3D-structure of synthetic peptides and molecular interactions in biomimetic media probed by NMR spectroscopy

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Linda Jütten
aus Viersen

Köln, 2021

Die vorliegende Arbeit wurde im Zeitraum von April 2017 bis April 2021 am Institut für Organische Chemie der Universität zu Köln unter der Anleitung von Dr. Dolores Diaz angefertigt.

Berichterstatter	Prof. Dr. Albrecht Berkessel
	PD Dr. Martin Breugst
Vorsitzende	Prof. Dr. Ines Neundorf
Beisitzerin	Dr. Heike Henneken
Tag der Disputation	08.06.2021

Meinen geliebten Menschen

Our responsibility is to do what we can,

learn what we can,

improve the solutions

and pass them on.

— Richard Feynman —

Danksagung

An erster Stelle danke ich Frau Dr. Diaz und Herrn Dr. Schlörer ganz herzlich für die Möglichkeit die sie mir gaben, meine Arbeit in ihrer Arbeitsgruppe anfertigen zu können. Ich habe bei ihnen in den letzten Jahren mehr gelernt, als ich jemals für möglich gehalten hätte.

Außerdem bedanke ich mich ganz herzlich bei Herrn Prof. Berkessel für die freundliche Übernahme meines Erstgutachtens und seine stete Unterstützung.

Herrn PD. Martin Breugst danke ich sehr herzlich für die Übernahme des Zweitgutachtens.

Frau Prof. Ines Neundorf danke ich nicht nur für die Übernahme des Vorsitzes, sondern auch sehr für das interessante Kooperationsthema und den wissenschaftlichen Austausch.

Ein besonderer Dank geht an meine ehemaligen Kolleginnen, Frau Katrin König und Frau Daniela Naumann. Sie beiden sind das Herz der Abteilung und ich hoffe Ihnen wird immer der nötige Respekt zuteil. Sie haben mich die letzten Jahren immer aufrichtig und helfend begleitet, das werde ich nie vergessen. Frau König, Sie sind eine bedingungslos mitfühlende Frau vor der ich einen sehr tiefen Respekt hege. Frau Naumann, ich hoffe wir werden uns noch oft über Strickvorschriften und Blumen austauschen. Bitte verlieren Sie niemals diese Studentennähe, denn diese wird von mehr Studenten geschätzt als Sie vermuten.

Ich möchte mich auch sehr bei Herrn Dr. Philip Hegemann bedanken für die konstante Wartung der, für mich essentiell notwendigen, Geräte und die netten Unterhaltungen.

Des Weiteren möchte ich mich bei all meine Freunden bedanken für die schöne und auch wissenschaftlich konstruktive Zeit die wir miteinander verbracht haben. Der "alte Arbeitskreis ’Goldis’": Falco Fox, Florian Wolf, Eric Brühning, Florian Dato. Es gab immer einen warmen Kaffee, einen guten Rat und eine stützende Schulter bei euch.

Arbeitskreis Griesbeck, ihr seid einfach zu viele! Wir haben so viel Zeit privat und bei der Arbeit verbracht. Ich schätze jede Konversation und Unterhaltung die ich mit euch führen durfte, allen voran: Moritz Vollmer und Seyma Boskus. Ich danke dem Arbeitskreis Berkessel/Breugst für konstruktive Unterhaltungen und Ratschläge. Dabei ist besonders Dr. Jonas König hervorzuheben, der nicht nur ein Kollege war, sondern mittlerweile ein sehr guter Freund. Einen besonderen Dank möchte ich auch an Prof. Schmalz und seinen Arbeitskreis wenden. Es war mir eine Freunde an Ihren Seminaren teil nehmen zu dürfen. Besonders möchte ich Dr. Hanna Sebode und Dr. Ömer Taspinar hervorheben, die immer für einen guten Rat zur Seite standen.

Ich möchte an dieser Stelle auch nicht meine engsten Date-Donnerstag-Freunde unterschlagen. Tim Michael Luigi Lippold ist nicht nur mein 'Haus-Freund', sondern hat mir auch in jeder schwierigen Situation zur Seite gestanden. Und wenn es nur die Sichtung meiner gefluteten Wohnung war. Katharina Susanne Eveline von Schoeler, du bist meine beste Freundin und auch gleichzeitig einer der faszinierendsten Menschen die ich kenne. Du bist nicht nur wahnsinnig intelligent sondern auch mitfühlend und bedingungslos hilfsbereit. Ein Dank geht an dieser Stelle auch an dein Helfersyndrom, denn auf dich kann ich mich immer verlassen. Ich hoffe wir werden es in 2021 noch auf einen Wein in ein Restaurant schaffen! Dominik Albat, du bist einer der interessantesten Menschen die ich je kennen lernen durfte und auch gleichzeitig einer der fokussiertesten, arbeitswütigsten, verrücktesten und liebenswertesten! Und der letzte Platz ist für dich reserviert, Lars Hemmersbach. Es ist mir immer wieder ein Genuss mit dir über einen guten Tropfen Wein, einen angesagten Gin oder die Zubereitung von schmackhaftem Essen zu referieren. Deine Bereitschaft mich bei der Fertigstellung meiner Arbeit zu unterstützen ist mit nichts aufzuwiegen. Ohne eure Hilfe hätte ich das niemals geschafft. Ihr seid nicht nur Freunde, sondern Familie.

Meiner Familie danke ich natürlich auch an dieser Stelle: Mama, du bist die inspirierendste, unterstützendste und selbstloseste Mama die es gibt! Ich könnte mir keine perfektere Mutter vorstellen! Ich danke auch meinem Bruder Alexander Jütten und meinen Vater Herbert Hürtgen für die konstante Unterstützung. Ich glaube, wir alle sind das, was man eine Familie nennt. Danke für alles.

Kurzzusammenfassung

Die Kernspinresonanzspektroskopie (engl. *Nuclear magnetic resonance*, abk. NMR) ist eine leistungsstarke Methode um Vorgänge biologischer Makromoleküle auf atomerer Ebene zu charakterisieren. Die Analyse von molekularen Strukturmerkmalen ermöglicht die spezifische Verknüpfung einer Struktur mit ihrer zugehörigen Aktivität. Die Erforschung derart komplexer Vorgänge begründete die Entwicklung der heutigen biomolekularen NMR.

Bakterien haben verschiedene Strategien entwickelt, um in einem Wirt zu überleben. Eine dieser Strategien ist die stetige Translation zu vorteilhaften orten im Wirtsorganismus. Ein bekanntes Beispiel ist das Protein PPEP-1, welches die translatorische Bewegung durch selektives Spalten von Prolin–Prolin Bindungen initiiert. Bisher war nicht bekannt ob die zu adressierenden Prolin–Prolin Bindung eine bevorzugte Konformation für eine effektive Spaltung einnehmen muss. Durch eingehende NMR spektroskopische Untersuchungen konnte eine eindeutig bevorzugte Konformation der Bindung identifiziert werden. Diese Erkenntnis trägt zu einem tieferen Verständnis der Überlebensstrategie bei. Somit könnten sich potentiell neue Ansatzpunkte für Behandlungsmöglichkeiten ergeben.

Die gezielte Behandlung von kanzerogenen Zellen unterliegt noch heute einem großen Forschungsinteresse. Zellpenetrierende Peptide könnten einen Lösungsansatz für die selektive Adressierung dieser darstellen, wobei der genaue Aufnahmemechanismus der Peptide bislang nicht hinreichend analysiert wurde. In dieser Arbeit konnte durch NMR spektroskopische Methoden gezeigt werden, dass bei der Wechselwirkung zwischen derartigen Peptiden mit Membranmimetika, das Peptid eine Konformationsänderung durchläuft. Dies stellt einen weiteren Ansatzpunkt in der Aufklärung des Aufnahmemechanismus dieser untersuchten Peptide dar.

Im Prozess der Proteinfaltung ist die Mechanismusaufklärung ebenfalls von großer Bedeutung. Fehlgefaltete Proteine sind oftmals die Ursache von verschiedenen Krankheiten. Die Rückfaltung der Proteine in ihre funktionsfähige Form ist indes Gegenstand aktueller Forschung. Kleine organische Moleküle können dabei als Faltungshelfer fungieren. Die

räumlichen Nähe dieser Moleküle zu kleinen peptidabgeleiteten Modelsystemen konnte mit Hilfe der NMR Spektroskopie bewiesen werden. Somit wurde im Rahmen dieser Arbeit ein Ansatz etabliert mit welchem der Wirkort von Faltungshelfern an einem Protein in Zukunft lokalisiert werden könnte.

Modelsysteme sind in der NMR Spektroskopie wichtig, um u.a. Wechselwirkungsmechanismen von kleinen Molekülen zu makromolekularen Systemen zu untersuchen. Die Untersuchung räumlicher Annäherungen der Interaktionspartner ist dabei eine der essentiellen Aspekte. Hierfür eignet sich besonders die Verwendung von Mizellen. Verschiedene Moleküle wurden in Hinblick ihrer Positionierung zu Mizellen mit einer Vielzahl an NMR spektroskopischen Methoden untersucht. Die Erkenntnisse dieser Untersuchungen wurden im Anschluss auf einen postulierten Reaktionsmechanismus übertragen.

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is a powerful method to characterise the processes of biological macromolecules at an atomic level. The analysis of molecular structure properties enables the specific linkage of a structure with its associated activity. The study of such complex processes motivated the development of today's biomolecular NMR.

Bacteria have developed different strategies to survive in a host. One of these strategies is the continuous translation to advantageous habitats in the host organism. A well-known example is the protein PPEP-1, which initiates translational movement by selectively cleaving proline-proline bonds. So far, it was not known whether the addressed proline-proline bond has to adopt a preferred conformation for effective cleavage. Through detailed NMR spectroscopic investigations, a clearly preferred conformation of the bond could be identified. This finding contributes to a deeper understanding of the survival strategy. Thus, potentially new starting points for treatment options could emerge.

The targeted treatment of carcinogenic cells is still of great research interest today. Cell-penetrating peptides could represent a solution for the selective addressing of these, whereby the exact uptake mechanism of the peptides has not yet been sufficiently analysed. In this work, NMR spectroscopic methods were used to show that the interaction between such peptides and membrane mimetics causes the peptide to undergo a conformational change. This represents a further approach in the elucidation of the uptake mechanism of these investigated peptides.

Mechanism elucidation is of great importance in the process of protein folding as well. Misfolded proteins are often the cause of various diseases. The refolding of proteins into their functional form is the subject of current research. Small organic molecules can act as folding helpers. The spatial proximity of these molecules to small peptide-derived model systems could be proven with the help of NMR spectroscopy. Thus, in the context of this work, an approach was established with which the site of action of folding helpers on a protein could be localised in the future.

Model systems are important in NMR spectroscopy to investigate, among other things, interaction mechanisms of small molecules to macromolecular systems. The investigation of spatial approximations of the interaction partners is one of the essential aspects. The use of micelles is particularly suitable for this purpose. Various molecules were investigated with regard to their positioning in micelles using a variety of NMR spectroscopic methods. The findings of these investigations were then transferred to a postulated reaction mechanism.

Contents

Abstract v

List of Abbreviations xv

I. GENERAL BACKGROUND AND THEORY

1. Introduction to the application of NMR in the field of biomolecules and life science 3

2. Application of NMR in the field of life science 7

2.1. Proteins, peptides, amino acids . . . 7
2.2. Secondary structure . . . 9
2.3. Basics of NMR . . . 11
2.4. Peptide structure elucidation by NMR spectroscopy . . . 14
  2.4.1. Sample requirements in BioNMR . . . 15
    2.4.1.1. pH value of the sample . . . 15
    2.4.1.2. The experimental temperature . . . 15
  2.4.2. Paramagnetic reagents . . . 17
2.5. The experimental toolbox for BioNMR . . . 17
  2.5.1. *TO*tal *C*orrelation *S*pectroscop*Y* – TOCSY . . . 17
    2.5.1.1. Assignment of the spin systems . . . 18
  2.5.2. *N*uclear *O*verhauser *e*ffect *S*pectroscop*Y* – NOESY . . . 19
    2.5.2.1. Sequential assignment . . . 23
  2.5.3. *Tr*ansfer *NOE* – trNOE . . . 24
  2.5.4. *S*aturation *t*ransfer *d*ifference – STD . . . 25
  2.5.5. *water-l*igand *o*bserved via *g*radient *s*pectroscop*y* – waterLOGSY . 26
  2.5.6. *D*iffusion-*o*rdered *S*pectroscop*Y* – DOSY . . . 27
2.6. From experimental data to 3D structure . . . 28

II. NMR-BASED CONFORMATIONAL STUDY OF PROLINE-RICH PEPTIDES

1. Background 33

1.1. Proline associated peptide bonds and their influence on peptide/protein folding . . . 33

1.2. Relevance of proline for the bacterium *Clostridium difficile* . . . 34

1.2.1. Mechanism of action of *C. difficile* to colonise the host intestine . . 34

1.2.1.1. PPEP-1 involvement in the motility of *C. difficile* 36

1.2.1.2. Structure of PPEP-1 . . . 36

2. Goals 39

3. Material and Methods 41

4. Results and Discussion 43

4.1. NMR assignment of proline-rich peptide sequences 4-6 . . . 44

4.1.1. Identification of spin systems . . . 45

4.1.2. Sequential assignment of the spin systems . . . 47

4.1.3. 3D structure calculation based on experimental NMR data . . . 49

4.1.4. NMR experiments on the recognition process . . . 50

4.2. An exploratory computational investigation . . . 52

5. Summary and Outlook 55

III. STRUCTURES OF PEPTIDES IN SOLUTION AND IN MEMBRANE MIMETIC MEDIUM

1. Background 59

1.1. Design of CPPs . . . 59

1.2. Internalisation mechanism and classification of CPPs . . . 61

2. Goals 65

3. Material and Methods 67

4. Results and Discussion **71**

4.1. Studied DKP-peptidomimetics 71

4.1.1. Resonance assignment: spin system identification 72

4.1.2. Sequential assignment of the DKP-sC18* peptides 73

4.1.3. Temperature Coefficients 75

4.2. Structural analysis 76

4.2.1. Molecular modeling studies 77

4.3. DKP-peptidomimetics in membrane mimetic agents 78

4.3.1. Resonance assignment of peptidomimetics in presence of micelles . 80

4.4. Sequential assignment of peptidomimetics 81

4.5. Structural analysis 82

5. Summary and Outlook **85**

IV. FIRST STEPS IN NMR SPECTROSCOPIC STUDIES OF INTER-MOLECULAR INTERACTION MECHANISMS OF OSMOLYTES

1. Background **89**

1.1. Folding mechanism and pathway 90

1.2. Misfolding and disease 91

1.2.1. Chaperons 91

1.2.2. The chemical chaperons: Osmolytes 92

1.2.2.1. Mechanism of osmolytes 93

1.3. Model systems used in BioNMR research 94

2. Goals **97**

3. Materials and methods **99**

4. Results and discussion **101**

5. Summary and outlook **109**

V. ^{1}H NMR STUDIES ON THE INTERACTIONS BETWEEN SMALL MOLECULES AND MICELLES IN WATER

1. Background 113

1.1. Membrane-mimicking media in BioNMR 113

1.2. Micelle formation 114

1.3. The Diels-Alder reaction 115

2. Goals 119

3. Materials and Methods 121

4. Results and discussion 123

4.1. Monitoring the formation of micelles in aqueous solution 123

4.2. Positioning of small molecules in a SDS micelle 125

4.2.1. Chemical shift monitoring 126

4.2.2. Diffusion studies on the mode of interaction 127

4.2.3. 1D inter-molecular NOE investigations 127

4.2.4. Investigations of binding properties using waterLOGSY 128

4.2.5. Titration with paramagnetic agents 130

4.3. Micellar catalysis of a Diels-Alder reaction – an exploratory NMR study . 132

4.3.1. The reaction mechanism 133

5. Summary and Outlook 137

VI. APPENDICES

A. Bibliography iii

B. Non-standard residue library files xxxiii

C. Structure chart of selected amino acids xxxix

D. Appendices Part I xli

D.1. Proton assignment of peptides **4-6** xlii

D.2. Summary tables of NMR-derived structure calculations of proline-rich peptides 4-6 xliii
D.3. NMR derived 3D structure ensembles xliv
D.4. Selected spectra of Proline-rich peptides 4-6 xlv

E. Appendices Part II **xlix**
E.1. Sample preparation of DKP peptidomimetics xlix
E.2. Proton assignment of DKP peptides li
E.3. Summary tables and structure ensembles of NMR-derived structure calculations of DKP-linked peptides lxi
E.3.1. Linear Peptides in aqueous medium lxi
E.3.2. Cyclic Peptides in aqueous medium lxiii
E.3.3. Linear Peptides in presence of SDS lxv
E.4. NMR spectra of studied DKP-sC18* peptidomimetics lxix

F. Appendices Part III **lxxv**
F.1. Proton and carbon assignment of the Trp-cage mini-protein lxxv
F.2. Summary tables of NMR-derived structure calculations of Trp-cage miniprotein lxxix
F.3. Selected analytics of ^{15}N TMAO lxxx

G. Spectra and tables of ^{1}H NMR studies of interaction on small molecules and micelles **lxxxi**

H. Contribution from others **xcvii**

List of Abbreviations

%	percent
1D	one dimensional
2D	two dimensional
Å	Angström
Abz	2-aminobenzoyl
approx	approximately
CAMP	cationic antimicrobial peptides
CD	circular dichroism
cf	confer
CPP	cell-penetrating peptide
cGV	cyclo-glycylvaline
cyc	cyclic
d	doublet
DOSY	diffusion ordered spectroscopy
DKP	diketopiperazine
Dnp	2,4-dinitrophenyl
e. g.	(exempli gratia) for instance

HIFP	hexafluoroisopropanol
i. e.	(id est) that is
J	couppling constant
lin	linear
K	kelvin
LPS	lipopolysaccharides
MC	monte-carlo
MD	molecular dynamics
mg	milligram
MHz	megahertz
mM	milli molar
ms	milli seconds
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser spectroscopy
NVF	*N*-vinylformamide
PBS	phosphate buffered solution
PPEP-1	proline-proline endopeptidase-1
ppb	parts per billion
ppm	parts per million
rf	radio frequency

rmsd	root-mean square deviation
SAR	structure activity relationship
sc	side chain
SD	stochastic dynamics
SDS	sodium dodecyl sulfate
STD	saturation transfer difference
TAT	trans-activator of transcription
TMAO	trimethylaminoxid
TOCSY	total correlation spectroscopy
TSP	trimethylsilylpropanoic acid
waterLOGSY	water-ligand observed via gradient spectroscopy
Xaa	arbitary amino acid but proline

Part I.

GENERAL BACKGROUND AND THEORY

1. Introduction to the application of NMR in the field of biomolecules and life science

The understanding of macroscopic phenomena in physics, chemistry and biology is nowadays based on the detailed description of their components at a microscopic level.
The chemist and biochemist *Louis Pasteur* changed medicine by making pioneering discoveries that led to the development of the first rabies and anthrax vaccines in 1870-1880s.[1] *Pasteur* was already convinced that the explanation of a phenotypic expression of a system must be sought on a molecular basis. He proved this by correlating the optical activity of tartaric acid with its asymmetry at the molecular level.[2] His conclusion contributed to a fundamental understanding of structures in organic chemistry.
Several milestones of science have one essential thing in common: a fundamental description at a molecular level was the key step to progress and for further development. The challenges in modelling these systems still remain in the structural elucidation of biochemically relevant molecules as well as in the quantitative analysis of the chemical composition of biological structures. Suitable analytical methods have been developed for investigations at the atomic level. In structural biology, X-ray crystallography and NMR spectroscopy are the two largest and most established methods. Until the early 1980s, protein structures at an atomic level were solved exclusively by X-ray-based methods. Crystallography yields an high resolution and is not limited by the molecular weight of the sample. The majority of published and deposited structures in data bases were resolved by crystallography.[3]
The basic requirement for crystallography is a crystal of very high quality. Obtaining a crystal from a protein is not always possible. In addition, crystals are a rigid view of a biological system which is dynamic in the living organism.[4] In contrast to crystallography, NMR spectroscopy offers the possibility to solve 3D macromolecular structures on an atomic level in solution. The investigations of biomolecules in solution opens new doors to the studies of life-like systems.

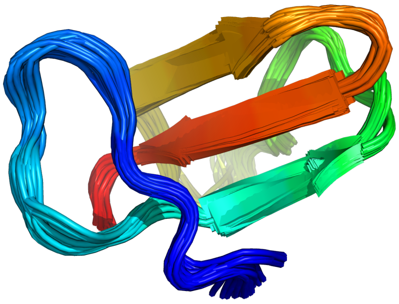


Figure 1.1.: 3D solution structure which was solved by NMR spectroscopy and hybrid distance geometry-dynamical simulated annealing (pdb-code: 2bds).[5]

NMR spectroscopy is nowadays deeply rooted in many scientific fields as chemistry, physics, medicine and biology. It found its beginning with the discovery of the nuclear resonance by *Bloch*[6] and *Purcell.*[7] This discovery was so innovative that both were rewarded with the Nobel prize in 1952 "for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith".[i] Since then, NMR spectroscopy has become increasingly relevant. The growing interest in NMR methods required not only rapid development of hardware and software, but also more sophisticated acquisition techniques. One of the most important achievements was the introduction of a second dimension of *Ernst*[8] in 1976, which laid the foundation for modern high-resolution 2D NMR spectroscopy, which was also awarded the Nobel Prize ("for his contributions to the development of the methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy.").[ii] 2D NMR quickly attracted interest in the biochemical field, where *Kurt Wüthrich*'s group first published a spectrum of a small globular protein.[9] With the transfer of NMR to biochemistry, it quickly became clear that the achievable sensitivity was not sufficient. The optimisation of the performance was realised by the introduction of isotopic labelling strategies of N and C nuclei, which previously could not be detected in biochemical experiments.

The employment of labelled molecules subsequently led to the establishment of a third dimension, making triple resonance experiments possible. The once small branch of NMR in biochemistry has developed within a few decades into a completely separate scientific field, called BioNMR. This reputation of BioNMR was awarded with the Nobel Prize for the establishment of the structure solution of unlabelled peptides to *Wüthrich* in 2002 "for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution"[iii]. Within the field of BioNMR, omics science has emerged in the last two decades. In omics, a certain distinction is made in the type of data collection, processing and evaluation. The evaluation process can be performed qualitative or quantitative. A quantitative evaluation is carried out, for example, with blood or cervical fluid, where the metabolomes are profiled using statistical methods. The interest in the entire composition of a solution without separating it in the laboratory is generally referred to as metabolomics.[10] Qualitative studies on the 3-dimensional peptide structures correspond to the field of proteomics. In addition to the exact structure solution, kinetic studies of protein interactions, screening of drug motifs and dynamic studies of protein folding are included in proteomics.[11]

---

[i] MLA style: The Nobel Prize in Physics 1952. NobelPrize.org. Nobel Media AB 2021. Mon. 15 Mar 2021. <https://www.nobelprize.org/prizes/physics/1952/summary/>

[ii] MLA style: Richard R. Ernst – Biographical. NobelPrize.org. Nobel Media AB 2021. Wed. 24 Mar 2021. <https://www.nobelprize.org/prizes/chemistry/1991/ernst/biographical/>

[iii] MLA style: The Nobel Prize in Chemistry 2002. NobelPrize.org. Nobel Media AB 2021. Thu. 11 Mar 2021. <https://www.nobelprize.org/prizes/chemistry/2002/summary/>

2. Application of NMR in the field of life science

The 3D structure evaluation of biological macromolecules by NMR involves many steps that have to be carried out to obtain a reliable structure. At first, biomolecules and their properties will be introduced as well as some basics of NMR spectroscopy. The part on the application of BioNMR describes aspects of sample preparation followed by the experimental tool box for 3D solution structure assignment. In this work unlabelled peptides were investigated, which have been only available in very low quantities. The description of the experiments used is expanded to include sophisticated methods such as saturation transfer and a water-ligand observed approach, which were implemented for this work. Experiments like the total correlation experiment have not received much attention here so far, though it is fundamental for the structure elucidation of polypeptide based molecules. Finally, 3D structure calculation derived from NMR experimental data is presented.

2.1. Proteins, peptides, amino acids

The general transition between the terminology 'protein–oligopeptide/polypeptide– peptide' is fluent. *Polypeptides* are roughly defined as a peptide chain of 10–50 amino acids, whereby oligopeptides consists of less than 10. *Peptides* can adopt certain secondary structures or form random coil, which depends on many factors as for instance specific amino acid sequence, charge of residues, pH value or solvent.[12]

In the primary structure, the specific amino acid as well as their connectivity within the peptide are considered, which are linked to each other via so-called peptide bonds. A peptide bond, as can be seen in Figure 2.1, is formed formally by the condensation of the carboxyl function of the preceding amino acid with the amino group of the succeeding one. By convention it is read from left to right, e. g. from the N- to the C-terminus.[12]

The three main torsion angles which belong to the peptide backbone N_i–C^{α}_i–C_i-O_i, are defined as ω_i ($\mathrm{C}^{\alpha}_{i-1}$–$\mathrm{C}_{i-1}$–$\mathrm{N}_i$–$\mathrm{C}^{\alpha}_i$), ϕ_i (C_{i-1}–N_i–C^{α}_i–C_i) and ψ_i (N_i–C^{α}_i–C_i–N_{i+1}), as is

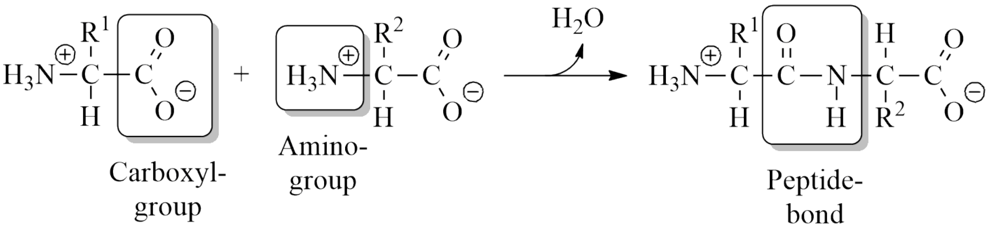

H3N
R1
C
O
H
+
Carboxyl-group
H3N
R2
Amino-group
H2O
Peptide-bond

Figure 2.1.: Formation of a peptide bond between two amino acids through a condensation reaction.[12]

schematically shown in Figure 2.2.[13]

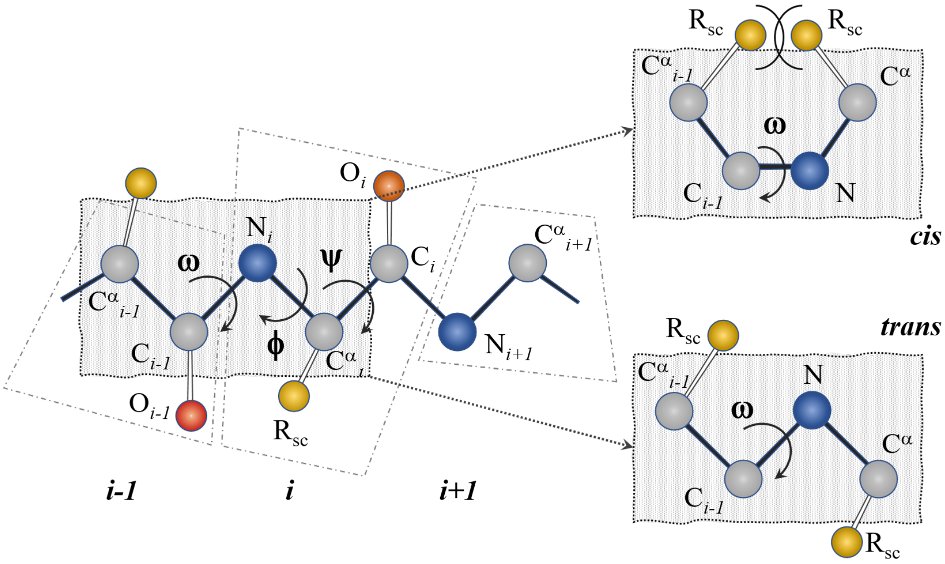

Rsc
Cα i-1
ω
Ci-1
N
Cα
cis
trans
Oi
Ni
ψ
Ci
Cα i+1
φ
Ni+1
Oi-1
i-1
i
i+1

Figure 2.2.: Schematic drawing of the definition of dihedral/torsion angles of a peptide backbone (left) with the definition of *cis*- (ω=0°) and *trans*-conformation (ω=180°) (R_{sc}: side chain of any amino acid but proline).[12]

The covalent peptide bond has a special characteristic, which is the delocalisation of nitrogens lone-pair electrons on the C–N bond, resulting in a partial double bond character. Because of this, the C–N flexibility is clearly hindered in its free rotation with a barrier of around 65–90 kJ mol^{-1}. Furthermore, this has the consequence that this planar appearing bond underlies a *cis*-(synperiplanar) and *trans*-(antiperiplanar conformation) isomerisation, with a clear preference fot the energetically favoured *trans*-conformation.[13]

2.2. Secondary structure

The secondary structure of a protein is defined by regularly repeating local structural elements. When considering these structural elements, the focus is particularly on the conformation of the peptide backbone. These regular structures are stabilised by intramolecular interactions such as hydrogen bonds. The most common secondary structures are α-helices, β-sheets and β-turns. The possibility of forming structural elements depends on the amino acid and the amino acid sequence. Proline in particular is known to be a helix-breaker and forms polyproline helices instead.[12]

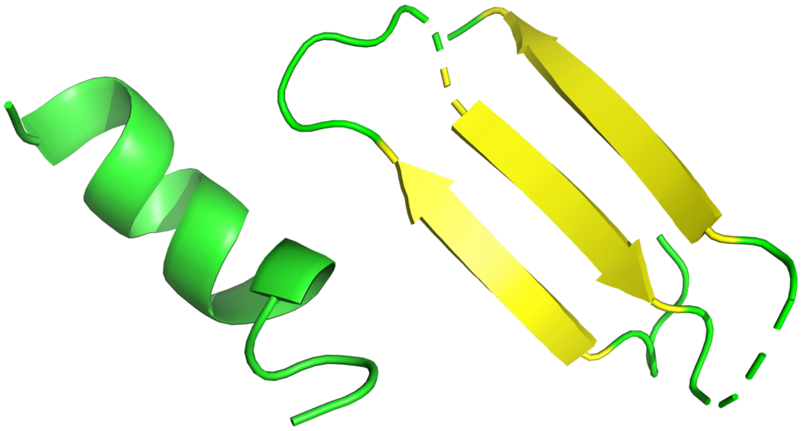


Figure 2.3.: Two frequently observed secondary structure elements: an α-helix (left) and β-sheet (right).(NMR derived structures are taken from the BMRB: α-helix: penetratin, DOI: doi:10.13018/BMR5290, pdb-code: 1kz2, pdb-code: 1KZ2 , β-sheet: part of the sequence doi:10.13018/BMR17375, pdb-code: 1axc)

NMR spectroscopy can be used to determine the secondary structure of peptides and proteins. One approach is to evaluate the coupling constants between ${}^3J_{H^\alpha H^\mathrm{N}}$ which are directly related to the backbone dihedral angle ϕ and differ for each structural element. Coupling constants in proteins, however, are hardly evaluable, among other things because of a too severe signal overlap.[14]

Another possibility is the use of structural information coming through space by internuclear interactions. Each structural element has an individual signal pattern between interacting protons. Figure 2.3 summarises the NOE pattern which is observed for several secondary structures.

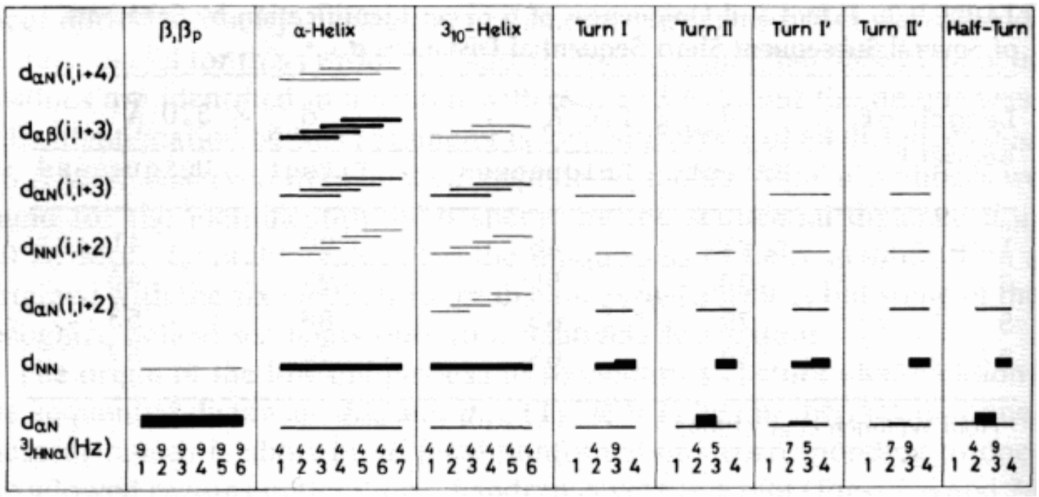

β,β_P
α-Helix
3_10-Helix
Turn I
Turn II
Turn I'
Turn II'
Half-Turn
dαN(i,i+4)
dαβ(i,i+3)
dαN(i,i+3)
dNN(i,i+2)
dαN(i,i+2)
dNN
dαN
³J_HNα(Hz)
9 9 9 9 9 9
1 2 3 4 5 6
4 4 4 4 4 4 4
1 2 3 4 5 6 7
4 4 4 4 4 4
1 2 3 4 5 6
4 9
1 2 3 4
4 5
1 2 3 4
7 5
1 2 3 4
7 9
1 2 3 4
4 9
1 2 3 4

Figure 2.4.: NOE pattern for different secondary structure elements. Thickness of lines shows how likely this spacial contact is.[15]

The dihedral angles of the peptide backbone can only adopt certain values, which are mainly dependent on the amino acid sequence. A conformational analyses of the angles of proteins can be performed with a Ramachandran plot. This representation is therefore an important way to describe secondary structure of proteins. Secondary structure is characterised by the free rotation of the dihedral angles ϕ and ψ of the peptide backbone. ω is the peptide bond, which is normally at 180°due to its partial double bond character. The rotation of ϕ and ψ is measured from –180°to 180 ° for each angle, as shown in Figure 2.5.[14]

Numerous angle combinations are not allowed because they are associated with steric hindrance. The energetically allowed conformational spaces of ϕ/ψ angles are shown as core regions. Each nucleus region therefore contains combinations of angles that are indicative of a particular secondary structure.[16]

The secondary structure is of fundamental importance for the function of a protein as it determines its biological function. This relation is called *s*tructure-*a*ctivity *r*elationship (SAR). On this basis, investigations towards the function of a biomolecule and their ligand plays a central role for the fragment-based drug discovery. The combination of 'SAR by NMR' and computational data enables a screening of huge ligand libraries of targeted proteins.[17]

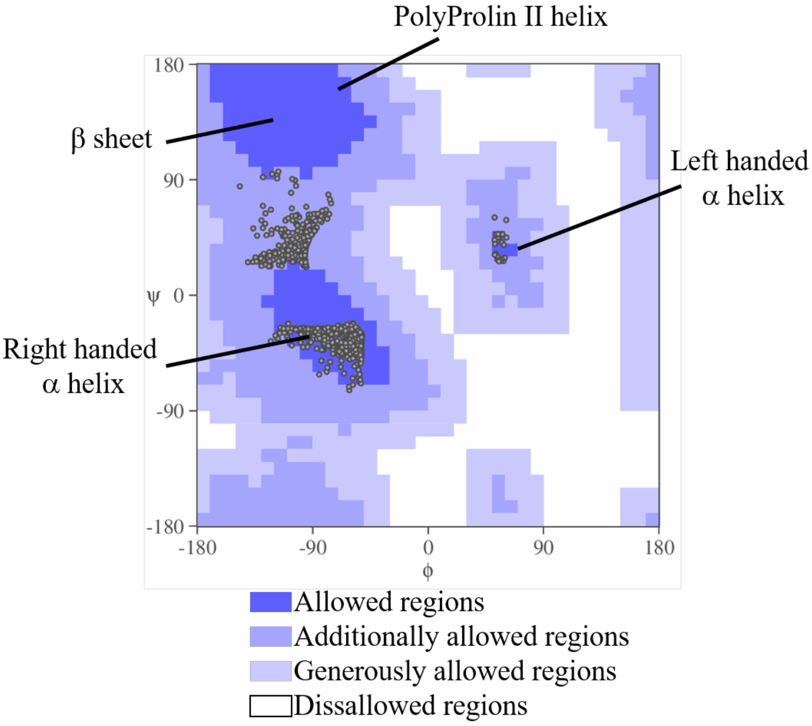


Figure 2.5.: Representation of a general ramachandran plot.

