.159	0.079	single
	R	0.305	0.153	single/ext.
	G	0.140	0.070	double
	S	0.180	0.090	single
	S	0.180	0.090	single
	P	0.159	0.079	single
	G	0.140	0.070	double/ext
	DG	0.291	0.145	single
	K	0.220	0.110	single
	L	0.166	0.083	double
	W	0.248	0.124	single
	Q	0.287	0.144	single
	A	0.146	0.073	double
	Y	0.216	0.108	double
	A	0.146	0.073	double
	D	0.193	0.097	single

Table V.3.3.: Tc10b-A2K-S20C(S-tBu)G

Sequenz N to C	DKYAQWLADGGPSSGRPPPC(StBu)G			
MW Peptide (g/mol)	2216.43			
Peptid Menge (g) / (mol)	0.2	0.090		
Resin Type	Wang			
Resin Capacity (mmol/g)	0.85			
Resin amount / (g)	0.106			
Aceticanhydride (mL)	0.426			
Pyridine (mL)	0.426			
	C to N	Double	Single	Annotation
4.9 eq HATU / g		0.336	0.168	
10 eq DIPEA / ml		0.314	0.157	
5 eq. amino acids (g)	G	0.268	0.134	double
	C	0.389	0.195	single
	P	0.304	0.152	single
	P	0.304	0.152	single
	P	0.304	0.152	single
	R	0.585	0.293	single/ext.
	G	0.268	0.134	double
	S	0.346	0.173	single
	S	0.346	0.173	single
	P	0.304	0.152	single
	G	0.268	0.134	double/ext.
	DG	0.558	0.279	double
	A	0.281	0.140	double
	L	0.319	0.159	single
	W	0.475	0.238	single
	Q	0.551	0.276	single
	A	0.281	0.140	double
	Y	0.415	0.207	single
	K	0.423	0.211	single
	D	0.371	0.186	single

Table V.3.4.: Tc10b-S20C-21G

<b>Sequence N to C</b>	<b>DAYAQLADGGPSSGRPPPCG</b>			
MW Peptide (g/mol)	2159.34			
Peptid Menge (g) / (mol)	0.2	0.093		
Resin Type	Wang			
Resin Capacity (mmol/g)	0.85			
Resin amount / (g)	0.109			
Aceticanhydride (mL)	0.438			
Pyridine (mL)	0.438			
	C to N	Double	Single	Annotation
4.9 eq HATU / g		0.345	0.172	
10 eq DIPEA / ml		0.323	0.161	
5 eq. amino acids (g)	G	0.275	0.138	double
	C	0.542	0.271	single
	P	0.313	0.156	single
	P	0.313	0.156	single
	P	0.313	0.156	single
	R	0.601	0.300	single/ext.
	G	0.275	0.138	double
	S	0.355	0.178	single
	S	0.355	0.178	single
	P	0.313	0.156	single
	G	0.275	0.138	double/ext.
	DG	0.573	0.287	double
	A	0.288	0.144	double
	L	0.327	0.164	single
	W	0.488	0.244	single
	Q	0.566	0.283	single
	A	0.288	0.144	double
	Y	0.426	0.213	single
	A	0.288	0.144	single
	D	0.381	0.191	single

Table V.3.5.: Tc10b - TcKKA

Sequence N to C	Ac-DKYAQWLADGGPSSGRPPPK			
MW peptide (g/mol)	2169.38			
Peptide amount (g/mmol)	0.4		0.184	
Resin Type	Wang			
Resin capacity (mmol/g)	0.85			
Resin amount (g)	0.217			
Capping: Pyridine (mL)	0.871			
Aceticanhydride (mL)	0.871			
	C to N	Double	Single	Annotation
4.9 eq. HATU (g)		0.687	0.343	
10 eq. DIPEA (mL)		0.642	0.321	
5 eq. amino acids (g)	K	0.864	0.432	single
	P	0.622	0.311	single
	P	0.622	0.311	single
	P	0.622	0.311	single
	R	1.196	0.598	single/ext.
	G	0.548	0.274	double
	S	0.707	0.353	single
	S	0.707	0.353	single
	P	0.622	0.311	single
	G	0.548	0.274	double/ext.
	DG	1.141	0.570	single
	A	0.574	0.287	single
	L	0.652	0.326	double
	W	0.971	0.485	single
	Q	1.126	0.563	single
	A	0.574	0.287	double
	Y	0.847	0.424	double
	K	0.864	0.432	double
	D	0.759	0.379	single





## 4. List of Chemicals

Dichloromethane and diethyl ether were distilled under reduced pressure before use. All other commercial chemicals were used without further purification.

**Table V.4.1.:** List of used chemicals and their purity and supplier.

Substance	Supplier
Acetonitrile Chromasolv gradient grade for HPLC	Sigma-Aldrich
Acetic anhydride, 99%	Acros
Ammonium hydrogen carbonate, 98%	Alfa Aesar
Azobenzene-4,4'-dicarboxylic Acid, 99%	TCI
Deuterium oxide, 99.9%	Euriso-top
Dichloromethane	BCD Chemie
Diethyl ether	BCD Chemie
Dimethylformamide, 99.5 %	Grüssing
DSS	Deutero
Fmoc-Ala-OH, 95.0 %	Aldrich Chemistry
Fmoc-Arg(Pbf)-OH, 99.15 %	ChemPep
Fmoc-Asp(OtBu)-(Dmb)Gly-OH, 96.3 %	ChemPep
Fmoc-Asp(OtBu)-OH, 99.33 %	ChemPep
Fmoc-Gln(Trt)-OH, 99.12 %	ChemPep
Fmoc-Gly-OH, 99.93 %	ChemPep
Fmoc-Leu-OH, $\geq 99.0$ %	Aldrich Chemistry
Fmoc-Lys(Boc)-OH, 99.07 %	ChemPep
Fmoc-Pro-OH, $\geq 99.0$ % (HPLC)	Aldrich Chemistry
Fmoc-Ser(tBu)-OH, 99.76 %	ChemPep
Fmoc-Trp(Boc)-OH, 99.10 %	ChemPep
Fmoc-Tyr(tBu)-OH, 99.2 %	ChemPep
HATU, 99.8 %	ChemPep
Immersion Oil Type B	Ted Pella Inc.
N,N'-Dicyclohexylcarbodiimid, 99%	Sigma-Aldrich
N,N-Diisopropylethylamine, $\geq 99$ %	Sigma-Aldrich
Piperidine (Peptide Grade), 99 %	abcr
Pyridine, 99.5 %	Grüssing
Sodium hydrogen phosphate, 99%	Grüssing
Sodium dihydrogen phosphate, 99%	Grüssing
Trifluoroacetic Acid, $> 99.0$ %	TCI
Triisopropylsilane, $> 98.0$ %	TCI
Wang-Resin, 200-400 mesh, 0.86 mmol/g	ChemPep
Water Chromasolv Plus for HPLC	Sigma-Aldrich

## List of Figures

Figure 1.1:	Exemplary secondary structure elements. . . . .	4
Figure 1.2:	Overview of the structural diversity of different AFPs obtained from multiple species. <sup>11–20</sup> . . . . .	5
Figure 1.3:	Different ice crystal surfaces, which AFP can interact with. . . . .	6
Figure 1.4:	The shape of an ice crystal in the presence of AFP Type 1. . . . .	7
Figure 1.5:	Schematic view of the ice binding of AFP Type 1 to the $20\bar{2}1$ ice surface and picture of a hexagonal bipyramidal crystal. . . . .	8
Figure 1.6:	Structure of AFP Type 1 HPLC6. . . . .	9
Figure 1.7:	Structure of the first Exendin-4 NMR-structure. . . . .	11
Figure 1.8:	Model of the Trp-cage NMR structures of the sequence Tc5b with the lowest energy (first model). . . . .	12
Figure 1.9:	Concept of G. A. WOOLLEY of azobenzene cross-linked to side chains. . . . .	15
Figure 2.1:	Concept of a photo-switchable AFP Type 1 by H. KOBARG. . . . .	17
Figure 3.1:	Model of the Trp-cage variant Tc10b. . . . .	21
Figure 3.2:	Two different concepts for a switchable Trp-cage. . . . .	22
Figure 3.3:	Concept for combining the Trp-cage with a biological active peptide sequence in two alternative approaches. . . . .	24
Figure I.1.1:	$^1\text{H}$ -NMR spectrum of Tc10b at 298 K and pH 3 with enlargement of the indole proton region. . . . .	29
Figure I.1.2:	CSD plot of $\text{H}\alpha$ -atoms in Tc10b at 298 K at pH 3 (black) and pH 7 (red). . . . .	31
Figure I.1.3:	$^1\text{H}$ NMR spectrum of the region of the $\text{W6H}\epsilon$ proton from 9.5 to 10.3 ppm of TcKKA at pH 3 and 298 K. . . . .	36
Figure I.1.4:	Normalized intensity of the folded and unfolded all <i>trans</i> (fast equilibrium, blue circles) and unfolded ( <i>cis</i> -Pro, red squares) conformations against temperature. . . . .	37
Figure I.1.5:	Enlarged $\text{W6H}\epsilon$ region of the 1D steady state tr-NOE of TcKKA at 298 K and pH 3 between $\delta = 9.2$ and 10.4 ppm. . . . .	38
Figure I.1.6:	Enlargement of the 1D steady state tr-NOE difference spectra of $\text{W6H}\epsilon$ in fast equilibrium between 9.60 and 9.85 ppm. . . . .	39
Figure I.1.7:	$^1\text{H},^1\text{H}$ NOESY spectrum of our Trp-cage variant TcKKA at 328 K and pH 3. Enlarged region of the $\text{W6H}\epsilon$ between $\delta = 9.80$ and 10.15 ppm. . . . .	40
Figure I.2.2:	CSD plot of the $\text{H}\alpha$ atoms of Tc5b at pH 3 (black) and pH 8 (red). . . . .	43