2.3. Basics of NMR

*N*uclear *m*agnetic *r*esonance spectroscopy, commonly referred to as NMR, is widely used for determining the structure of (in)organic molecules and is already well established in the structure elucidation of peptides and proteins. This method provides not only information about the raw structure of a molecule but also about reaction states, dynamics over a wide range of time scale, interaction processes on an atomic basis and also screening of potential ligands is possible in the field of drug development. In order to gain a deeper understanding of the used method it is necessary to introduce some basic physical principles.[4,18]

Nuclear magnetic resonance spectroscopy is a technique using the interaction of the magnetic moment of an atomic nucleus and an external magnetic field. The magnetic moment $\vec{\mu}_N$ of an atomic nucleus is proportional to its nuclear spin $\vec{I}$ and given by

$$\vec{\mu}_N = \gamma_N \vec{I}, \tag{2.1}$$

where γ_N denotes its characteristic gyromagnetic ratio. It follows that this interaction can only be observed for nuclei with a non-zero nuclear spin.[4,19]

If one considers a classical magnetic moment $\vec{mu}$ in a magnetic field $\vec{B}$, its energy E depends on the magnitude and relative orientation of these two vector quantities and is mathematically described by

$$E = -\vec{\mu} \cdot \vec{B}. \tag{2.2}$$

Analogously, in the quantum mechanical system, the energy of an NMR-active atomic nucleus also depends on its specific magnetic moment and the external magnetic field. Thereby, this external field is influenced by other nearby atomic nuclei as well as the electronic structure of the environment. Therefore, NMR spectroscopy allows for conclusions in terms of spatial structures.[19,20]

The magnetic field to which the nucleus is exposed creates a preferred direction in an otherwise isotropic space. This so-called quantisation axis along the magnetic field is usually defined as the z-axis of the Cartesian coordinate system. A nuclear energy level that would itself be degenerate thus splits into different energy levels. For an atomic nucleus with nuclear spin quantum number I, this results in $2I + 1$ different states quantised along the magnetic field direction, to which an additional magnetic quantum number m is assigned. The quantum mechanical uncertainty of the x- and y-direction of the nuclear spin can be visually associated with the precession motion of a classical angular momentum. Similarly, in quantum mechanics, the expectation value of the nuclear spin vector rotates around the quantisation axis at an angular velocity ω_0, which is given by

$$\omega_0 = -\gamma_N B \tag{2.3}$$

and is called Larmor frequency.[19,20]

A spin-half nucleus, as it is for hydrogen for example, has two different energy states in a magnetic field B. One of those states is a lower energy state with $m = +\frac{1}{2}$ (α or 'spin-up'), which is aligned with the external field. The other one is a higher energy state with $m = -\frac{1}{2}$ (β or 'spin-down'), which is opposed to the field. For the gyromagnetic ratio γ of the proton this results in an energy difference of

$$\Delta E = \gamma \frac{h}{2\pi} B. \tag{2.4}$$

In Figure 2.6 the spatial quantisation of a spin-half nucleus (red) and a spin-one nucleus (green) are shown.[4,19,21]

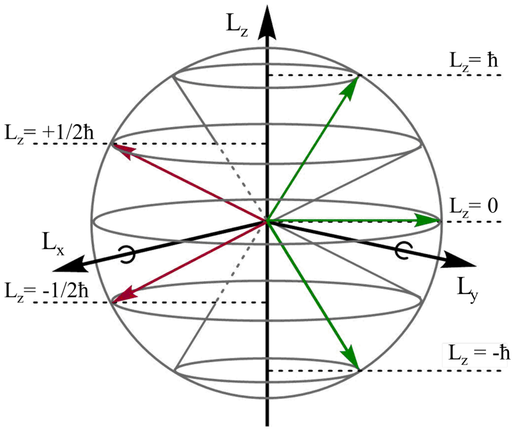


Figure 2.6.: Illustration of the spatial quantisation for spin-half (left side of the sphere) and spin-one (right side of the sphere) nuclei. The values $-I$ to $+I$ for three states are given in integer steps for spin-one nuclei.[19]

Given an ensemble of such two level systems, the population difference in thermal equilibrium is described by the Boltzmann distribution. According to this, the ratio of the populations of the ^{1}H system is given by

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{k_B T}}, \tag{2.5}$$

where k_B denotes the Boltzmann constant and T the temperature. This population difference can be changed by electromagnetic irradiation at a radio frequency. Resonance occurs when the frequency of the electromagnetic field ν matches the Larmor frequency of the nucleus. As a result of the absorption the magnetic moment of the $+\frac{1}{2}$-state populates into the energetically higher $-\frac{1}{2}$-state.[22]

Subsequently, the system returns to the equilibrium state with relaxation times characteristic for the respective nuclei. It is usually this relaxation that is detected in NMR spectroscopy. Through the appropriate choice of excitation pulses in terms of frequency and amplitude, as well as their temporal variation, a large number of NMR experiments can be realised. With the help of these pulses, it is possible to examine different properties of the observed system.[23]

2.4. Peptide structure elucidation by NMR spectroscopy

NMR spectroscopy has developed rapidly over the last decades. In BioNMR a multitude of established ’omics’ sciences have developed. Each ’omics’ field has its own requirements for sample preparation and data acquisition, and differs greatly in the interpretation of NMR spectroscopic data. In particular different are the two ’omic’ fields metabolomics and proteomics. Metabolomics as newest ’omics’ family member, deals with complete sets of low weighted compounds in e.g. blood, cerebrospinal fluid or urine. Those samples are evaluated quantitatively mainly by statistical techniques so that the contained small compounds can be identified.[10] In contrast, proteomics cares for the 3D structure of a particular biochemical compound itself. Evaluation is carried out by the use of 1D and 2D correlation experiments, whereby the size of the macromolecule plays a crucial role for the choice of experiment. Peptides with roughly more than 40 amino acids require a selective labelling scheme which is well established in BioNMR. Hence, many protocols for the labelling of NMR insensitive, but interesting nuclei, ^{13}C and ^{15}N, are provided.[24] Unlabelled peptides are evaluated by the use of 1D proton and 2D homonuclear correlation experiments and a number of NOE based approaches. The general procedure for the 3D structure solution of unlabelled peptides is shown in Figure 2.7

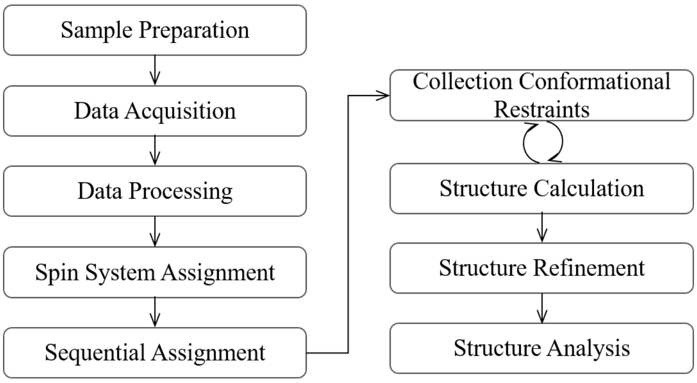

Sample Preparation
Data Acquisition
Data Processing
Spin System Assignment
Sequential Assignment
Collection Conformational Restraints
Structure Calculation
Structure Refinement
Structure Analysis

Figure 2.7.: Flow chart of structure elucidation process by NMR spectroscopic methods.[25]

In the following, the manifold tool kit for the structure elucidation process of unlabelled peptides is described.

2.4.1. Sample requirements in BioNMR

High quality peptide samples are essentially necessary in structural biology and drug discovery. The basic requirement for the evaluability of a peptide spectrum is a high signal quality. An optimal signal shows a high intensity and separation of signals. The acquisition of an optimal signal shape can be difficult for macromolecular systems. Thus, it is necessary to approximate a high signal quality by optimising different factors in the sample preparation process as well as in the NMR experimental conditions.

In the field of BioNMR, work is carried out almost exclusively in aqueous medium.

Water as a measurement medium makes some non-standard NMR methods necessary in data acquisition. Several parameters must be calibrated to achieve a clean water suppression which does not affect nearby signals. Another factor which must be considered is the experimental temperature. Changes in temperature have a noticeable effect on signals distribution and the development of correlation signals. Sample preparation requires optimal conditions which have to be tuned for a component of interest. Relevant conditions are, among others, the composition of the aqueous medium including pH, buffer and ions.

2.4.1.1. pH value of the sample

The chemical shift, splitting and sharpness of exchanging protons are dependent on the charge in a molecule. The charge can be influenced by protonation or deprotonation of functional groups by acids or bases. The acidity or basicity of an aqueous solution is specified by the pH scale. On this basis, the pH scale can be related to the exchanging H^N protons of amino acids, as it is shown Figure 2.8:[21,26]

It can be seen that the exchange rate for amide protons, which are part of a peptides backbone, show a minimum at an acidic pH value of 3.0. Unfortunately, pH 3.0 is very far from realistic physiological conditions, so that measurements are usually taken at a slightly higher pH value.[21,26]

2.4.1.2. The experimental temperature

One factor which limits the work of biomolecules in water is the experimental temperature, which should not be lower than 4 °C. Below this temperature there is the risk that the sample will freeze. A frozen sample is not beneficial for the quality of the spectra and

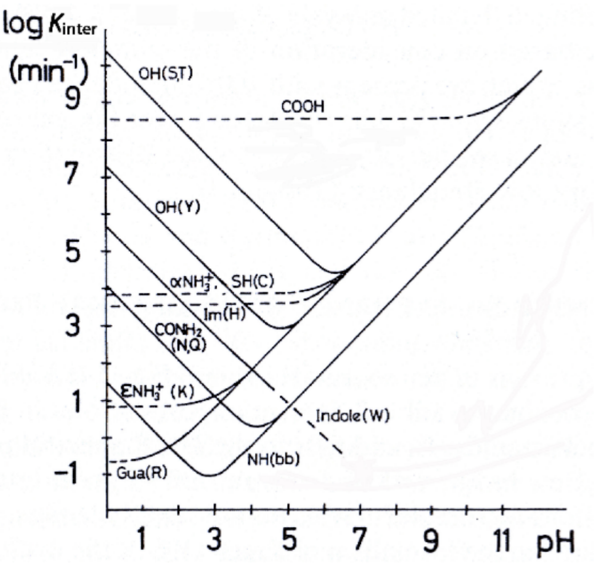


Figure 2.8.: Effekct of pH on exchanging protons with logK_{inter} is plotted as function of the pH value. The NH protons of the backbone (bb) show the minimum at pH 3.0.[26]

also, could damage the biomolecule itself. If experiments underneath 4 °C are required, the addition of salts can lower the freezing point, but may also affect the biomolecules 3D structure. The general need of measurements at lower temperatures is reasoned in the interest of slowing down the mobility of a molecule. A slowed movement leads to improved signal shapes and more distinct signals in 2D correlation experiments.[21,27]

The experimental temperature is not only important for the signal appearance and the survival of the biomolecule; it can also give rise to the existence of hydrogen bonds between amide groups and an acceptor group.

The exchanging protons of amide groups H^N show a particularly high temperature dependence. A temperature induced change in the chemical shift of amides groups is significantly high when they are not involved in hydrogen bonds. Hydrogen bond formation causes a decreased exchange rate and hence, a decrease in the temperature depended chemical shift change.[28,29] The mathematical relation thereof is the so-called temperature coefficient $\frac{\Delta\delta_{HN}}{\Delta T}$. This coefficient is calculated by plotting the chemical shift as a function of temperature for each amide proton. The slope of the least square fit of all points is the temperature coefficient $\frac{\Delta\delta_{HN}}{\Delta T}$ in $^{ppb}/_{K}$. A value above $-4.6\,^{ppb}/_{K}$ indicates an 85%[28] probability of the existence of a hydrogen bond. This probability increases to 93%[28] in a range of -4 to -1 $^{ppb}/_{K}$.[28,29]

2.4.2. Paramagnetic reagents

The use of auxiliary paramagnetic reagents such as manganese, gadolinium, europium or praseodymium compounds is a long-established method in structural biology to study positioning of peptides on membranes. These shift reagents are used to induce line broadening and a shift of signal resonances. Thereby is line broadening the result of shortened relaxation times of the affected nuclei. The signal shift is caused by an interaction through space between the magnetic dipole of the anisotropic paramagnetic reagents unpaired electron and the solvent accessible nuclei spins.[30]

The degree of influence depends on the proximity to the paramagnetic nucleus, which is located in the solution. Consequently, protons that are not in contact with water experience less pertubation than nuclei that are exposed to the solvent.[31]

The experimental approach involves the titration of with paramagnetic stock solution, i. e. the concentration of paramagnetic compound is successively increased in the sample. This successive increase in the concentration of paramagnetic nuclei is primarily intended to prevent an excess that would lead to the damaging of the sample. In addition, a gradual increase in shift and broadening can be observed by following the titration spectra, which allows the correct performance of the experiment to be verified.

2.5. The experimental toolbox for BioNMR

2.5.1. ***TO*tal *C*orrelation *S*pectroscop*Y* – TOCSY**

The identification of unique spin systems of amino acid residues of a polypeptide chain is performed by the use of 2D TOCSY experiments. TOCSY represents an alternative to the more common COSY experiment for the detection of homo-nuclear proton correlation based on scalar couplings. In general, in a 2D COSY correlation of only geminally or vicinally coupled protons are observed. In contrast, the total correlation experiment facilitates the structure elucidation of larger biomolecules since protons that make up a spin system are resolved. Every amino acid within a polypeptide chain has its own spin system since no scalar coupling is present across an amide bond. Magnetisation is transferred along a spin system by the use of a so-called spin-lock. The length of the applied spin-lock defines the number of bonds that are correlated.[20,23]

A rather short mixing time of 20–25 ms yields a COSY-like spectrum due to a single-step transfer between two adjacent protons. Increasing the mixing period, thus, offers the possibility of a multiple step transfer – the rule of thumb hereby is: one more step per additional 20 ms mixing time. The application of a maximal mixing time, which is usually around 120 ms, enables the assembly of an entire spin system. If the selected mixing time is too long, the signal disappears completely as consequence of transverse relaxation. The signal intensity itself depends on prevailed coupling constants, where a large one gives a more intense peak and vice versa.[20,23]

2.5.1.1. Assignment of the spin systems

In Figure 2.9 a schematic 1D proton spectrum with chemical group's resonance positions that are expected for macromolecules in an aqueous sample is displayed.

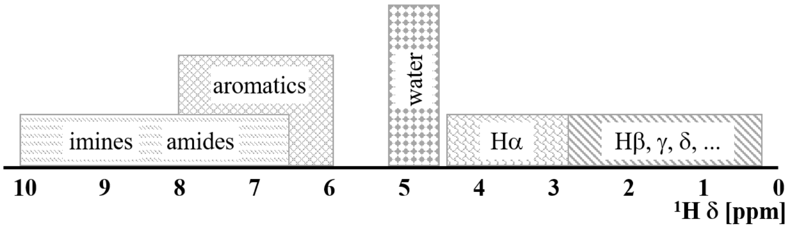


Figure 2.9.: Illustrative representation of typical ^{1}H chemical shift ranges where protons from peptides or proteins resonate in a 1D spectrum.[26]

Every amino acid contains one amide proton, except of proline. Thus, the number of amide resonances is in a direct relation to the moieties in a polypeptide chain. The starting point for the backbone assignment is the so-called fingerprint region where the H^{N}–H^{α} cross-peaks are located (~6.5 ppm–8.5 ppm in F2 and ~0.0–5.0 ppm in F1). In Figure 2.10 an exemplary peptide chain is shown, consisting of only three amino acids residues for sake of simplicity. Those amino acids are connected through peptide bonds, which always interrupt an amino acids spin system. The right side of this Figure shows the expression of those three amino acids in a 2D TOCSY spectrum in the fingerprint region.

The spin system can be built up from from the H^{N}–H^{α} backbone resonance onto the side chain protons. Proline is a secondary amine because its side chain is connected to the backbone nitrogen forming a cyclic structure. Thus, for proline assignment only the aliphatic region below 5.0 ppm is considered. For every amino acid, random coil chemical shift values have been determined by *Wüthrich* and *Bundi*.[32] They determined for every

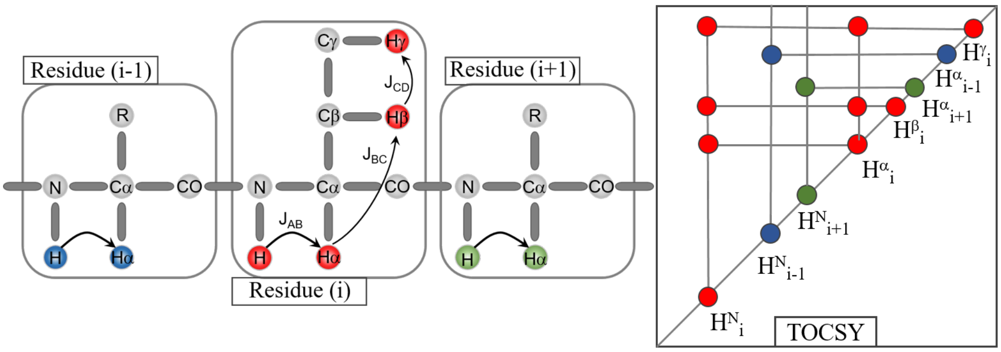


Figure 2.10.: Strategy for spin system assignment through J-coupled nuclei. The arrow shows amide proton as a starting point and goes to adjacent α- and β-protons (R: side chain). On the right side TOCSY signals are displayed.[3]

proton of an amino acid random coil chemical shift values. Therefore, the amino acid of interest was up into a short tetra-peptide chain (i. e. GGXA, pH 7.0, 35°C), to ensure the absence of secondary structure.[32]

2.5.2. *N*uclear *O*verhauser *e*ffect *S*pectroscop*Y* – NOESY

The NOE effect is a phenomenon when a strong radio frequency fields is applied on one nucleus, and a NOE signal of a second, spatial close nucleus can be observed. This spatial contact is caused by dipole-dipole interactions, respectively direct through-space magnetic interaction.[20]

The NOE experiment can be described illustratively using a two-spin system. In this system, the spins carry the denotations I and S, which derive historically from '*i*nteresting' (enhanced) and '*s*ource' (perturbed)). Both are considered as being spatially close enough to share dipolar coupling (D_{IS}), though a scalar coupling is excluded, i. e. $J_{IS}=0$. For this two-spins exist four states $\alpha\alpha, \alpha\beta, \beta\alpha$ and $\beta\beta$, which are arranged according to the energetic population probability in the energy diagram 2.11 below. Basically, they are only energy states that are in exchange with each other and the solvent. Their population equilibrium state is described according to the Boltzmann distribution.[23]

Due to the similarity of $\alpha\beta, \beta\alpha$ energy states they are assumed to be equal at equilibrium. The $\beta\beta$ state is most unlikely and $\alpha\alpha$ is most favourable due to its lowest energy, represented by an excess of nuclei. The NOE experiment essentially involves a 180° selective pulse, which is calibrated to match the frequency of the source spin S. This leads to a saturation of all four population states. After the rf pulse is switched off, the system

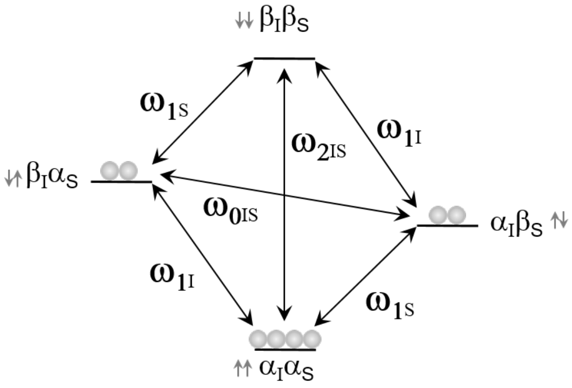


Figure 2.11.: Energy-level diagram of a two spin-system. ω with concomitant number (0-2) indicates transition probabilities along the arrows. Equilibrium state of population is shown as grey balls.

tries to strive back into its initial equilibrium state. Four of the six pathways are an ω_1 transition which is also known as the spin lattice relaxation or T_1. ω_1 transitions do not give a NOE signal but are the only allowed ones according to quantum mechanics. This means that this transition is always in competition with the other two.[20]

ω_0 and ω_2 are described as zero quantum and double quantum transitions according to the sum of flipped spins. Those two are the responsible for NOE signals of the two neighbouring nuclei, *I* and *S*.[23]

The NOE intensity of the *I* spin depends on the pathway followed to reach the ground state. In the ω_2 transition, the population is increased by a double spin into the ground state, resulting in a "positive NOE" signal. The ω_0 transition describes a the spin flip of both, but each one in opposite directions. This results in a decrease in population and thus "negative NOE". Figure 2.12 shows this described signal attenuation, with dashed line indicating the original height of the signal *I*. The dominant mechanism for small molecules is the ω_2 transition whereas for large molecules ω_0 transition is observed.[14]

The choice of transition, as can be further deduced from this Figure, depends on the molecular size. On the left side of the Figure 2.12 the development of the NOE intensity in dependence of the molecular tumbling rate is shown. The tumbling rate describes how long it takes a molecule to rotate around itself by 1 rad. Small molecules, thus, spin much faster than large molecules such as peptides or proteins. At a tumbling rate of around 1 Hz rad^{-1}, corresponding to a molecular weight of approximately 800-1200 Da, the NOE undergoes a zero crossing, i.e. no signal can be detected. An experiment which

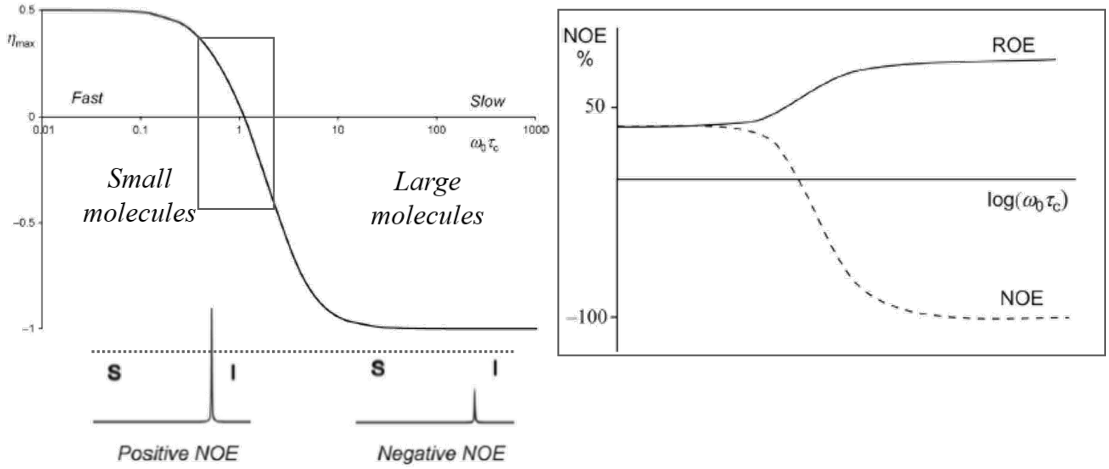


Figure 2.12.: Left: NOE enhancement η plotted as function of dimensionless molecular tumbling $\omega_0\tau_c$ rate. Emphasised are the size and NOE intensities which are expected for molecular sizes. The box highlights the zero crossing of the NOE intensity for which another approach is necessary, shown on the right: ROE of homonuclear two-spin system as function of molecular tumbling in comparison to transient NOE curve.[23]

only develops positive intensities and thus, the absence of NOE can be overcome, is the rotating frame spectroscopy (ROESY).[23]

The measurement of NOE signals is a general principle which can also be applied to hetero nuclei.

The NOE intensity and sign is crucially defined through the gyromagnetic ratio, which is defined value for every nucleus. Thus, for every hetero nucleus the maximal NOE enhancement η_{max} changes in comparison to proton NOE, as visualised in Figure 2.13.[26]

Fluorine, phosphorus and hydrogen have not only a natural abundance close to 100% but also have spin $^1/_2$ and a high positive gyromagnetic ratio. Nitrogen, on the other hand, is highly present in biomolecules and shows the lowest natural abundance beneath a gyromagnetic ratio with an opposite sign. Hence, sensitivity is in a range of 10% of proton nuclei that causes a s/n which is 300 fold higher for proton at same field strength. Isotopic labelling is inevitable and NOE signals could be cancelled out due to non-advantageous relaxation. Optimal cross-peak intensities differ from homo-nuclear proton spectra since relaxation happens in a different time range (e. g. carbon relaxation is much slower).[23] The properties of the nuclei of interest in BioNMR for hetero-nuclear NOE are shown in Table 2.1.

The use of NOE experiments opens up many possibilities in the field of BioNMR. Mainly

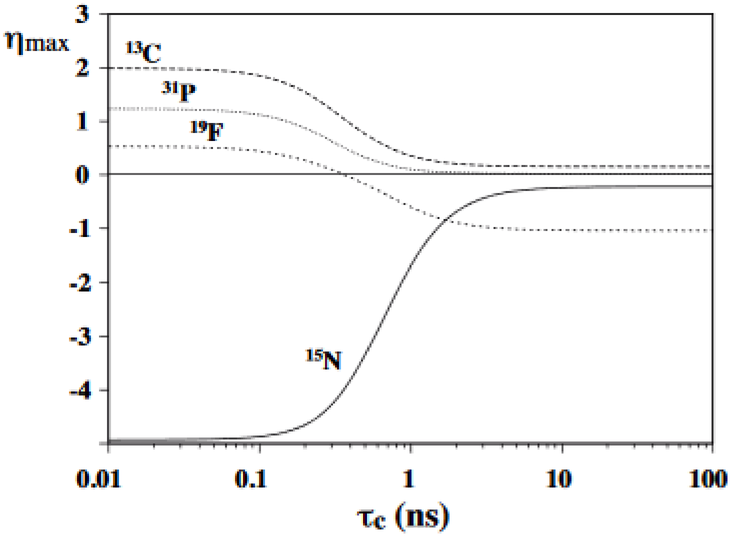


Figure 2.13.: Heteronuclear enhancement η as function of the molecular tumbling rate τ_c.[15]

Table 2.1.: Properties selected nuclei.[23]

Nucleus	Nuclear spin I	Gyromagnetic ratio $\gamma\ 10^{-6}(T \cdot s)^{-1}$	Natural abundance (%)	Relative sensitivity
^{1}H	$^{1}/_{2}$	267.52	99.99	1.0
^{2}H	1	41.07	0.012	1.2
^{13}C	$^{1}/_{2}$	67.28	1.07	$1.76 \cdot 10^{-4}$
^{14}N	1	19.34	99.63	
^{15}N	$^{1}/_{2}$	-27.126	0.37	$3.85 \cdot 10^{-6}$
^{19}F	$^{1}/_{2}$	251.8	100	0.83
^{31}P	$^{1}/_{2}$	108.39	100	$6.63 \cdot 10^{-2}$

important is the directly relation of the spatial distance and the intensity of the NOE signal – the closer the more intense and vice versa. This relation forms the basis for the evaluation of structural features in macromolecules.[23]

The experimental setup requires a proper calibration of the parameters. If mixing times and delays between the pulses are not chosen correctly, either no spectrum at all or false positive results, as a consequence of an oversaturated system, will be obtained. Also, the instrument may be damaged through an overheating of the probe head due to inappropriate strong rf pulses.[23]

2.5.2.1. Sequential assignment

The sequential assignment can not be achieved alone by the use of *J*-coupling based spectra as COSY and TOCSY. 2D NOESY spectra are used for the connection of spin systems. The most intense and, thus, relevant signals for sequentiation in a 2D NOESY spectrum is the H_i^N and H_{i-1}^{α} NOE, as emphasised with a arrows in Figure 2.14.[33]

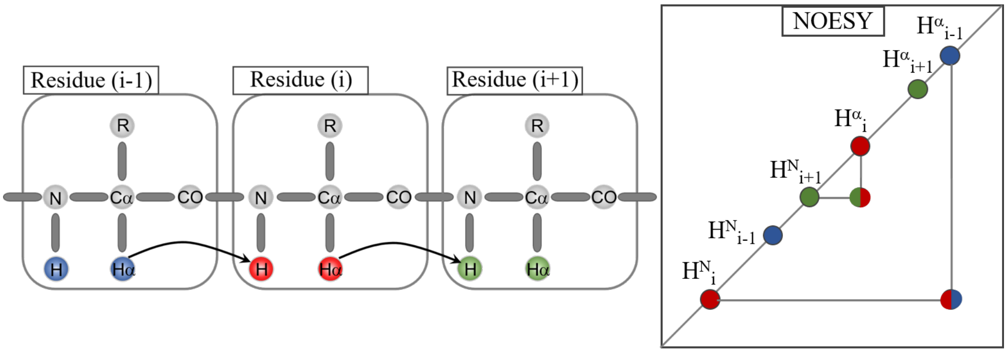


Figure 2.14.: Connection of the spin systems is mainly perfored through the identification of NOE signals of adjacent H_i^N and H_{i-1}^{α} protons (R: side chain). On the right side the exemplary NOSY signals are displayed.[3]

The ability of peptide sequences to exhibit secondary structure elements, causes additional spatial closeness of backbone and side chain protons between non-adjacent amino acids.[14] A survey of sequential NOE of several structure elements is displayed in Figure 2.4.

Depending on the distance between two protons, NOEs are divided into short, medium and long range NOEs. In a conventional α-helix, short range NOEs are usually detected for H^N–H^{α} contacts. Further, short distances between backbone amide protons, d_{NN}, just as α-protons of *i* to β-protons of *i+3*, $d_{\alpha\beta\ (i,\ i+3)}$.[14]

Medium range NOEs are mostly found in connection to signals related to spacial proximity involving (*i+2*) to (*i+4*), whereas a connection of amino acids, which are five residues apart, are defined as long range.[14]

In comparison, β strand conformations are characterised by short sequential $d_{\alpha N}(i,i+1)$ distances just as very close $d_{\alpha\alpha}(i,j)$. $d_{NN}(i,i+1/i,j)$ NOEs are very weak up to not present, in comparison to the former described helical arrangement.[14]

Differentiation between intra-residual NOE (spins related to same spin system) and sequential or inter-residual NOE (neighbouring spin network) must be performed by a careful comparison of NOESY and TOCSY spectra spectra.

One method to evaluate the structure through the shift of a signal in a spectrum is the *c*hemical *s*hift *i*ndex (abbr. CSI) technique. The CSI enables the direct identification of secondary structure elements with an accuracy of up to 80–85%[34] for H^α chemical shifts. Conventionally, secondary structure is determined in biology with the established *c*ircular *d*ichroism (CD) approach, which analyses the global protein secondary structure. In contrast, the CSI method is able to reveal information about secondary structure on a residual level.[3]

In Figure 2.15 an exemplary CSI diagram is shown.

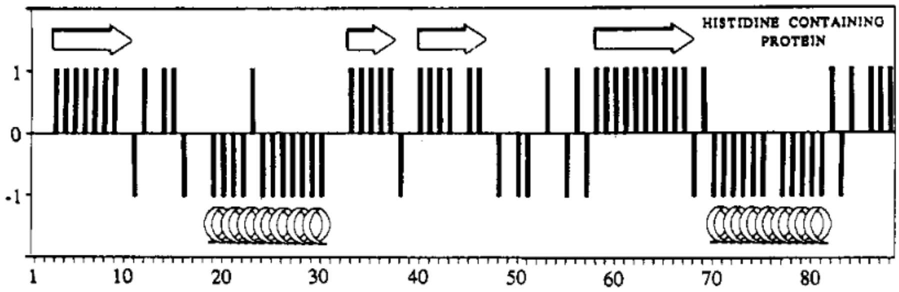


Figure 2.15.: Bar graph of chemical shift index of H^α-protons. Arrows and loops are indications for the kind of fold, β-turn and α-helices, respectively. y-axis: CSI value, x-axis: residue number.[35]

An α-helical structure is assumed when four or more amino acids are in a row with a –1 value. If three or more adjacent residues show a +1 value residues they are said to be part of a β-strand.[35]

2.5.3. *Tr*ansfer *NOE* – trNOE

The approach has broad applicability for the identification of binding of small molecules to a receptor. Key of the process is the change in correlation time τ_c of free vs bound ligands.[36]

Molecules differ in their NOE sign depending on their size. Thus, small ligands and large biomolecules have opposite NOE signs which provide information about the binding status of a ligand.[23,27]

Key point of this effect is that a small molecule which is bound to a macromolecule behaves in its motional properties like its huge complex-partner, i. e. undergoes a NOE phase-change. In consequence, during the resistance time of the small compound on the surface of the macromolecule, NOE is developed more rapid due to the adaption of the macromolecules large correlation time τ_c. The sign of the NOE is thereby transferred from the huge onto the small molecule. The small molecule dissociates into the solution and is detected there.[23,27]

Sample preparation plays a crucial role in the experimental setup. Tr-NOE depends on the free and bound ligand fractions. An optimal ratio between ligand to macromolecule is 10–30:1 is optimal. If a too large ligand excess is present, change in NOE could be remained unnoticed due to cancellation with ligand in an opposite binding status.

In summary, tr-NOESY offers the possibility to detect binding activity very quickly through the sign and size of NOE signals. Further, the build-up rate differs in bound vs unbound state, so that maximum intensity is achieved in a range of 50-100 ms in contrast to unbound, small molecules, which needs 4-10 times longer.[36]

2.5.4. *S*aturation *t*ransfer *d*ifference – STD

STD is a NOE based approach to study interactions between macromolecules (receptor) and small to medium sized compounds (ligands) in solution.

STD provides an experimental setup to identify ligand epitopes involved in interaction processes and allows mapping of the ligand protons involved.[37,38]

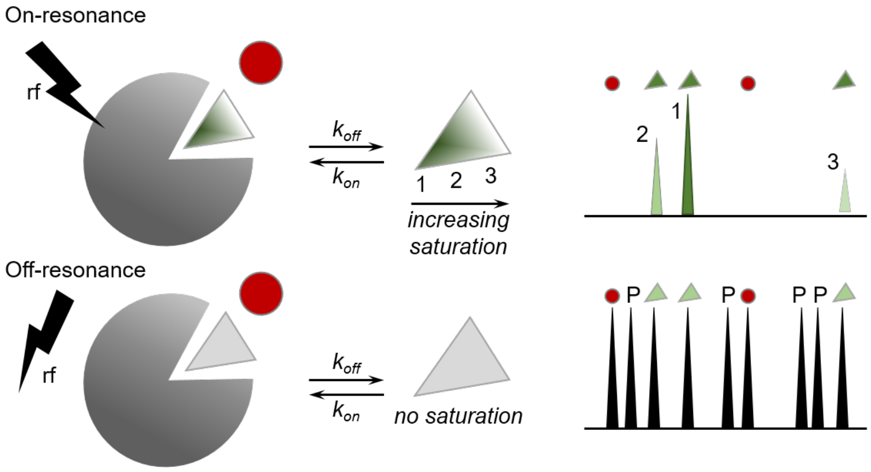


Figure 2.16.: Schematic representation of saturation transfer difference (STD) NMR experiment. Colour gradient indicates saturation state of ligand from black: highly saturated to white: not saturated. Non-binding ligands are not detected in 1D proton spectrum as shown on right hand side. In the off-resonance experiment, selective saturation is set far from all signals (around 30.000 ppm) to detect blank spectrum with all involved components as shown right.[31]

In this experiment, the macromolecule is selectively irradiated in a frequency range where no ligand signals are observed, which is usually around –0.5 ppm. The ligand bound to the saturated protein appears saturation by spin diffusion. By dissociation, the ligand carries this saturation into the solution, where it is detected.[38]

The interaction strength of protons of a ligands is of particular interest in the area of drug design. The distinction between the interaction strength of ligand protons is called epitope mapping which is achieved by the evaluation of the obtained signal intensities.[39]

2.5.5. *water*-*l*igand *o*bserved via *g*radient *s*pectroscop*y* – waterLOGSY

The waterLOGSY experiment is a NOE based experiment which is mostly applied to distinguish binding and non-binding of a ligand to a receptor.[23] The benefit of waterLOGSY is that it was proven to be more sensitive as STD.[40,41]

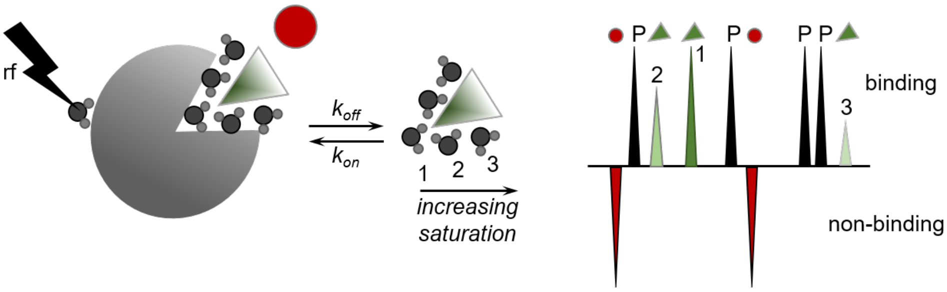


Figure 2.17.: Schematic representation of waterLOGSY experiment with 1D proton trace on right hand side. After selective saturation on water signal, magnetisation is transferred from bulk water to protein to ligand. Non-binders (circles) show opposite signal intensities to binders (triangles).[31]

In comparison to the STD approach, in waterLOGSY experiments the selective irradiation is performed on bulk water. Hence, magnetisation is transferred to the ligand bound to the protein. Ligands that are located unbound in the aqueous solution show also a transfer of magnetisation from the bulk water. Differentiation of both binding situations happens now through the NOE sign of the small compound, following the same principle as already described.[42–44]

An exemplary waterLOGSY spectrum for a model system consisting of BSA:Trp:Suc is shown in Figure 2.18

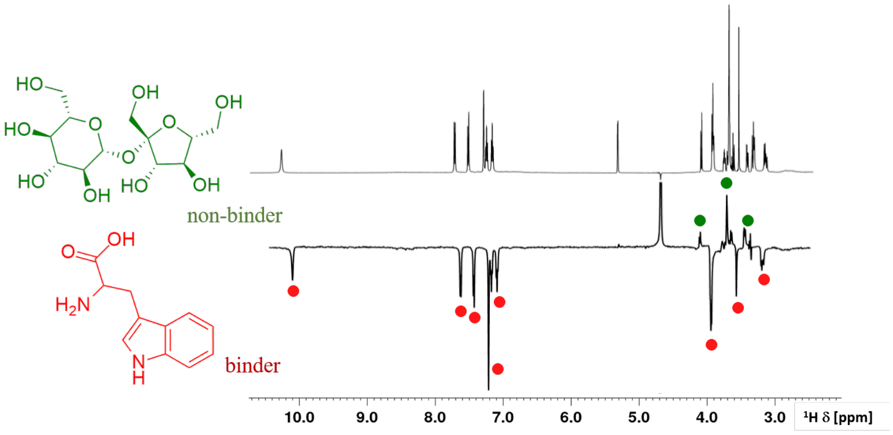


Figure 2.18.: waterLOGSY spectrum of BSA:Trp:Suc model sample (600 MHz, mixing time: 0.8 s, 298 K, pH 6.08, H_2O/D_2O 9:1, BSA:Trp:Suc 1:100:100).

The signals which belong to the binder Trp show in the aromatic region an intense positive NOE, whereas Suc signals show, as non-binder, the opposite signal intensity. Exchanging protons (here: indole H^N from Trp) should not be considered for the evaluation since they are in a direct exchange with the bulk water.[23]

2.5.6. *D*iffusion-*o*rdered *S*pectroscop*Y* – DOSY

Diffusion-ordered NMR spectroscopy enables the spectroscopic separation of multicomponent mixtures or oligomeric states based on an altered diffusional behaviour of the compounds. The diffusion coefficient, thereby, is described by the viscosity of the medium η, temperature T, the Boltzmann constant k_B and the hydrodynamic radius r_S. This relation is summarised in the so called *Stokes-Einstein* equation 2.6:[23]

$$D = \frac{k_B T}{6\pi\eta r_s} \tag{2.6}$$

This Formula emphasises an inversely proportional correlation between molecular size and diffusion coefficient. Hence large molecules or complexes are expected to have a small diffusion coefficient and small molecules are estimated to show a higher one.[23,45]
In Figure 2.19 the general principle of a DOSY is schematically displayed.
Signals of molecules that translate fast through the solution cannot be completely refocused, so the signal intensity is low. Molecules that move slowly through the solution show a high signal intensity.
In the experimental setup, two 1D DOSY experiments are acquired at different gradient

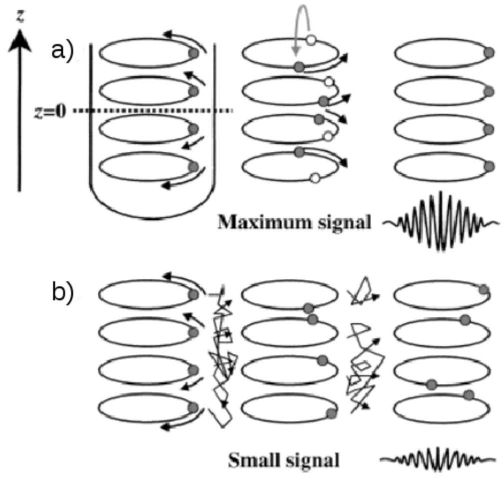


Figure 2.19.: Effect of the absence (a) and presence (b) of diffusion on the phase shift and signal intensity.[45]

strength. By the comparison of the 1D experiments, a signal decay of $\tilde{5}\%$ signal intensity indicates an optimal condition. In general, the experiment can be calibrated by tuning the values of diffusion time Δ and the applied gradient length δ, is thereby used for the adjustment of the experiment.[46]

2.6. From experimental data to 3D structure

The calculation of a 3D structure based on NMR spectroscopic data is performed by using a combination of simulated annealing and *m*olecular *d*ynamics (MD) in torsion angle space.[47,48] The calculation protocol is outlined in Figure 2.20.

Essential parameters that are required for the calculation are proton chemical shifts, NOE intensities and the sequence itself. By providing the amino acid sequence the software defines the type of amino acid, bond lengths, bond angles and chiralities. In addition, a library is stored to parametrise the amino acids of the sequence. This file contains the Cartesian coordinates, atom type, nomenclature, dihedral angles definitions and the standard geometry of each amino acid.[14] In biochemical compounds, often non-standard residues are included in peptide or protein sequences. 'Non-standard compounds' are not available in the library and have to be parametrised.[49]

The calculation protocol starts with the generation of a set of random structures. This is the basis for the torsion angle dynamics using a simulated annealing technique. The simulated annealing allows the biomolecular conformer to overcome local minima of a

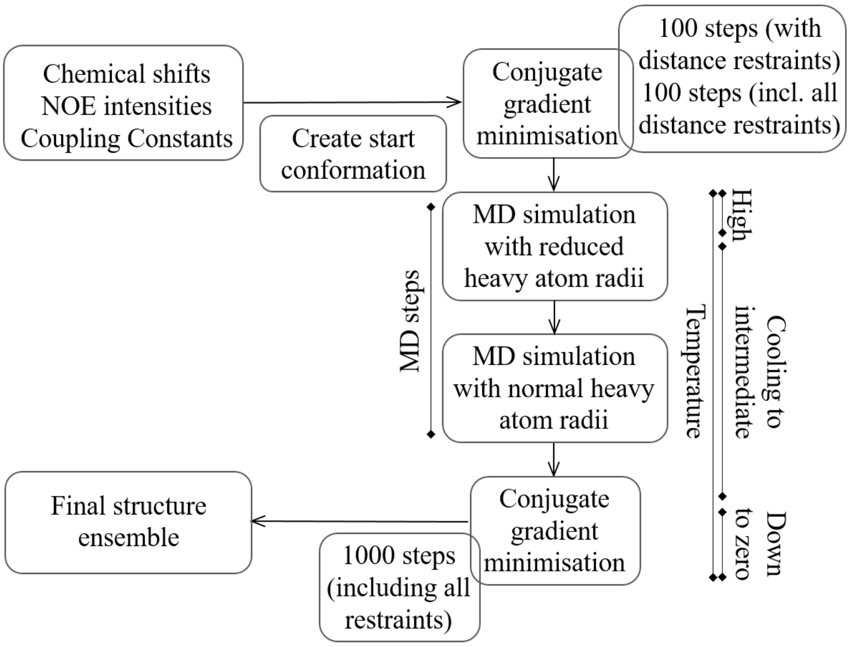

Chemical shifts
NOE intensities
Coupling Constants
Create start conformation
Conjugate gradient minimisation
100 steps (with distance restraints)
100 steps (incl. all distance restraints)
MD steps
MD simulation with reduced heavy atom radii
MD simulation with normal heavy atom radii
Conjugate gradient minimisation
1000 steps (including all restraints)
Final structure ensemble
High
Cooling to intermediate
Down to zero
Temperature

Figure 2.20.: Standard simulating annealing protocol of NMR restrained structure calculation software.[14]

target function to approximate the global optimum, which is the structure with the best agreement of experimental data. Therefore, the system is added enough kinetic energy to avoid falling into a local minimum, as would be the case with simple energy minimisation.[47,48]

The MD simulation step only corrects bond angles based on NMR restraints by free rotations around single bonds. This provides an efficient way to handle biomolecules in a comparatively short computational time. Peptide bonds are fixed, reducing the degrees of freedom and the associated computational time. The MD steps in the calculation of the 3D structure differ from a classical MD simulation. A classical MD simulation describes a system with a full force field, taking into account additional parameters such as van der Waals interactions and electrostatic interactions. In the protein structure calculation, MD is used to optimise the target function. The final cooling of the system leads to several low-energy conformers, which ultimately constitute the NMR ensemble.[14,47,48]

Traditionally, in the area of BioNMR, this ensemble is always shown when the structure is presented. The reason for this is that molecules in solution are always subject to a certain degree of flexibility. The method of NMR provides a picture of the average situation of the structures in solution. The quality of the structure ensemble is indicated by a parameter called $r$oot-$m$ean $s$quare $d$eviation (rmsd), which is determined by aligning all the individual structures along the backbone atoms N, C_α and C. The smaller the rmsd value, the higher populated is only one conformation.[27]

Part II.

NMR-BASED CONFORMATIONAL STUDY OF PROLINE-RICH PEPTIDES

1. Background

1.1. Proline associated peptide bonds and their influence on peptide/protein folding

Proline itself is a very special building block in polypeptide chains in comparison to the other 19 proteinogenic amino acids. The side chain of proline is covalently linked to its amide nitrogen at the Cδ, resulting in a cyclic structure, a pyrrolidine loop, and a secondary amine, which is a tertiary amine within a peptide chain. This causes the backbone dihedral angles to be restricted to $\phi = -75°$ and $\psi = 145°$. As a consequence, the non-protonated nitrogen cannot be present as a hydrogen donor anymore. Thus, secondary structures are stabilised or destabilized differently and proline acts as a so-called α-helix or β-sheet disruptor. The introduction of proline enables slightly different secondary structures such as the well established right-handed polyproline type I- and left-handed II-helix (abbr. PPI and PPII).[12,13,50]

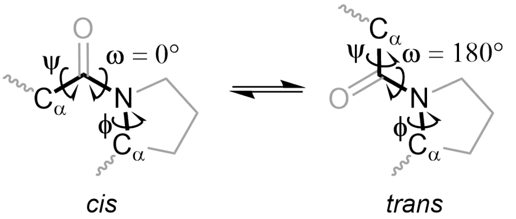


Figure 1.1.: *Cis/trans*-conformers of Pro–Xaa bonds with labeles on relevant dihedral angles.

The *cis/trans*-proportion of the Xaa–Pro has an energy difference of about 2.5 kcal mol^{-1} with a rotational barrier of 20 kcal mol^{-1}. Thus, the *trans*-conformer is only favored by 0.5 kcal mol^{-1} over the *cis*-form resulting in around 30% of the *cis*-form in equilibrium.[50–52]

The total ratio of *cis/trans*-isomers is significantly influenced by size of the peptide chain, structure elements as well as polarity of the solvent. Smaller chains tend to show a higher ratio of *cis*-isomer, so do Xaa–Pro bonds in polar solvent due to electrostatic interactions. Therefore, the total rate of interconversion is higher in less polar solvents.[52]

1.2. Relevance of proline for the bacterium *Clostridium difficile*

C. difficile is a gram-positive, spore-forming, anaerobic bacterium, which is one of the most common pathogens that cause antibiotic-associated diarrhoea infections worldwide. Symptoms range from mild diarrhoea to toxic megacolon, which can lead to perforation of the intestine, sepsis and death. Infection is particularly commonly observed in hospitalised patients, since bacterial participation in gut flora is increased from 2-5% (general non-hospitalised population) to 20%.[53] A high percentage of 31% of patients who have undergone antibiotic treatment, which alters the gut flora and enables *C. difficile* to flourish, were infected.[54]

The first-line treatment, nowadays, includes only the interruption of antibiotics treatment to allow the colonic microflora to recover and also treat symptoms such as dehydration.[55,56] A more targeted treatment is still pending.

Thus, this disease has gained in relevance due to high motility rates. A high level of multi-drug resistance to antibiotics and a high recurrence rate, the complicate therapy and treatment.

Hypervirulent strains are known to secrete multiple factors into the extracellular milieu, such as increased amounts of various toxins (16-23 fold of toxin A and B),[53] including the two glycosyltransferase toxins: enterotoxin A (TcdA) and cytotoxin B (TcdB). Both are mainly responsible for *C. difficile* virulence, since they are able to inactivate the glucosyltransferase and thus cause damage to the gut tissue.[57] The general consequences are extensive inflammation, increased permeability of the intestinal mucosa and necrosis, which emphasises the need for new therapeutics.[57–59]

1.2.1. Mechanism of action of *C. difficile* to colonise the host intestine

The process of colonisation of the gut through a bacterium is a highly complex process with adhesion of a bacterium to epithelium mucosa, invasion of the host immune system, motility, toxin production and overcoming mucosal barriers.[60] Bacteria must always be able to adapt to the rapidly changing environmental medium to find the optimal breeding ground for their growth and survival. Some bacteria are able to swim to advantageous areas and settle down there for a self-defined period of time. If the environment changes, it is possible to react quickly and move on through tight control of so-called flagellar motors/proteins which can also be found under the excreted compounds of *C. difficile*.[61] *C. difficile* is able to move through the intestines almost unhindered and flourish in areas which are advantageous for its continued existence by adhere to the intestine wall. When the environment deteriorates it cuts off the anchor-proteins and develops a more advantageous area.

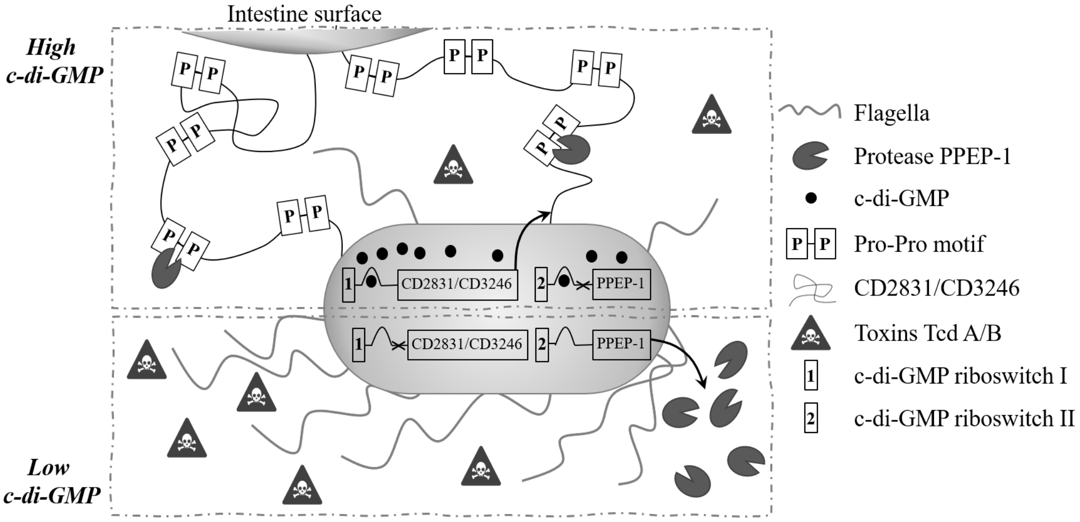


Figure 1.2.: Schematic representation of *C. difficiles* modes of action: adhesion (top) and motility (bottom). High c-di-GMP level causes an higher expression of CD2831 and CD3246 surface anchors by interaction with type II riboswitches. Low levels of bacterium messenger leads to an upregulation of toxins Tcd A and B as well as enhanced flagella biosynthesis for its motility. Missing interaction of c-di-GMP further causes higher release of *P*ro-*P*ro *e*ndo*p*eptidase-1 (abbr. PPEP-1), which has a unique specificity for Pro-Pro bonds.[62–65]

This colonisation process is regulated by a bacterial second messenger, namely Bis-(3',5')-*c*yclic-*di*meric-*g*uanosine-*m*ono*p*hosphate (abbr. c-di-GMP), which is responsible for a 'swim or to stick' mechanism of a bacterium. It controls the biofilm formation and the toxin production, but above all it regulates negatively its flagellar swimming motility and positively the expression of putative adhesins.[60,62,66]

C-di-GMP works with two classes of riboswitches of *C. difficile*, type I and type II. High intra-cellular concentration of c-di-GMP leads to an up-regulated expression of protein putative surface adhesins CD2831, which is equipped with a type II riboswitch. This protein is responsible for effective attachment of the bacterium to the host. Another gene of the bacterium carries a type I riboswitch, which interestingly encodes the metalloprotease *poly*proline*e*ndo*p*eptidase-1 PPEP-1, that is activated at low c-di-GMP levels. This protease is able to recognise and cleave selectively the putative surface adhesin CD2831 and CD3246. Cleavage regulates and limits the total adhesion rate of the bacterium to the intestinal wall of the host. In summary, at low concentration levels of c-di-GMP, the expression of PPEP-1, as well as TcdA/B toxins and flagella are up-regulated — whereas high c-di-GMP concentrations lead to an up-scaled encoding of PPEP-1 substrates CD2831 and CD3246.[60,62,67]

1.2.1.1. PPEP-1 involvement in the motility of *C. difficile*

PPEP-1 is the gene product of CD2830 which is secreted under well defined conditions by the pathogen bacterium *C. difficile*. The zinc-dependent metalloprotease takes its name from its unique, outstanding ability to cut selectively between Pro–Pro peptide bonds. Several peptidases are already known to cleave between Pro-Xaa bonds and, more rarely, Xaa–Pro, but there are very few representatives for the selective choice of Pro–Pro bonds. PPEP-1 preferred recognition motive was determined as Asn^3–Pro^4 ↓ Pro^5–Val^6–Pro^7–Pro^8, with the arrow pointing down indicating the scissile bond. For an effective cleavage, asparagine is required as preceding residue to the proline of the scissile bond (Pro^4–Pro^5).[62,67,68] Proteins CD2831 and CD3246 are equipped with multiple cleavage sides, i. e. CD2831: 3 sites and CD3246: 6 sites, which all are cleaved when those peptides are in contact with PPEP-1.[69]

1.2.1.2. Structure of PPEP-1

PPEP-1 is involved in the pathogenicity of *C. difficile*, it selectively cuts Pro–Pro peptide bonds of the bacterial surface anchor-peptide sequences. Therefore, it is of high interest to know some of its structural features. In Figure 1.3 the general structure of PPEP-1 is displayed in cartoon representation.

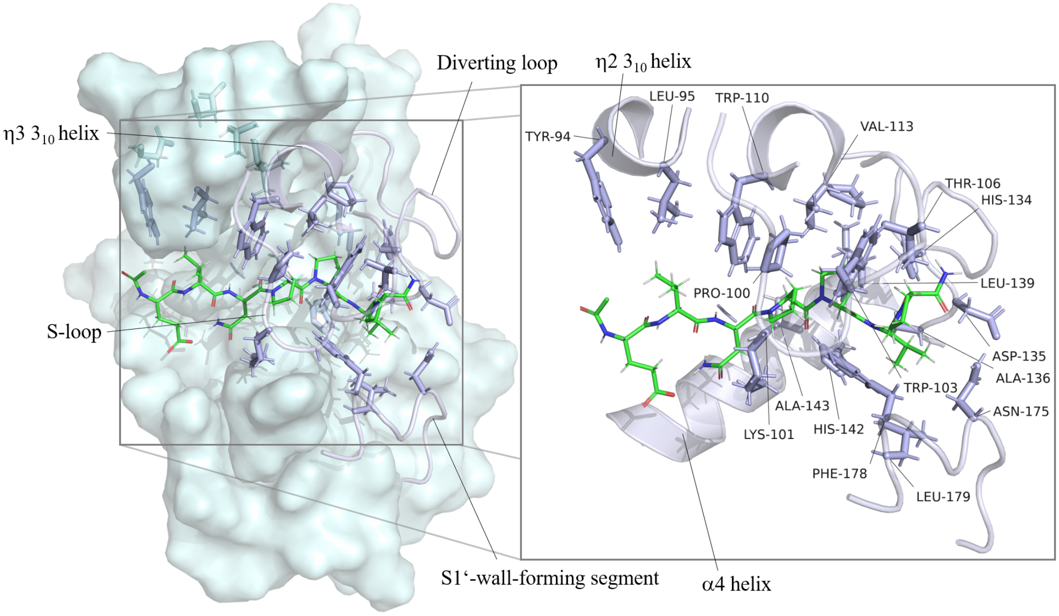

η3 3_{10} helix
Diverting loop
η2 3_{10} helix
S-loop
S1‘-wall-forming segment
α4 helix
TYR-94
LEU-95
TRP-110
VAL-113
THR-106
HIS-134
LEU-139
PRO-100
ASP-135
ALA-136
TRP-103
ASN-175
ALA-143
LYS-101
HIS-142
PHE-178
LEU-179

Figure 1.3.: Representation of PPEP-1 with short hepta-peptide **1** (Ace–GVNPPVP–NH$_2$) bound to active side. All residues involved in recognition of the substrate are represented in cartoon for backbone atoms and side chains with sticks. Only domains important for protein-ligand complexion are emphasized. The enlargement (right) has all those amino acids labeled. The peptide is drawn in stick representation, colored by atom type (N: blue, O: red, C: green, H: white)(PDB-code: 5a0x).[68]

2. Goals

Infections with *C. difficile* are still very difficult to treat. The classical clinical approaches do not offer any specific targeted medication. The discovery of how the bacterium appears to move autonomously or how it settles in the intestine could be an important breakthrough, as this could be a point of attack to develop a targeted therapy. Although not all secreted compounds of *C. difficile* have been investigated or their structure and function could be clarified, PPEP-1, which preferentially cuts the surface adhesions peptides exclusively between Pro–Pro bonds, offers hereby a starting point.[62–65]

In the past crystal structure studies have shown that the Pro–Pro scissile bond is always present in an all-*trans* configuration within the active side.[62,68] Since proline undergoes a less hindered *cis*/*trans* isomerisation in comparison to Xaa–Xaa bonds, it would be conceivable that a significant percentage of these important binding sites are also present in the *cis*-conformation when free movement is possible in solution. Crystal structures allow many important conclusions to be drawn about a highly complex system, such as relative and absolute conformations and structure-function relationships. Nevertheless, they are comparatively rigid so that highly dynamic processes have to be analysed by other methods. These include CD, UV- and fluorescence detection, where indirect tertiary elements and binding modes in solution are investigated. Additionally, NMR spectroscopy represents the optimal method for direct analysis of solvent processes and evaluation of 3D solution structures, through ligand or protein view approaches, receptively. Here, it is of high interest if the active side of PPEP-1 is specifically selective for only one isoform, so that substrates already needs to be in a specific pre-conformation or if the active side does not favour one isomer and forces the Pro–Pro bond into the required shape itself.

Consequently, NMR spectroscopy will be used to elucidate the structure of the proline-rich peptide sequences **4-6** displayed in Table 2.1. It will be clearly determined if a *cis*/*trans* isomerism of the Xaa–Pro/Pro-Xaa or Pro–Pro bonds of peptides **4-6** is present in solution. These results will then be compared with the previous work on the shorter sequences **1-3**.

Table 2.1.: Investigated peptide sequences. Differing residues are outlined in a box. All sequences were synthesized and kindly provided by the *Baumann*-group (Biochemical Insitute of the University of Cologne)

#	sequence	short
1	Ace–G^1–V^2–N^3–$\boxed{\text{P}^4\text{–P}^5}$–V^6–P^7–NH$_2$	PP-pep
2	Ace–G^1–V^2–N^3–$\boxed{\text{A}^4\text{–P}^5}$–V^6–P^7–NH$_2$	AP-pep
3	Ace–G^1–V^2–N^3–$\boxed{\text{P}^4\text{–A}^5}$–V^6–P^7–NH$_2$	PA-pep
4	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{P}^5\text{–P}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-PP-Dnp
5	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{A}^5\text{–P}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-AP-Dnp
6	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{P}^5\text{–A}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-PA-Dnp

Further, NOE based approaches as STD and waterLOGSY are performed with short proline-rich sequences **1-3** together with an inactive mutant of protein PPEP-1. This shall give an insight into the dynamic interaction process between the individual residues of the ligand and the active side of the protein. On the basis of these studies, a comparison can then be made between the crystal structure data and NMR investigations.

Finally, computer studies will be conducted to evaluate the 3D solution structure of the long proline-rich sequences (in comparison to shorter hepta-peptides). Additional docking calculations will further support the experimental results and possibly explain them in more detail.

3. Material and Methods

General. All reagents (NaH_2PO_4, Na_2HPO_4, TSP, NaN_3) were purchased from *Sigma Aldrich* and used without further purification. The peptides were purchased from GenScript with a purity of at least 90% and gently provided by the *Baumann group* (Biochemical Institute, University of Cologne).

NMR sample preparation. A phosphate-buffered saline stock solution (PBS 100 mM, pH 7) containing the chemical shift reference compound (i. e. *T*rim*e*thyl*s*ilylpropanoic acid abbr. as TSP) as well as the preservative for biological contamination (sodium azide) was used to prepare the NMR samples. For all applied NMR experiments a sample consisting of an amount of 1-3 mM peptide was dissolved in 180 μl of PBS and 20 μl of H_2O/D_2O, 90:10 in a 3 mm NMR tube to give a final concentration of around 3.28 mM.

NMR experiments and assignment. NMR spectra were recorded on a Bruker Avance II spectrometer(^{1}H frequency of 600 MHz) equipped with a triple-resonance high-resolution probe (TBI). All NMR data were acquired and processed using Topspin software (Bruker).[46] The transmitter frequency was set on the HDO/H_2O signal, and the d_4-TSP resonance was used as chemical shift reference (^{1}H $\delta_{TSP} = 0$ ppm). For assignment, 2D homonuclear spectra (i. e. 2D ^{1}H , ^{1}H total correlated spectroscopy (TOCSY) and ^{1}H , ^{1}H Nuclear Overhauser spectroscopy (NOESY)) were recorded using standard Bruker pulse sequences, including excitation sculpting with gradients for solvent suppression.

The temperature dependence of the amide protons were calculated from the shift of the amide protons within the range of 283 – 308 K. The complete assignment of the backbone and side chain ^{1}H resonances was performed using standard sequential assignment procedures according to the methodology developed by *Wüthrich*[26] and analysed using CARA software.[70]

Structure calculation. Three-dimensional structures were determined by the standard protocol of the CYANA program (version 2.1),[49] using seven cycles of combined automated NOESY assignment and structure calculations followed by a final structure calculation. Since the peptides contain non-standard residues, the corresponding libraries for CYANA were built using the PyMol software. PyMol was also used to visualise three-dimensional structures.[71] For each CYANA cycle, 100 randomized conformers and the standard simulated annealing schedule were used. The 20 conformers with the lowest final score were retained for analysis and passed on to the next cycle. Weak restraints on ϕ/ψ torsion-angle pairs and on side-chain torsion angles between tetrahedral carbon atoms were applied temporarily during the high temperature and cooling phases of the simulated annealing schedule in order to favour the permitted regions of the Ramachandran plot and staggered rotamer positions, respectively. The list of upper-distance bonds for the final structural calculation consists of unambiguously assigned upper-distance bonds and does not require the possible swapping of diastereotopic pairs. Root-mean-square deviation (RMSD) values were calculated using CYANA for superpositions of the backbone N, Cα and CO atoms; the heavy atoms over the whole peptide or the cyclic fragment. To obtain the rmsd of a structure represented by a bundle of conformers, all conformers were superimposed upon the first one and the average of the rmsd values between the individual conformers and their average coordinates was calculated.

Docking studies. Docking studies were performed using AutoDock vina software compiled on a Linux system.[72] AutoDockTools was used for the preparation of ligands, the protein, AutoGrid points and for the visualisation of the docked ligands. Ligands and receptor were prepared as follows: hydrogen atoms and partial charges were added to the molecules, water molecules and non-polar hydrogens were deleted. All dihedral backbone bonds of the ligands were allowed to rotate. The docking calculations were performed with only one side of the PPEP-1 dimer which was extracted from the crystal structure. The grid centre was established by centring the grid box on the active side of the protein. The spacing of the grid maps was 0.375 Å.

4. Results and Discussion

Investigations have already been performed on unlabelled, shorter hepta peptides[i] (cf Table 4.1 with the focus on the same Pro–Pro core motif as the longer peptides **4-6**. The Pro–Pro core motif is the preferred cutting position of PPEP-1. For the peptides **1-3**, multiple isoforms in solution were detected. The main isoform of peptide **1** was determined with the aid of 2D NOESY experiments as an all-*trans*-configuration, which was present in an excess of 88% compared to all other minor isoforms present in solution. The ratio could not be determined for peptides **2** and **3** since the signal overlap was too strong. The 3D solution structure was calculated by the NMR restraints proton and carbon chemical shifts and proton inter-nuclear distances, which were extracted from 2D NOESY experiments. The major configuration of all three peptides in solution **1-3** was determined as *trans* for the Pro–Pro peptide bond.[67]

Table 4.1.: Peptide sequences that were extensively investigated previously. They were kindly provided by the *Baumann*-group (Biochemical Insitute of the University of Cologne)

#	**sequence**	**short**
1	$\text{Ace-G}^1\text{-V}^2\text{-N}^3\text{-}\boxed{\text{P}^4\text{-P}^5}\text{-V}^6\text{-P}^7\text{-NH}_2$	PP-pep
2	$\text{Ace-G}^1\text{-V}^2\text{-N}^3\text{-}\boxed{\text{A}^4\text{-P}^5}\text{-V}^6\text{-P}^7\text{-NH}_2$	AP-pep
3	$\text{Ace-G}^1\text{-V}^2\text{-N}^3\text{-}\boxed{\text{P}^4\text{-A}^5}\text{-V}^6\text{-P}^7\text{-NH}_2$	PA-pep

The NMR chemical shift data of this paper are available from the BiologicalMagnetic Resonance Data Bank under BMRB accession numbers **1**: 27863, PDB ID: 6r9z, **2**: 27864 and **3**: 27862 ,

---

[i] *partly unpublished* L. Jütten, Master Thesis, University of Cologne, 2017.

4.1. NMR assignment of proline-rich peptide sequences 4-6

In Table 4.2 the proline-rich peptide sequences are displayed that are investigated in the scope of this chapter. They differ from the heptapeptide sequences **1-3** in their length, amino acid sorting and substituents. To conduct biochemical investigations as absorption measurements/fluorescence-based assays, the two modifications in residues Asp^1 and Lys^{11}, which are displayed in Figure 4.1, were included.

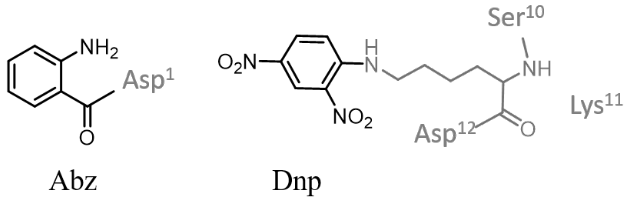


Figure 4.1.: Structural formula of 2-aminobenzoyl (Abz) and 2,4-dinitrophenyl (Dnp).

The 13 amino acid long proline-rich sequences are linked to an Abz-moiety, which acts as a so-called fluorophore, while the Dnp is linked to the terminal ϵ-group of the Lys^{11} side chain and act as the fluorescence quencher.

Table 4.2.: Investigated peptide sequences. The motif of interest is outlined in a box.

#	**sequence**	**short**
4	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{P}^5\text{–P}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-PP-Dnp
5	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{A}^5\text{–P}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-AP-Dnp
6	Abz–D^1–V^2–V^3–N^4–$\boxed{\text{P}^5\text{–A}^6}$–V^7–P^8–P^9–S^{10}–K(Dnp)^{11}–D^{12}–D^{13}–NH$_2$	Abz-PA-Dnp

The 1D proton spectra of the peptides **4-6** are displayed in Figure 4.2. Distinct additional signal sets are assumed to be present due to a proline induced *cis*/*trans*-interconversion of the proline-associated peptide bond in solution.

If multiple isoforms coexist and do not interconvert faster than the NMR time scale, the signals are not averaged and thus, separately detectable.

By a closer inspection of the amide area (between around 7.5-9.0 ppm) those additional resonances are revealed (marked with an asterix). Peptide **1** shows four additional signal sets in the range of 7.9-8.1 ppm. Peptide **2** has three well separated minor signals at around 8.0 ppm and also one for the alanine side chain at ~1.3 ppm is observed. No evidence of the coexistence of isoforms was observed for the peptide with the Pro–Ala (**6**) residues.

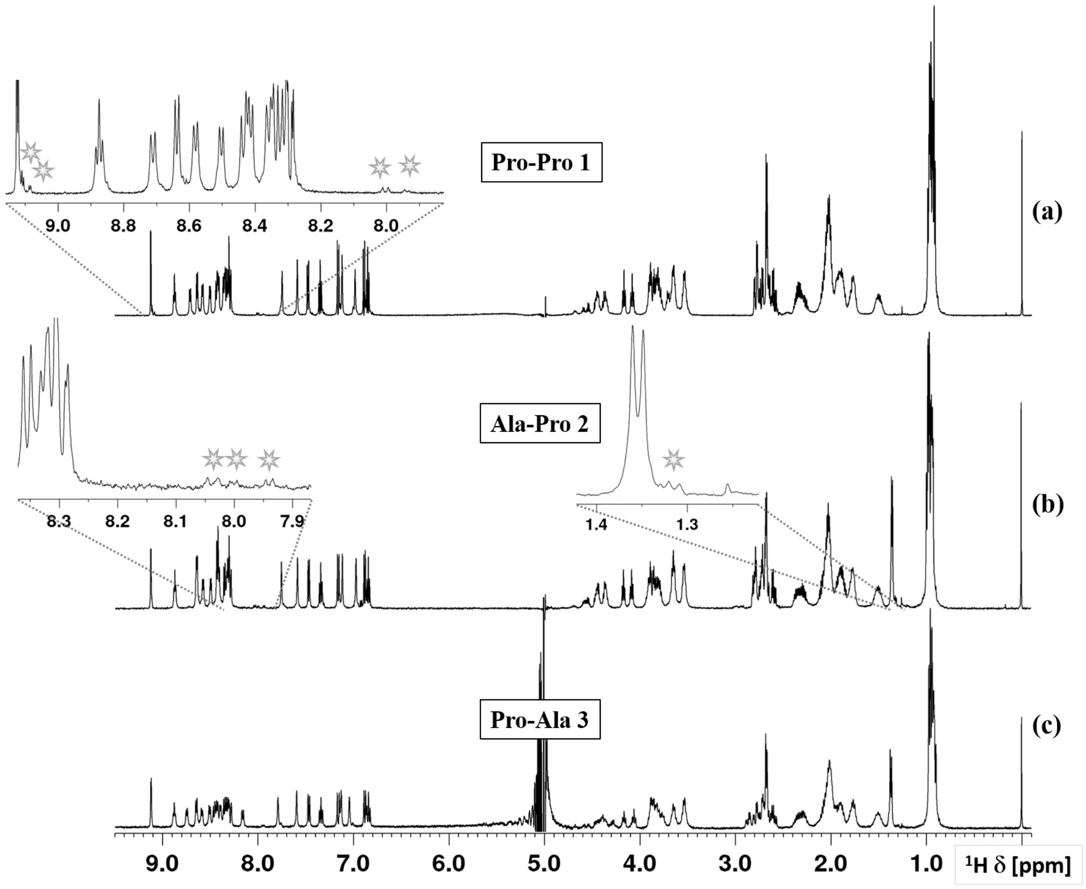


Figure 4.2.: ^{1}H NMR spectrum of peptides **4-6**. The regions where additional signal sets are detected are enlarged and marked with asterics (**a, b**: 600 MHz/**c**: 500 MHz, PBS (H_2O/D_2O, 9:1), 283 K.

The attempt to determine a ratio between main form to (multiple) subform(s) was unsuccessful. Consequently, it can only be said that for two of the three sequences more than one isoform is present in solution, the main form being present in a clear excess – specifically which one this is will be resolved later in this chapter with the help of 2D NOESY spectra.