Figure I.2.3:	Concept for a switchable Trp-cage with azobenzene in a push position between the N- and C-terminus. . . . .	45
Figure I.2.4:	Distance distribution of 4,4'-diiodoacetamide azobenzene, 4,4'-dibromomethyl azobenzene and of Tc5b-L2C-S20C. . . . .	46
Figure I.2.5:	The two major conformations of Tc5b-L2C-S20C observed in the MD simulation. . . . .	47
Figure I.2.7:	<sup>1</sup> H NMR spectra of Tc5b-L2C-S20C at 298 K and different pH values. . . . .	49
Figure I.2.8:	MALDI-ToF spectra of the crude reaction solution of Tc5b-L2C-S20C with the WOOLLEY linker and of the crude reaction solution of Tc5b-L2C-S20C with the 4,4'-dibromomethyl azobenzene. . . .	52
Figure I.2.9:	Progress of the linking reaction between Tc5b-L2C-S20C and the WOOLLEY linker was monitored by HPLC. . . . .	53
Figure I.2.10:	Reaction progression of Tc5b-L2C-S20C and 4,4'-dibromomethyl azobenzene monitored by HPLC. . . . .	54
Figure I.2.11:	Enlarged <sup>1</sup> H NMR spectra between 9.5 and 10.4 ppm (indole proton region) of Tc5b-L2C-S20C cross-linked with 4,4'-dibromomethyl azobenzene (top) in <i>trans</i> (black) and <i>cis</i> (red) at 298 K and pH 3. . . . .	55
Figure I.3.1:	<sup>1</sup> H NMR spectrum of TcKKA at 298 K and pH 3 (left). . . . .	60
Figure I.3.2:	CSD plot of the H $\alpha$ atoms of TcKKA at pH 3. . . . .	61
Figure I.3.3:	The fraction unfolded $\chi_U$ plotted against the temperature for N-terminal $\alpha$ -helix (black squares) and the cage loop (red circles). . .	63
Figure I.3.4:	Distance distribution of 4,4'-dibenzoic acid azobenzene and Tc10b-A2K-K8A-S20K (TcKKA). . . . .	65
Figure I.3.5:	Representative structure of Tc10b-A2K-K8A-S20K (TcKKA) of the MD simulation. . . . .	66
Figure I.3.6:	HPLC chromatogram of the reaction mixture of TcKKA with LOMANT'S REAGENT prior purification with a detection wavelength $\lambda = 220\text{ nm}$ . . . . .	67
Figure I.3.7:	<sup>1</sup> H-NMR spectrum of TcKKA linked with LOMANT'S REAGENT at 298 K and pH 3 . . . . .	68
Figure I.3.8:	CSD plot for the H $\alpha$ -atoms of TcKKA linked with LOMANT'S REAGENT at pH 3. . . . .	69
Figure I.3.9:	Representative model of TcKKA-LOMANT of MD simulations. . .	70
Figure I.4.1:	HPLC chromatogram of the reaction mixture of TcKKA with azobenzene prior purification with a detection wavelength of 220 nm. . .	73

Figure I.4.2:	UV/vis spectra of TcKKA-azo after irradiation to the PSS at $\lambda = 450 \text{ nm}$ (blue spectrum) and $\lambda = 365 \text{ nm}$ (red spectrum) at 298 K and pH 3. . . . .	75
Figure I.4.3:	UV/vis spectra of TcKKA linked with azobenzene at pH 3. . . . .	76
Figure I.4.4:	Ratio of <i>trans</i> (black) and <i>cis</i> (red) linked TcKKA-azo. . . . .	77
Figure I.4.5:	CD spectra of TcKKA linked with azobenzene at pH 3 and 298 K after irradiation at different wavelengths. . . . .	79
Figure I.4.6:	$^1\text{H}$ NMR spectrum of TcKKA linked with azobenzene in the <i>trans</i> state after irradiation at $\lambda = 450 \text{ nm}$ at 298 K and pH 3. . . . .	80
Figure I.4.7:	CSD plot of the $\text{H}\alpha$ atoms for TcKKA-azo at pH 3 with the azobenzene in its <i>trans</i> configuration. . . . .	81
Figure I.4.8:	$^1\text{H}$ NMR spectrum of TcKKA linked with azobenzene in the <i>cis</i> state at 298 K and pH 3 with magnification of indole proton region. . . . .	83
Figure I.4.9:	CSD plot of the $\text{H}\alpha$ atoms of TcKKA-azo at pH 3 with the linker in its <i>cis</i> configuration. . . . .	84
Figure I.4.10:	Possible models of TcKKA-azo with the azobenzene linker in <i>trans</i> (left) and <i>cis</i> configuration (right) based on the results of the spectroscopic data. . . . .	86
Figure II.1.1:	Concept for combining the Trp-cage with a biological active peptide sequence in two alternative approaches. . . . .	91
Figure II.1.2:	Concept for the design of a chimera protein of a Trp-cage with one repeat of the AFP Type 1 HPLC6. . . . .	92
Figure II.1.3:	Sequence pattern of the used ice binding motif including the N-terminal capping sequence (in brackets) and view along the helix (from N to C) of the chimera protein with marked orientations (blue lines) of the ice binding surface. . . . .	93
Figure II.1.4:	$^1\text{H}$ NMR spectrum of AFP-Tc-7 at 274 K and pH 3 with enlargement of the indole proton region around $10 \text{ ppm}$ . . . . .	96
Figure II.1.5:	$^1\text{H}$ NMR spectrum of AFP-Tc-7 at 274 K and pH 7.9 in 0.1 M $\text{NH}_4\text{HCO}_3$ buffer with enlargement of indole proton region around $10 \text{ ppm}$ . . . . .	97
Figure II.1.6:	CSD plot of $\text{H}\alpha$ -atoms in AFP-Tc-7 pH 3 (top) and pH 7.9 (bottom) at 274 K. . . . .	98
Figure II.1.7:	$^{13}\text{C}\alpha$ CSD plot of the helical region from residue 2' to 8 for AFP-Tc-7 pH 3 (left) and pH 7.9 (right) at 274 K. . . . .	101
Figure II.1.8:	CD spectra of Tc10b (hollow circles) and AFP-Tc-7 (filled circles) recorded at 274 K and pH 7.9. . . . .	103
Figure II.1.9:	$^1\text{H}$ NMR spectra of AFP-Tc-6 at 274 K and pH 3 with enlargement of the indole proton region. . . . .	104