4.1.1. Identification of spin systems

Proton chemical shift assignment was performed as described in the experimental part. Since sample availability was low, spectra acquisition was challenging and time consuming. Nevertheless, increasing the number of scans and increments in F1 lead to a well resolved 2D TOCSY spectrum. The Fingerprint region of the peptide **4** is displayed in

Figure 4.3

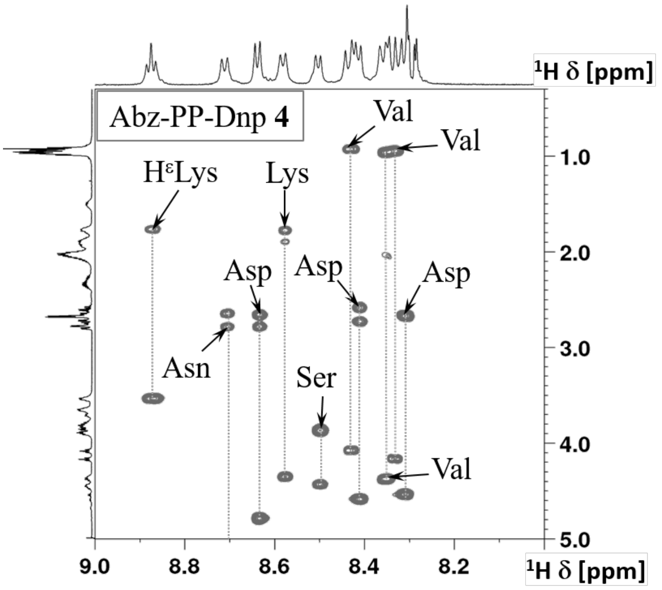


Figure 4.3.: 2D TOCSY details of the fingerprint region of peptide **4**. Identified spin-systems are emphasised with vertical dashed lines (600 MHz, 100 ms mixing time, 283 K).

A variation in the mixing time from 30 ms–120 ms further allowed the step-by-step build-up of the individual spin systems. The spin-system build up enabled a more precise correlation of the amino acids spin system length and thus, their unambiguous assignment.

Methyl groups of valine and alanine side chains were identified although the overlap was significantly below 1.0 ppm. Serine could be identified due to its two cross signals of the coupling between $\mathrm{H^N}$–$\mathrm{H^\alpha}$ and $\mathrm{H^N}$–$\mathrm{H^\beta}$. Lysine has the longest side chain of all the residues included in this sequence. Thus, the high number of vertical cross-peaks related to the amide proton at approx. 8.6 ppm could be clearly assigned to lysine residue. Asparagine and aspartic acid have an identical spin system, with two resonances at approx. 2.5 ppm, which corresponds to the coupling of $\mathrm{H^N}$ to the diastereotopic $\mathrm{H^\beta}$ protons. The signal shifted to the far left could be assigned to the $\mathrm{H^\epsilon}$ proton of the lysine side chain. Its triplet splitting is caused by the scalar coupling to to two neighbouring $\mathrm{H^\delta}$ protons. This high shift is expected for the $\mathrm{H^\epsilon}$ proton due to its deshielded location close to the electron poor 2,4-nitro-substituted aromat.

In general, the signal pattern of the 2D TOCSY spectra are very similar due to the high agreement of the three peptide sequences. The influence of the alanine exchange at position five or six nevertheless causes a signal shift of some amino acids. In peptide **5**, alanine itself is shifted at about 8.3 ppm, in peptide **6** to about 8.4 ppm. The most obvious difference is that in peptide **6** a valine residue is shifted under the influence of alanine to the right to less than 8.2 ppm, whereas in the original sequence **4** all valine amide signals are clearly at 8.3 ppm or higher (two valines overlap at approx. 8.3 ppm). In peptide **5**, a valine amide signal is also shifted from 8.3 ppm to 8.4 ppm, overlapping then finally with the valine signalat the most left.

4.1.2. Sequential assignment of the spin systems

α-Protons are very close in space to the preceding amide proton with commonly <5Å. This closeness causes a high NOE signal intensity, which makes it particularly possible to identify the signal. Thus, initial sequential assignment includes the identification of $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ NOE cross-peaks, as it is displayed in Figure 4.4, together with the related 2D TOCSY spectrum.

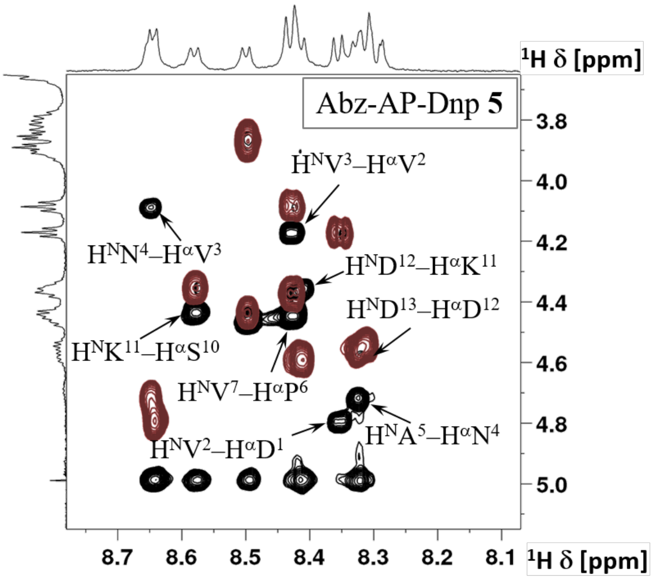


Figure 4.4.: 2D NOESY excerpts (black) of the fingerprint region of peptide **5** superimposed with related TOCSY spectra (red) excerpts. NOE cross-peaks of $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ are emphasised and labelled (500/600 MHz, 200-300 ms mixing time, 283 K).

The missing amide proton of proline disrupts the sequential assignment of neighbouring amino acids through the $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ NOE cross signals.

The signal pattern for Pro–Pro adjacent residues is analysed instead in the NOE connectivity between α- and δ-protons. NOE cross-peaks for *cis*- or *trans*-configured Xaa–Pro bonds are also presented by the dotted arrows in Figure 4.5:

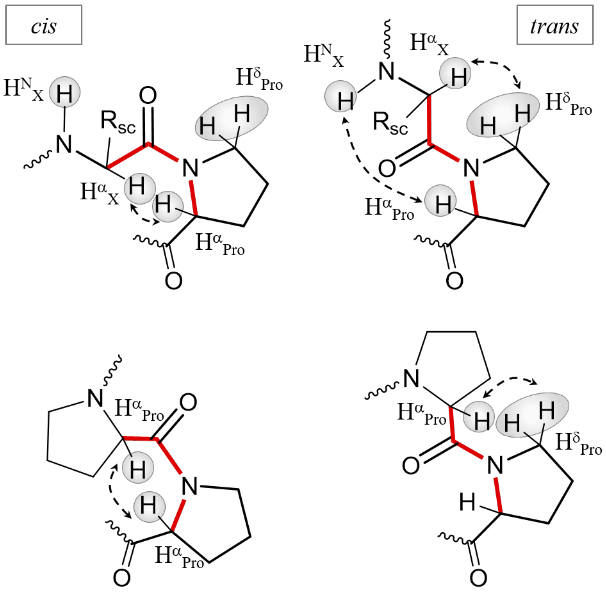


Figure 4.5.: Expected NOE contacts in Xaa-Pro *cis* (left) and *trans* (right) configured peptide bonds.

As a consequence of the expected NOE signals, identification of each proline residue was realised by NOE signals in the region between 3.5-5.0 ppm.

In Figure 4.6 the correlated NOE signals of the Pro–Pro and Xaa–Pro residues for peptides **4-6** are presented. The signal pattern shows that a *trans*-configuration for all proline involved bonds is preferred.

A comparison with the previous work on proline-rich heptapeptides **1-3** shows, that the same experimental outcome was achieved.

All proline involved bonds of the six investigated peptides show an all-*trans*-configuration for the main isoform. The signals of minor isoforms could not be assigned and identified unambiguously.

Nevertheless, those additional signal sets are originated by an angle variation of the ω-bonds of proline. This angle variation indicates that minor isoforms without an all-*trans* configuration are simultaneously present in solution.

During the evaluation of the 2D NOESY spectra, signals in the amide-amide region were observed. The assignment of those signals showed that some amide protons from neighbouring amino acids are very close to each other ($<$5Å). Amide-amide NOE signals are usually observed in secondary structure elements. To find more indications for secondary structure, the amide-aliphatic region was inspected more closely. However, in the finger-

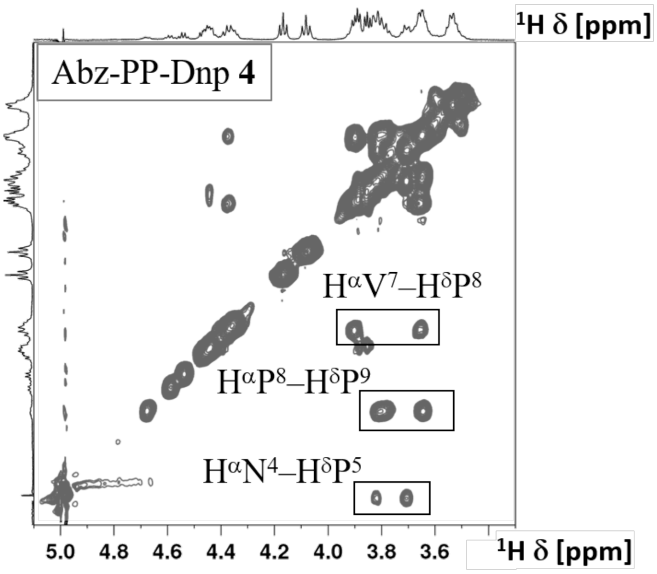


Figure 4.6.: 2D NOESY excerpt of Pro–Pro bonds. Indicated are spacial close contacts, which clearly reveal an all-*trans* conformation for the ω-angles is present (500/600 MHz, 200-300 ms mixing time, 283 K).

print region no characteristic signals were detected. Figure 4.7 shows the excerpts of the amide-amide region of peptides **4**. In those excerpts can be seen, that the number and identity of the NOE correlations differ in all three peptides.

The main information that can be gathered from the 2D NOESY spectra is that peptides **4-6** tend to form spatial proximity between neighbouring amino acids, which is also observed in secondary structure elements. These signals are definitely not observed in the shorter variants **1-3**. But overall, no higher structure can be identified, as further characteristic signals are missing.

4.1.3. 3D structure calculation based on experimental NMR data

The 3D solution structure was calculated based on NMR experimental restraints using proton chemical shifts, NOE internuclear distance and partly, scalar coupling constants. Peptides **4-6** have two modified residues within the sequence (cf Figure 4.1). Asp^1 has Abz attached and the Lys^{11} side chain is modified with Dnp. These modifications are non-standard residues for the structure calculation software. To be able to carry out the 3D structure calculation, those non-standard residues had to be parametrised and implemented into the software library.

20 final energy minimised structures were computed with least violations against experi-

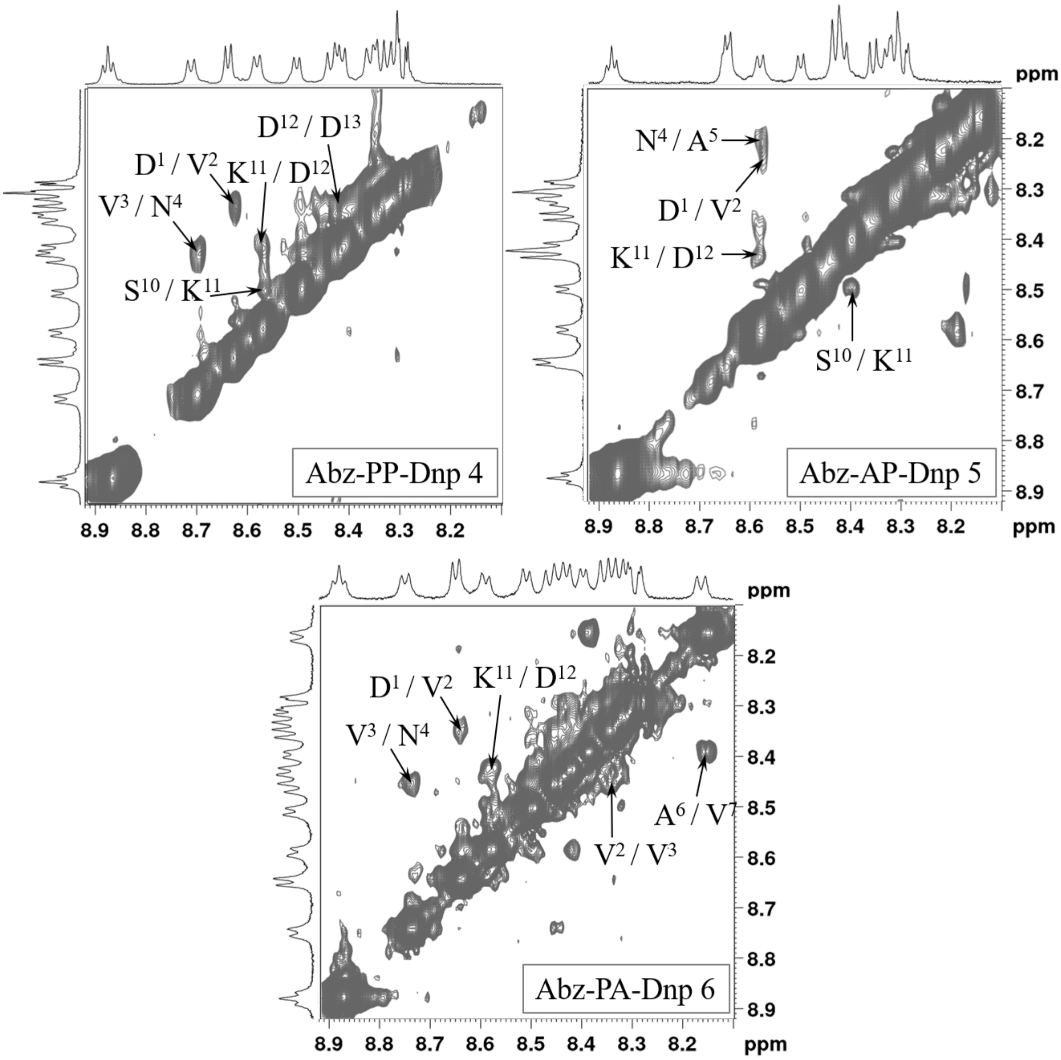


Figure 4.7.: 2D NOESY excerpt of amide-amide region of peptide **4** (600 MHz, 200-300 ms mixing time, 283 K).

mental restrains. In Figure 4.8 the structure ensemble of the 10 energy lowest structures is shown:

The structural calculation showed no evidence of secondary structure. This is consistent with the conclusion drawn from the 2D NOESY experiments. Thereof, the three peptides **4-6** are present in a random coil conformation in solution .

4.1.4. NMR experiments on the recognition process

NMR experimental studies on binding events between peptides **1-3** and PPEP-1 were carried out with STD experiments. The inactive form of the protein, apo PPEP-1, was used. The inactive apo is necessary to prevent the peptide sequence from being cut by

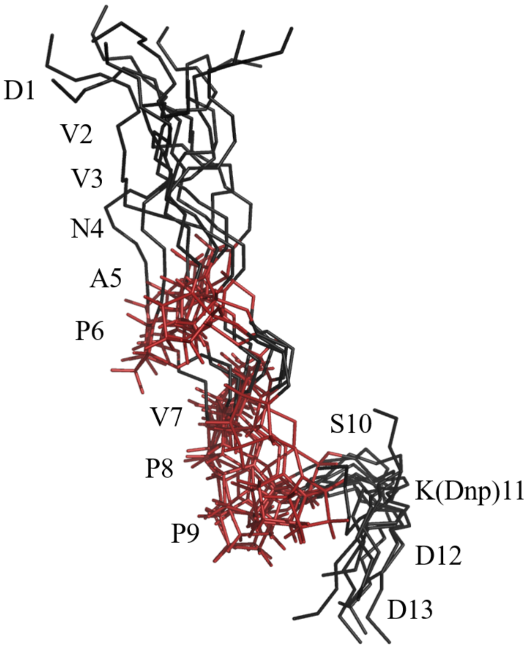


Figure 4.8.: Superimposition of 10 energy minimised 3D solution structures along backbone atoms (N, C_α, C, O) from residue 2 to 14 (terminal caps Abz and NH_2 groups were excluded in alignment process).

the protein.

In Figure 4.9 the STD amplification factors of peptides **1-3** are plotted as function of the saturation time.

The identification of the interacting protons was strongly hindered by the signal overlap of the proline protons. For the proline protons, the amplification factors were partially combined because separation was not possible.

An epitope mapping (the signal with the highest STD amplification factor, i. e. β1-protons of prolines, were set to 100%), which is shown in Figure 4.10, was generated from the calculated amplification factors. All combined β- and γ-protons of the three proline residues interact most strongly. The interaction mode of the proline protons is also in agreement with the observed interaction in the crystal structure. In the crystal structure it is evident that the side chain of the prolines are involved in either an aliphatic-aromatic network (Pro^4 and Pro^7) or sourrounded by a hydrophobic pocket (Pro^5).[73]

STD experiments with peptides **2** and **3** were not insightful because no signals could be obtained. The success of an STD measurement depends significantly on the dissociation properties of the ligand. Both modified sequences with one alanine each in the recognition

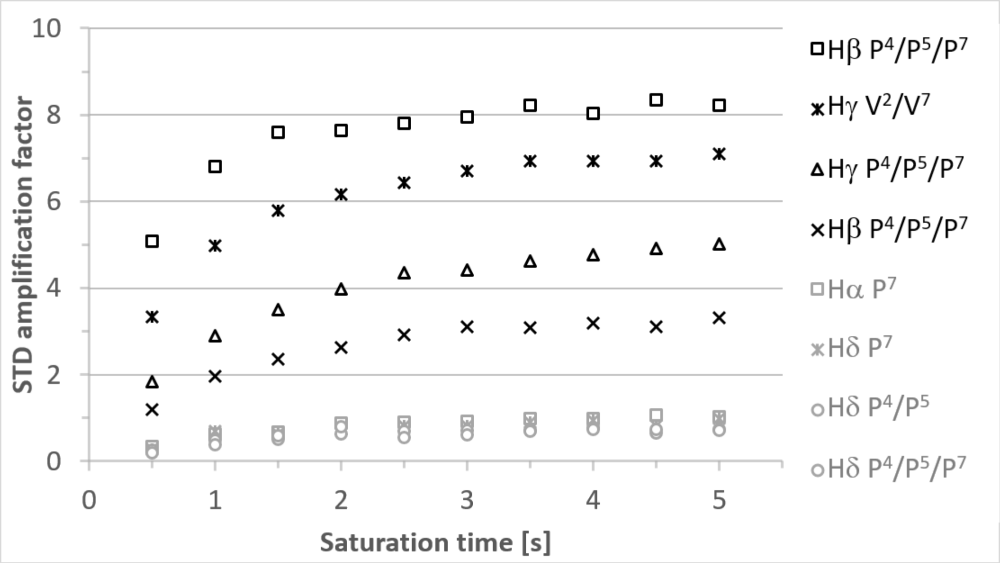


Figure 4.9.: STD amplification factor as function of applied saturation times. In black are all build-up curve with an amplification factor larger than 1, in grey all other curves belonging to proline moieties. Amide signals and all other protons of the heptapeptide are left out for the sake of clarity (600 MHz, 298 K).

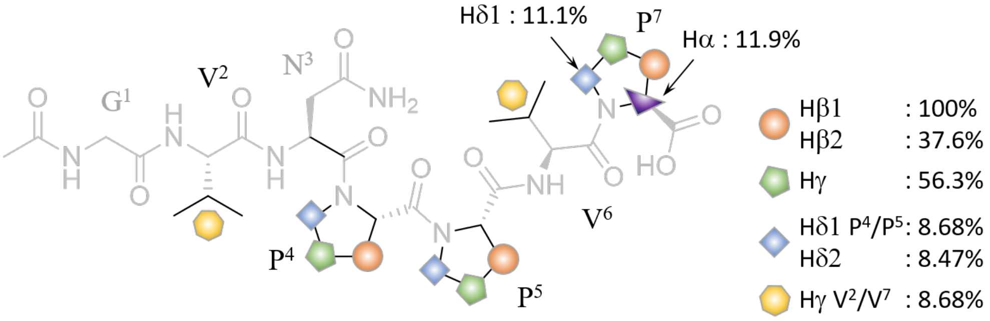


Figure 4.10.: Epitope mapping graphic of the relevant part which is located within the binding pocket during interaction.

segment have a reduced Michaelis-Menten kinetic.[67] This lowered selectivity between ligand and protein leads to an insufficient saturation of the ligands during its interaction with the protein and consequently, to the absence of STD signals.

4.2. An exploratory computational investigation

Docking studies are often used to screen broad ligand libraries before testing their performance at the laboratory scale. These preliminary tests for binding properties save time and resources. Computer-aided calculations with large molecules are usually carried out with approximated rigid structures with few degrees of freedom.[74,75] The aim here is to find out whether a reasonable result can be found with the full rotatability of a hepta peptide.

In Figure 4.11 analysis of docking studies are displayed. As it can be seen for all of the three heptapeptides, multiple interactions can occur. The free binding energy for PP-pep **1** amounts to –7.4 (kcal/mol) as is the case for AP-pep **2** and PA-pep **3**.

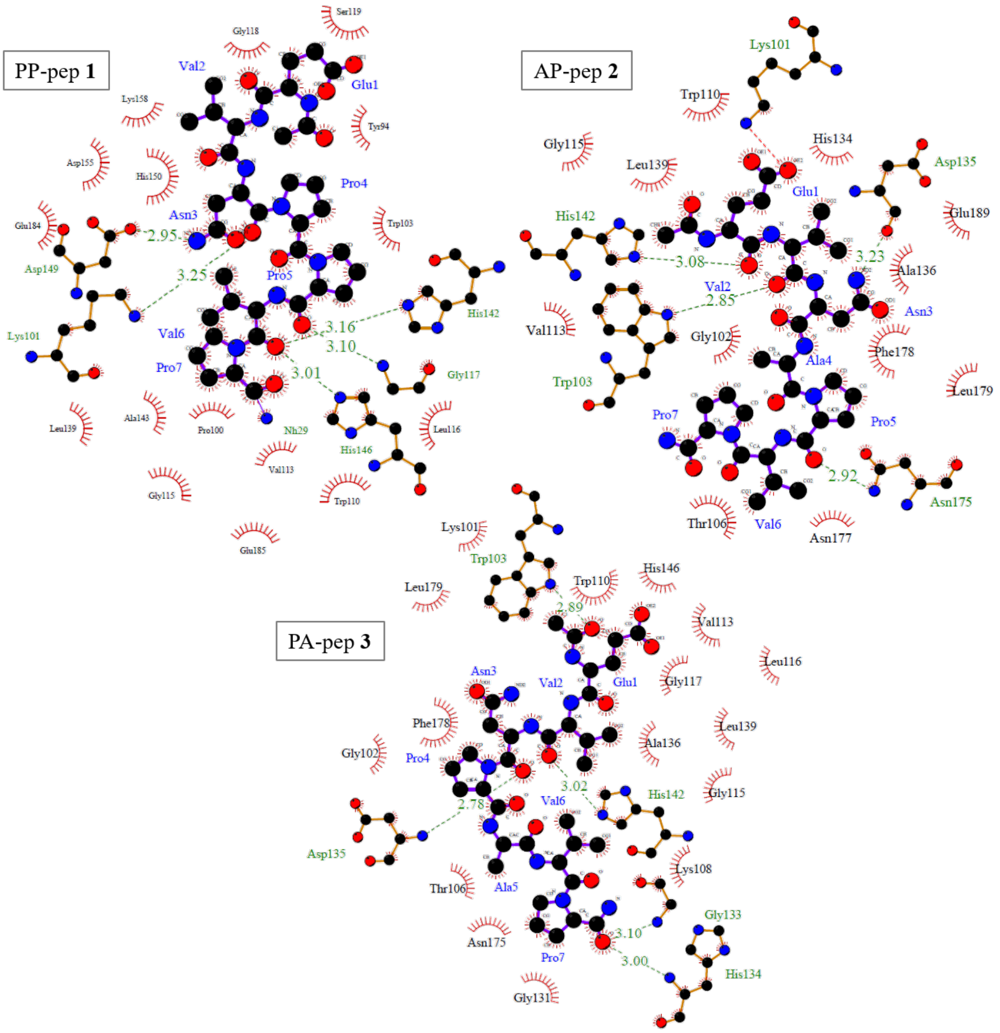

PP-pep 1
Ser119
Gly118
Val2
Glu1
Lys158
Tyr94
Asp155
His150
Pro4
Glu184
Asn3
2.95
Trp103
Asp149
3.25
Pro5
His142
Lys101
Val6
3.16
3.10
Pro7
3.01
Gly117
Leu139
Ala143
Pro100
Nh29
His146
Leu116
Val113
Gly115
Trp110
Glu85
AP-pep 2
Lys101
Trp110
Gly115
His134
Leu139
Asp135
Glu1
His142
Glu189
3.08
3.23
Ala136
Val2
2.85
Asn3
Val113
Gly102
Phe178
Ala4
Trp103
Leu179
Pro7
Pro5
2.92
Asn175
Thr106
Val6
Asn177
Lys101
Trp103
Leu179
Trp110
His146
2.89
PA-pep 3
Val113
Leu116
Asn3
Val2
Glu1
Gly117
Phe178
Leu139
Gly102
Pro4
Ala136
3.02
His142
Gly115
2.78
Val6
Asp135
Lys108
Thr106
Ala5
Gly133
Asn175
3.10
Pro7
3.00
His134
Gly131

Figure 4.11.: Results of the docking studies displayed with the Ligplot$^+$ application (green dashed lines: hydrogen bonds, red beams with semicircle: residues involved in hydrophobic contacts that do not belong to the ligand – black spheres of ligand with beams: atoms belonging to ligand which are the corresponding to hydrophobic contacts)

The proline side chain is significantly involved in interaction processes which is important for the recognition process on the protein. Further, multiple hydrogen bonds are formed within the binding pocket, which involve the protein residues His^{142} in all three cases. The importance of His^{142} for the binding interactions is also observed in the crystal structure. Surrounding residues involved in hydrophobic interactions are in all three situations, similar.

The docking experiments with so many torsion angles respectively degrees of freedom prove that these calculations are too demanding for the AutoDock vina software. Only some interactions could be identified that were also found with NMR studies and X-ray crystallography.

5. Summary and Outlook

Three proline-rich peptides were investigated with NMR spectroscopical methods in solution. Each of these peptides consisted of 13 amino acids and differed in their core motif, which is important for the function of the *C. difficile* bacterium. The specificity of the PPEP-1 protein, which is secreted by the bacterium to selectively cut between two proline bonds, was investigated by using peptides and varying one of the two prolines with alanine. The determination of the solution structures by NMR spectroscopy and further docking studies were to provide a deeper insight into the mechanism. Finally, the results were to be compared with a previous work on proline-rich hepta peptides, which already showed significant differences in their 3D structure due to inserted alanine residues. For this purpose a complete assignment of the spin systems was performed with 2D TOCSY experiments. The amino acids were then successfully assigned to the sequence using 2D NOESY experiments. Clear NOE signals showed that the peptide components are in a *trans* conformation. Substructures were present in such a small percentage that they could not be further identified. Based on experimental data and the 3D solution structure was calculated.

STD experiments with the hepapeptides showed that this approach is limited for these systems. Unfortunately, the differentiation of the individual proline signals was complicated due to signal overlap no absolute values were determined for individual protons in the epitope mapping. Furthermore, STD spectra could not be evaluated for peptides **2-3**, because no clear STD signals were detected. An unfavourable dissociation constant might be the reason for a short binding interaction with in the proteins active side and thus, no build-up of STD. Multicomponent experiments with two peptides together with the PPEP-1 protein were also performed, but could not be evaluated unambiguously, even though alanine shows well separated resonances. Computer chemical docking calculations finally showed a general agreement of the interacting residues in the active side of the protein with the respective residues of the peptides. In summary, the 3D solution structures are all in good agreement with the crystal structures. Interaction experiments and docking studies supported the presence of the *trans* configuration.

Part III.

STRUCTURES OF PEPTIDES IN SOLUTION AND IN MEMBRANE MIMETIC MEDIUM

1. Background

Cell membranes are natural selective barriers that protect living cells from the surrounding environment. The passage of substances through them is largely dependent on their lipophilicity and charge. Novel, effective therapeutics include large, hydrophilic and thus, often cell-impermeable molecules. Their applicability is therefore limited by their penetration characteristics through a cell membrane. Consequently, an approach for drug delivery into the cell to their point of action is required to overcome this limitation.[76,77]

In 1988, two working groups discovered simultaneously that the *t*ransactivating *t*rans-*a*ctivator of *t*ranscription (abbr. TAT) protein of *h*uman *i*mmunodeficiency *v*irus 1 (abbr. HIV-1) was taken up from various cell lines. On this basis, the first *c*ell-*p*enetrating *p*eptide (abbr. CPP) TAT, and later the prominent penetratin, were discovered.[78–80]

CPPs are able to translocate themselves, and facilitate translocation of small components up to large DNA fragments, into cells without causing membrane damage. Bioactive cargoes are thereby associated on the CPP through either a covalent linkage or non-polar interactions.[81]

One broad field of ongoing research is cancer therapy as it was shown that CPPs highly accumulate in cancer tissue. Conventional treatment approaches are mostly limited by the toxicity of the applied chemotherapy to healthy cells. Thus, a main difficulty is to increase the selectivity of cell types, i. e. delivery of a drug to the cancer cells while leaving healthy cells unaffected. Effective linkage and targeted release of chemotherapeutics through the use of CPPs might help to exceed the described limitations.[82,83] Consequently, CPPs are a promising area of research for the potential application of toxic or cell-impermeable drugs, which would harm or even not penetrate a cell membranes on an independent basis.[84]

1.1. Design of CPPs

Peptides as pharmaceutics represent a unique class of potential therapeutics which has evolved in the last decades. The first developed therapeutic was insulin in diabetics, followed by oxytocin and vasopressin as synthetics and later the bioactive peptides of the venoms of arthropods and cephalopods. The first extracted bioactive peptides were the trigger to an increasing isolation of peptides and proteins as potential therapeutics.[85] Native bioactive peptide therapeutics suffer from degradation by proteases, which lead to a decreasing effectiveness of the bioactive compounds.[85]

Among numerous strategies to minimise or stop rapid biological degradation, cyclisation is an emerging concept. The cyclic structure rigidifies the peptide and also forces the positively charged side chains of the lysine and arginine moieties into an optimal spatial distribution for the interaction with anionic membranes.[86,87]

A rough distinction is drawn between two types of cyclisation: head-to-tail (peptide bond between C-terminus and N-terminus) and side-chain-to-side-chain cyclisation. A representative example for the latter one is the triazole-bridged CPP. It is based on the linear sC18 peptide, which is descended on the CAP18 sequence. This was thought to stabilise β-turn, -hairpin or α-helical structure. Translocation into different cell lines, optimised stability against proteolysis and non-toxicity were proven with those cyclic peptides by biochemical investigations, as well as using multiple NMR spectroscopic techniques.[88]

The sCAP18 sequence (18 kDa) is a member of the group of *c*ationic *a*nti*m*icrobial *p*eptides (CAMPs), which can be found in living mechanisms. CAMPs are able to inactivate bacterial pathogens very quickly by neutralising the *lip*opoly*s*accharides (LPS), which are the major cause for gram-negative bacteria induced septic shocks. Their cationic or amphiphilic character enables a binding of these negatively charged LPS and the subsequent transfer out of the cell.[89,90] Studies show that the highly cationic fragment at the C-terminal domain, C18, adopts an α-helix conformation when in contact with artificial lipid membranes[88] which is thought to be responsible for the antimicrobial properties. NMR spectroscopic studies demonstrated that the influence of membrane-mimetic media, i.e. 30% (v/v) TFE, induces a rigid helical structure in the CAP18 fragment in which the cationic groups are optimally arranged to favour membrane interactions.[91] The C18 fragment describes a sequence which is very similar to the TAT protein fragment.[92] This sequence similarity led to the assumption that a shorter variant of the CAP18 peptide, called sC18, develops similarly good CPP properties as the TAT protein.[93] Meanwhile, the sC18* sequence is preferably used, which is four amino acids shorter at the C-terminal end than the sC18 structure. Despite the shortened sequence, there is hardly any difference in the cell permeability.[94]

The chemical properties of 2,5-*dik*eto*p*iperazine (DKP) range from resistance to proteolysis up to mimicking pharmacophoric groups. The instrumentalisation of those properties by incorporating 2,5-DKP scaffolds as non-natural amino acid component into a CPP sequence opens up many possibilities for therapeutics.[95]

DKP scaffolds itself are important biological active compounds that bind to a wide variety of receptors and show antitumoural, antiviral, antifungal and antibacterial properties. In nature DKPs are found e. g. produced by degradation in processed food and biosynthesised by different organisms, mostly embedded in more complex structures. Synthetically it is easy accessible by condensation of two α-amino acids.[95,96]

One key fact for the 2,5-DKP motif is their rigid six-membered, heterocyclic structure that can be controlled stereochemically in up to four positions, thus providing a larger pool of structure diversity.[97]

Studies on the use of cyclic tripeptides (Arg–Gly–Asp (RGD)), which are cyclised through the incorporation of a DKP scaffold into the sequence backbone, are recently conducted in the area of cancer therapy.[98] Further studies on those modified RGD sequences showed that DKP can also act as structure inducer, e. g. forces conformational rigidity through intra-molecular cyclisation and hydrogen bonds. DKPs can thus be classified in the group of peptidomimetics, which are known for their structure-inducing property and the associated, facilitated interaction with membranes.[99]

1.2. Internalisation mechanism and classification of CPPs

CPPs are short amino acid sequences (up to 30 residues) that are usually positively charged.[76] They can be divided into three classes based on their individual *i)* origin, *ii)* conformation and *iii)* physico-chemical properties.

i) Penetratin and the TAT protein are representatives for the 'origin-based' category, since both are protein-derived. *ii)* The conformation of CPPs can be linear or cyclic, whereby the number of linear forms is predominant. *iii)* The physico-chemical character of CPPs is broadly divided into cationic, amphiphilic and hydrophobic. Thereby, the net positive charge at physiological pH is always optimal to interact excellently with a negatively charged cell membrane surface. Lysines and arginines are naturally found and preferentially used (in synthetic sequences) amino acids in CPPs. The number and position of strongly positively charged residues influences the grade of the cationic character.[84]

The exact internalisation mechanism and thus the transport of cargoes into a cell is still not clear. The consensus are these three pathways: direct penetration, endocytosis-mediated entry and formation of a transitory structure.[76,87] It is assumed that translocation is not exclusively described by one single mechanism. Rather, the 'experimental' conditions seem to influence the pathway. In general, it is considered that properties such as CPP concentration and charge have a high impact. Thus, it was reported that higher peptide-to-cell ratios favour endocytosis pathway or the direct penetration to occur.[100]

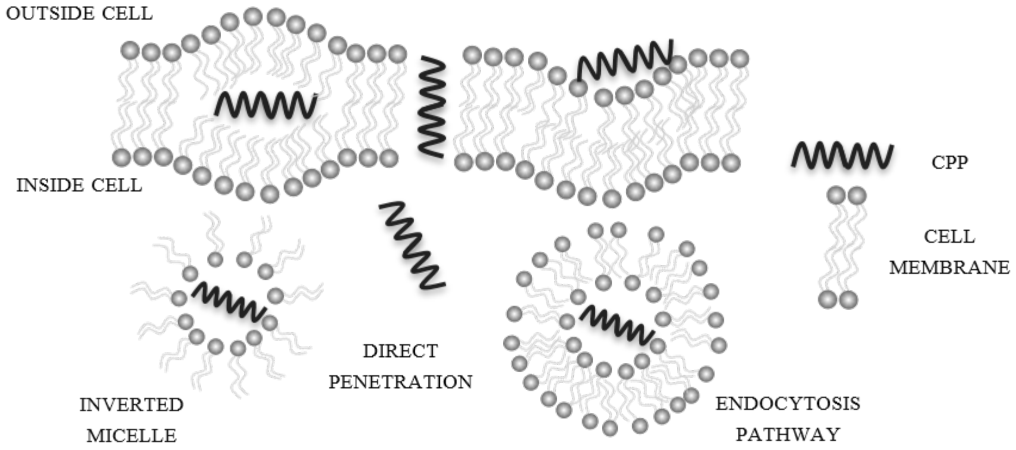


Figure 1.1.: Three considered mechanisms of CPP translocation. Through formation of transitory structure (left), direct penetration (middle) and endocytosis pathway (right).[76]

Direct penetration is an energy-independent process involving several processes. The first step is always an electrostatic interaction of the positively charged CPP with the negatively charged surface of a membrane, which is then disrupted and destabilised.[81]

In addition to direct penetration, cellular internalisation is also achieved by the second, energy-dependent *endocytosis* mechanism. Through cellular ingestion, the plasma membrane is pressed into the cell interior, allowing the substance to enter the cell.[76]

The third translocation mechanism is based on the formation of *inversed micelles*. They are aggregates where the hydrophobic tail groups point into the surrounding medium and the hydrophilic head groups into the interior where the CPP is located. This mechanism generally consists of three steps: the trapping of the CPP in the hydrophilic micelle, the interaction of the micelle with the membrane so that the inverted micelle is destabilised, and then the release of the CPP into the cytoplasm. High positively charged peptides such as the TAT peptid are barely able to form out the necessary hydrophobic interactions with the micelle and thus, are not translocating through this pathway.[76]

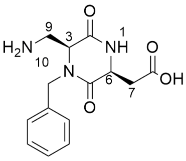
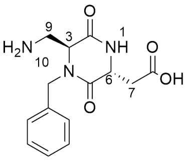
Another parameter that must not be neglected for the efficiency of translocation of CPPs is the way, in which they interact with cell membranes: the secondary structure of the peptides.[87] As already mentioned, the interaction of CPPs with membranes or membrane components is mainly based on electrostatic interactions. It has been found that CPPs often adopt a α-helical or β-turn like structure as a result of these membrane-peptide interactions.[87,101] In aqueous solution CPPs such as the TAT protein are entirely unstructured. The 'active form' for cell penetration is triggered by the approach and/or interaction with the membrane.[102,103] The 3D solution structure of the TAT peptide has already been studied in detail by NMR spectroscopy.[104] Secondary structure during the effective interaction is thought to be crucial for the effectiveness of internalisation for the final cell penetration. Consequently, a more detailed investigation of the secondary structure of CPPs in cell membrane presence is one of the most important objects of investigation.[87,101]

2. Goals

CPPs have the ability to pass through cell membranes on the basis of a special amino acid sequences. It has been shown that cyclic peptides with the sC18* sequence equipped with a DKP scaffold show an increased proteolytic stability, higher cellular uptake rates and the evidence to transport cargoes inside a cell by crossing membranes effectively.[105] The detailed mechanism of internalisation of sC18*-derived peptides is still under debate. A penetration through the adoption of a secondary structure, in which charged amino acids are arranged in an optimal position for an effective interaction with a membrane, is known to date. In the context of this work, several cell-penetrating peptides will be investigated for their 3D solution structure by NMR spectroscopy. The peptides were each equipped with two different DKP scaffolds, which only differ in one stereogenic centre. This altered stereogenic information was to be investigated with regard to its influence on the 3D structure.

In Table 2.1 the molecular formula of the DKP motifs, short name description and the peptides sequence are listed:

Table 2.1.: Peptide sequences including amino acid numbering studied in the scope of this work.

DKP scaffold	short	Sequence
DKP1: *3S, 6S* 	**1** linear DKP1-sC18*	DKP1-G^{1}LRKR^{5}LRKFR^{10}NK^{12}
	2 linear DKP3-sC18*	DKP3-G^{1}LRKR^{5}LRKFR^{10}NK^{12}
DKP3: *3S, 6R* 	**3** cyclic DKP1-sC18*	DKP1-G^{1}LRKR^{5}LRKFR^{10}NK^{12}
	4 cyclic DKP3-sC18*	DKP3-G^{1}LRKR^{5}LRKFR^{10}NK^{12}

The complete assignment of the peptide sequences is first carried out in aqueous medium.

On the basis of the obtained experimental restraints, 3D structure calculations are performed to examine the peptides for their structural features. Molecular modelling is performed in addition to the NMR experiments to examine the structures for secondary structure in the best possible way. The interaction with a membrane is subsequently investigated by using membrane mimetics (i. e. SDS). The assignment of the structures is thus based on a peptide that is in an active interaction. The structure calculation based on the experimental data should ultimately provide information about whether a conformational change in the structure can be observed, as this may indicate effective penetration.

3. Material and Methods

General. All compounds used for the preparation of the buffer (i. e. sodium phosphate mono- and dibasic) as well as the deuterated water were purchased from Sigma-Aldrich and d_{25}-SDS from Cambridge Isotopes Laboratories. DKP scaffolds and peptide sequences were synthesised and kindly provided by the research group of Prof *Neundorf*.

NMR samples. DKP-peptides (i. e. cyclic and linear DKP-sCAP18 derivatives) were dissolved in 200 μl PBS buffer (pH* 6.08, H_2O/D_2O, 9:1, final concentration ca. 1.3 mM, 3 mm NMR tubes). The concentrations for perdeuterated SDS (d_{25}-SDS) samples resulted in a peptide:SDS molecule ratio of 1:20-100.

NMR experiments and assignment. Spectra were recorded on a Bruker Avance II 600 spectrometer (^{1}H frequency of 600 MHz) equipped with a triple resonance high-resolution probe (TBI) using Topspin software (Bruker). The transmitter frequency was set on the HDO/H_2O signal, and the resonance of d_4-TSP (3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt) was used as chemical shift reference (^{1}H $\delta_{\text{TSP}} = 0$ ppm).

One and two-dimensional spectra were acquired by standard pulse sequences using solvent suppression sequence. Two-dimensional homonuclear experiments at specified temperatures (i. e. TOCSY and NOESY) were recorded for assignment and structural calculations. The spectra were processed with Topspin software (Bruker).[106] For the measurement of diffusion coefficients, data acquisition and analysis were carried out using Bruker Topspin software (Bruker BioSpin, Germany). The DOSY experiments were performed using a standard pulse sequence from the Bruker library including water suppression (*stebgp1s19*). The diffusion delay Δ was set at 100 ms, the gradient strength, g (1 ms) was linearly incremented in 32 steps from 2% to 98% of its maximum value, and a longitudinal eddy currents delay of 5 ms and eight/sixteen scans were recorded for each experiment. Processing was achieved using 8K points in the F2 dimension and 32 points in F1. An exponential window function with 3 Hz line broadening was applied in the F2 dimension prior to Fourier transformation. After baseline correction, the diffusion dimension was processed with the DOSY processing program (Bruker TopSpin software 3.5). A logarithmic scaling was applied in the diffusion axis, and a noise sensitivity factor

of 4 and line width factor of 1 were used. The fitting of the diffusion dimension in the 2D DOSY spectra was obtained using a single exponential fit.
The processed data were analysed by using CARA[70] software. Peak lists for the TOCSY and NOESY spectra were generated by interactive peak picking. NOESY cross-peak volumes were determined by the automated peak integration routine implemented in CARA. Complete assignment of the backbone and side-chain ^{1}H resonances was performed using standard sequential assignment procedures, according to the methodology developed by Wüthrich.[15]

Structure calculation. Three-dimensional structures were determined by the standard protocol of the CYANA program (version 2.1),[49] using seven cycles of combined automated NOESY assignment and structure calculations followed by a final structure calculation. Since the cyclic peptides contain non-standard residues, the corresponding libraries for CYANA were built using the PyMol software. PyMol was also used to visualise three-dimensional structures.[71] For each CYANA cycle, 100 randomized conformers and the standard simulated annealing schedule were used. The 20 conformers with the lowest final score were retained for analysis and passed on to the next cycle. Weak restraints on ϕ/ψ torsion-angle pairs and on side-chain torsion angles between tetrahedral carbon atoms were applied temporarily during the high temperature and cooling phases of the simulated annealing schedule in order to favour the permitted regions of the Ramachandran plot and staggered rotamer positions, respectively. The list of upper-distance bonds for the final structural calculation consists of unambiguously assigned upper-distance bonds and does not require the possible swapping of diastereotopic pairs. Root-mean-square deviation (RMSD) values were calculated using CYANA for superpositions of the backbone N, Cα and CO atoms; the heavy atoms over the whole peptide or the cyclic fragment. To obtain the rmsd of a structure represented by a bundle of conformers, all conformers were superimposed upon the first one and the average of the rmsd values between the individual conformers and their average coordinates was calculated.

Molecular modelling. All calculations were run by using the Schrödinger suite of programs (http://www.schrodinger.com) through the Maestro graphical interface.[107] Molecular mechanic studies on the DKP-containing cyclic peptidomimetics were carried out using MacroModel from Schroedinger software suite.[108] Initial conformations were energy minimised using conjugate gradient method[109] with the OPLS_2005 force field.[110] The convergence criterion was set to 0.05 kJ mol-1 Å-1 on the energy gradient. An unconstrained study for conformational preferences was then performed starting from this structure using the mixed-mode Metropolis Monte Carlo/Stochastic Dynamics (MC/SD) simulation method[111] with the same force field and convergence criterion. MC/SD simulations were performed at 300 K with the framework of MacroModel release 2019-2. Side-chain dihedral angles were defined as internal coordinate degrees of freedom in the Monte Carlo part of the algorithm. A time step of 0.5 fs was used for the stochastic dynamics (SD) part of the algorithm for 5 ns of simulation time. 51 structures were generated and energy minimized for analysis. To select starting geometries for conformational search, clustering of these results by using the average linkage method on the basis of atomic rmsd was performed.

4. Results and Discussion

4.1. Studied DKP-peptidomimetics

NMR experimental studies were performed on a series of linear and cyclic DKP-sC18* peptidomimetics (see Table 2.1) that were synthesised in the laboratories of *Prof. Ines Neundorf.* The peptides studied were each equipped with a DKP scaffold that had either *cis*-configured (DKP1) or *trans*-configured (DKP3) substituents (cf Table 2.1).

In Figure 4.1 1D proton traces of linear and cyclic DKP3-equipped peptides are shown.

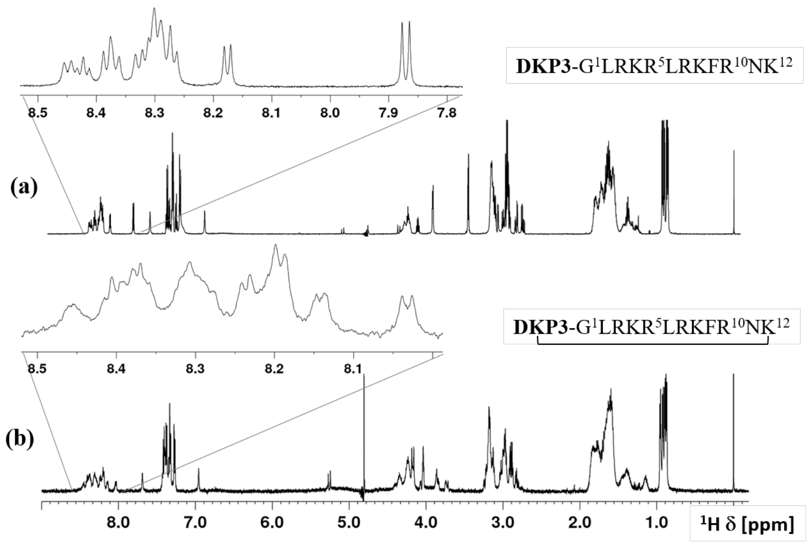


Figure 4.1.: 1D ^{1}H NMR spectra of the peptides (a) linear DKP3-sC18* and (b) cyclic DKP3-sC18*. The enlargements shows the region of amide and aromatic protons. (peptide concentration ca. 1.3 mM, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

In both spectra the signal splitting is very similar due to the identical amino acid sequence. The severe overlap of signals is mainly caused by the repetitive amino acids arginine and lysine. The enlargement of the amide regions of both peptides shows that signals of the linear peptide extend over a wider frequency range and show, additionally, a more defined signal splitting than in the cyclic analogue. This large signal dispersions indicate a lack in secondary structure in both peptides.

4.1.1. Resonance assignment: spin system identification

The identification of the proton signals of the individual amino acids was carried out using 2D TOCSY spectra at different temperatures and mixing times.

In Figure 4.2 the H^N-H^α chemical shift region of both cyclic peptidomimetics are displayed.

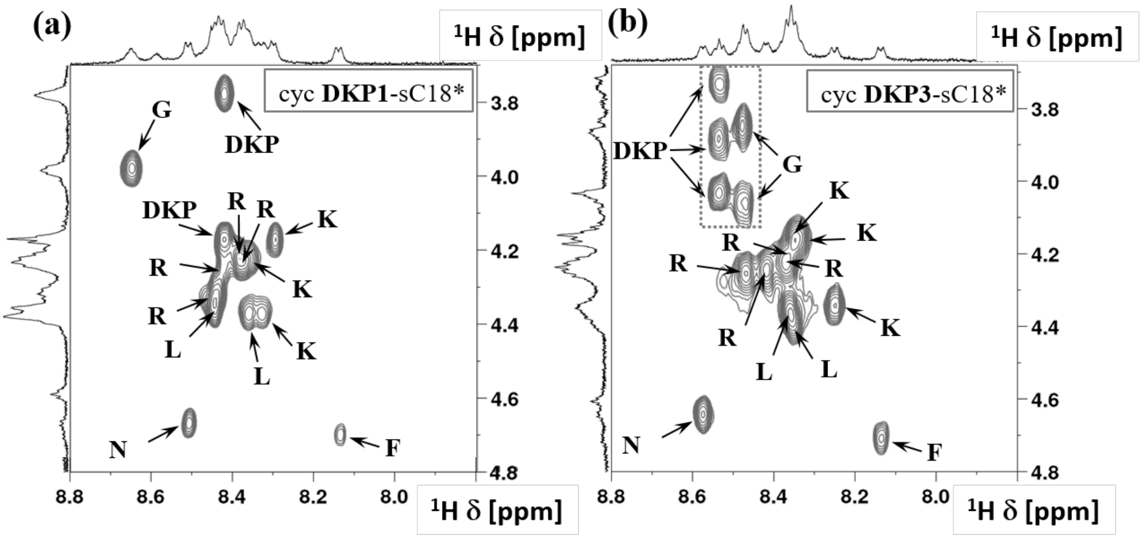


Figure 4.2.: Excerpts of the fingerprint region of 2D TOCSY spectra of (a) cyclic DKP1-sC18* and (b) cyclic DKP3-sC18*. Diastereotopicity of the methylene protons of glycine and DKP3 are emphasised in a box (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 80-120 ms, 600 MHz).

In the fingerprint region, the classic signal patterns of the amino acids Leu, Asp, Gly and Phe could be identified clearly. The signals from Arg and Lys, on the opposite, were more difficult to identify due to their very similar spin systems and thus, strong overlap in the narrow interval of 8.2–8.6 ppm.

In both cyclic peptides, the amide proton of Phe^9 is consistently shifted to the lowest frequency.

The total correlation spectrum reveals a distereotopic splitting of the methylene protons of glycine and the DKP3 scaffold. However, this splitting cannot be observed for the same protons of the cyclic DKP1-sC18* peptide. Consequently, the stereochemistry of the DKP scaffold peptide definitely influences the conformations of the cyclic DKP-sC18* peptides.

4.1.2. Sequential assignment of the DKP-sC18* peptides

Sequential assignment of all identified spin systems was performed with 2D NOESY spectra acquired at various mixing times. The identification of NOE cross-peaks was hampered by the sever overcrowding of the finger print region of all four peptides. Nevertheless, the assignment could be realised through the NOE cross-peaks between adjacent $H^{\alpha}_{(i)}$—$H^{N}_{(i+1)}$ cross-peaks. In Figure 4.3 illustrates the assignment for a few of the well-separated signals of the 2D NOESY spectrum of DKP3-sC18* peptide.

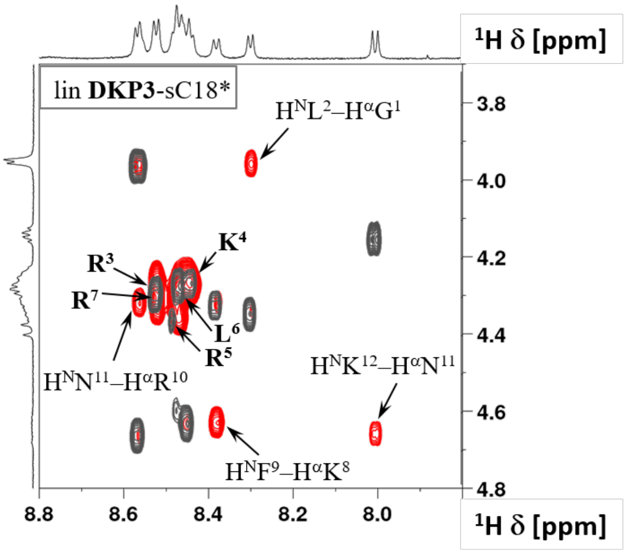


Figure 4.3.: Sequential assignment of linear DKP3-sC18* peptide. For the sake of clarity, labelling is only shown for NOE signals that do not overlap. Assignment of signals in severe overcrowded area are indicated (red: 2D NOESY, black: 2D TOCSY, 50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz).

Inter-residual NOE cross-peaks were identified that confirmed the cyclised form of the peptide. Furthermore, the conformation of the linkers itself could be verified through observed intera-residual NOE signals within the DKP scaffold, that are schematically drawn in Figure 4.4.

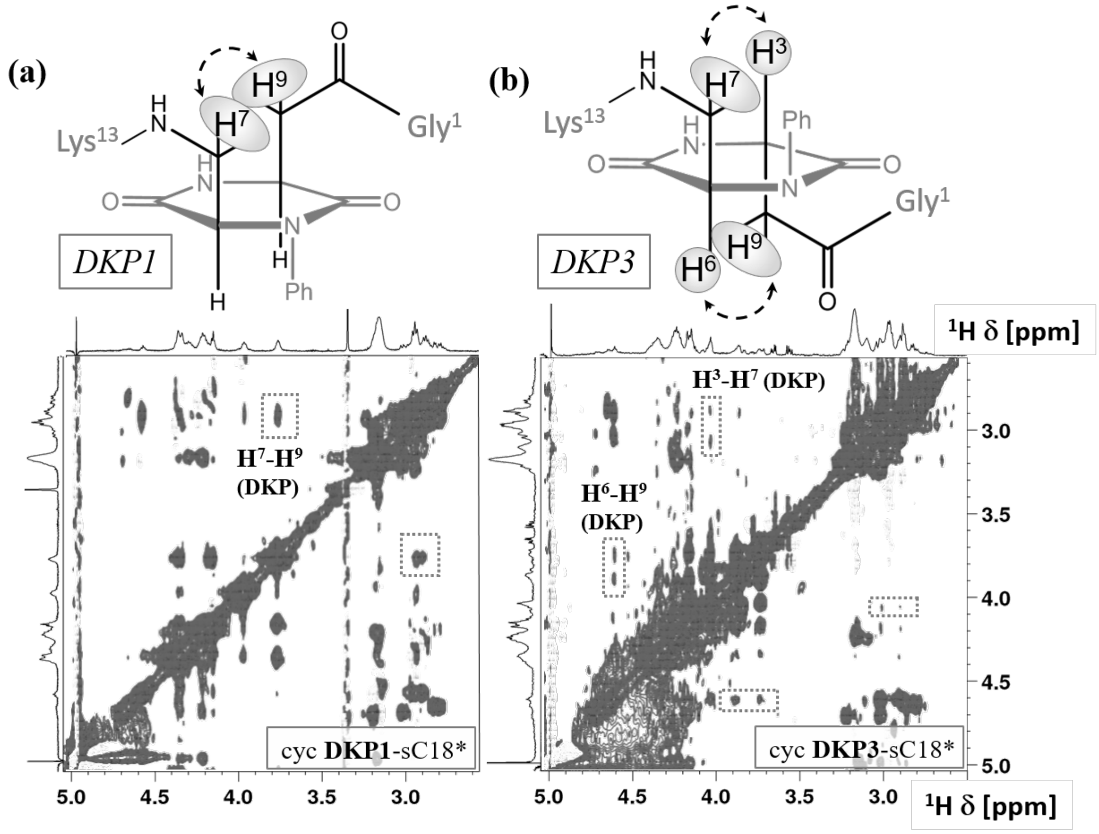


Figure 4.4.: Top: Expected intra-residual NOE (dashed lines) in the scaffolds, that is drawn in Haworth projection. Bottom: intra-residual NOE cross-signals of both DKP scaffolds within the aliphatic region of 2D NOESY spectra (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz).

A clear NOE signal between H^7–H^9 protons of DKP1 confirmed the *cis*-configuration in comparison to their absence in *trans*-configured DKP3-scaffold. *Trans*-configuration of DKP3 could be verified through NOE between H^6–H^9 as well as H^3–H^7 protons.

During the assignment process, some inter-residual NOE signals were recognised in the amide-amide region at 283 K, that were absent at 298 K. The observed inter-residual NOE signals are shown in Figure 4.5.

Neither the number nor the identity of the detected NOE contacts of peptides **3** and **4** are coincident.

This dissimilarity indicates that the structure preference in solution is different for both cyclic peptides and thus, demonstrates that the stereochemistry of the scaffold is determinant for the peptide's conformation. The appearance of those signals at a lower temperature suggest that the cyclic structures tend to exhibit a secondary structure at lower temperatures. These conformers are averaged at higher temperatures but de-

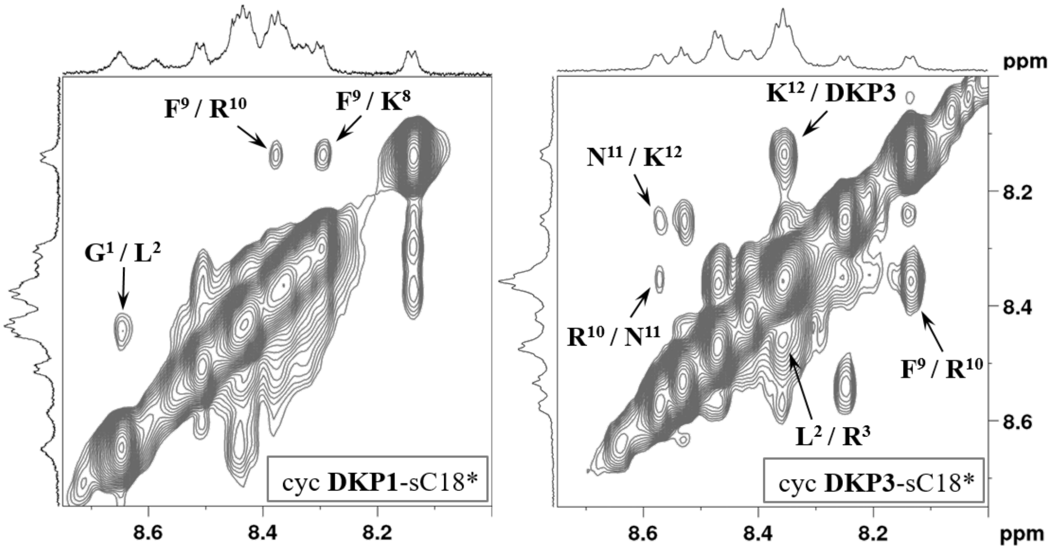


Figure 4.5.: Excerpt of 2D NOESY spectra of both cyclic peptides cyc DKP1-sC18* (left) and cyc DKP3-sC18* (right). Inter-residual NOE signals are labelled (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 200 ms, 600 MHz).

tectable when conformational interconversion is slowed down.

The amide-amide signals of cyclic peptide DKP3-sC18* are observed for four adjacent amino acids, namely Phe^{9}–Arg^{10}–Asn^{11}–DKP3. This continuous connection along four amino acids is indicative for a β-turn structure. However, the severe signal overlapping hampered an identification of further characteristic inter-residual signals.

4.1.3. Temperature Coefficients

The temperature coefficient correlates the chemical shift variation and temperature. Low temperature coefficients are taken as indicative for the presence of hydrogen bonds. The existence of hydrogen bonds thereby may correlate with stabilisation of secondary structure.[112,113] In Table 4.1 the determined temperature coefficients of all four peptide's amide backbone protons are shown. Calculation of those values was performed by the determination of the linear square fit of the slope of the chemical shift plotted as function of the temperature.

The mostly very negative temperature coefficients obtained for the four peptides suggest an absence of hydrogen bonds for most of the amino acids. However the restriction of the temperature coefficient values especially for smaller and cyclic peptides can be doubted.[112]

Table 4.1.: Temperature coefficient $\frac{\Delta\delta}{\Delta T}\left[\frac{ppb}{K}\right]$ of amide protons of all four peptide sequences determined for an experimental temperature range of 283 K–298 K. Values which might indicate an involvement in an hydrogen bond are emphasised in boxes (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 600 MHz).

Residue	**linear**		**cyclic**	
	DKP1-sC18*	DKP3-sC18*	cDKP1-sC18*	cDKP3-sC18*
Gly^1	–6.27	–8.33	$\boxed{-5.80}$	–6.93
Leu^2	–7.33	–7.47	–8.53	–8.53
Arg^3	–10.8	–8.40	–5.87	–7.07
Lys^4	–10.2	–10.4	–8.80	–8.40
Arg^5	–9.33	–6.60	$\boxed{-3.73}$	–7.07
Leu^6	–10.3	–10.0	–11.3	–10.6
Arg^7	–9.73	–12.2	–7.47	$\boxed{-5.73}$
Lys^8	–9.67	–10.7	–8.07	–10.7
Phe^9	–9.33	–9.53	–6.33	–6.80
Arg^{10}	–6.40	$\boxed{-6.00}$	$\boxed{-5.87}$	$\boxed{-5.33}$
Asn^{11}	–6.60	–7.13	–7.80	–8.07
Lys^{12}	–7.93	–8.40	–7.47	–7.20

4.2. Structural analysis

The calculation of the 3D solution structures of the peptides free in solution was performed for both temperatures, 283 K and 293 K, on the basis of proton chemical shifts and NOE signal intensities.

The structure ensemble of the linear peptides shows for all structures a random coil conformation. The random coil structures are caused by the high flexibility of the linear peptides in solution.

The ensemble of the 20 energy minimised structures of the cyclic DKP3 peptidomimetic is presented in Figure 4.6.

As can be seen, the structure ensemble indicates an uniform turn between the residues Phe^9–DKP3. This β-turn like motif confirms the observation in the amide region of the 2D NOESY spectra. An examination of the formed hydrogen bonds in the most

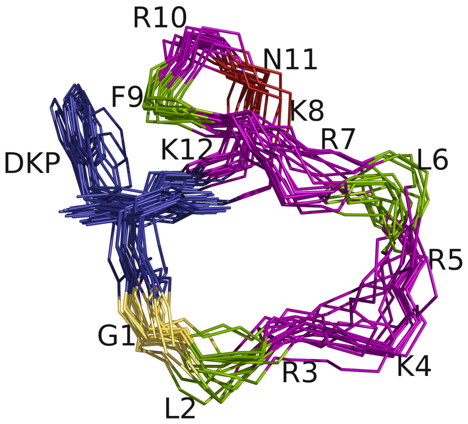


Figure 4.6.: Superimposition of 20 energy minimised structures of cyclic **DKP3**-sC18* peptide calculated with NMR experimental data evaluated from spectra acquired at 283 K (residue colours: violett: charged, green: hydrophobic, yellow: non-polar, red: polar uncharged, blue: DKP scaffold).

populated conformation at 283 K shows hydrogen bond appearance between Lys^8 O – H^N Asn^{11}, Lys^8 O – H^N Lys^{12} and Arg^5 O – H^N Arg^7. The hydrogen bond pattern is shwon in Figure 4.7.

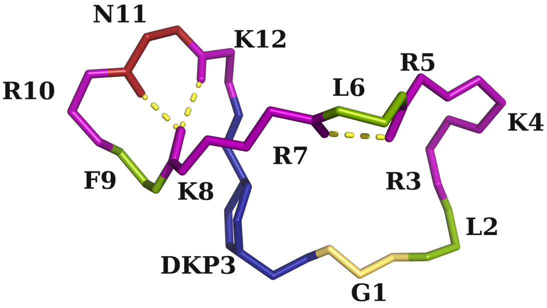


Figure 4.7.: Hydrogen bond interactions (yellow dashes) derived from NMR experimental data collected at 283 K.

In the represented structure, it is strongly evident that the hydrogen bonds stabilise a β-turn. The tendency of β-turn formation in cyclic peptides that containing a DKP scaffold has been described in literature,[114] supporting the obtained experimental data.

4.2.1. Molecular modeling studies

To further support the hypothesis of secondary structure, unrestrained mixed-mode Metropolis Monte Carlo/Stochastic Dynamics (MC/SD) molecular modelling studies were performed starting from random conformations.

In Figure 4.8 an overview of the obtained hydrogen bond pattern from the simulations in comparison to the NMR derived structure, is displayed (in appendix, the same representation of *cis*-configured peptide is shown).

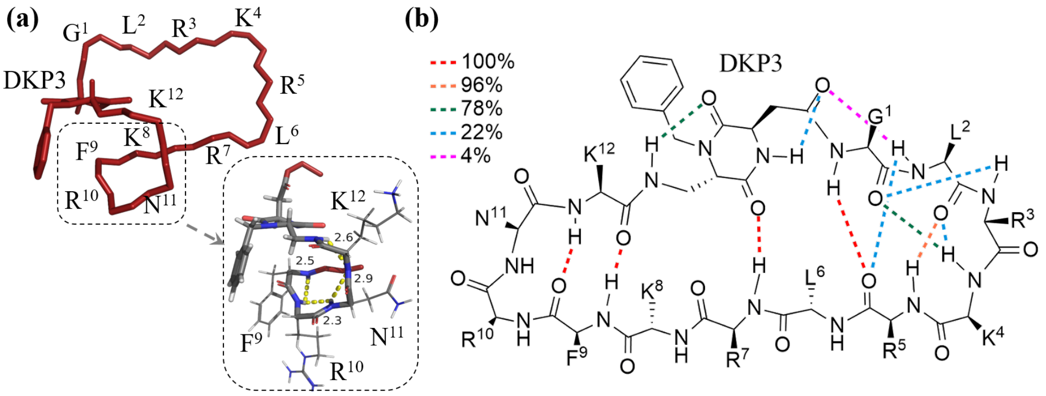


Figure 4.8.: (a) Backbone trace of energy lowest structure calculated based on NMR experimental data at 283 K. (b) Preferred intra-molecular hydrogen bond pattern (dotted lines) obtained through MC/SD simulations. Colour code of the dotted lines is indicated in the legend.

During all simulations, a minimal β-pin structure was formed between residues Phe^9–Lys^{12}, including bonding/back-bonding of the donor and acceptor part of these amino acids. Furthermore, the amide-group of the DKP3-scaffold itself forms a hydrogen bond to the amide proton of Arg^7 in 100% of the generated structures.

As has already been described in the literature[115] is the oxygen of the amide-group likely to act as hydrogen bond donor within the chain. This DKP-oxygen participation further promotes turn-like structures.

Between the amino acids Gly^1 and Arg^5 the hydrogen bond propensity varies between the structures. This variation in inter-residual hydrogen bonds supports the assumption of several simultaneously present conformers in solution. Furthermore, a variation in the hydrogen bond propensity indicates a higher flexibility in this peptide part, in contrast to the more bound residues Phe^9 to DKP3.

4.3. DKP-peptidomimetics in membrane mimetic agents

Investigations on the interaction of the CPPs with a membrane mimetic was performed by the use of water solutions containing anionic SDS micelles.

In Figure 4.9 the 1D proton spectrum of linear DKP1-sC18* peptide and its cyclic analogues in SDS-micellar medium are shown. The signal shape of amide protons is significantly broadened and they are wider dispersed in the amide region (ca. 7.6–9.0 ppm). This signal changes suggest a strong interaction between peptides and micelles.

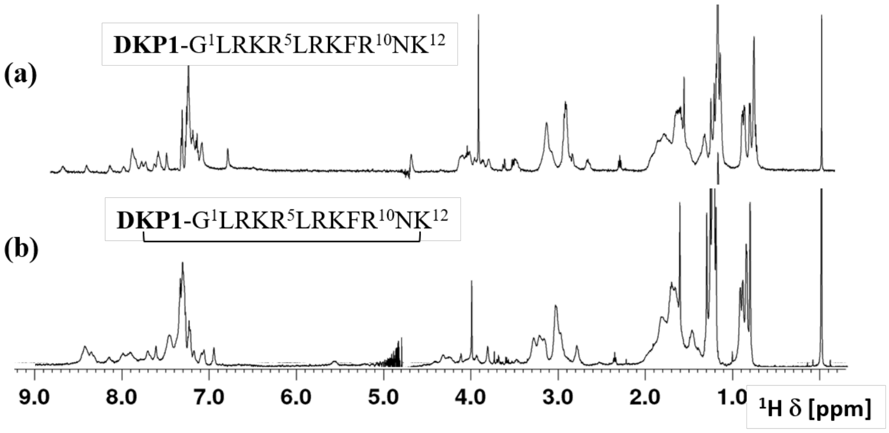


Figure 4.9.: 1D ^{1}H NMR spectra of linear and cyclic peptides in presence of SDS micelles (a) lin DKP1-sC18*, (b) cyc DKP1-sC18* (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 500 MHz cryoprobe or 600 MHz).

This interaction between micelles and peptide was confirmed through diffusion coefficient measurements, shown in Figure 4.10.

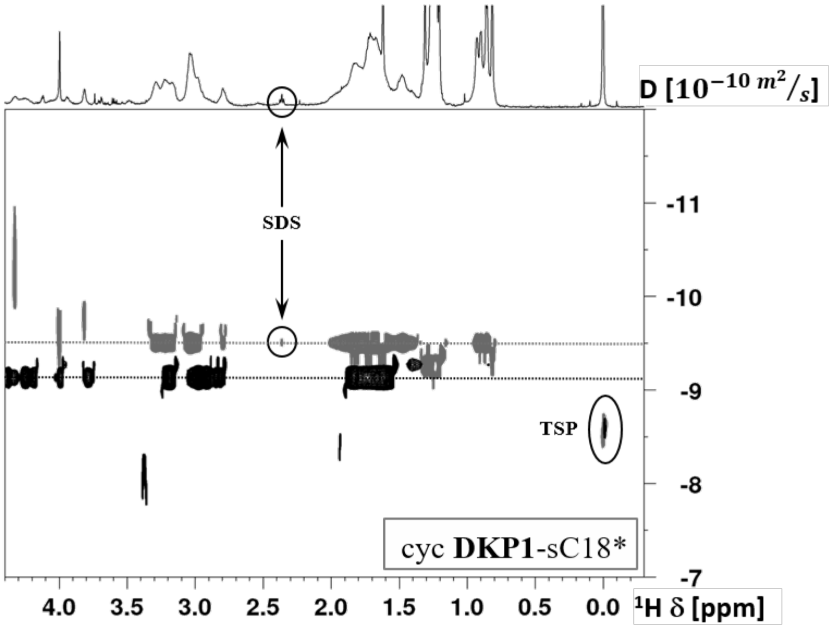


Figure 4.10.: DOSY NMR spectra of cyclic DKP1-sC18* in aqueous solution (black trace) in comparison to their presence in SDS micelles (gray trace). TSP is superimposed and taken as internal reference. 1D projection shows peptide together with micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

Remaining proton signals from perdeuterated micelles can be identified at 2.3 ppm. The diffusion coefficients of the peptide in aqueous medium in comparison to membrane mimetic agent correlate to an efficient interaction.

4.3.1. Resonance assignment of peptidomimetics in presence of micelles

The identification of the peptide's spin systems was performed through the evaluation of 2D TOCSY spectra. The assignment process was hindered by a poor signal resolution. A whole spin system identification could be accomplished for both linear peptidomimetics and the cyclic DKP1-sC18* peptide. In Figure 4.11 the fingerprint region of the linear DKP1-sC18* peptide is shown.

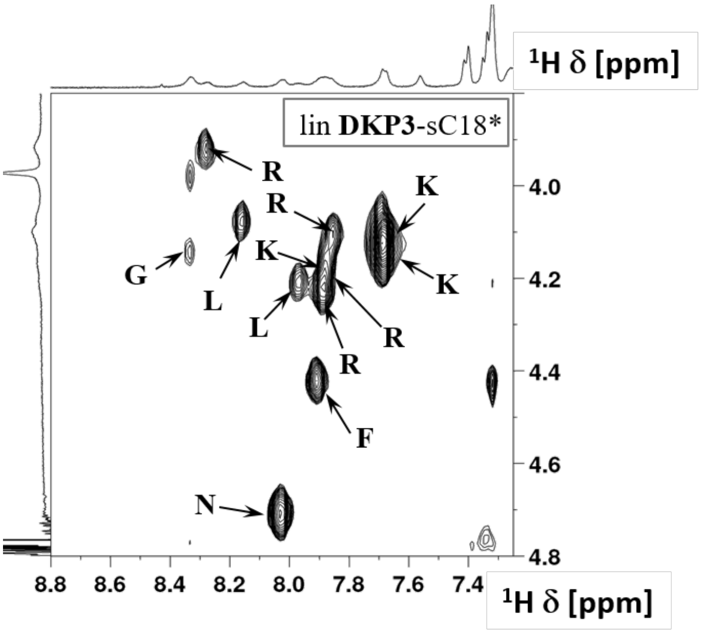


Figure 4.11.: 2D TOCSY spectrum of an excerpt of the finger print region of linear DKP3 peptidomimetic in presence of micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

The amide resonances are significantly shifted to a lower frequency range (~7.6–8.4 ppm) through the interaction with micelles (without micelles: 8.0–8.6 ppm). The spin systems of Asn, Phe, Leu and Gly could be identified through their classical pattern. The identification of the Lys and Arg spin systems was slightly facilitated in micellar solution compared to the peptide in water due to the more segregated signals.

4.4. Sequential assignment of peptidomimetics

The connection of the identified spin systems was achieved through the evaluation of 2D NOESY spectra.

Figure 4.12 shows part of the fingerprint region of the linear DKP3-sC18* peptide.