Figure II.1.10:	$^{13}\text{C}\alpha$ CSD plot of the complete helical region from residue 2' to 8 of AFP-Tc-6 pH 3 (left) and pH 7.9 (right) at 274 K. . . . .	105
Figure II.1.11:	CD spectra of Tc10b (hollow circles) and AFP-Tc-6(filled circles) recorded at 274 K and pH 7.9. . . . .	106
Figure II.1.12:	$^1\text{H}$ NMR spectra of TcAFP5b at 274 K and pH 3. . . . .	108
Figure II.1.13:	$^{13}\text{C}\alpha$ CSD plot of the helical region from residue 2' to 8 for AFP-Tc-5 pH 3 (left) and pH 8 (right) at 274 K. . . . .	108
Figure II.1.14:	CD spectra of Tc10b (hollow circles) and AFP-Tc-5(filled circles) recorded at 274 K and pH 7.9. . . . .	110
Figure II.1.15:	$^1\text{H}$ -NMR spectra of AFP-Tc-4 at 274 K and pH 3 with enlargement of the indole proton region. . . . .	111
Figure II.1.16:	$^{13}\text{C}\alpha$ CSD plot of the complete helical region from residue 2' to 8 for AFP-Tc-4 pH 3 (left) and pH 8 (right) at 274 K. . . . .	112
Figure II.1.17:	CD spectra of Tc10b (hollow circles) and AFP-Tc-4(filled circles) recorded at 274 K and pH 7.9. . . . .	113
Figure II.1.18:	$^1$ NMR spectra of AFP-Tc-7 (top, left), AFP-Tc-6 (top, right), AFP-Tc-5 (bottom, left) and AFP-Tc-4 (bottom, right) between 9.5 ppm and 10.5 ppm (region of the indole proton). . . . .	114
Figure II.1.19:	Models of AFP-Tc-4 (top) and AFP-Tc-5 (bottom) with the secondary structure presented in cartoon style. . . . .	116
Figure II.1.20:	Pictures of reference ice crystals. . . . .	119
Figure II.1.21:	Ice crystals during the growth in the presence of AFP-Tc-4 (first column), AFP-Tc-5 (second column), AFP-Tc-6 (third column) and AFP-Tc-7. . . . .	120
Figure III.1.1:	Possible models of TcKKA-azo with the azobenzene linker in <i>trans</i> (left) and <i>cis</i> configuration (right) based on the results of the spectroscopic data. . . . .	124
Figure III.2.1:	Sequence pattern of the used ice binding motif including the N-terminal capping sequence and view along the helix axis. . . . .	127
Figure III.3.1:	Fusion concept for stabilizing helical AMP sequences, exemplarily shown for the 77 residue long protein Amoebapore A. . . . .	131
Figure III.3.2:	Concept for a serial connection of two Trp-cage modules (blue and red) for anti parallel $\alpha$ -helical dimers (green). . . . .	132
Figure III.3.3:	Cartoon representation of the zinc-finger structure. . . . .	133
Figure IV.1.1:	Solvent profile of the short (left) and long analytical method (right). . . . .	140
Figure IV.1.2:	Exemplary profile of a preparative run. . . . .	141
Figure V.1.1:	HPLC chromatogram of purified Trp-cage Tc5b. . . . .	156
Figure V.1.2:	Mass spectrum of the purified Trp-cage Tc5b. . . . .	157

Figure V.1.3:	$^1\text{H}$ -NMR spectrum of Tc5b at 298 K and pH 7 with enlargement of the indole proton region. . . . .	158
Figure V.1.4:	HPLC chromatogram of purified Trp-cage Tc5b-L2C-S20C. . . . .	161
Figure V.1.5:	$m/z$ : $[M + H]^+$ calcd for $[\text{Tc5b} - \text{L2C} - \text{S20C} + \text{H}]^+$ 2176.5, found 2176.9. . . . .	162
Figure V.1.6:	RMSD of the $\text{C}\alpha$ atoms during the MD simulation of Tc5b-L2C-S20C. . . . .	163
Figure V.1.7:	MALDI-ToF spectrum of the fraction collected with a $t_R = 17.8$ min, containing cross-linked Tc5b-L2C-S20C with the WOOLLEY linker. . . . .	164
Figure V.1.8:	$^1\text{H}$ NMR spectra between of Tc5b-L2C-S20C cross-linked with the WOOLLEY-linker (top) in <i>trans</i> (black) and <i>cis</i> (red) at 298 K and pH 3. . . . .	165
Figure V.1.9:	Enlarged $^1\text{H}$ NMR of the indole and amide region of the purified cross-linked Tc5b-L2C-S20C with the WOOLLEY linker in <i>trans</i> (bottom/black) and <i>cis</i> (top/red) at 298 K and pH 3. . . . .	166
Figure V.1.10:	MALDI-ToF spectrum of cross-linked Tc5b-L2C-S20C with 4,4'-dibromomethyl azobenzene linker. . . . .	167
Figure V.1.11:	$^1\text{H}$ NMR spectra of Tc5b-L2C-S20C cross-linked with 4,4'-dibromomethyl azobenzene (top) in <i>trans</i> (black) and <i>cis</i> (red) at 298 K and pH 3. . . . .	167
Figure V.1.12:	Total ion chromatogram of crude Tc10b (top) and purified (bottom). . . . .	168
Figure V.1.13:	$^1\text{H}$ -NMR spectrum of Tc10b at 298 K and pH 7 with enlargement of the indole proton region. . . . .	169
Figure V.1.14:	MALDI-ToF spectrum of the crude Tc10b-S20C-21G (left) and purified (right). . . . .	173
Figure V.1.15:	HPLC chromatogram of the crude Tc10b-S20C-21G (left) and purified (right). . . . .	173
Figure V.1.16:	$^1\text{H}$ NMR of Tc10b-S20C-21G at pH 3 and 298 K (top) and pH 7 (bottom). . . . .	174
Figure V.1.17:	MALDI-ToF spectrum of the crude Tc10b-A2K-S20C(S-tBu)-21G (left) and purified (right). . . . .	178
Figure V.1.18:	HPLC-chromatogram of the crude Tc10b-A2K-S20C(S-tBu)-21G (left) and purified (right). . . . .	178
Figure V.1.19:	Total ion chromatogram of crude Tc10b-A2C(S-tBu)-S20K (top) and purified (bottom). . . . .	181
Figure V.1.20:	Total ion chromatogram of crude Tc10b-S20K (top) and purified (bottom). . . . .	184

Figure V.1.21: Total ion chromatogram of crude TcKKA (top) and purified (bottom). . . . .	186
Figure V.1.22: RMSD of the $C\alpha$ atoms during the MD simulation of TcKKA. . .	196
Figure V.1.23: Total ion chromatogram of crude TcKKA linked with LOHMANNT'S REAGENT. . . . .	197
Figure V.1.24: RMSD of the $C\alpha$ atoms during the MD simulation of TcKKA-Lomant. . . . .	199
Figure V.1.25: Total ion chromatogram of crude TcKKA linked with azobenzene (top) and purified (bottom). . . . .	200
Figure V.1.26: RMSD of the $C\alpha$ atoms during the MD simulation of TcKKA-azo in <i>cis</i> configuration of the <i>P</i> and <i>M</i> enantiomer. . . . .	203
Figure V.1.27: Total ion chromatogram of crude 15er reference peptide (top) and purified (bottom). . . . .	204
Figure V.1.28: CD spectrum of 15er reference peptide at 274 K and pH 3. . . .	205
Figure V.1.29: $^1\text{H-NMR}$ spectra of 15er at 274 K and pH 3 (left) or pH 7.9 (0.1 M $\text{NH}_4\text{HCO}_3$ buffer) (right). . . . .	205
Figure V.1.30: Total ion chromatogram of crude 15er reference peptide (top) and purified (bottom). . . . .	209
Figure V.1.31: CD spectrum of 15erYW reference peptide at 274 K and pH 3. .	210
Figure V.1.32: $^1\text{H-NMR}$ spectra of 15erYW at 274 K and pH 3 (left) or pH 7.9 (0.1 M $\text{NH}_4\text{HCO}_3$ buffer) (right). . . . .	210
Figure V.1.33: HPLC chromatogram of purified AFP-Tc-4, synthesized by S. Rothemund. . . . .	214
Figure V.1.34: $m/z$ : $[M + H]^+$ calcd for $[\text{TcAFP4}]^+$ 2447.5, found 2448.4. . .	215
Figure V.1.35: $^1\text{H NMR}$ spectrum of AFP-Tc-4 at 274 K and pH 7.9 with enlargement of the indole proton region. . . . .	216
Figure V.1.36: $\text{H}\alpha$ CSD plot of AFP-Tc-4 at pH 3 (top) and pH 7.9 (bottom) at 274 K. . . . .	223
Figure V.1.37: $\text{H}\alpha$ CSD plot of AFP-Tc-4 at pH 3 (top) and pH 7.9 (bottom) at 298 K. . . . .	224
Figure V.1.38: HPLC chromatogram of purified AFP-Tc-5, synthesized by S. Rothemund . . . . .	239
Figure V.1.39: $m/z$ : $[M + H]^+$ calcd for $[\text{TcAFP5}]^+$ 2476.5, found 2477.2. . .	240
Figure V.1.40: $^1\text{H NMR}$ spectrum of AFP-Tc-5 at 274 K and pH 7.9 with enlargement of the indole proton region. . . . .	241
Figure V.1.41: $\text{H}\alpha$ CSD plot of AFP-Tc-5 at pH 3 (top) and pH 7.9 (bottom) at 274 K. . . . .	248
Figure V.1.42: $\text{H}\alpha$ CSD plot of AFP-Tc-5 at pH 3 (top) and pH 7.9 (bottom) at 298 K. . . . .	249