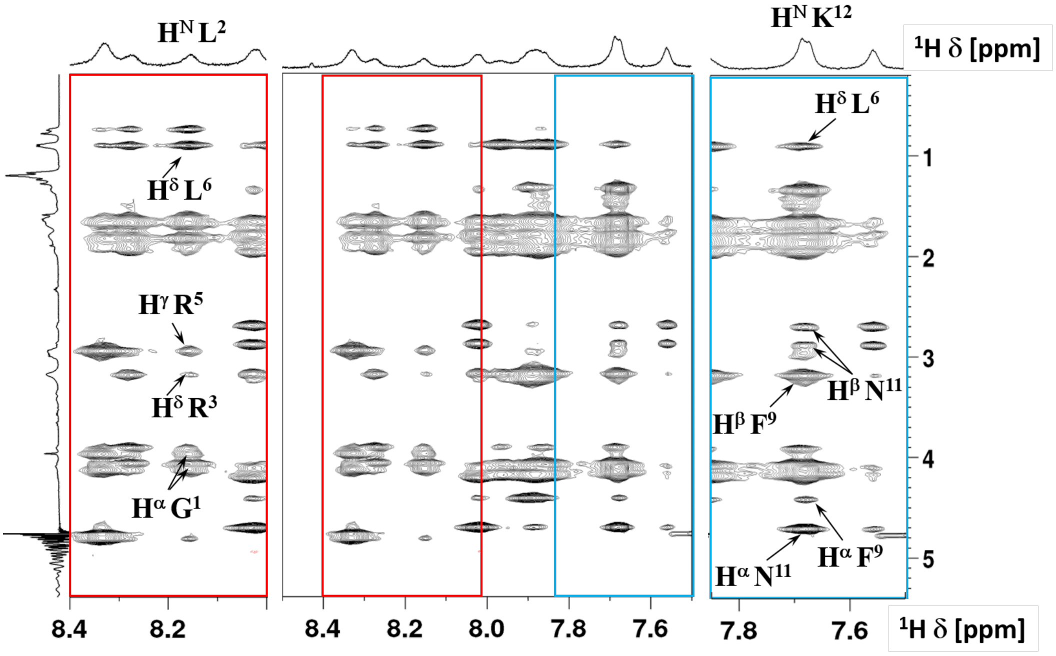


Figure 4.12.: 2D NOESY of the fingerprint region of linear DKP3-sC18* peptide (middle). Two exemplary excerpts of inter-residual contacts in an α-helix are shown (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

In the 2D NOESYs, it was quickly noticed that the classic signal pattern of an helical structure was formed. The spatial contacts of Leu^2 and Lys^{12} are shown as examples.

In the amide region, clear NOE contacts were detected for all peptides. Figure 4.13 shows the H^N-H^N NOE connectivities of the cyclic DKP1-equipped peptidomimetic.

Several inter-residual contacts are observed for. It is particularly interesting that interactions are detected along the four neighbouring amino acids Arg^{10}–DKP1 amide protons. The same spatial contacts were previously observed for the cyclic DKP3-sC18* peptide at low temperature. Thus, this signal pattern suggests that a β-turn may also be present, but induced by the interaction with a membrane mimetic.

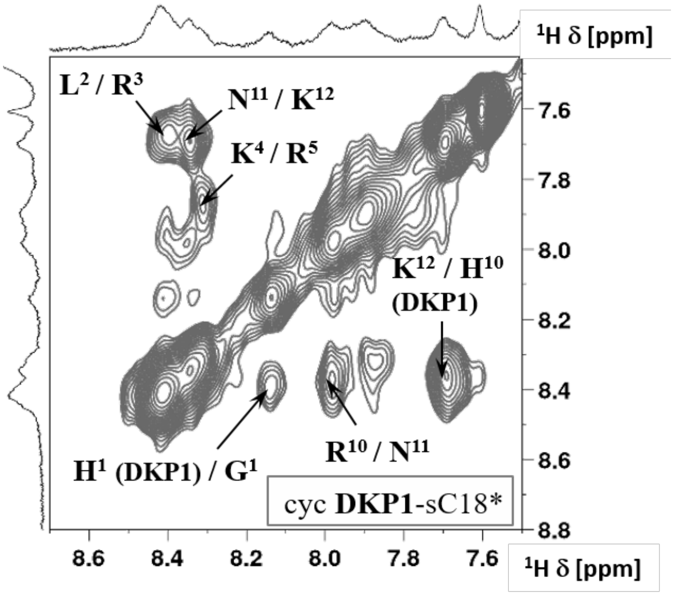


Figure 4.13.: 2D NOESY amide-amide region of DKP1-sC18* peptide (middle). Two exemplary excerpts of inter-residual contacts in an α-helix are shown (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

4.5. Structural analysis

The calculation of the 3D structure of the linear peptides derived from NMR data acquired in a micellar medium, confirms the assumption that the peptides adopt an α-helix. Based in the amide and α-proton chemical shifts, a CSI plot was calculated. The CSI plot confirms the existence of an helical structure as can be seen in the bottom in Figure 4.14. Furthermore, the localisation of the amino acids in an α-helix is be visualised by creating a so-called helical wheel, which is presented in the same Figure.[i]

The effective interaction of the SC18*-derived motifs might occur through the electrostatic interactions of the highly positively charged side chains of Arg and Lys amino acids with the anionic membrane surface. A favoured positioning of Arg and Lys residues in one side might facilitate an interaction with a micelle, which could lead to cell penetration. Thus, the exact positioning of Arg and Lys amino acids during interaction with the micelles was of interest. Inthe linear peptidomimetics, the localisation of the Arg and Lys amino acids is evenly distributed around the helical wheel. This observation is in agreement with the obtained inter-residual contacts from the NOE spectra.

---

[i] The helical wheel was build by using the pepwheel application provided on the emboss webside: https://www.bioinformatics.nl/cgi-bin/emboss/pepwheel

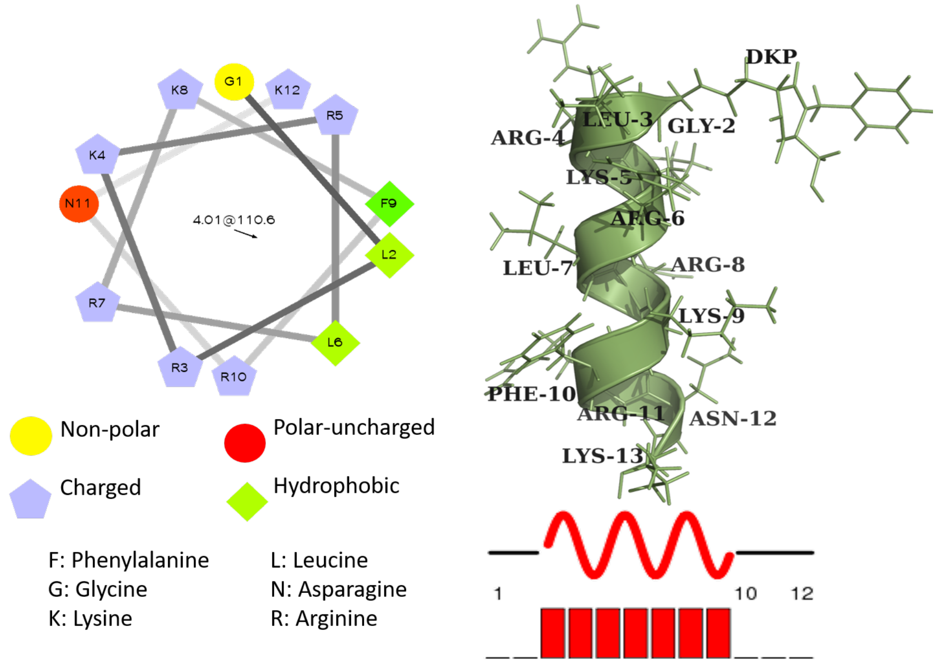

G1
K8
K12
R5
K4
N11
F9
L2
4.01@110.6
R7
L6
R3
R10
Non-polar
Polar-uncharged
Charged
Hydrophobic
F: Phenylalanine
G: Glycine
K: Lysine
L: Leucine
N: Asparagine
R: Arginine
DKP
LEU-3
GLY-2
ARG-4
LYS-5
ARG-6
LEU-7
ARG-8
LYS-9
PHE-10
ARG-11
ASN-12
LYS-13
1
10
12

Figure 4.14.: The helical wheel (left) of the sC18* peptide sequence together with a cartoon presentation of the α-helix (right) that was calculated derived from NMR spectroscopic restraints. The CSI (bottom) was generated based in H^N and H^α chemical shift of the DKP1-sC18* peptide.[116]

The 3D structure calculation of the cyclic DKP1-sC18* was successfully carried out. In Figure 4.15 the superimposition of backbone traces of 10 lowest energy minimised structures is displayed (a) together with an illustration of the side chain arrangement of the charged amino acids Lys and Arg.

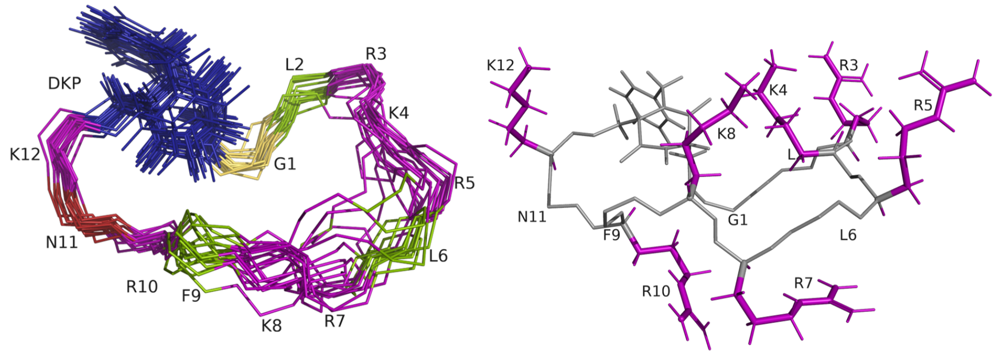

DKP
L2
R3
K4
K12
G1
R5
N11
L6
R10
F9
K8
R7
K12
R3
K4
R5
K8
L
N11
G1
F9
L6
R10
R7

Figure 4.15.: Backbone superimposition of 10 energy-minimised structures derived from experimental data (a) and an illustration of location of charged side chains of arginine and lysine residues (violet)(b).

The NMR data derived structure shows that five of seven charged side chains are located on one side of the cyclic peptide. This arrangement of side chains suggests that cyclic peptidomimetics interact with a micellar surface through this preferred conformation.

5. Summary and Outlook

In this part of the thesis the structure of four DKP-sC18* dipeptidomimetics were investigated by NMR spectroscopy. All sequences were identical in their primary amino acid sequence but differ in two points: linear vs cyclic structure and stereogenic property of the introduced DKP scaffold.

The linear sC18* derived peptides exhibit a random coil structure in aqueous medium, whereas the cyclic analogues at an experimental temperature of 283 K show amide-amide signals. Those amide signals suggest the existence of a structural family which is characterised by a secondary structure. Molecular modelling studies confirmed that the existence of multiple hydrogen bonds is very likely.

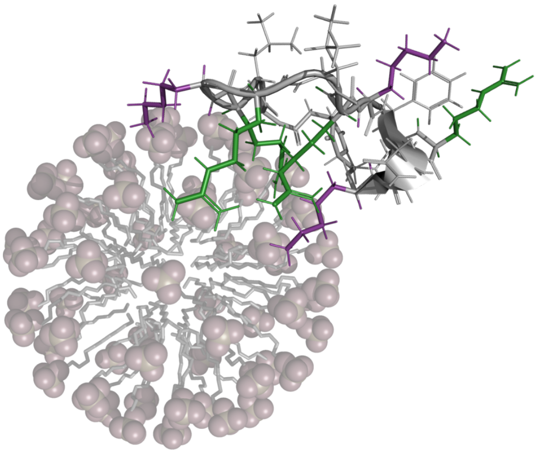


Figure 5.1.: Charged side chains of cyclic DKP-sC18* peptidomimetics are mostly arranged on one side of the peptide providing a highly efficient interaction with the anionic surface of a micelle.

In presence of a membrane mimetic medium the linear peptidomimetics undergo a conformational switch to an α-helical structure. This was clearly shown through the classical 2D NOESY signal pattern, the subsequent NMR derived 3D structure calculations as well as the calculated CSI plot. The helical wheel plot revealed an equal distribution of highly charged Lys and Arg around the helix. Structural analysis was also achieved of the cyclic DKP1-sC18* peptide. The inter-residual amide-amide signals along the four adjacent amino acids Arg^{10}–Asp^{11}–Lys^{12}–DKP1(H^{10}) in the 2D NOESY spectrum suggest that a β-turn element is exhibited under the interaction with micelles.

Part IV.

FIRST STEPS IN NMR SPECTROSCOPIC STUDIES OF INTER-MOLECULAR INTERACTION MECHANISMS OF OSMOLYTES

1. Background

Proteins are the most abundant molecules in organisms. They perform a variety of vital functions that make life possible. These magnificent molecular tools can serve as enzymes, enable cell movement, transport metabolites, catalyse chemical reactions or recognise signalling substances.[117] The protein translation and synthesis occurs via an mRNA template at ribosomes. The polypeptide chain of a protein, which has just been built on such a ribosome, is not able to perform any function at this point. To enable a proper function, a protein must adopt its special, unique structure. This active form is called the native fold.

A functional organism depends on correct folding, whereas disease is caused by incorrect folding. The folding pathway itself is unpredictable and highly complex. It crosses in (some cases) multiple intermediate conformations.[117]

During folding, a polypeptide can form an almost infinite number of different tertiary and quaternary structures on the basis of only the primary amino acid sequence, as shown by pioneering studies of *Anfinsen*.[118]

Protein folding is a process that occurs spontaneously in a cell. The start of folding is usually cotranslational, which means that the N-terminus of the polypeptide chain is already folding while the C-terminus is still being translated at the ribosome.[119] Factors that contribute significantly to folding are the formation of intra-molecular hydrogen bonds, van der Waals forces, the decrease of the proteins entropy and hydrophobic interactions. A hydrophobic interaction describes the effect in which solvent exposed hydrophobic side chains collapse into a proteins core. Hence, this hydrophobic collapse causes the burial of hydrophobic residues in the interior of the folded protein. The loss of solvent exposed hydrophobic moieties causes an striking entropy decrease of the protein. The decrease in entropy through the proteins adaption of a more ordered state causes an increase in entropy of the overall system. Consequently, the hydrophobic collapse is favoured to a high degree by this increased system entropy and therefore play a key role in protein folding.[120]

Folding does not only depend on the amino acid primary structure itself and the related interactions but also on factors such as severe pH values, extreme temperatures and pressures, high salt levels, presence of cofactors or the presence of molecular chaperons.[121]

1.1. Folding mechanism and pathway

Folding can never be accomplished in a single step, which is why different conformations are defined on an energy map of folding as shown in Figure 1.1. On the 'healthy side' of protein folding towards the native structure, intra-molecular interactions dominate the pathway. Due to the intra-molecular interactions, the folding process is thought to cross several funnel-shaped potential energy surfaces[117] instead of a single step pathway.[122] Intermediate conformations as well as partially folded ones are in kinetically favoured stages, which can be overcome by an appropriate energy expenditure or the help of chaperones.[117]

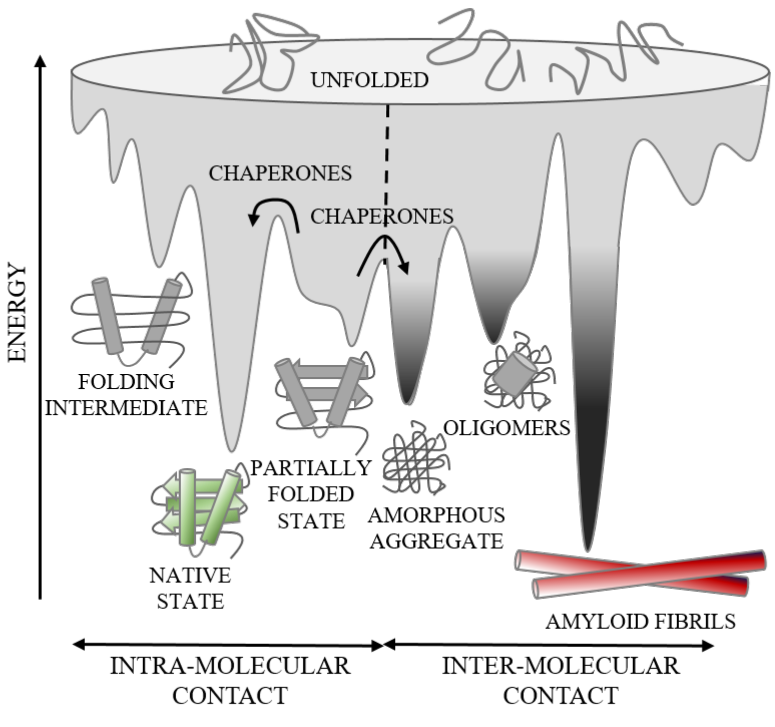

UNFOLDED
ENERGY
CHAPERONES
CHAPERONES
FOLDING INTERMEDIATE
PARTIALLY FOLDED STATE
NATIVE STATE
AMORPHOUS AGGREGATE
OLIGOMERS
AMYLOID FIBRILS
INTRA-MOLECULAR CONTACT
INTER-MOLECULAR CONTACT

Figure 1.1.: Energy landscape of protein folding into its native structure (left) and into misfolded structural elements (right). Also shown are kinetically favourable states in which the protein can be trapped on its folding pathway. Chaperones can assist to overcome free energy barriers so that a transition to more energetically favourable conformations can take place.[117]

Folding on the 'unhealthy side' towards a non-native structure, an increased population of strongly kinetically trapped intermediates is passed. Those kinetic highly populated states are denoted as oligomers or amorphous aggregates. The global energy minimum of non-native structures are amyloid fibrils. Fibrils occur through the formation of intermolecular interactions of perpendicular arranged β-strands. These highly ordered structures are characterised by hydrophobic moieties which are exposed to the solvent. Highly solvent exposed hydrophobic residues contradict the hydrophobic effect and thus further provoke a fibril forming aggregation process.[123,124] Once a protein is in an intermediate or the final fibril state a return to the native structure is very energy demanding. Refolding is only possible through a strong rearrangement of the primary sequence, which requires in particular the break of the described intermolecular interactions.[117]

1.2. Misfolding and disease

Misfolding of proteins is the fundamental driver of many diseases such as Alzheimer's disease, Creutzfeldt-Jakob's or cystic fibrosis. These diseases are associated with metastable proteins such as α-synuclein or the tau protein. Tau protein and α-synuclei tend to form fibrillar aggregates through intermolecular interactions (cf Figure 1.1). In the best scenario, a misfolded protein merely loses its function; in the worst it develops toxicity. Misfolded proteins can trigger further misfolding or even protein aggregation. The latter leads to degenerative diseases and even cell death. Consequently, at all stages in the life cycle of a protein, synthesis–folding–degradation, must be a tightly controlled intra-cellular balance. This balance is permanently challenged by the two stress factors: genetics (mutations) and environment.[125]

1.2.1. Chaperons

Nature has developed many ways to achieve a correct folding. Some proteins fold spontaneously and independently. Some other proteins, especially large and/or more complex ones, need assistance to find their native structure. Increased complexity and size trigger the tendency to misfold or form aggregates. Nature developed a complex system consisting of many different chaperone classes that serve as folding assistants, for maintenance and regulation of the degradation of misfolded proteins. One of the best studied chaperone class is probably the ones of the heat shock proteins (Hsp). Hsps are huge multicomponent molecular machines which are categorised according to their molecular weight (e. g. Hsp70, Hsp90). The activity of each chaperone class is initiated by factors such as very high temperatures (e. g. Hsp), oxidative stress or the accumulation of cell-damaging substances (e. g. urea).[117,126] Another class of chaperones are the chemical chaperones, also known as osmolytes. Structurally, osmolytes differ from supramolecular chaperones mainly in size and complexity. They are small organic or inorganic compounds that are also able to mediate the correct, native fold without determining it nor being part of the final functional structure. The distinction between chaperones and osmolytes in their application is essentially that macromolecular chaperones are adapted to proteins, but osmolytes always exert the same function on different proteins.[117]

1.2.2. The chemical chaperons: Osmolytes

A multitude of environmental factors permanently affect organisms and, as a consequence, their cells. The native folded proteins, are therefore constantly exposed to harmful influences. To ensure persistent preservation of a proteins native form, cells are equipped with many little helpers/supporting auxiliaries, called osmolytes. Osmolytes are generally divided into osmoprotecting/compatible and destabilising/counteracting, depending on their effect onto a protein structure.[121,127]

In Figure 1.2 a schematic presentation of the chaperone assisted protein folding, starting from translation, is shown. Osmolytes are divided into six categories: carbohydrates including small sugars, polyols and derivatives, amino acids and derivatives and methylamines. The most prominent member of the denaturing osmolyte class is urea. Generally the occurrence of the members of these categories is observed in different organisms. For example, taurine is common in sea animals, carbohydrates in some plants and fungi as well as sugars and polyols in organisms that need to protect themselves from freezing, such as amphibians. Several osmolytes like glycine betaine, glycerophosphocholine, sorbitol, inositol and taurine, occur in human kidney cells to act as antagonists to urea.[128,129]

The general influence of the osmolyte on a protein is universal and independent of the protein. – i. e. urea and guanidine hydrochloride (GuHCl) always have a denaturing effect, while the majority of osmolytes (e.g. methylamines, sugars, polyols) always act as a protectant.[132] The general occurrence of an osmolyte dependents on the diet, the metabolism and thus on the environment of an organism.[128]

Probably the most prominent naturally occurring example of destabilising and protective

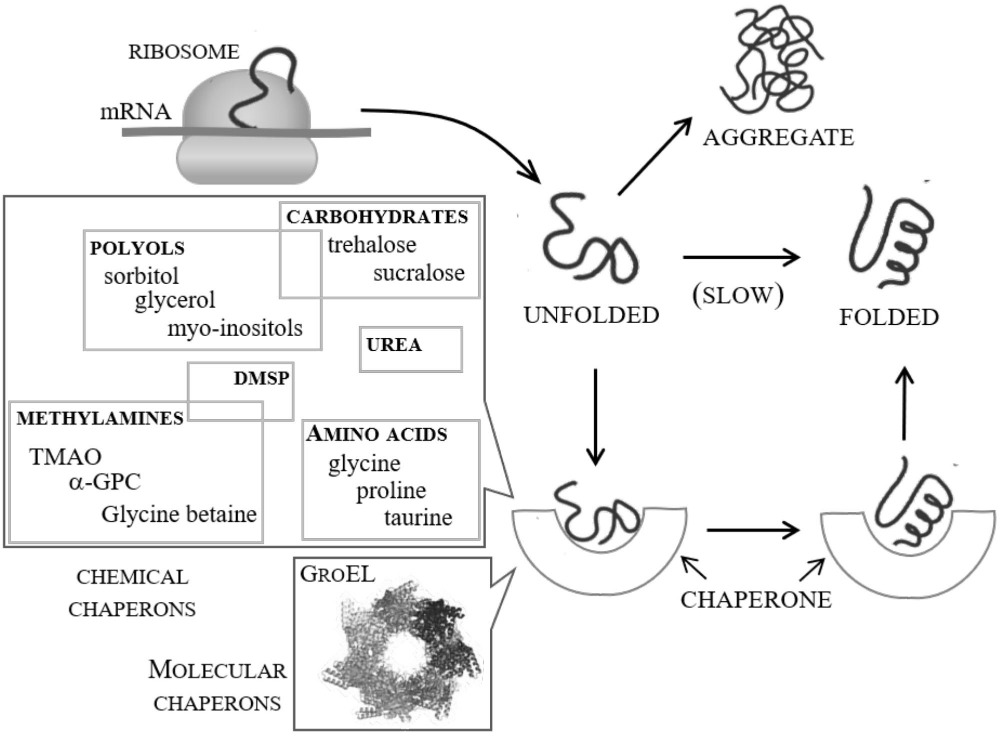

RIBOSOME
mRNA
AGGREGATE
CARBOHYDRATES
trehalose
sucralose
POLYOLS
sorbitol
glycerol
myo-inositols
UREA
DMSP
METHYLAMINES
TMAO
α-GPC
Glycine betaine
AMINO ACIDS
glycine
proline
taurine
UNFOLDED
(SLOW)
FOLDED
CHEMICAL CHAPERONS
GROEL
MOLECULAR CHAPERONS
CHAPERONE

Figure 1.2.: Schematic drawing of the protein biosynthesis pathway with possibly involved chemical or molecular chaperones (exemplary shown with GroEL complex).[130,131]

osmolyte pairing is the interplay between urea and *tri*m*etyhl*a*mine N-*o*xide (TMAO). TMAO has become known through its increased occurrence in deep-sea fish, where it accumulates in cells as an antifreeze agent. Deep-sea fished accumulate TMAO and urea as osmoprotectant from the increased (underwater) pressure. Therefore, TMAO is found to be always in an excess to urea, so that proteins are retained in their native fold. In a disbalance with urea being in an excess, proteins are perturbed and lose their structure and thus, their function.[128,133–135] It has been shown that TMAO not only protects proteins, but also helps proteins to return to their native structure when they were already in a misfolded state caused by urea.[136,137]

1.2.2.1. Mechanism of osmolytes

In recent decades, many attempts have been made to understand the exact molecular mechanism of the effect of osmolytes on the folding process of proteins. Nevertheless, the pathway is still controversial because it could not be resolved clearly. Various theories, such as that some osmolytes interact directly through e. g. hydrogen bonds with the protein backbone or side chains are discussed. In the mechanism studies it has been emerged that osmolytes significantly alter the water structure around a protein.[138] In Figure 1.3 the effect of TMAO and urea on a protein, and its hydration layer, is shown. Destabilising osmolytes such as urea and GuHCl appear to interact in a favoured way with protein backbones and side chains. This seems to be enabled through less strong depletion of the hydration layer of the protein in comparison to stabilising osmolytes.[140,141] Perturbants are consequently thought to accumulate close to protein's backbone and side chains. This close location to the protein surface provokes a partly unfolding which results in an increased exposure of the hydrophobic moieties to water. This increased exposure results, finally, into an unfolding cascade.[142–145] Stabilising osmolytes such as TMAO seem to be more strongly excluded from the proteins hydration layer, so that the bulk water becomes a better solvent than the water closely localised on the protein surface. As consequence, no direct interaction between osmolyte and protein can occur. It is proposed that compatible osmolytes cause a higher organisation of water molecules around a protein. This simultaneously induces an enhanced formation of hydrogen bonds among water molecules, leading to a whole hydrogen-bonding network in the proteins hydration layer. Consequently, less water molecules are able to interact with the protein leading to a preservation of the native folded form.[142–149]

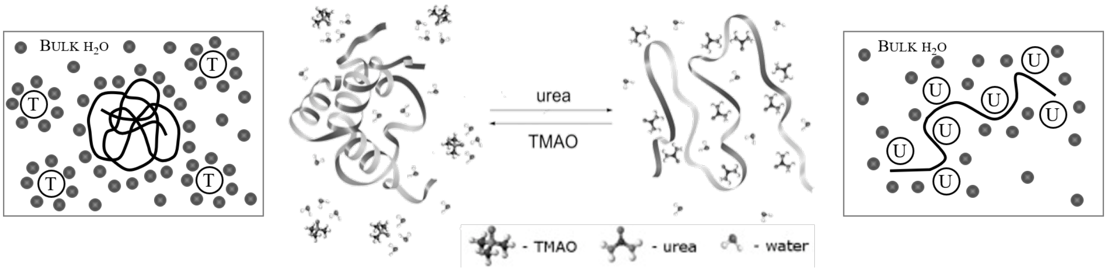


Figure 1.3.: Protein folding and unfolding induced by TMAO and urea, respectively. TMAO is excluded from the hydration layer (left) and keeps the protein in its native form, whereas urea interacts more tightly with the protein backbone to cause an unfold of the protein.[128,139]

1.3. Model systems used in BioNMR research

The study of protein folding mechanisms using solution NMR spectroscopy has become a central element.[150] Although a variety of NMR experiments have now been developed to study macromolecular components, it is still preferable to work with small and simplified model systems which meet the minimum requirements of the system to be approximated. Since *Anfisen* demonstrated that a protein can spontaneously fold into its native structure, an increasing number of model systems have been developed and studied. Two representative ones are the cyclic dipeptide system, which has already been sufficiently used by *Gerig* to study interactions and hydration properties of TMAO and cyclo-alanylglycine. The other model system is the rather more complex mini-protein, which is denoted as tryptophan cage (abbr. Trp-cage). Trp-cage proteins exhibit in its folded form both, secondary structure and tertiary structure elements, giving the possibility to track folding and unfolding.

In 2001, in *Neidigh*'s laboratories an 18 amino acid long peptide segment was discovered that is part of the exendin-4 peptide. Exendin-4 itself folds spontaneously and is, with its 39 amino acid long chain, the shortest known.[151] In 2002, the same group developed the synthetic 20 amino acid long analogue peptide , which was named Trp-cage after *Baruna* and *Andersen*.[152,153] This fascinating peptide forms an α-helix, a 3_{10}-helix and a PPII-helix on these only 20 amino acids. Because of this large number of structural elements, this peptide is also referred to be a mini-protein. The spontaneous folding process is cooperative and is driven by the encapsulation of the indole moiety of Trp^6 out of the hydrophobic pocket of the proline rings Pro^{17} and Pro^{18}. Folding from random coil to native structure proceeds, analogously, by burying this indole group and the associated hydrophobic collapse of the entire sequence.[153]

A very small peptide model was used in 2005 by *Gerig* and coworker to investigate interactions between TMAO – water – *cyclo*-alanylglycine.[154] This small dipeptide is very suitable for NMR spectroscopic investigations since it fulfils both, low molecular weight and direct recognisable signals. The investigations carried out were mainly self-diffusion, relaxation measurements and inter-molecular NOE. The latter, in particular, is very difficult to acquire in many terms. The complexity of detecting inter-molecular NOE makes the use of such basic small systems crucial for testing the concepts conceived.

2. Goals

Misfolded proteins are known to lose their natural function or even cause severe diseases. The return to native protein folding has been observed under the action of small molecules such as TMAO. The mechanism under which protective osmolytes operate is still not clearly understood. Therefore, a deeper understanding regarding the interaction mechanism of osmoyltes on proteins is of great interest in current research.

The role of both osmolytes and solvation (water–protein interactions) on protein folding/unfolding through the detection of inter-molecular homo- and hetero-nuclear NOE should be investigated

The two previously described model systems, Trp-cage and a cyclic dipeptide, *c*yclic-*g*lycinyl*v*aline cGV, were chosen to investigate the mechanism. The Trp-cage's spontaneous folding and unfolding, caused by the action of TMAO and urea, makes it an ideal test system for NMR spectroscopy. Nevertheless, with its structural features and size, the Trp-cage is a rather complex model system for a project in its early stages.

For the first steps in this project, the small dipeptide cGV is chosen as a simplified peptide model. With its low molecular weight and signals that are easily recognisable in a spectrum, it fulfils the main criteria for an NMR spectroscopic test system.

The acquisition of inter-molecular hetero-nuclear NOE is challenging, so inter-molecular homo-nuclear NOE should be acquired first. The system should then be extended to a readily detectable hetero nucleus such as fluorine, where *h*exa*i*so*f*luoro*p*ropanol (HIFP) is an appropriate choice.

If the general experimental setup for the acquisition of inter-molecular hetero NOE is established, the step to the more complex model systems should be considered.

3. Materials and methods

General. All compounds used for the preparation of the buffer (i. e. sodium phosphate mono- and dibasic) as well as the deuterated water were purchased from Sigma-Aldrich. ^{15}N labelled TMA·HCl was purchased from Cambridges Isotopes Laboratories, Inc. The synthesis of ^{15}N isotopically labelled TMAO was performed according to the synthesis of *A. J. Pearson*.[155] The Trp-peptide sequence was synthesised and gently provided by the group of Prof *Neundorf*.

NMR samples. The Trp-cage mini-protein was dissolved in 200 μl PBS buffer (pH* 7.0, H_2O/D_2O, 9:1, final concentration ca. 1.0 mM, 3 mm NMR tubes). cGV samples contained a concentration of 120 mM and 2 M TMAO and 2 M HIFP, respectively.

NMR experiments and assignment. Spectra were recorded on a Bruker Avance 400 (^{1}H transmitter frequency of 400 MHz), an Avance III 499 spectrometer (^{1}H frequency of 499 MHz) equipped with a cryoprobe triple resonance high-resolution probe (TCI prodigy) or on a Bruker Avance II 600 spectrometer (^{1}H frequency of 600 MHz) equipped with a triple resonance high-resolution probe (TBI) using Topspin software (Bruker). The transmitter frequency was set on the HDO/H_2O signal, and the resonance of d_4-TSP (3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt) was used as chemical shift reference (^{1}H $\delta_{\mathrm{TSP}} = 0$ ppm). One and two-dimensional spectra were acquired by standard pulse sequences using solvent suppression sequences. Two-dimensional homonuclear experiments at specified temperatures (i. e. TOCSY and NOESY) were recorded for assignment and structural calculations. The spectra were processed with Topspin software (Bruker).

The processed data were analysed by using CARA[70] software. Peak lists for the TOCSY and NOESY spectra were generated by interactive peak picking. NOESY cross-peak volumes were determined by the automated peak integration routine implemented in CARA. Complete assignment of the backbone and side-chain ^{1}H resonances was performed using standard sequential assignment procedures, according to the methodology developed by Wüthrich.[15]

ESI-MS. ESI-MS were measured In an Agilent 1100 Series LC/MSD instrument with a

G1312A binary pump, G1313 autosampler and G1956 mass selective detector.

Structure calculation. Three-dimensional structures were determined by the standard protocol of the CYANA program (version 2.1),[49] using seven cycles of combined automated NOESY assignment and structure calculations followed by a final structure calculation. For each CYANA cycle, 100 randomized conformers and the standard simulated annealing schedule were used. The 20 conformers with the lowest final score were retained for analysis and passed on to the next cycle. Weak restraints on ϕ/ψ torsion-angle pairs and on side-chain torsion angles between tetrahedral carbon atoms were applied temporarily during the high temperature and cooling phases of the simulated annealing schedule in order to favour the permitted regions of the Ramachandran plot and staggered rotamer positions, respectively. The list of upper-distance bonds for the final structural calculation consists of unambiguously assigned upper-distance bonds and does not require the possible swapping of diastereotopic pairs. Root-mean-square deviation (RMSD) values were calculated using CYANA for superpositions of the backbone N, Cα and CO atoms; the heavy atoms over the whole peptide or the cyclic fragment. To obtain the rmsd of a structure represented by a bundle of conformers, all conformers were superimposed upon the first one and the average of the rmsd values between the individual conformers and their average coordinates was calculated.

4. Results and discussion

In this chapter the initial steps that have been taken on this topic will be described. At this early stage of the investigation of inter-molecular interactions between a model peptide and osmolytes, the model system had to be investigated under our own laboratory conditions. The conditions chosen were based on the deposited structure (BMRB entry 5292).[153] Figure 4.1 shows the 1D proton spectrum of the Trp-cage mini-protein.

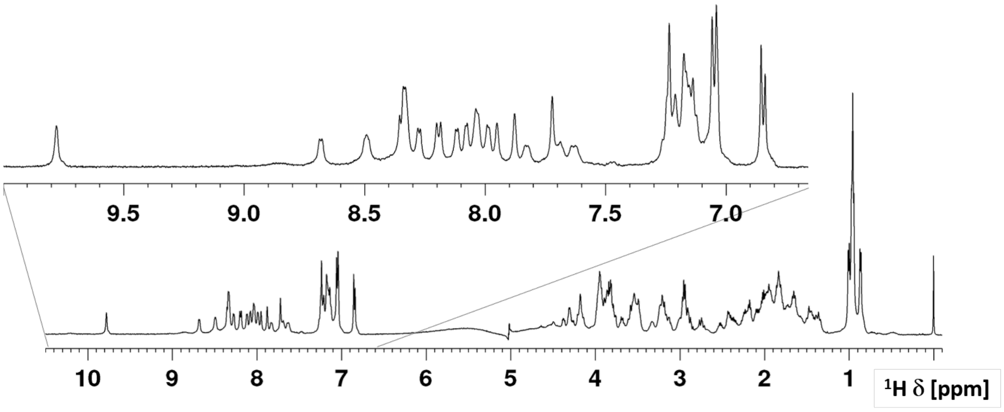


Figure 4.1.: ^{1}H NMR spectrum of the Trp-cage mini-protein (peptide concentration ca. 1.3 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz)

In the amide region, the signals are well dispersed and barely exhibit a splitting. The shape of the signals and the distribution over a rather large range of 7.5–8.8 ppm suggests that the protein has folded spontaneously under the chosen conditions.[153]

Figure 4.2 shows the TOCSY detail of the aliphatic region where this signal shift from Pro^{18} is indicated with red dashed lines.

The correlation of Pro^{12} (blue lines) is shown as an example to demonstrate the extent to which the β-proton of Pro^{18} is shifted. Pro^{18} is located in a hydrophobic pocket and interacts with the Trp^6 indole ring through space. As consequence, the β-proton of Pro^{18} is affected by a ring current. This ring current shift is a unique characteristic of the folded form of the Trp-cage. Usually, the shift of proline β-protons are expected at around 2.0 ppm.

In the 2D NOESY spectrum 4.3 some selected NOE signature signals for the folded state of the mini protein are shown.

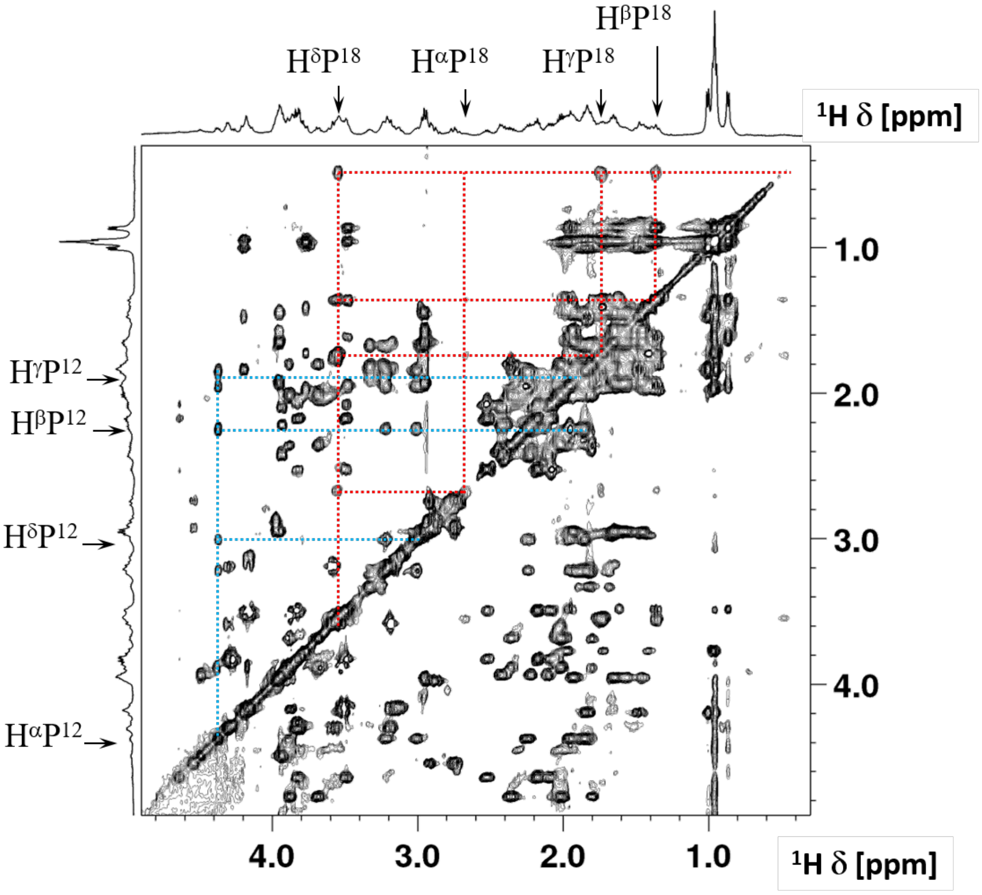


Figure 4.2.: 2D TOCSY NMR spectrum of the Trp-cage mini-protein of the aliphatic region (peptide concentration ca. 1.3 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz)

The hydrophobic pocket consists mainly of proline residues which are in an inter-residual interaction with the Trp-indole moiety. In a), the correlations between the $H^{\epsilon}1$ of Trp^6 to H^{α} of Pro^{17} and Pro^{18} are clearly visible. In addition, NOE signals to the residue Arg^{16}, which is not part of the hydrophobic pocket, are labelled. In section b), the inter-nuclear proximity between two protons of the indole group to several proline residues is shown. In Figure 4.4 is given the computed Ramachandran plot on the basis of NMR restraints. The conformations which are expected for the secondary structure elements of the Trp-cage mini-protein can be found in the distinct regions.

The 3D solution structure shown in Figure 4.5 was calculated by the inter-nuclear distances derived from the NOE cross-peak intensities.

The structure calculated is clearly folded. The indole residue (red) of the tryptophan is embedded inside the hydrophobic pocket (green residues). In addition, an α-helix is formed between the amino acids Leu^2 and Asp^9, as well as a 3_{10}-helix across the acids

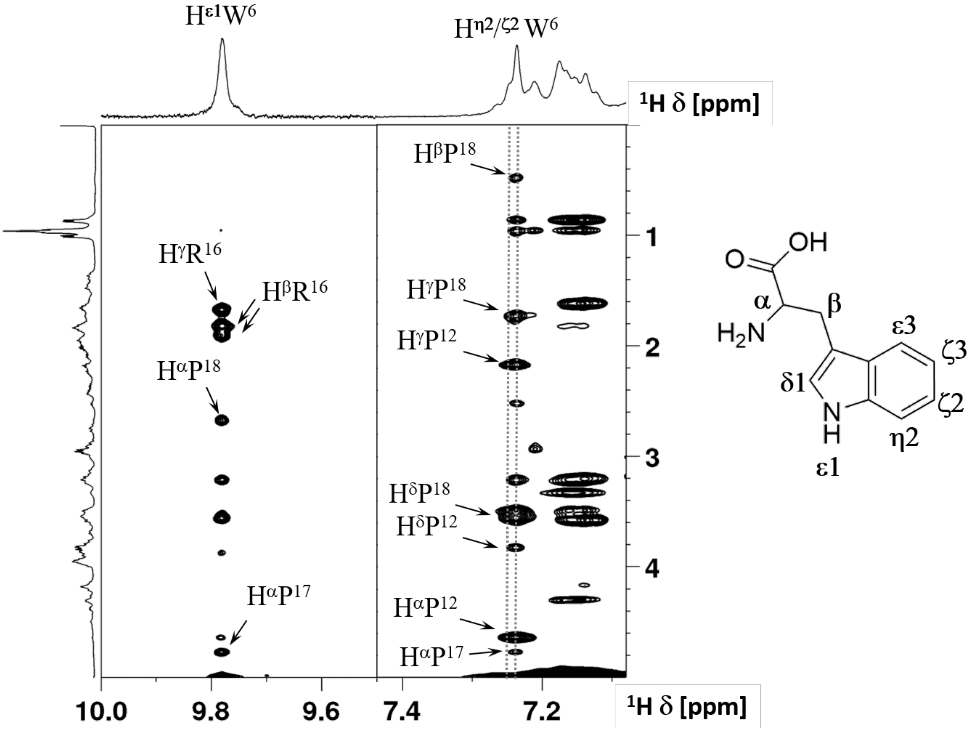


Figure 4.3.: 2D NOESY NMR excerpts of the Trp-cage with a) correlation from aromatic $\epsilon 1$ protons and b) correlation of aromatic $\eta 2$ and $\zeta 2$ protons to proline residues of the hydrophobic pocket. Nomenclature of the Trp residue is adopted from the CYANA software standard library (peptide concentration ca. 1.0 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz)

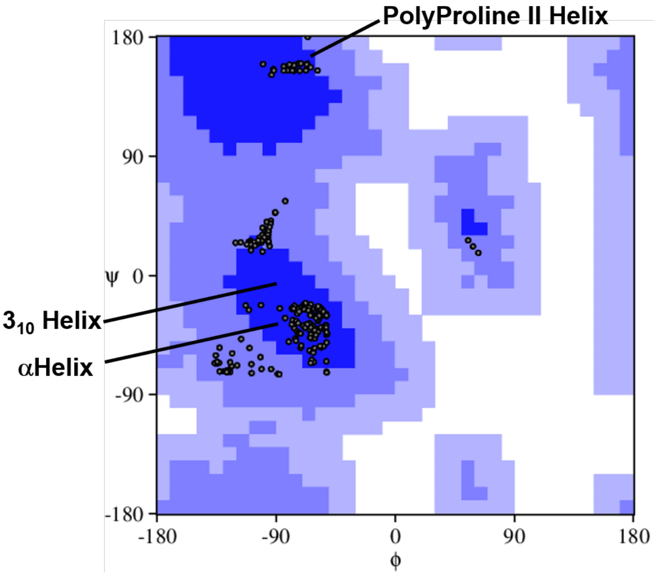


Figure 4.4.: Ramachandran diagram of the Trp-cage mini-protein, derived from NMR experimental data.

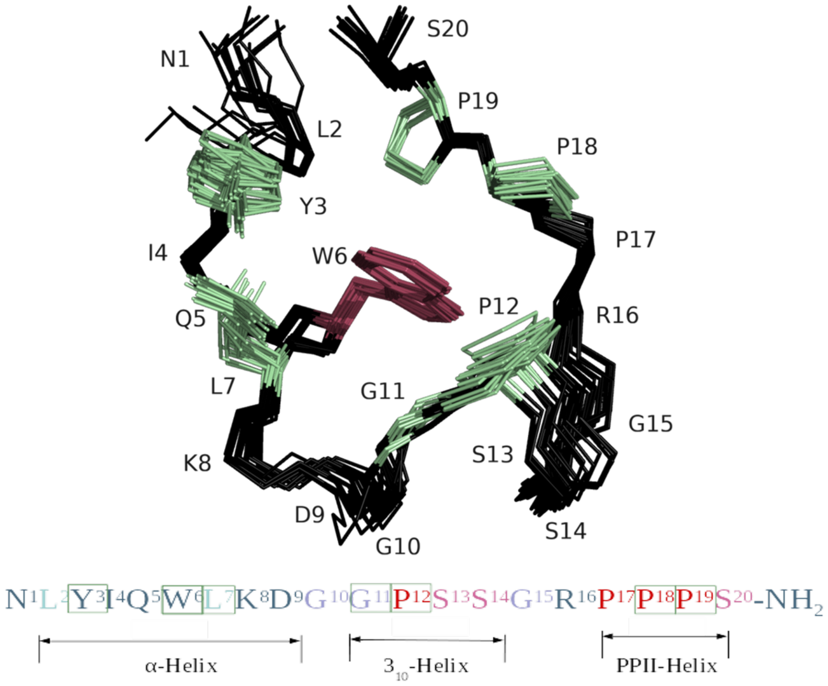

S20
N1
P19
L2
P18
Y3
P17
I4
W6
Q5
P12
R16
L7
G11
G15
S13
K8
D9
S14
G10
N1L2Y3I4Q5W6L7K8D9G10G11P12S13S14G15R16P17P18P19S20-NH2
α-Helix
3₁₀-Helix
PPII-Helix

Figure 4.5.: 20 energy minimised structures calculated based on NMR experimental restraints. Red: Trp side chain, green: residues that form the hydrophobic pocket.

Gly^{11} to Ser^{14} and a PPII-helix from the three prolines Pro^{17}–Pro^{19}.

The first test systems and the experimental set-up were to remain as simple as possible in the initial attempts. For the simplicity of the peptide component, the cyclic dipeptide cGV was used. The preliminary study of whether a spatial proximity between the dipeptide and a small molecule can be measured was carried out by homo-nuclear NOE experiments. For these NOE measurements, a sample was prepared containing the dipeptide and 1 M TMAO. Figure 4.6 shows three 1D spectra of the NOE measurements in the range of 0.5–3.5 ppm, together with the 1D proton spectrum of the cGV alone in aqueous solution.

The three excerpts from the 1D NOE experiments show that under selective irradiation of the TMAO signal, the signals of the terminal methyl groups of valine build up with increasing mixing time. This signal build-up gets lost with longer mixing time. No further proton of cGV showed inter-molecular NOE response. Thus, TMAO is in spatial proximity to the H^γ protons of valine.

The inter-molecular NOE between TMAO and cGV shows that the peptide system is suitable for testing the concept. The next step was to measure hetero-nuclear NOE to

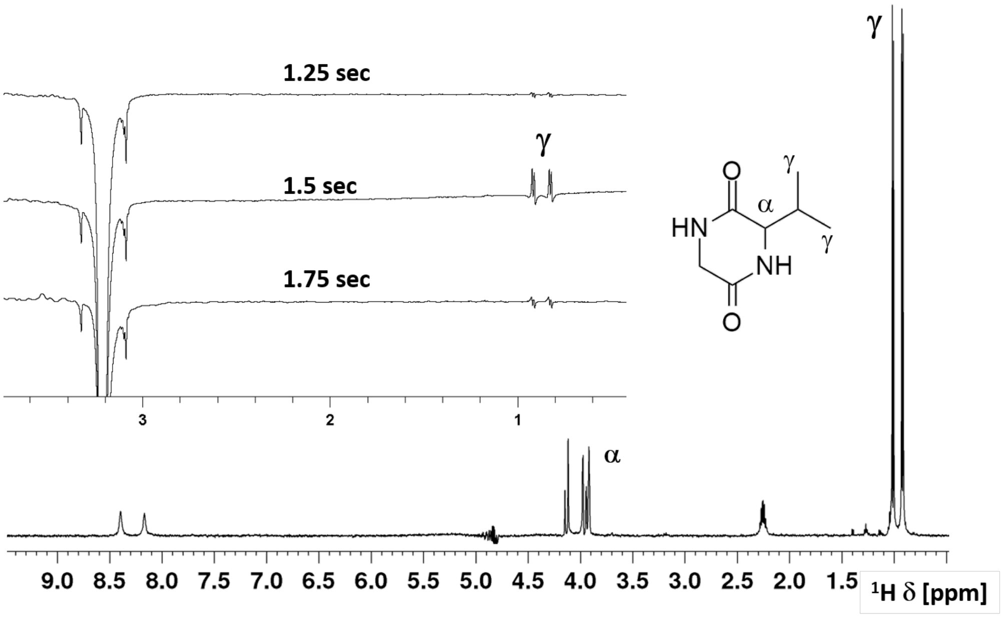


Figure 4.6.: 1D NOE spectra at different mixing times of TMAO and cGV. Irradiation was applied by a selective pulse on the proton resonance of TMAO. The 1D proton spectrum shows the dipeptide in water (H_2O/D_2O 9:1, 298 K, 400 MHz).

the dipeptide. However, the acquisition of HOE from nitrogen is not feasible without a ^{15}N isotopic labelling. The experimental set-up of the instrument to perform HOE experiments, though, was to be tested before carrying out expensive isotopic labelling. Since fluorine is very suitable for NOE experiments, as it is similar in its sensitivity compared to hydrogen, H–F NOE measurements were performed with HIFP. In this HOESY experiment, shown in Figure 4.7, selective irradiation was applied to the HIFP fluorine signal.

In the 1D HOE spectra, it can be seen that the H^{γ}-protons of valine build up with increasing mixing time. The maximum signal intensity is reached at a mixing time of 1.0 sec and disappears with increasingly longer mixing times. The observed HOE shows that HIFP is, as well as TMAO, spatially close to the same protons of the dipeptide.

In order to detect unambiguous NOE, the acquisition of 1D NOE must always be carried out through several experiments with varying mixing times. This approach is less sensitive compared with a 2D NOESY experiment.[23] Therefore, the previously confirmed inter-molecular NOE between HIFP and cGV was also measured in a 2D experiment, which is shown in Figure 4.8.

In this 2D experiment, it can be observed that a second NOE correlation is formed

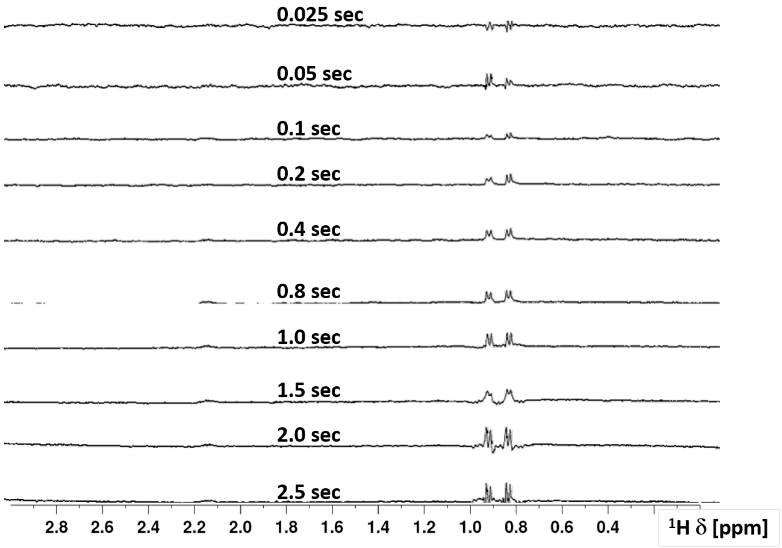


Figure 4.7.: $^1H^{19}F$ 1D HOE NMR spectra of cGV with HIFP together with the full 1D proton spectrum of cGV. Irradiation was set on the flourine signal of HIPF (^{19}F: 76.33 ppm). The mixing time for each 1D spectrum is indicated in the respective spectrum (H_2O/D_2O 9:1, 298 K, 400 MHz).

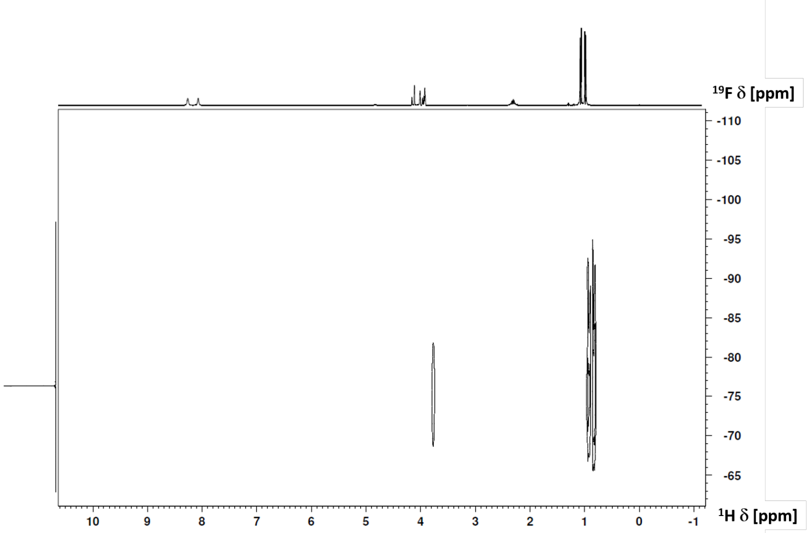


Figure 4.8.: $^1H^{19}F$ 2D HOESY NMR spectrum of cGV with HIFP (H_2O/D_2O 9:1, 298 K, 400 MHz).

between HIFP and the α-proton of cGV.

The difficulty in detecting H–N NOE is the negative gyromagnetic ratio of nitrogen,which causes the NOE to evolve under a zero crossing (see image in the introduction). This zero crossing can manifest either in an extremely attenuated NOE intensity or a complete absence.

As the first test system for the acquisition of ^{15}N hetero-nuclear NOE, a sample was chosen in which nitrogen isotopically labelled urea and methanol in DMSO were present. Figure 4.9 shows the 2D HOE.

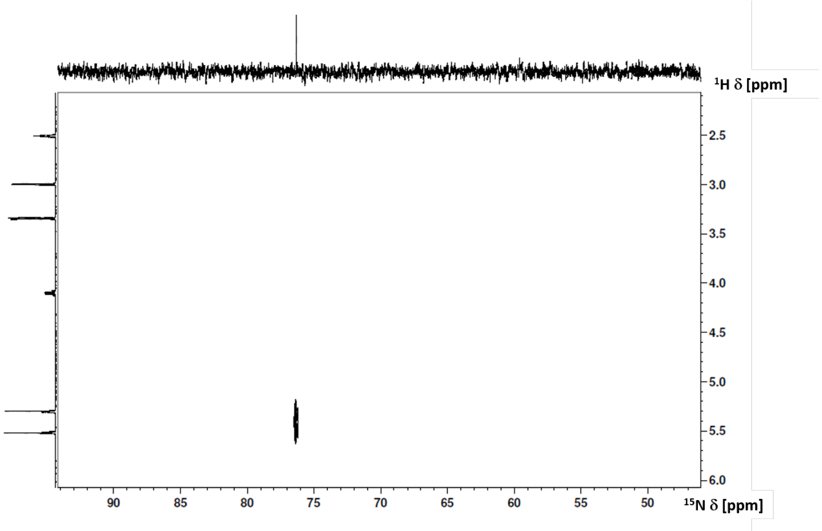


Figure 4.9.: $^{15}N^1H$ 2D HOESY NMR spectrum of labelled urea in presence of methanol (DMSO-d_6, 298 K, 400 MHz).

It is evident that an intra-molecular HOE could be captured of the osmolyte urea. This measurement clearly proves that the acquisition of ^{15}N wa successfully implemented. The successful implementation of the 2D HOESY experiment paves the way for further studies on H–N HOE with ^{15}N labelled TMAO.

5. Summary and outlook

In this chapter, interactions between osmolytes and model peptides were investigated by the detection of inter-molecular NOE. For this purpose, two model peptides which fulfil different requirements, were chosen, namely cGV and Trp-cage.[153]

In a first step, full resonance assignment and 3D solution structure calculation of the Trp-cage min-protein was reproduced. Test of the proposed concept and instrumental conditions were settled with the simpler model peptide cGV. Initial homo-nuclear experiments showed that an inter-molecular NOE of TMAO to the methyl groups of cGV can be observed. The confirmation of the spatial proximity between TMAO and cGV led to the subsequent instrumental testing of the HOESY structure. For this test, HIFP was selected because fluorine is more sensitive in NMR than nitrogen. In this HOESY experiment, it was found that HIFP is also spatially close to the same methyl groups as TMAO. The acquisition of intra-molecular NOE of the labelled osmolyte urea has opened the way for further extended experiments on the acquisition of inter-molecular ^{15}N–^{1}H HOE.

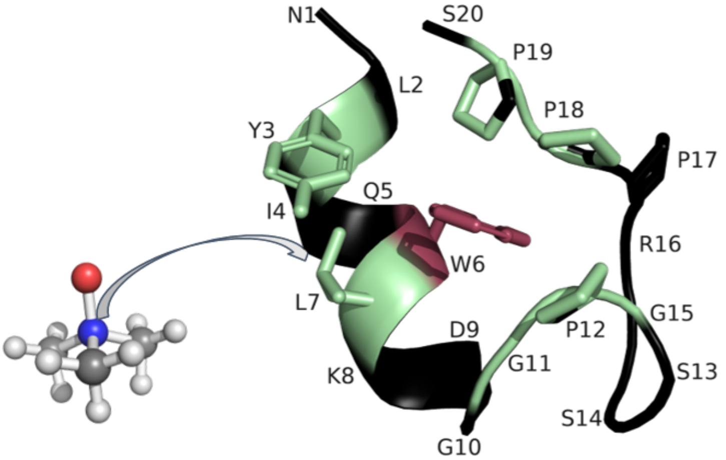


Figure 5.1.: Visual representation of the hetero-nuclear H–N NOE between the Trp-cage mini-protein and isotopically labelled TMAO (arrow: envisioned HOE, O: red, N: blue, C: grey, H, white).

Since the concept and experimental set-up could be confirmed in our laboratories, the synthesis of isotope labelled TMAO was carried out.

With these initial studies, a basis for further hetero-nuclear NOE experiments between peptides and osmolytes has been established. The main focus for ongoing research is the acquisition of inter-molecular H–N HOE, which might be more difficult than for fluorine. As schematically represented in Figure 5.1, the next step would be to perform HOESY experiments with the Trp-cage mini-protein and the isotope labelled TMAO in order to obtain information about the sites of action of TMAO.

Part V.

^{1}H NMR STUDIES ON THE INTERACTIONS BETWEEN SMALL MOLECULES AND MICELLES IN WATER

1. Background

Surfactants are amphiphiles with the ability to self-assemble into supramolecular systems like micelles. These supramolecular structures have already been established for NMR spectroscopic investigations. For example, they are used in BioNMR for 3D structure investigations of peptides on membranes as represent a suitable model due to their properties. A surfactant molecule possess both, a hydrophilic head group and a hydrophobic tail. These properties combined in one molecule enable it to interact with polar as well as apolar substrates.[156] On this basis, the use of micellar solutions offer a variety of practical application in life-related areas and in the research field of green chemistry. Micelles are nowadays well established as e. g. foam stabiliser in the firefighting, general detergents, emulsifiers in cosmetics and pharmacy and in biochemistry for protein denaturation and solubilization of membrane proteins.[157] Green chemistry stands for the principle of reducing or eliminating harmful and hazardous substances from a whole life circle of a chemical product to the greatest possible extent. In this context, water as a reaction medium and the use of catalytic reagents are some of the solutions for a more sustainable chemistry.[158] A prominent example of the use of water as a reaction medium is the Diels Alder reaction. The pioneers Diels and Alder already carried out the reaction in an aqueous medium. However, the substrates employed are often non-polar and thus poorly soluble in water. The solubility problem can be overcome by using surfactants as catalysts.[159]

1.1. Membrane-mimicking media in BioNMR

Biological membranes are highly complex systems found in all cell types (pro and eukaryotic). Membranes form the protective barrier to the outside environment of the cell. The general composition of all biomembranes is a phospholipid bilayer. A variety of membrane proteins and carbohydrates are embedded in this bilayer or are present on the surface.[12] This complexity of a real cell membrane is a challenge for the investigation of different processes on and in cells with NMR spectroscopy. Consequently, simpler and smaller models are required, that can be adapted to individual study priorities.[160]

Models which are used in BioNMR always consist of self-assembling surfactant molecules in aqueous solution. The surfactants are divided into two groups: detergents, which form micelles, and lipids, which form bilayer membranes. The formed model membrane depends on the surface active agent chosen. The surfactants differ in their charge density, chain length and type of head group. Figure 1.1 shows three membrane models used in BioNMR.

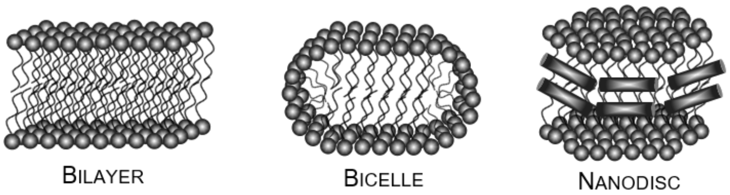


Figure 1.1.: Membrane mimetic types that are used in BioNMR spectroscopy.[160]

The choice of membrane model depends on the physico-chemical characteristics necessary for the type of study. For example studies on membrane proteins are preferably carried out in more complex models. Studies of the specific interaction between peptides and membranes, on the other hand, tend to be carried out on micelles.[161]

A classic membrane system that is used for interaction studies is the *s*odium *d*odecyl *s*ulphate (SDS) micelle. SDS consists of a hydrophilic sulphate head group that is attached to the hydrophobic tail. Micelle formation occurs, under standard conditions, in water at a *c*ritical *m*icelle *c*oncentration (cmc) of 8.2 mM.[162] The preparation of SDS micelles is rather simple compared to more complex membrane systems, which explains their frequent use in NMR studies.[163]

1.2. Micelle formation

Micelle formation is dependent on two opposing forces: hydrophobicity and hydrophilicity. The hydrophobic effect takes place at the hydrocarbon-water interface, pushing the monomer to associate. The hydrophilic interaction occurs at the head groups which are exposed to the solvent. SDS monomers assemble, under standard conditions (rt, water as solvent), to spherical micelles.[164,165] In Figure 1.2 a schematic structure of an SDS micelle is shown.

It consists of a hydrophobic carbon chain in its inner and a sodium sulfate head-group on its outer layer. Since SDS is negatively charged, counter ions accumulate close to

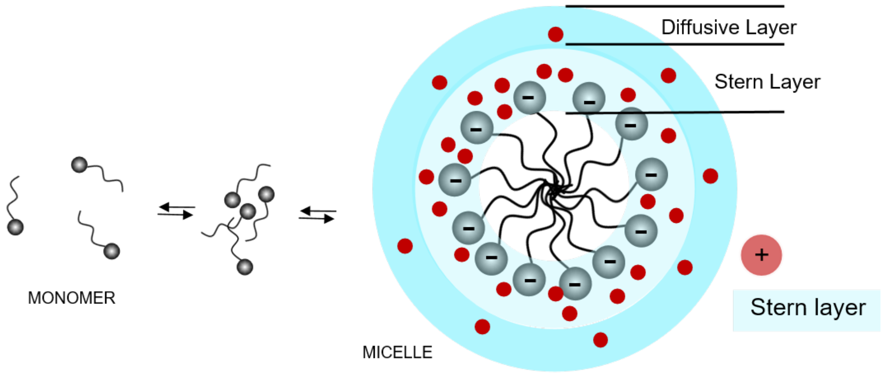


Figure 1.2.: Simplified illustration of a SDS micelle with emphasise on double layer, consisting of a positive charged surface and the contrary charged *Stern* layer. The diffusive layer constitutes the area where the charge decreases from high (double layer) to low (bulk solvent).[165,166]

the micelle surface. At the interface between water and micelle surface. This is called the diffusive layer, which itself consists of an inner layer, the so-called *Stern* layer and the outer diffusive layer. The *Stern*-layer comprises the anionic head groups of SDS and the directly attached counterions. Whereby the diffusive layer contains counterions and hydrating water.[167]

1.3. The Diels-Alder reaction

The Diels-Alder reaction was first described by *Otto Diels* and *Kurt Alder* in 1928.This reaction describes the formation of two carbon-carbon bonds through a [4+2] cyclo addition of conjugated dienes and dienophiles to form substituted six-membered rings with a good control over regiochemistry. The reaction is a representative for pericyclic reactions with a concerted mechanism, where four π-electrons of the diene react with two π-electrons of the dienophile (origin of name [4+2]) in one single, cyclic transition state. Thus, the driving force of a Diels-Alder reaction is the formation of these two σ-bonds under "loss" of two π-bonds, as single bonds are energetically more favoured than double-bonds. In Figure 1.3 the *f*rontier *m*olecule *o*rbital (FMO) description of a Diels-Alder cyclo addition is shown.[168]

Two variants are distinguished: the classical one with a normal and the inverse reaction with an inverse electron demand. In a *normal* Diels-Alder reaction, the electron-rich π_2 HOMO of the diene undergoes a reaction with the electron-poor π^* LUMO of the dienophile. Therefore, the diene is substituted with electron donating groups so the

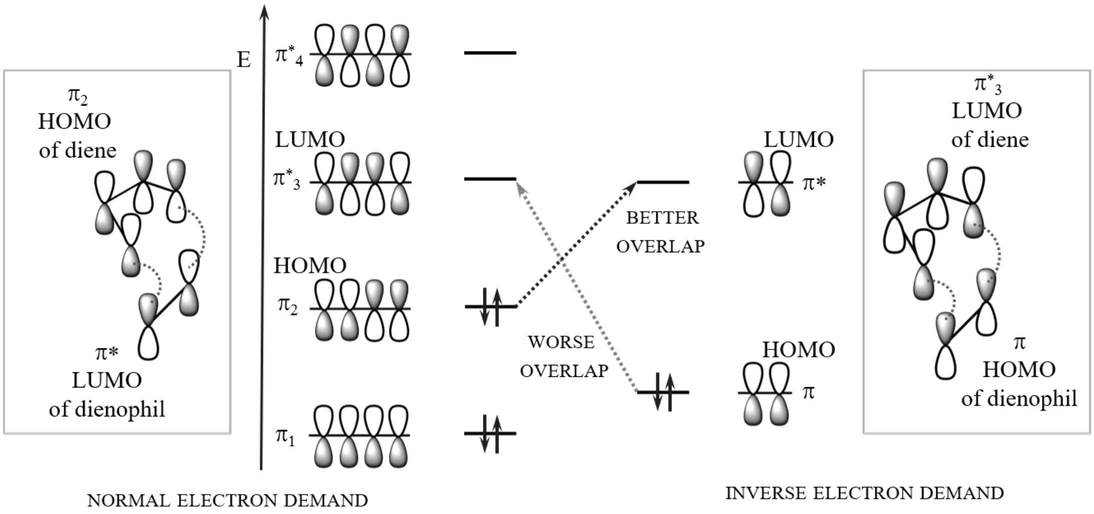


Figure 1.3.: Molecule orbital presentation of the normal electron demand Diels-Alder reaction (left side) and the reverse version (right side). How good orbitals of HOMO and LUMO overlap is determined through the size of the energy gap, which is larger (gray dashed arrow) for the reverse reaction in opposite to the normal one (black dashed arrow). The symmetry of the orbitals during a reaction is emphasised on the outside for both cases.[168]

reacting HOMO is higher in energy. In contrast, the dienophile is usually attached with an electron-withdrawing group, that shall lower the π^*-LUMO. As can be seen, the symmetry predefines, that the dien has one node to match the node-symmetry of the dienophile. For the *inverse* type of the reaction the conditions are chosen exactly opposite. To bring HOMO and LUMO energetically closer, so that they are in the most favored molecule orbital interaction of the diene or the dienophile, electron donating groups are attached to the diene and electron withdrawing groups to the dienophile. Thus the empty ψ_3 LUMO is lowered whereby the occupied π HOMO is raised.[168]

The Povarov reaction is an aza-Diels-Alder variant and can thus be described as the Diels-Alder reaction of imines. The Povarov reaction is an elegant way to synthesise *t*etra*h*ydro*q*uinoline (THQ) derivatives, which attract attention of synthetic and medical chemists alike.[169] THQ subunits can be found in nature and are shown to have multiple advantageous properties as being anti-viral, -biotic and -tumourous.[170] Some experimental conditions are either expensive, harmful or require a high energy cost.[171] Thus, Lewis acids like $BF_3 \cdot OEt_2$ are thereby used for activation of the imine, as invented by Povarov himself.[172–174] Further, Diels-Alder reactions are known to work in neat to almost every solvent condition, since no ionic transition state is crossed during the reaction. For the Povarov reaction, organic solvents must be used until now, resulting in a lot of undesired waste. Therefore, finding more environmentally green reaction conditions is of great interest nowadays.[171]

N-arylimines (Schiff's base) undergo thereby a [4+2] cyclo addition with the activated alkenes to yield heterocyclic compounds, as quinoline alkaloid derivatives that constitute a major class of heterocycles.[175] The dienophilic compound, i. e. the activated alkene, has to be equipped with an electron donating group, according to the previous descriptions of inverse Diels-Alder reactions.

2. Goals

SDS micelles are a suitable medium for a variety of NMR spectroscopic investigations ranging from interaction studies between peptides and membrane mimetica up to solvation studies of small compounds in micellar medium. A deeper understanding of the interactions between micelles and small molecules contributes to several fields, such as the detergent industry or green chemistry.

In this chapter, the investigations on the mode of interaction of small organic molecules with SDS micelles are discussed. For this purpose, first, the micelle formation is monitored by following proton chemical shift deviations and translational diffusion experiments.
A deeper insight into the positioning of the small compounds is gained through internuclear 1D NOE experiments. Hydration studies of the compounds in micellar medium are carried by the use of waterLOGSY experiments.

Micelles have also found their way into the catalysis of organic reactions in the context of green chemistry. Micellar catalysis has already been successfully tested to several reaction types.[156] An example is the Diels-Alder reaction. An exploitative study of a SDS promoted hetero Diels-Alder reaction was investigated using NMR spectroscopy.

3. Materials and Methods

General. All compounds used for the preparation of the buffer (i. e. sodium phosphate mono- and dibasic) as well as the deuterated water and small organic molecules for studies and reference were purchased from Sigma-Aldrich. *N*-vinylformamide was distilled under reduced pressure. All other compounds were used without further purification.

NMR samples. A stock solution of 225 mM concentration was diluted to the required concentration. In order to ensure that the micelle/monomer preparation always turns out the same, an extensive analysis was carried out: SDS was dissolved, treated in a ultrasonic bath and examined by NMR spectroscopy (^{1}H , DOSY) in the following time steps [min]: 0, 1, 5, 15, 30. In addition, each concentration was dissolved once, treated in an ultrasonic bath directly for 30 min and measured. Based on the same experimental approach, all SDS solutions were prepared by simply dissolving SDS without the use of an ultrasonic bath. Small organic molecules were added into the existing SDS solution via an Hamilton syringe (2.0–2.5 μL) to give a total concentration of additive compound of 47 mM.

Inspection of reactants. The reactants, 4-aminobenzoic acid and *N*-vinylformamide, were analysed prior to the reaction. The reactants were analysed separately at pH 1 and pH 7 in water and in a 12 mM SDS aqueous solution.

4-aminobenzoic acid (**1**) shows in H_2O/D_2O (9:1), pH 7 two aromatic signals at H_{ar} 6.920/7.860. At pH 1, the signals shift to H_{ar} 7.453/8.100. Carbon chemical shifts were acquired with a phase sensitive HSQC. The chemical shift difference between the aromatic carbons at pH 7 is $\Delta\delta_C$15.5 ppm and at pH 1 $\Delta\delta_C$ 7.8 ppm. Compound **1** is poorly soluble in water.