Figure V.1.43: HPLC chromatogram of purified TcAFP5-KK, synthesized by S. Rothemund . . . . .	264
Figure V.1.44: $m/z$ : $[M + H]^+$ calcd for $[TcAFP5 - KK]^+$ 2574.7, found 2575.7.265	
Figure V.1.45: CD spectra of Tc10b (hollow circles) and AFP-Tc-5-KK (filled circles) at 274 K and pH 3. . . . .	266
Figure V.1.46: $^1\text{H-NMR}$ spectra of AFP-Tc-5-KK at 274 K and pH 3 (left) or pH 7.9 (0.1 M $\text{NH}_4\text{HCO}_3$ buffer). . . . .	266
Figure V.1.47: HPLC chromatogram of purified AFP-Tc-7, synthesized by S. Rothemund . . . . .	268
Figure V.1.48: $m/z$ : $[M + H]^+$ calcd for $[TcAFP6]^+$ 2731.9, found 2732.4. . .	269
Figure V.1.49: $\text{H}\alpha$ CSD plot of AFP-Tc-7 at pH 3 (top) and pH 7.9 (bottom) at 298 K. . . . .	276
Figure V.1.50: HPLC chromatogram of purified AFP-Tc-6, synthesized by S. Rothemund . . . . .	279
Figure V.1.51: $m/z$ : $[M + H]^+$ calcd for $[TcAFP7]^+$ 2702.9, found 2703.8. . .	280
Figure V.1.52: $^1\text{H-NMR}$ spectrum of AFP-Tc-6 at 274 K and pH 7.9 with enlargement of the indole proton region. . . . .	287
Figure V.1.53: $\text{H}\alpha$ CSD plot of AFP-Tc-6 at pH 3 (top) and pH 7.9 (bottom) at 274 K. . . . .	288
Figure V.1.54: $\text{H}\alpha$ CSD plot of AFP-Tc-7 at pH 3 (top) and pH 7.9 (bottom) at 298 K. . . . .	289
Figure V.1.55: $^1\text{H}$ and $^{13}\text{C}$ NMR of 4,4'-bis(2,5-dioxopyrrolidin-1yl)azobenzene .	292
Figure V.2.1: UV/vis spectrum of Tc10b at 298 K and pH 7.9. . . . .	294
Figure V.2.2: UV/vis spectrum of TcKKA at 298 K and pH 7.9. . . . .	294
Figure V.2.3: UV/vis spectrum of 15erYW reference peptide at 298 K and pH 7.9.294	
Figure V.2.4: UV/vis spectrum of 15erYW reference peptide at 298 K and pH 7.9.295	
Figure V.2.5: UV/vis spectrum of AFP-Tc-4 at 298 K and pH 7.9. . . . .	295
Figure V.2.6: UV/vis spectrum of AFP-Tc-5 at 298 K and pH 7.9. . . . .	295
Figure V.2.7: UV/vis spectrum of AFP-Tc-5-KK at 298 K and pH 7.9. . . . .	296
Figure V.2.8: UV/vis spectrum of AFP-Tc-6 at 298 K and pH 7.9. . . . .	296
Figure V.2.9: UV/vis spectrum of AFP-Tc-7 at 298 K and pH 7.9. . . . .	296



## List of Schemes

Scheme 1.1:	Scaffold of amino acids and the peptide bond with structure defining torsion angles $\phi$ , $\psi$ and $\omega$ highlighted in red. . . . .	3
Scheme 1.2:	Mechanism of the <i>trans</i> to <i>cis</i> isomerization of azobenzene. . . .	14
Scheme I.1.1:	Scaffold of X-Pro. . . . .	34
Scheme I.1.2:	Schematic view of the folding and unfolding equilibrium with the downstream equilibrium of the <i>cis</i> -Pro isomerization of the Trp-cage. . . . .	35
Scheme I.2.1:	Reaction scheme of cross-linking of Cys with iodo or bromo functionalized linker. . . . .	41
Scheme I.2.2:	Reaction scheme of cross-linking of Cys with iodo- or bromo-functionalized linker. . . . .	51
Scheme I.2.3:	Possible side reactions of TCEP (top, left) with the reactive group of the WOOLLEY linker. . . . .	56
Scheme I.3.1:	Reaction scheme of cross-linking of Lys with N-hydroxysuccinimide ester functionalized linker. . . . .	59
Scheme I.3.2:	Scaffold of an azobenzene based photochromic linker with NHS functionalized reactive groups for cross-linking (left) and LOMANT'S REAGENT (right). . . . .	64
Scheme I.3.3:	Numbered scaffold of LOMANT'S REAGENT. . . . .	70
Scheme I.4.1:	Scaffold of the switching of azobenzene linked to the Trp-cage. The <i>cis</i> -TcKKA-azo (right) can exhibit an additional helical chirality ( <i>P</i> or <i>M</i> ), induced by the linker. . . . .	74
Scheme I.4.2:	Numbered scaffold of the azobenzene incorporated in the Trp-cage. . . . .	82
Scheme I.4.3:	Numbered scaffold of the azobenzene in <i>cis</i> incorporated in the Trp-cage. . . . .	85
Scheme III.1.1:	Scheme of NHS functionalized diazocine. . . . .	126
Scheme III.1.2:	Scaffold of spiropyran. . . . .	126
Scheme IV.1.1:	Used resins for SPPS. . . . .	137
Scheme IV.1.2:	Coupling-mechanism with HATU and DIPEA. . . . .	138
Scheme IV.1.3:	Mechanism of the removal of the Fmoc protecting group using piperidine. . . . .	139
Scheme IV.1.4:	Mechanism of the base catalyzed aspartimide formation of the Asp residue. . . . .	142

## List of Tables

Table 1.1:	Categories for the 20 naturally occurring amino acids based on their side chains propensity to be in contact with water. . . . .	3
Table 1.2:	Sequence in one letter code of AFP Type 1 HPLC6. . . . .	9
Table 1.3:	Sequence comparison of GLP-1 and Exendin-4 represented in one letter code. . . . .	11
Table I.1.1:	Numbered Tc10b sequence in one letter code. . . . .	29
Table I.1.2:	CSD for selected protons in the structure representing the cage fold of Tc10b at 298 K at pH 3 and pH 7. . . . .	33
Table I.2.1:	The sequences of Trp-cage variants Tc5b and Tc10b presented in one letter code, which are used in work. . . . .	41
Table I.3.1:	Trp-cage sequence TcKKA for NHS based cross-linking. . . . .	60
Table I.3.2:	CSDs for selected protons in the cage structure representing the cage fold of TcKKA at 298 K at pH 3. . . . .	61
Table I.3.3:	Melting points ( $T_M$ ) of the helix and the cage loop of TcKKA, Tc10b and Tc10b-K8A at acidic conditions (pH 3 for TcKKA and pH 2.5 for Tc10b and Tc10b-K8A, NEIDIGH <sup>43</sup> ). . . . .	63
Table I.3.4:	CSDs for selected protons representing the fold of TcKKA-LOMANT at 298 K at pH 3. . . . .	69
Table I.4.1:	CSDs for selected protons in the cage structure representing the cage fold and the helix of TcKKA-azo with the linker in its <i>trans</i> configuration at 298 K at pH 3. . . . .	82
Table I.4.2:	CSDs for selected protons in the cage structure representing the cage loop and helix of TcKKA-azo in <i>cis</i> at 298 K at pH 3. . . . .	84
Table II.1.1:	Sequences of the four different AFP-Tc- $X^n$ variants. . . . .	94
Table II.1.2:	Sequence of AFP-Tc-7 and Tc10b. . . . .	96
Table II.1.3:	CSDs for selected protons in the chimera AFP-Tc-7 representing the fold of the cage loop and chimera helix at 274 K and pH 3 and 7.9. . . . .	99
Table II.1.4:	Average chemical deviations ( $\overline{\Delta\delta}$ , determined by the $^{13}\text{C}_\alpha$ shifts) and fraction folded ( $\chi_F$ ) for the AFP-Tc-7 at 274 K and pH 3 and 7.9. . . . .	102
Table II.1.5:	Sequence of AFP-Tc-6 and Tc10b. . . . .	104

Table II.1.6:	Sum of the CSDs for the selected protons, which represent the cage loop and its fraction folded for AFP-Tc-6 at 274 K and pH 3 and 7.9. . . . .	106
Table II.1.7:	Sequence of AFP-Tc-5 and Tc10b. . . . .	107
Table II.1.8:	Sum of the CSD for the selected protons to represent the cage loop and its fraction folded for AFP-Tc-5 at 274 K and pH 3 and 7.9. . . . .	109
Table II.1.9:	Sequence of AFP-Tc-4 and Tc10b. . . . .	111
Table II.1.10:	Sum of the CSD for the cage loop and its fraction folded for AFP-Tc-4 at 274 K and pH 3 and 7.9 . . . . .	112
Table II.1.11:	Fraction of the ASA of the Trp side chain and the free folding energy of AFP-Tc-4 and AFP-Tc-5. . . . .	115
Table II.1.12:	Percentage of <i>cis</i> -Pro unfolded conformations present in the four AFP-Tc chimeras and their helical content (CD) and fraction folded of the helix (NMR) $\chi_F^{helix}$ at 274 K and pH 7.9. . . . .	117
Table II.1.13:	Sequences of 15 amino acids reference. . . . .	118
Table III.1.1:	Summary of the fraction folded $\chi_F$ of the helix and the cage loop of Tc10b and TcKKA (unlinked and linked with 4,4'-NHS-azobenzene) at pH 3 and 298 K. . . . .	124
Table IV.1.1:	Isolated yields of the peptide synthesis. . . . .	141
Table IV.1.2:	ESI-MS of the hexa-peptide ATD-GSK, ATDGSK and ATEGSK. . . . .	143
Table IV.3.1:	Sum of CSDs for selected protons in the helical region representing the helix fold of Tc10b at 280 K at pH 7. <sup>43</sup> . . . . .	150
Table IV.3.2:	Sum of CSDs for selected protons in the cage structure representing the cage fold of Tc10b at 280 K at pH 7. <sup>43</sup> . . . . .	150
Table V.1.1:	Nomenclature of the AFP-Tc chimera proteins used in this dissertation and in the measurements. . . . .	155
Table V.1.2:	Chemical shifts of Tc5b at 298 K and pH 3. . . . .	159
Table V.1.3:	Chemical shifts of Tc5b at 298 K and pH 8. . . . .	160
Table V.1.4:	Sum of CSDs for selected protons in the helical region representing the helix fold of Tc10b at 280 K at pH 7. <sup>43</sup> . . . . .	169
Table V.1.5:	Sum of CSDs for selected protons in the cage structure representing the cage fold of Tc10b at 280 K at pH 7. <sup>43</sup> . . . . .	169
Table V.1.6:	<sup>1</sup> H chemical shifts of Tc10b at 298 K and pH 3. . . . .	170
Table V.1.7:	<sup>1</sup> H-chemical shifts of Tc10b at 298 K and pH 8. . . . .	171
Table V.1.8:	<sup>15</sup> N and <sup>13</sup> C-chemical shifts of Tc10b at 298 K and pH 3. . . . .	172
Table V.1.9:	<sup>1</sup> H chemical shifts of Tc10b-S20C-21G at 298 K and pH 3. . . . .	175
Table V.1.10:	<sup>15</sup> N and <sup>13</sup> C-chemical shifts of Tc10b-S20C-21G at 298 K and pH 3. . . . .	176
Table V.1.11:	<sup>1</sup> H-chemical shifts of Tc10b-S20C-21G at 298 K and pH 7. . . . .	177

Table V.1.12:	$^1\text{H}$ -chemical shifts of Tc10b-A2K-S20C-21G at 298 K and pH 3. . .	179
Table V.1.13:	$^1\text{H}$ -chemical shifts of Tc10b-A2K-S20C-21G at 298 K and pH 7. . .	180
Table V.1.14:	$^1\text{H}$ -chemical shifts of Tc10b-A2C(S-tBu)-S20K at 298 K and pH 3. . .	182
Table V.1.15:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of Tc10b-A2C(S-tBu)-S20K at 298 K and pH 3. . . . .	183
Table V.1.16:	$^1\text{H}$ -chemical shifts of Tc10b-20K at 298 K and pH 3. . . . .	185
Table V.1.17:	$^1\text{H}$ -chemical shifts of TcKKA at 298 K and pH 3. . . . .	187
Table V.1.19:	$^1\text{H}$ -chemical shifts of TcKKA at 278 K and pH 3. . . . .	189
Table V.1.20:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of TcKKA at 278 K and pH 3. . . . .	190
Table V.1.21:	Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. . . . .	191
Table V.1.22:	Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. . . . .	192
Table V.1.23:	Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. . . . .	193
Table V.1.24:	Chemical shifts, CSD and fraction unfolded for TcKKA at pH 3 from 274.2 K to 338 K. . . . .	194
Table V.1.25:	$^1\text{H}$ -chemical shifts of TcKKA at 274 K and pH 7.9. . . . .	195
Table V.1.26:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of TcKKA at 274 K and pH 7.9. . . . .	196
Table V.1.27:	$^1\text{H}$ -chemical shifts of TcKKA-Lomant at 298 K and pH 3. . . . .	198
Table V.1.28:	$^1\text{H}$ -chemical shifts of TcKKA-Azo in <i>trans</i> at 298 K and pH 3. . . . .	201
Table V.1.29:	$^1\text{H}$ -chemical shifts of TcKKA-Azo in <i>cis</i> at 298 K and pH 3. . . . .	202
Table V.1.30:	$^1\text{H}$ chemical shifts of 15er reference peptide at 274 K and pH 3. . . . .	206
Table V.1.31:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of 15er at 274 K and pH 3. . . . .	206
Table V.1.32:	$^1\text{H}$ -chemical shifts of 15er reference peptide at 274 K and pH 7.9. . . . .	207
Table V.1.33:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of 15er at 274 K and pH 7.9. . . . .	208
Table V.1.34:	$^1\text{H}$ -chemical shifts of 15erYW reference peptide at 274 K and pH 3. . . . .	211
Table V.1.35:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of 15erYW at 274 K and pH 3. . . . .	211
Table V.1.36:	$^1\text{H}$ -chemical shifts of 15erYW reference peptide at 274 K and pH 7.9. . . . .	212
Table V.1.37:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of 15erYW at 274 K and pH 7.9. . . . .	213
Table V.1.38:	$^1\text{H}$ -chemical shifts of TcAFP4b at 274K and pH 3. . . . .	217
Table V.1.39:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-4 at 274 K and pH 3. . . . .	218
Table V.1.40:	$^1\text{H}$ -chemical shifts of TcAFP4b at 298K and pH 3. . . . .	219
Table V.1.41:	$^1\text{H}$ -chemical shifts of TcAFP4b at 274K and pH 7.9. . . . .	220
Table V.1.42:	$^1\text{H}$ -chemical shifts of TcAFP4b at 274K and pH 7.9. . . . .	221
Table V.1.43:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-4 at 274 K and pH 8 . . . . .	221
Table V.1.44:	$^1\text{H}$ -chemical shifts of TcAFP4b at 298K and pH 7.9. . . . .	222

Table V.1.45:	CSDs for selected protons in the cage structure representing the cage fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9.	225
Table V.1.46:	CSDs for selected protons in the helical region representing the helix fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9.	225
Table V.1.47:	CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP4b at 298 K and 274 K and at pH 3 and 7.9. . . . .	225
Table V.1.48:	CSDs for the $^{13}\text{C}\alpha$ atoms in the helix representing the helix fold of AFP-Tc-4 at 274 K and at pH 3 and 7.9. . . . .	226
Table V.1.49:	$^1\text{H}$ -chemical shifts of TcAFP5b at 274 K and pH 3. . . . .	242
Table V.1.50:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-5 at 278 K and pH 3. . .	243
Table V.1.51:	$^1\text{H}$ -chemical shifts of TcAFP5b at 298K and pH 3. . . . .	244
Table V.1.52:	$^1\text{H}$ -chemical shifts of TcAFP5b at 274 K and pH 7.9. . . . .	245
Table V.1.53:	$^1\text{H}$ -chemical shifts of TcAFP5b at 274 K and pH 7.9. . . . .	246
Table V.1.54:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-5 at 274 K and pH 7.9. .	246
Table V.1.55:	$^1\text{H}$ -chemical shifts of TcAFP5b at 298 K and pH 7.9. Shifts marked with an asterisk are averaged. . . . .	247
Table V.1.56:	CSDs for selected protons in the cage structure representing the cage fold of TcAFP5b at 298 K and 274 K and at pH 3 and 7.9.	250
Table V.1.57:	CSDs for selected protons in the helical region representing the helix fold of AFP-Tc-5 at 298 K and 274 K and at pH 3 and 7.9.	250
Table V.1.58:	CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP5b at 298 K and 274 K and at pH 3 and 7.9. . . . .	251
Table V.1.59:	CSDs for the $^{13}\text{C}\alpha$ atoms in the helix representing the helix fold of AFP-Tc-5 at 274 K and at pH 3 and 7.9. . . . .	251
Table V.1.60:	$^1\text{H}$ -chemical shifts of AFP-Tc-5-KK at 274K and pH 3. . . . .	267
Table V.1.61:	$^1\text{H}$ -chemical shifts of TcAFP6b at 274K and pH 3. . . . .	270
Table V.1.62:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-7 at 274 K and pH 3. . .	271
Table V.1.63:	$^1\text{H}$ -chemical shifts of TcAFP6b at 298 K and pH 3. . . . .	272
Table V.1.64:	$^1\text{H}$ -chemical shifts of TcAFP6b at 274 K and pH 7.9. . . . .	273
Table V.1.65:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-7 at 274 K and pH 3. . .	274
Table V.1.66:	$^1\text{H}$ -chemical shifts of TcAFP6b at 298 K and pH 7.9. . . . .	275
Table V.1.67:	CSDs for selected protons in the cage structure representing the cage fold of TcAFP6b at 298 K and 274 K and at pH 3 and 7.9.	277
Table V.1.68:	CSDs for selected protons in the helical extension representing the N-terminal helix fold of TcAFP6b at 274 K and at pH 3 and 7.9.	277
Table V.1.69:	CSDs for the $^{13}\text{C}\alpha$ atoms in the helix representing the helix fold of AFP-Tc-7 at 274 K and at pH 3 and 7.9. . . . .	278

Table V.1.70:	$^1\text{H}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 3. . . . .	281
Table V.1.71:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 3. . .	282
Table V.1.72:	$^1\text{H}$ -chemical shifts of AFP-Tc-6 at 298 K and pH 3. . . . .	283
Table V.1.73:	$^1\text{H}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 7.9. . . . .	284
Table V.1.74:	$^{15}\text{N}$ and $^{13}\text{C}$ -chemical shifts of AFP-Tc-6 at 274 K and pH 7.9. .	285
Table V.1.75:	$^1\text{H}$ -chemical shifts of AFP-Tc-6 at 298 K and pH 7.9. . . . .	286
Table V.1.76:	CSDs for selected protons in the cage structure representing the cage fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9.	290
Table V.1.77:	CSDs for selected protons in the helical region representing the helix fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9.	290
Table V.1.78:	CSDs for selected protons in the helical extension representing the N-terminal helix fold of AFP-Tc-6 at 298 K and 274 K and at pH 3 and 7.9. . . . .	290
Table V.1.79:	CSDs for the $^{13}\text{C}\alpha$ atoms in the helix representing the helix fold of AFP-Tc-6 at 274 K and at pH 3 and 7.9. . . . .	291
Table V.2.1:	Extinction coefficient, corresponding absorption and resulting concentrations of the solutions for CD spectroscopy. . . . .	293
Table V.2.2:	Observed and theoretical values of the mean residue ellipticity at 222 nm of AFP-Tc-7 to AFP-Tc-4 and the resulting helical content.	293
Table V.3.1:	Tc10b . . . . .	297
Table V.3.2:	Tc10b-S20K . . . . .	298
Table V.3.3:	Tc10b-A2K-S20C(S-tBu)G . . . . .	299
Table V.3.4:	Tc10b-S20C-21G . . . . .	300
Table V.3.5:	Tc10b - TcKKA . . . . .	301
Table V.4.1:	List of used chemicals and their purity and supplier. . . . .	303

## Bibliography

- (1) M. H. Ultsch, W. Somers, A. A. Kossiakoff and A. M. de Vos, *J. Mol. Biol.* **1994**, *236*, 286–99.
- (2) D. Lawson, P. Artymiuk, S. Yewdall, J. Smith, J. Livingstone, A. Treffry, A. Luzzago, A. Levi, P. Arosio, G. Ceareni, C. Thomas, W. Shaw and P. Harrison, *Nature* **1991**, *349*, 541–44.
- (3) J. C. Skou, *Biochim. Biophys. Acta* **1957**, *23*, 394–401.
- (4) A. Radzicka and R. Wolfenden, *Science* **1995**, *267*, 90–93.
- (5) C. M. Metallo and M. G. ander Heiden, *Molec. Cell* **2013**, *49*, 388–398.
- (6) C. M. Waters and B. L. Bassler, *Annu. Rev. Cell Dev. Bio.* **2005**, *21*, 319–346.
- (7) *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 49–76.
- (8) S. Hovmoller, T. Zhou and T. Ohlson, *Acta Cryst. D* **2002**, *58*, 768–776.
- (9) P. F. Scholander, L. van Dam, J. W. Kanwisher, H. T. Hammel and M. S. Gordon, *J. Cell. Physio.* **1957**, *49*, 5–24.
- (10) A. L. Devries and D. E. Wohlschlag, *Science* **1969**, *163*, 1073–1075.
- (11) F. Sicheri and D. S. Yang, *Nature* **1995**, *375*, 427–31.
- (12) E. Liepinsh, G. Otting, M. M. Harding, L. G. Ward, J. P. Mackay and A. D. Haymet, *Eur. J. Biochem.* **2002**, *269*, 1259–66.
- (13) A. H.-Y. Kwan, K. Fairley, P. I. Anderberg, C. W. Liew, M. M. Harding and J. P. Mackay, *Biochemistry* **2005**, *44*, 1980–1988.
- (14) W. Gronwald, M. C. Loewen, B. Lix, A. J. Daugulis, F. D. Sönnichsen, P. L. Davies and B. D. Sykes, *Biochemistry* **1998**, *37*, 4712–4721.
- (15) D. S. Yang, W.-C. Hon, S. Bubanko, Y. Xue, J. Seetharaman, C. L. Hew and F. Sicheri, *Biophys. J.* **1998**, *74*, 2142–2151.
- (16) Y.-C. Liou, A. Tocilj, P. Davies and Z. Jia, *Nature* **2000**, *406*, 322–324.
- (17) C. Li, X. Guo, Z. Jia, B. Xia and C. Jin, *J. Biomol. NMR* **2005**, *32*, 251–256.
- (18) C. P. Garnham, R. L. Campbell and P. L. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 7363–7367.
- (19) A. J. Middleton, C. B. Marshall, F. Faucher, M. Bar-Dolev, I. Braslavsky, R. L. Campbell, V. K. Walker and P. L. Davies, *J. Mol. Biol.* **2012**, *416*, 713–724.
- (20) T. Sun, F.-H. Lin, R. L. Campbell, J. S. Allingham and P. L. Davies, *Science* **2014**, *343*, 795–798.
- (21) M. B. Dolev, I. Braslavsky and P. L. Davies, *Annu. Rev. Biochem.* **2016**, *85*, 515–542.
- (22) L. L. C. Olijve, K. Meister, A. L. DeVries, J. G. Duman, S. Guo, H. J. Bakker and I. K. Voets, *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 3740–3745.
- (23) C. A. Knight, *Nature* **2000**, *406*, 249–51.

- (24) P. A. Cziko, A. L. DeVries, C. Evans and C.-H. C. Cheng, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 14583–88.
- (25) R. M. Fournay, C. L. Hew, S. B. Joshi and G. L. Fletcher, *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **1984**, *78*, 791–796.
- (26) A. Chakrabartty and C. L. Hew, *Eur. J. Biochem.* **1991**, *202*, 1057–63.
- (27) S. N. Patel and S. P. Graether, *Biochem. Cell Biol.* **2010**, *88*, 223–9.
- (28) J. Baardsnes, L. H. Kondejewski, R. S. Hodges, H. Chao, C. Kay and P. L. Davies, *FEBS Lett.* **1999**, *463*, 87–91.
- (29) P. Dalal and F. D. Sonnichsen, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1276–84.
- (30) K. S. Park, W. S. Jung, H. J. Kim and S. Y. Shin, *Bull. Korean Chem. Soc.* **2010**, *31*, 3791–3793.
- (31) S. Patel and S. Graether, *Protein Science* **2010**, *19*, 2356–2365.
- (32) J. Houston, M. E., H. Chao, R. S. Hodges, B. D. Sykes, C. M. Kay, F. D. Sonnichsen, M. C. Loewen and P. L. Davies, *J. Biol. Chem.* **1998**, *273*, 11714–8.
- (33) C. R. Underwood, P. Garibay, L. B. Knudsen, S. Hastrup, G. H. Peters, R. Rudolph and S. Reedtz-Runge, *J. Biol. Chem.* **2010**, *285*, 723–30.
- (34) R. Malhotra, L. Singh, J. Eng and J. P. Raufman, *Regul. Pept.* **1992**, *41*, 149–56.
- (35) J. W. Neidigh, R. M. Fesinmeyer, K. S. Prickett and N. H. Andersen, *Biochemistry* **2001**, *40*, 13188–200.
- (36) B. Thorens, A. Porret, L. Bühler, S.-P. Deng, P. Morel and C. Widmann, *Diabetes* **1993**, *42*, 1678–1682.
- (37) S. Al-Sabah and D. Donnelly, *Protein Pept. Lett.* **2004**, *11*, 9–14.
- (38) A. A. Young, B. R. Gedulin, S. Bhavsar, N. Bodkin, C. Jodka, B. Hansen and M. Denaro, *Diabetes* **1999**, *48*, 1026–34.
- (39) J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat. Struct. Biol.* **2002**, *9*, 425–30.
- (40) B. Barua and N. H. Andersen, *Letts. Pept. Sci.* **2001**, *8*, 221–226.
- (41) D. V. Williams, A. Byrne, J. Stewart and N. H. Andersen, *Biochemistry* **2011**, *50*, 1143–52.
- (42) S. H. Gellman and D. N. Woolfson, *Nat. Struct. Biol.* **2002**, *9*, 408–10.
- (43) B. Barua, J. C. Lin, V. D. Williams, P. Kummmler, J. W. Neidigh and N. H. Andersen, *Protein Eng. Des. Sel.* **2008**, *21*, 171–85.
- (44) J. C. Lin, B. Barua and N. H. Andersen, *J. Am. Chem. Soc.* **2004**, *126*, 13679–84.
- (45) A. Byrne, D. V. Williams, B. Barua, S. J. Hagen, B. L. Kier and N. H. Andersen, *Biochemistry* **2014**, *53*, 6011–21.
- (46) J. W. Pitera and W. Swope, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587–92.



- (47) F. Ding, S. V. Buldyrev and N. V. Dokholyan, *Biophys. J.* **2005**, *88*, 147–55.
- (48) M. R. Bunagan, X. Yang, J. G. Saven and F. Gai, *J. Phys. Chem. B* **2006**, *110*, 3759–63.
- (49) Y. Cote, G. G. Maisuradze, P. Delarue, H. A. Scheraga and P. Senet, *J. Phys. Chem. Lett.* **2015**, *6*, 1082–6.
- (50) J.-P. Sauvage and V. Amendola, *Molecular Machines and Motors*, Springer, Berlin/Heidelberg/New York, **2001**, vol. 99.
- (51) R. Klajn, *Chem. Soc. Rev.* **2014**, *43*, 148–184.
- (52) K. Fujimoto, M. Amano, Y. Horibe and M. Inouye, *Org. Lett.* **2006**, *8*, 285–7.
- (53) G. Hartley, *Nature* **1937**, *140*, 281.
- (54) S. Yesodha, C. Sadashiva and N. Tsutsumi, *Prog. Polym. Sci.* **2004**, *29*, 45–74.
- (55) C. Lagrasta, I. Bellobono and M. Bonardi, *J. Photoch. Photobio. A* **1997**, *110*, 201–205.
- (56) H. Bandara and S. Burdette, *Chem. Soc. Rev.* **2012**, *41*, 1809–25.
- (57) J. Magee, W. Shand and H. Eyring, *J. Am. Chem. Soc.* **1941**, *63*, 677–88.
- (58) D. Curtin, E. Grubbs and C. McCarty, *J. Am. Chem. Soc.* **1966**, *88*, 2775–86.
- (59) G. Haberhauer and C. Kallweit, *Angew. Chem. Int. Ed.* **2010**, *49*, 2418–21.
- (60) Z. Mahimwalla, K. G. Yager, J.-i. Mamiya, A. Shishido, A. Priimagi and C. J. Barrett, *Polym. Bull.* **2012**, *69*, 967–1006.
- (61) D. G. Flint, J. R. Kumita, O. S. Smart and G. A. Woolley, *Chem. Biol.* **2002**, *9*, 391–397.
- (62) R. S. Ritterson, K. M. Kuchenbecker, M. Michalik and T. Kortemme, *J. Am. Chem. Soc.* **2013**, *135*, 12516–9.
- (63) M. Banghart, K. 3, Borges, E. Isacoff, T. Dirk and R. H. Kramer, *Nat. Neurosci.* **2004**, *7*, 1381–1386.
- (64) A. A. Beharry and G. A. Woolley, *Chem. Soc. Rev.* **2011**, *40*, 4422–37.
- (65) E. Chen, J. R. Kumita, G. A. Woolley and D. S. Kliger, *J. Am. Chem. Soc.* **2003**, *125*, 12443–9.
- (66) G. A. Woolley, *Acc. Chem. Res.* **2005**, *38*, 486–93.
- (67) A. Muller, H. Kobarg, V. Chandrasekaran, J. Gronow, F. D. Sonnichsen and T. K. Lindhorst, *Chemistry* **2015**, *21*, 13723–31.
- (68) D. C. Burns, F. Zhang and G. A. Woolley, *Nat. Protoc.* **2007**, *2*, 251–258.
- (69) O. Shimomura, F. H. Johnson and Y. Saiga, *J. Cell. Physio.* **1962**, *59*, 223–239.
- (70) T. Panavas, C. Sanders and T. R. Butt, *SUMO Fusion Technology for Enhanced Protein Production in Prokaryotic and Eukaryotic Expression Systems*, ed. H. D. Ulrich, Humana Press, Totowa, NJ, **2009**, pp. 303–317.
- (71) J. McCoy and E. La Ville, *Expression and Purification of Thioredoxin Fusion Proteins*, John Wiley and Sons, Inc., **2001**.
- (72) B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **1959**, *31*, 526–35.

- (73) P. Rovo, V. Farkas, O. Hegyi, O. Szolomajer-Csikos, G. K. Toth and A. Perczel, *J. Pept. Sci.* **2011**, *17*, 610–9.
- (74) R. M. Culik, S. Annavarapu, V. Nanda and F. Gai, *Chem. Phys.* **2013**, *422*, 131–134.
- (75) A. Bundi and K. Wüthrich, *Biopolymers* **1979**, *18*, 285–97.
- (76) D. S. Wishart, B. D. Sykes and F. M. Richards, *Biochemistry* **1992**, *31*, 1647–1651.
- (77) D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, *J. Biomol. NMR* **1995**, *5*, 67–81.
- (78) A. P. Joseph, N. Srinivasan and A. G. De Brevin<sup>1</sup>, *Amino Acids* **2012**, *43*, 1369–1381.
- (79) D. E. Stewart, A. Sarkar and J. E. Wampler, *J. Mol. Biol.* **1990**, *214*, 253–260.
- (80) M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.* **1991**, *218*, 397–412.
- (81) C. Grathwohl and K. Wuthrich, *Biopolymers* **1981**, *20*, 2623–2633.
- (82) S. Ravindranathan, J.-M. Mallet, P. Sinay and G. Bodenhausen, *J. Magn. Reson.* **2003**, *163*, 199–207.
- (83) J. R. Kumita, O. S. Smart and G. A. Woolley, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3803–8.
- (84) M. Brinkley, *Bioconjugate Chem.* **1992**, *3*, 2–13.
- (85) A. J. Lomant and G. Fairbanks, *J. Mol. Biol.* **1976**, *104*, 243–61.
- (86) E. Merino and M. Ribagorda, *Beilstein J. Org. Chem.* **2012**, *8*, 1071–1090.
- (87) J. García-Amorós and D. Velasco, *Beilstein J. Org. Chem.* **2012**, *8*, 1003–1017.
- (88) A. Patgiri, A. L. Jochim and P. S. Arora, *Acc. Chem. Res.* **2008**, *41*, 1289–300.
- (89) J. W. Taylor, *Biopolymers* **2002**, *66*, 49–75.
- (90) J. Baardsnes, L. H. Kondejewski, R. S. Hodges, H. Chao, C. Kay and P. L. Davies, *FEBS Lett.* **1999**, *463*, 87–91.
- (91) W. Gronwald, H. Chao, D. V. Reddy, P. L. Davies, B. D. Sykes and F. D. Sönnichsen, *Biochemistry* **1996**, *35*, 16698–16704.
- (92) D. Wishart and B. Sykes, *J. Biomol. NMR* **1994**, *4*, 171–180.
- (93) P. Lou and R. L. Baldwin, *Biochemistry* **1997**, *36*, 8413–21.
- (94) L. Willard, A. Ranjan, H. Zhang, H. Monzavi, R. F. Boyko, B. D. Sykes and D. S. Wishart, *Nucleic Acids Res.* **2003**, *31*, 3316.
- (95) A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross and R. Y. Tsien, *Trends Biochem. Sci.* **1995**, *20*, 448–455.
- (96) N. Pertaya, C. B. Marshall, C. L. DiPrinzio, L. Wilen, E. S. Thomson, J. Wettlaufer, P. L. Davies and I. Braslavsky, *Biophys. J.* **2007**, *92*, 3663–3673.
- (97) J. Vivian, *Biophys. J.* **2001**, *80*, 2093–2109.
- (98) O. Hecht, N. A. van Nuland, K. Schleinkofer, A. J. Dingley, H. Bruhn, M. Leippe and J. Grötzinger, *J. Biol. Chem.* **2004**, *279*, 17834–17841.

- (99) R. F. Epand, G. Wang, B. Berno and R. M. Epand, *Antimicrob. Agents Chemother.* **2009**, *53*, 3705–14.
- (100) K. Lee, S. Y. Shin, K. Kim, S. S. Lim, K. S. Hahm and Y. Kim, *Biochem. Biophys. Res. Commun.* **2004**, *323*, 712–9.
- (101) M. Lee, G. Gippert, K. Soman, D. Case and P. Wright, *Science* **1989**, *245*, 635–637.
- (102) R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.
- (103) Y. Chen, C. T. Mant and R. S. Hodges, *J. Chromatogr. A* **2007**, *1140*, 112–20.
- (104) M. Mergler, F. Dick, B. Sax, P. Weiler and T. Vorherr, *J. Pept. Sci.* **2003**, *9*, 36–46.
- (105) M. Mergler, F. Dick, B. Sax, C. Stahelin and T. Vorherr, *J. Pept. Sci.* **2003**, *9*, 518–26.
- (106) M. Mergler and F. Dick, *J. Pept. Sci.* **2005**, *11*, 650–7.
- (107) W. J. van Woerkom and J. W. van Nispen, *Int. J. Pept. Protein Res.* **1991**, *38*, 103–13.
- (108) K. Sakurai, T. M. Snyder and D. R. Liu, *J. Am. Chem. Soc.* **2005**, *127*, 1660–1661.
- (109) R. Harris, E. Becker, S. Cabral de Menezes, P. Granger, R. Hoffmans and K. Zilm, *Pure Appl. Chem.* **2008**, *80*, 59–84.
- (110) A. Shaka, C. Lee and A. Pines, *J. Magn. Reson.* **1988**, *77*, 274–93.
- (111) T. L. Hwang and A. J. Shaka, *J. Magn. Reson. A* **1995**, *112*, 275–279.
- (112) K. Wüthrich, *Biopolymers* **1983**, *22*, 131–38.
- (113) M. Mayer and B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
- (114) N. Gux and M. C. Peitsch, *Electrophoresis* **1997**, *18*, 2714–23.
- (115) Schrödinger, LLC, “Schrödinger Release 2016-4: Maestro,” 2016.
- (116) Schrödinger, LLC, “Schrödinger Release 2016-4: MacroModel,” 2016.
- (117) E. Polak and G. Ribiere, *R.I.R.O.* **1969**, *3*, 35–43.
- (118) J. Banks, H. Beard, Y. Cao, A. Cho, W. Damm, R. Farid, A. Felts, T. Halgren, D. Mainz, J. Maple, R. Murphy, D. Philipp, M. Repasky, L. Zhang, B. Berne, R. Friesner, E. Gallicchio and L. R.M., *J. Comp. Chem.* **2005**, *26*, 1752–80.
- (119) P. Cattanei and M. Persico, *Phys. Chem. Chem. Phys.* **1999**, *1*, 4739–43.
- (120) Schrödinger, LLC, “The PyMOL Molecular Graphics System, Version 1.8,” 2015.
- (121) B. J. Kuipers and H. Gruppen, *J Agric Food Chem* **2007**, *55*, 5445–51.

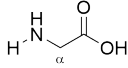
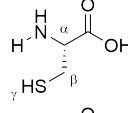
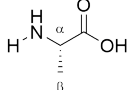
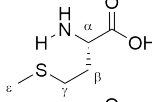
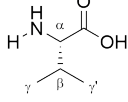
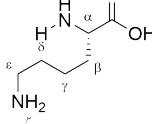
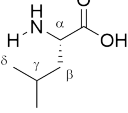
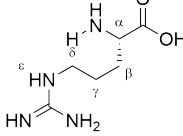
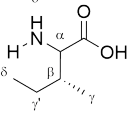
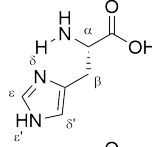
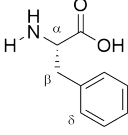
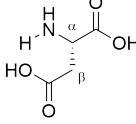
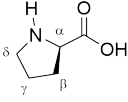
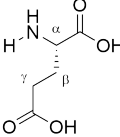
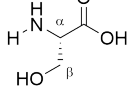
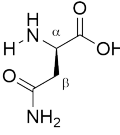
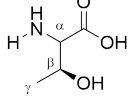
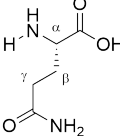
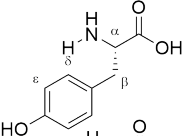
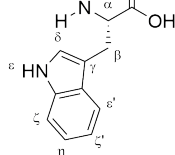


## Abbreviations

$\Delta\delta$ sum of the chemical shift deviation	<b>IBP</b> ice binding protein
$\chi_F$ fraction folded	<b>IR</b> infrared
$\delta$ chemical shift	<b>MALDI</b> matrix assisted laser desorption ionization
$\lambda$ wavelength	<b>MD</b> molecular dynamics
<b>Acn</b> acetonitril	<b>MS</b> mass spectrometry
<b>AFGP</b> antifreeze glycoprotein	<b>NHS</b> N-hydroxysuccinimide ester
<b>AFP</b> antifreeze protein	<b>NMR</b> nuclear magnetic resonance
<b>AMP</b> antimicrobial peptide	<b>NOE</b> nuclear Overhauser effect
<b>ASA</b> accessible surface area	<b>NOESY</b> nuclear Overhauser effect spectroscopy
<b>CD</b> circular dichroism	<b>PBS</b> Phosphate buffered saline
<b>COSY</b> correlation spectroscopy	<b>PLRP-S</b> poly(styrene-divinylbenzene)
<b>CSD</b> chemical shift deviation	<b>PSS</b> photo stationary state
<b>DCC</b> N,N'-dicyclohexylcarbodiimide	<b>RMSD</b> root-mean-square deviation of atomic positions
<b>DCM</b> dichloromethane	<b>RP</b> reverse-phase
<b>DIPEA</b> N,N-diisopropylethylamine	<b>SPPS</b> solid phase peptide synthesis
<b>DMB</b> 1,3-dimethoxy benzene	<b>T<sub>M</sub></b> melting point
<b>DMF</b> dimethylformamide	<b>t<sub>R</sub></b> retention time
<b>DMSO</b> dimethyl sulfoxide	<b>TCEP</b> tris-(2-carboxyethyl)-phosphin
<b>ESI</b> electron spray ionization	<b>TFA</b> trifluoroacetic acid
<b>Fmoc</b> Fluorenylmethyloxycarbonyl	<b>TH</b> thermal hysteresis
<b>GFP</b> green fluorescent protein	<b>TIS</b> Triisopropylsilane
<b>GLP-1</b> glucagon-like-peptide-1	<b>TOCSY</b> total correlation spectroscopy
<b>HATU</b> 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate	<b>ToF</b> time of flight
<b>HOBt</b> Benzotriazol-1-ol	<b>UV</b> ultra violet
<b>HPLC</b> high pressure liquid chromatography	<b>UV/vis</b> ultra violet and visible light
<b>HSQC</b> hetero single quantum coherence	



## List of the 20 Proteinogenic Amino Acids

Structure	Name	Abbreviations	Structure	Name	Abbreviations
	Glycine	Gly, G		Cysteine	Cys, C
	Alanine	Ala, A		Methionine	Met, M
	Valine	Val, V		Lysine	Lys, K
	Leucine	Leu, L		Arginine	Arg, R
	Isoleucine	Ile, I		Histidine	His, H
	Phenylalanine	Phe, F		Aspartic Acid	Asp, D
	Proline	Pro, P		Glutamic Acid	Glu, E
	Serine	Ser, S		Asparagine	Asn, N
	Threonine	Thr, T		Glutamine	Gln, Q
	Tyrosine	Tyr, T			
	Tryptophan	Trp, W			