NMR experiments and assignment. The DOSY experiments were performed using standard pulse sequence from Bruker library including water suppression (stebgp1s19). The diffusion delay Δ was set at 100 ms, the gradient strength, g (1 ms) was linearly incremented in 32 steps from 2% to 98% of its maximum value, and a longitudinal eddy currents delay of 5 ms and 32 scans were recorded for each experiment. Processing was achieved using 64K points in the F2 dimension and 32 points in F1. An exponential window function with 1 Hz line broadening was applied in the F2 dimension prior to Fourier transformation. After baseline correction, the diffusion dimension was processed with the DOSY processing program (Bruker TopSpin software 3.5). A logarithmic scaling was applied in the diffusion axis, and a noise sensitivity factor of 4 and line width factor of 1 were used. The fitting of the diffusion dimension in the 2D DOSY spectra was obtained using a single exponential fit.

4. Results and discussion

4.1. Monitoring the formation of micelles in aqueous solution

Micelle formation of SDS in water occurs above a concentration of 8.2 mM under standard conditions.[162] A distinction of aggregation states by the use of NMR spectroscopic methods forms the basis of the investigations which are presented in this chapter. To demonstrate the properties of SDS in water, 1D proton and 2D DOSY experiments of SDS at three different concentrations were performed. The three different concentrations were chosen to be below (3 mM – monomeric units), at (8.2 mM – aggregates) and above cmc (30 mM – micelles). In Figure 4.1 the 1D proton spectra of those samples are shown.

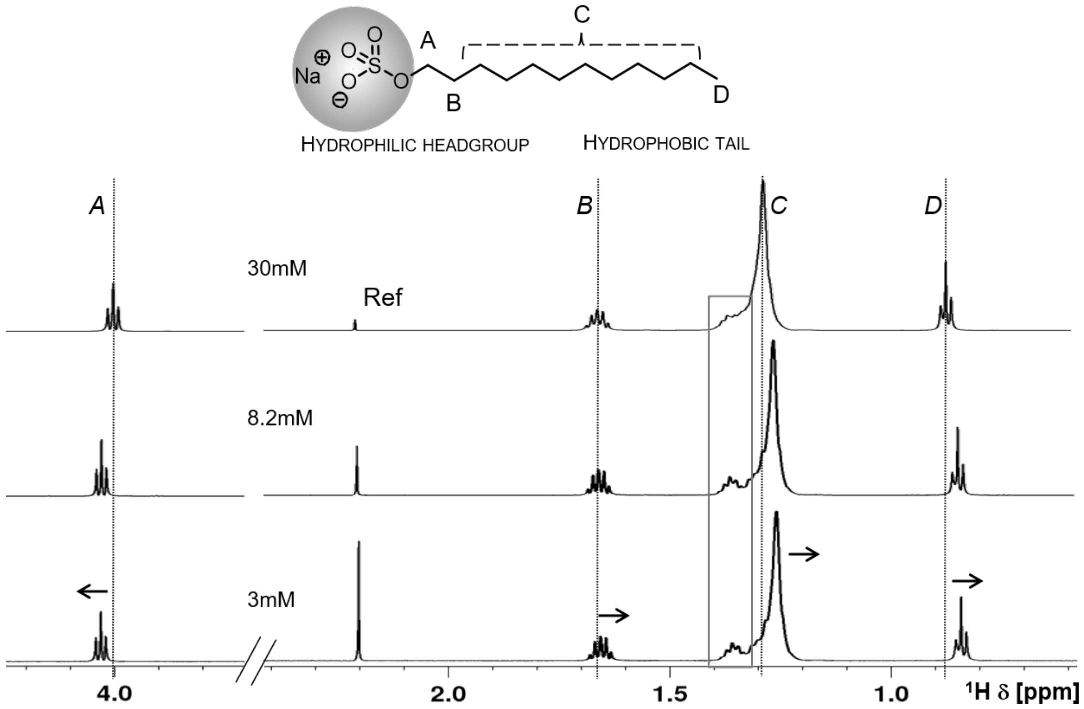


Figure 4.1.: 1D NMR trace of SDS at different concentrations: below (3 mM), at (8.2 mM) and above cmc (30 mM). Reference is an external capillary with acetone in water (H_2O/D_2O 9:1, 298 K, 600 MHz).

All three proton traces are aligned to the external reference containing acetone in water (capillary, 2.22 ppm). An external reference was required because SDS is assumed to interact, as consequence of its zwitterionic nature, with almost every compound. The comparison of the 1D proton spectra shows a clear correlation between the chemical shift of SDS and its concentration. The dashed lines are drawn starting from the signal middle of the spectrum of the highest concentration. This facilitates the recognition that the signals drift away from the centre of the spectrum at lower concentrations (marked with arrows). At a SDS concentration of 3 mM, at which individual SDS molecules are present in solution, a shoulder clearly separates from the intense signal *C* of the hydrocarbon chain (marked by the box). The shift of the tail group to the lower frequency range is caused by the stronger shielding of the hydrophobic chain by the hydrophobic core. In contrast, the head group becomes less shielded in a micelle due to a higher solvent exposure.

The aggregation status of SDS is not only distinguishable by ^{1}H proton spectra, but also through their diffusion properties. 2D DOSY experiments provide the possibility to resolve the diffusion coefficient of components in the second dimension F1. The differentiation through diffusion is possible mainly through the hydrodynamic radius of the molecules. This hydrodynamic radius is in a close relation to other contributing factors, as described by the Stokes-Einstein equation.

The aggregation of single SDS units into micelles enables the differentiation of those states by the use of DOSY. One SDS micelle is assumed to consists of around 60 SDS molecules and thus, should diffuse much slower than an individual SDS molecule. The diffusion measurements of the three SDS concentrations are shown in Figure 4.2.

As can be seen, all three concentrations are clearly distinguishable in the diffusion dimension. The diffusion coefficients of the micelles are much smaller as the ones of the individual SDS molecules, as expected. The grey diffusion trace originates from the 8.2 mM solution. A SDS concentration, which is exact at the cmc, shows a the distortion in the diffusion dimension. This distortion might be reasoned in the partial formation of lower weighted SDS aggregates.

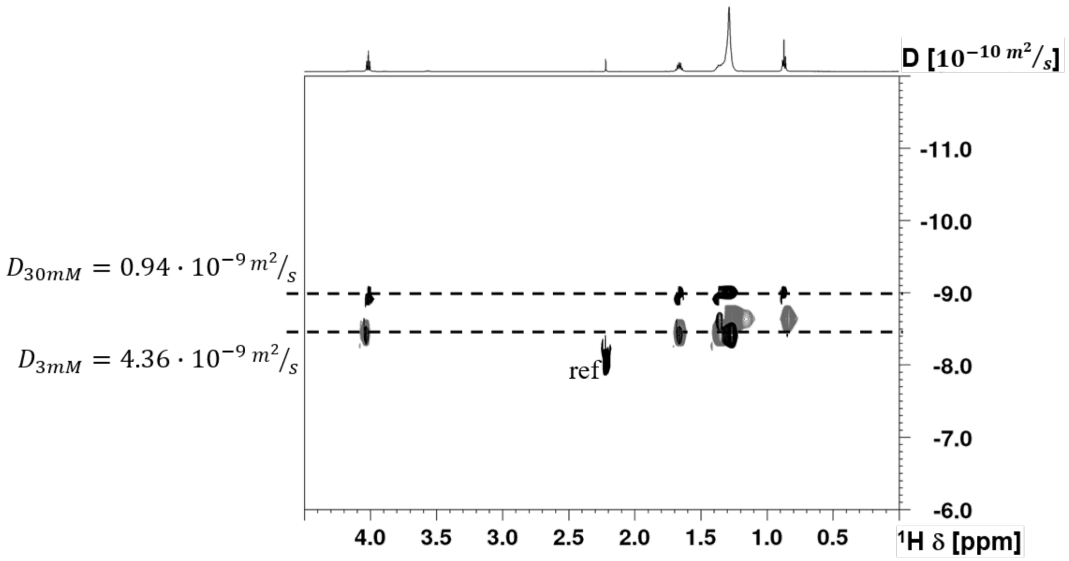


Figure 4.2.: 2D DOSY spectra of three peptide concentrations. Lowest (3 mM) and highest (30 mM) are shown in black traces. Diffusion coefficients are written on related dotted dashed lines. 8.2 mM diffusion dimension in displayed in grey. All spectra are referenced with an external acetone/water capillary, which was set to a diffusion coefficient of –8.1 log(m^2/s) (H_2O/D_2O 9:1, 600 MHz, 298 K, NS 32, 64k (f1) 32 (F2), linear ramp (2–98%), d1 1.5 s, d20 70–120 ms).

4.2. Positioning of small molecules in a SDS micelle

The positioning of molecules in SDS micelles were studied by the application of various ^{1}H NMR spectroscopic approaches. For these NMR investigations, small organic molecules were selected, which had to fulfil a few requirements. One requirement was that the signals of the organic molecule did not overlap with the signals from SDS. The distinct signal positions are particularly important for precise performance of 1D NOE experiments. These 1D NOE experiments were used to measure inter-molecular NOE. To achieve most accurate results, the model system should be simple. Optimally, all protons are magnetically equivalent, which would result in one distinct singlet in the proton spectrum. Furthermore, all small molecules should be similar in structure to be able to assume a similar hydrodynamic radius. Comparability in hydrodynamic radius is particularly important for the diffusion measurements. The ultimate selection criterion was the molecules polarity and hence, its associated miscibility with water.[22] The spectrum from polar to apolar should be covered, in order to fully exploit the zwitterionic property of SDS. Polar groups should either remain in the solvent or interact with the polar head group of SDS. Apolar groups should interact with the hydrophobic tail of SDS and thus, penetrate into the micelle. four organic molecules that are presented in Figure 4.3 were chosen as model systems:[i]

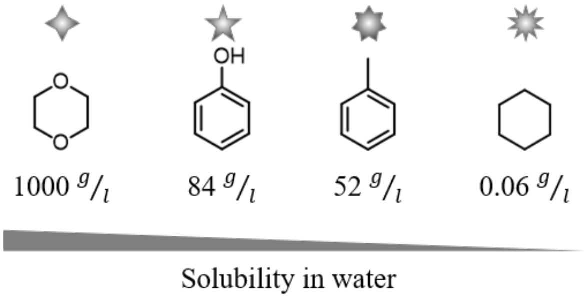


Figure 4.3.: Solubility of the four compounds under investigation.

4.2.1. Chemical shift monitoring

One approach to study the interactions of small organic molecules with SDS micelles is the deviation of the proton chemical shift. In Figure 4.4 all changes in the chemical shifts of SDS, induced by the addition of the additive, are summarised schematically.

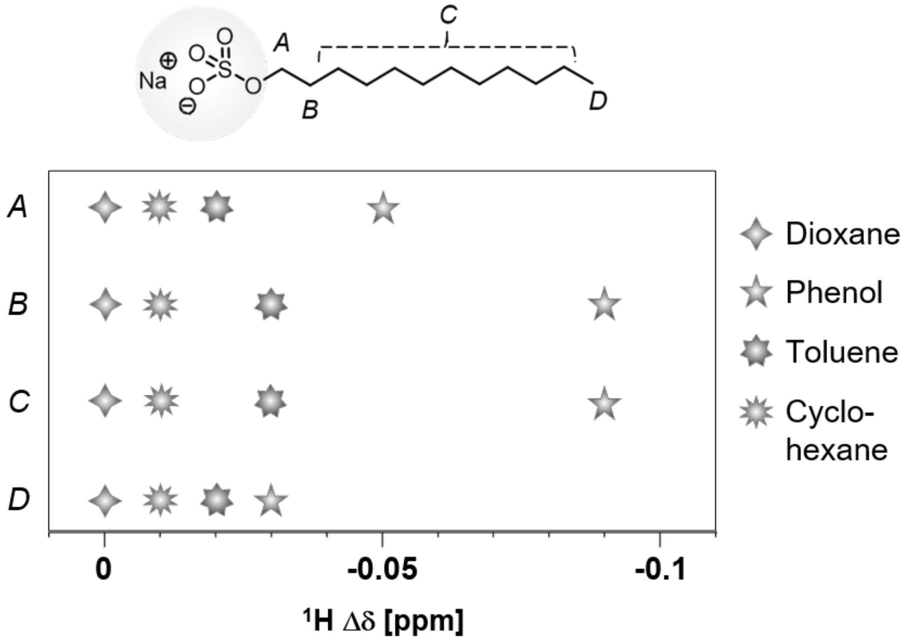


Figure 4.4.: The change in the chemical shift of the SDS micelle signals among addition of a small organic molecule (30 mM SDS, 50 mM additive, H_2O/D_2O 9:1, 600 MHz, 298 K. All related ^1H NMR spectra are included in the appendix).

The chemical shift movement of the SDS micelles is in almost all spectra uniformly to the lower frequency region. No movement of the micelle signals was observed among addition of dioxane. The overall chemical shift migration to lower frequencies might be an indication for an effective incorporation of cyclohexane, phenole and toluene into the micellar core. The largest chemical shift differences are observed among phenole and toluene under influence of micelles. Toluene and phenol have both an aromatic moiety in common. This aromatic group might cause a through space shielding of the spatial close SDS tail groups which are located inside the hydrophobic core.[176]

---

[i] all miscibility values for 20°C taken from gestis substance database gestis.itrust.de

4.2.2. Diffusion studies on the mode of interaction

The investigation of the diffusion properties of SDS in interaction with the small organic molecules are summarised in Table 4.1.

As can be seen from Table 4.1, only the addition of phenol causes a reduction of the diffusion coefficient of the micelles.

Phenol's unique property compared, to the other three additives, is the hydroxyl group. Phenol has a pK_a value of 10.0, which means that the phenolate ion is formed rather simply in aqueous solution under the release of the hydroxy proton. This proton might arrange in the Stern-layer as counter ion of the negatively charged micelle surface. Those proton counterions cause a decrease of the charge density of the micelle and a related weakening of the repulsive interactions between the SDS head groups.[177]

4.2.3. 1D inter-molecular NOE investigations

The localisation of the small organic molecules was to be determined by the detection of inter-molecular NOE. Therefore, 1D NOE experiments were carried out in which the additive itself and various groups of the SDS were selectively irradiated.

Figure 4.5 shows the spectra obtained for the dioxane in SDS micelles.

When irradiating dioxane, groups *A* and *B* show the most intense NOE signal. In contrast to the selective irradiation on *D*, the irradiation on *B* shows that dioxane must be spatially adjacent to *B*. It can therefore be assumed that dioxane is located near the head group.

The 1D inter-molecular NOE experiments showed no preferential orientation for the positioning of toluene, phenol and cyclohexane, i. e. core or close to head group. Thus, it all additives, except dioxane, are located in the hydrophobic interior of the micelle.

Table 4.1.: Experimentally determined diffusion coefficients. Values are relative to the external standard acetone/water which was set to -8.1 $log(^{m^2}/_s)$. Coefficients belong always to bold emphasized compounds. D values are all in $[\cdot 10^{-9}\ ^{m^2}/_s]$ (ΔD: relative deviation in reference to pure component, CMC: critical micellar concentration, Dxn: dioxane, Tln: toluene, cyH: cyclohexane).

		$<$ **CMC**		$>$ **CMC**	
	D $[^{m^2}/_s]$	D $[^{m^2}/_s]$	ΔD [%]	D $[^{m^2}/_s]$	ΔD [%]
pure SDS	—	4.36	—	0.94	—
pure **Dxn**	5.15	—	—	—	—
SDS + Dxn	—	2.59	41	1.40	49
SDS + ***Dxn***	—	6.25	21	7.93	54
pure **PhOH**	6.89	—	—	—	—
SDS + PhOH	—	3.59	21	0.74	21
SDS +***PhOH***	—	5.78	16	4.12	40
pure **Tln**	7.94[*]	—	—	—	—
SDS + Tln	—	3.30	24	1.41	50
SDS + ***Tln***	—	7.69	3.0	3.30	58
pure **cyH**	4.70[**]	—	—	—	—
SDS + cyH	—	3.33	24	1.42	51
SDS + ***cyH***	—	7.85	67	1.42[***]	70

[*] Measured in three days old sample where one signal set was left

[**] Value taken from singlet which was together with a broad blurred signal, that was not detectable in diffusion dimension

[***] Same diffusion coefficient is assumed for micelle and cyH because they do not split up up in diffusion dimension

4.2.4. Investigations of binding properties using waterLOGSY

The binding properties of the four small organic molecules to SDS micelles have been investigated using the waterLOGSY method. During the waterLOGSY experiments, it was noticed that the signals of the small molecules, while in the presence of micelles, did not show the expected waterLOGSY response. Instead, distorted signals were always obtained. However, in the reference measurements of the small molecules in water no distorted signals were observed.

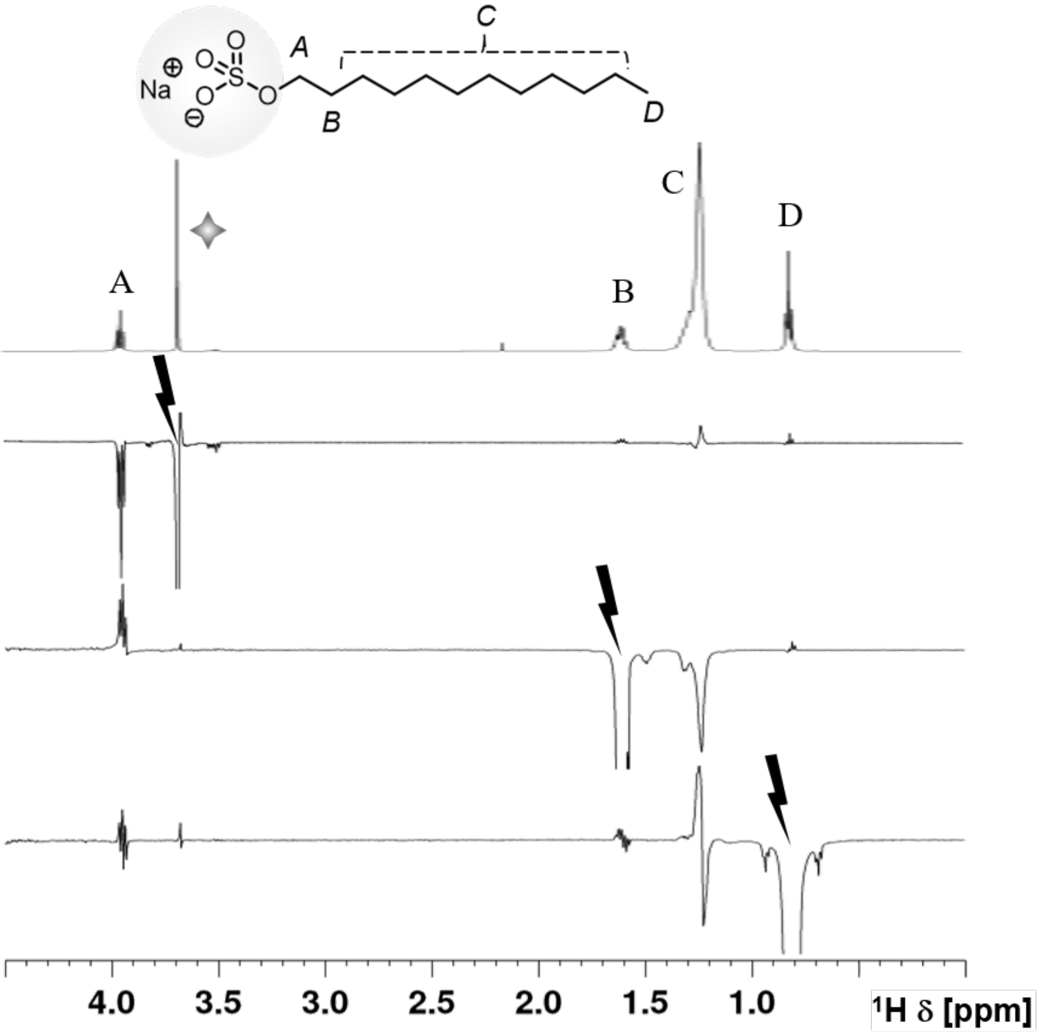


Figure 4.5.: 1D NOESY spectra of dioxane in presence of SDS micelles. The selective irradiated signal in indicated by a flash H_2O/D_2O 9:1, 500 MHz, mixing time: 0.9 sec, 298 K).

Figure 4.6 shows the waterLOGSY spectrum of cyclohexane in micellar medium as well as the reference spectrum in water and the proton spectrum:

The obtained signal of cyclohexane does not correspond to a clear binding or non-binding situation to the micelle, as was explained earlier in the introduction. Consequently, the waterLOGSY experiment cannot be evaluated in the conventional way. If the component under investigation is not in contact with water, it cannot give a NOE response in the waterLOGSYs experiment. Thus, in this particular case, the experiment can be used to distinguish whether a component is located inside the hydrophobic micelle or in solution.

Toluene and phenol show the same behaviour in the waterLOGSY spectrum as cyclo-

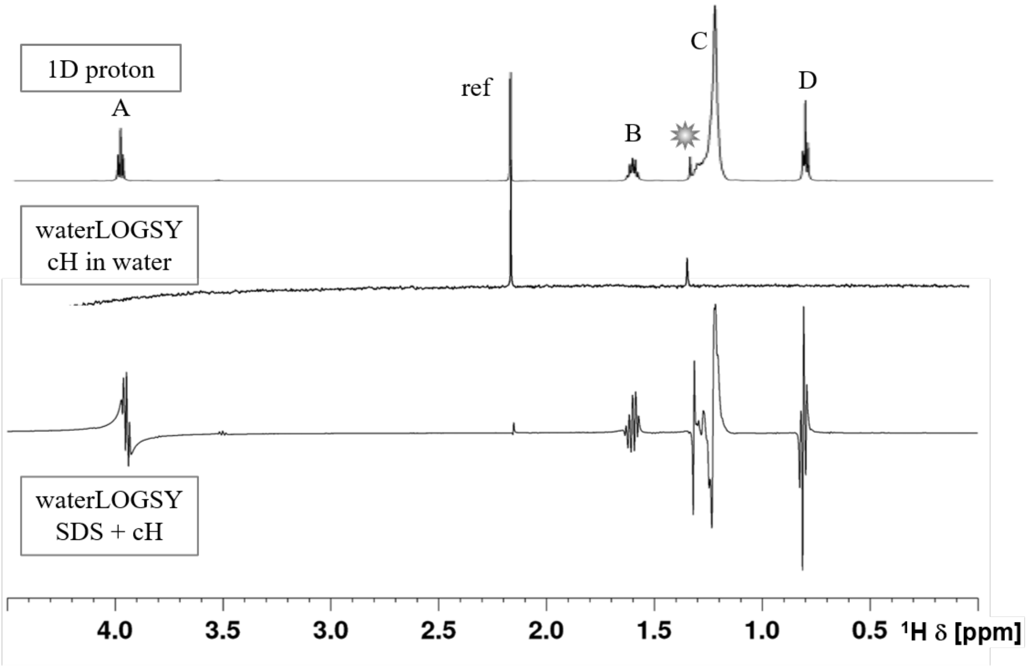


Figure 4.6.: waterLOGSY (bottom) and 1D proton (top) spectrum of the sample including cyclohexane and micelles (H_2O/D_2O 9:1, 500 MHz, mixing time: 0.8 sec, 298 K).

hexane. Thus, it can be assumed for toluene that it is present in the micelle core.

4.2.5. Titration with paramagnetic agents

The experiments carried out clearly showed that toluene, cyclohexane and phenol are located inside the micelles. Based on the previously described waterLOGSY experiments, the exact penetration depth of dioxane could not yet be clarified. It remains to be determined if dioxane is attached to the micellar surface (exposed to solvent) or close to the head group inside the micelle.

To investigate a possible accumulation of dioxane on the surface, titration experiments were carried out with the paramagnetic shift reagent $MnCl_2$. For this purpose, a bi-component mixture of cyclohexane and dioxane was prepared. Cyclohexane serves as a reference, as it is proven to be located in the micellar core. The 1D proton NMR spectra of the titration is shown in Figure 4.7.

The 1D proton spectra with increasing manganese ion concentration clearly shows that the signals of all three components gradually lose signal sharpness, splitting and intensity. This very uniform signal change of the proton signals during the titration suggests that dioxane is located inside the micelle. Localisation of dioxane on the micelle surface would

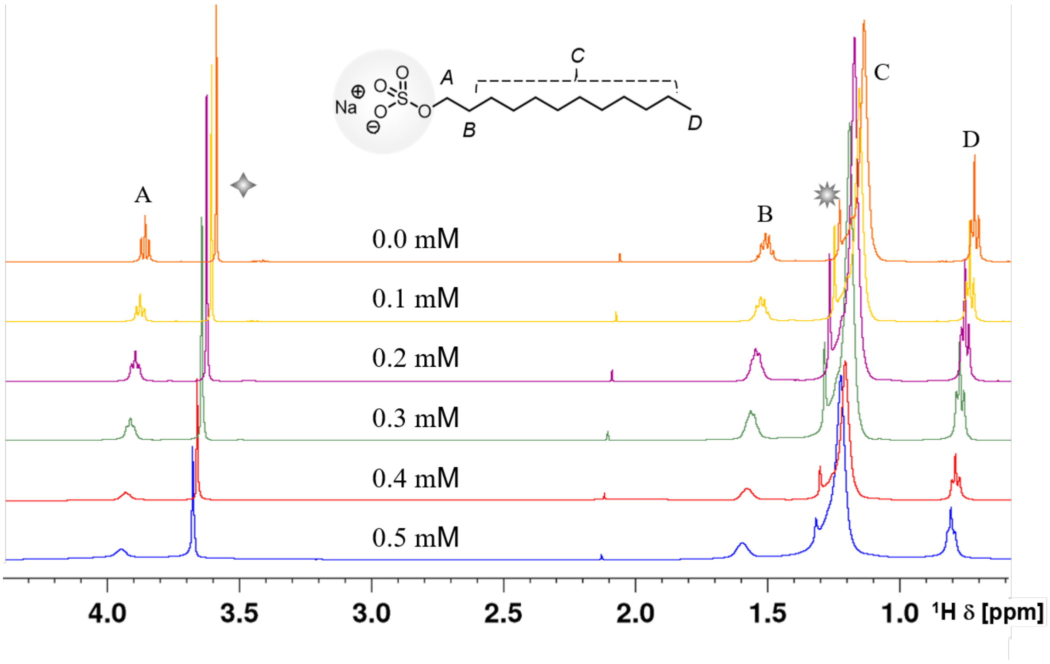


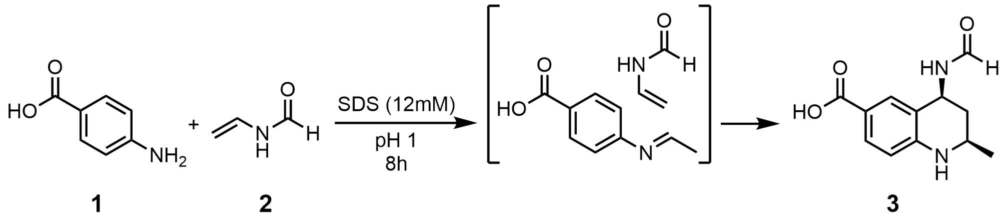
Figure 4.7.: 1D proton spectra of the $MnCl_2$ titration experiment of the sample including dioxane and SDS micelles. $MnCl_2$ concentration is indicated for every spectrum (H_2O/D_2O 9:1, 500 MHz, 298 K).

have led to a comparatively more rapid loss of signal.

Based on the experimental results, it can be concluded that dioxane is localised inside the micelle.

4.3. Micellar catalysis of a Diels-Alder reaction – an exploratory NMR study

An exploratory investigation of micelles as a catalyst in an organic reaction was carried out with a hetero Diels-Alder reaction. A variety of substrates and optimal reaction conditions (i. e. pH value, SDS concentration) are described in this publication.[171] The amine substrate used was mostly an aniline derivative, which is known for its toxicity. 4-aminobenzoic acid (**1**) was thereby the less harmful selection so that it was chosen together with *N*-*v*inyl*f*ormamide (**2**) (abbr. NVF) as reaction system. One factor in the choice of system was the reaction time. Reactions that take place very rapidly are only detectable through the use of special NMR methods. Therefore, 8 hours reaction time is appropriate as a test system. Best reaction conditions were reported at pH 1.0–2.5 at an SDS concentration of 12 mM at ambient temperature. The reaction in Scheme 4.1 was chosen as a suitable reaction system for the NMR spectroscopic investigations.



Scheme 4.1.: Reaction scheme of chosen model system.[171]

Prior to the performance of the reaction, the behaviour of the substrates at micelle concentration and pH was investigated. NVF **2** in the presence of 12 mMSDS showed in the proton spectra 4.8 that it does not rearrange into the active imine form under neutral pH (step 1). In contrast, at pH 1 it is evident that the rearrangement takes place. This can be seen through the appearance of protons 3b and 4b. However, blank measurements with NVF in water without SDS have shown that this rearrangement also takes place without micelles being pH-dependent (cf. NVF in water spectrum in appendix G.19) The three signals which are enlarged in the top proton trace belong to the exchanging NH groups (cf Figure 4.8). This signal form is well known and investigated in relation to guadinium moieties.[178]

The differentiation between amide and imide form of substrate **2** is also possible in a 2D HSQCed spectrum in the carbon dimension. The C–H correlation spectrum provides a

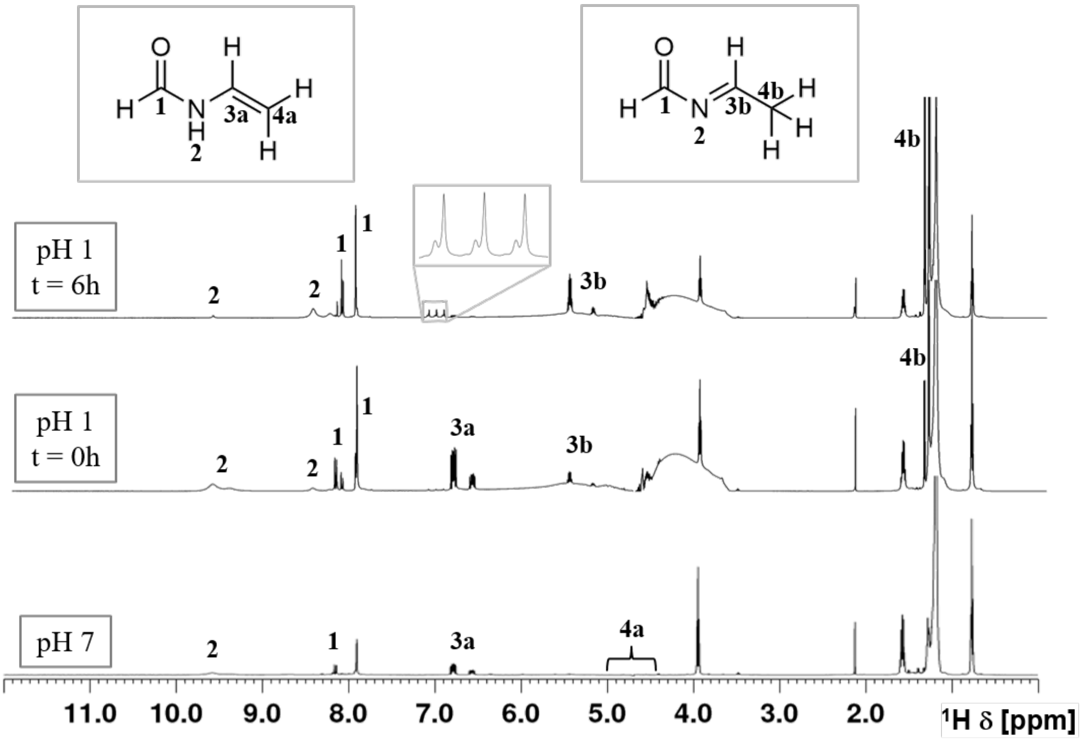


Figure 4.8.: 1D proton spectra of substrate **2** at pH 1 and 7. At pH 7 (bottom trace), the amide form (left) is predominant. At pH 1, the amide form begins to rearrange to the imine form (right). At t=0h (middle trace), H3a signals of both forms are detected, whereas at t=6h (top trace) conversion to imine form is passed almost completely (H_2O/D_2O 9:1, 298 K, 600 MHz).

clear overview of the exact signal location of the protons H3a (pH 1) and H4a (pH 7). The signal of SDS is not affected upon pH change.

Diffusion experiments with NVF in micelle solution show that SDS drastically reduces its diffusion coefficient at lower pH from $5.36 \cdot 10^{-9}\ ^{m^2}/_{s}$ to $2.22 \cdot 10^{-9}\ ^{m^2}/_{s}$ (cf. appendix). In the studies of the interaction of phenol and SDS a decrease in the diffusion coefficient was already observed. This decrease in diffusion velocity was explained by the dissociated protons of phenol's hydroxy group. In this case, the charge density of the micelle is reduced by the increased amount of protons at pH 1, which results in an extreme reduction in diffusion.

4.3.1. The reaction mechanism

The postulated reaction mechanism (cf Figure 4.9) of the micelle catalysed Provarov reaction starts with the amide that approaches the *Stern* layer of the micelle where it rearranges to form the imide. This activated imide form can now react with the carboxylic acid-functionalised aniline. After the cleavage of acetamide, the reaction with the second equivalent of vinylamide proceeds. The next step is the cyclisation reaction and dearomatisation of the aniline component. The driving force for the final product formation is a rearomatisation, which takes place through electron rearrangements. It is assumed that the amine group of the functionalised aniline is coordinated to the negatively charged micelle surface during the entire reaction.

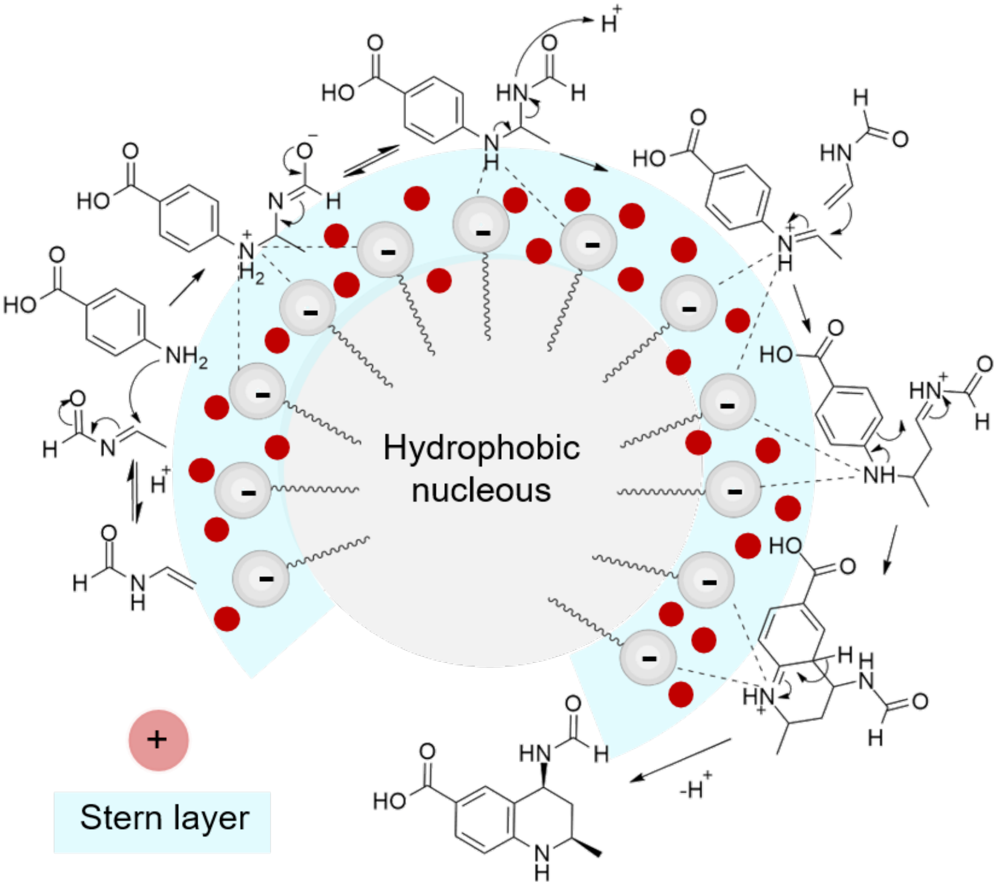


Figure 4.9.: Proposed reaction mechanism by Kouznetsov *et al.*.[171]

Two experimental setups were employed: (*a*) for an exact reproduction, the reaction was carried out under laboratory conditions. At defined intervals, samples were taken from the flask and measured; (*b*) All reaction components were added to an NMR tube according to the protocol. The reaction took place in the NMR tube without constant stirring. Figure 4.10 shows the samples taken from the reaction under laboratory conditions.
At the beginning of the reaction, it is clearly visible that the reaction solution is very milky due to barely dissolved starting material (left). After the reaction has been completed, the solution clears up and turns slightly orange (right) with yellowish precipitate.

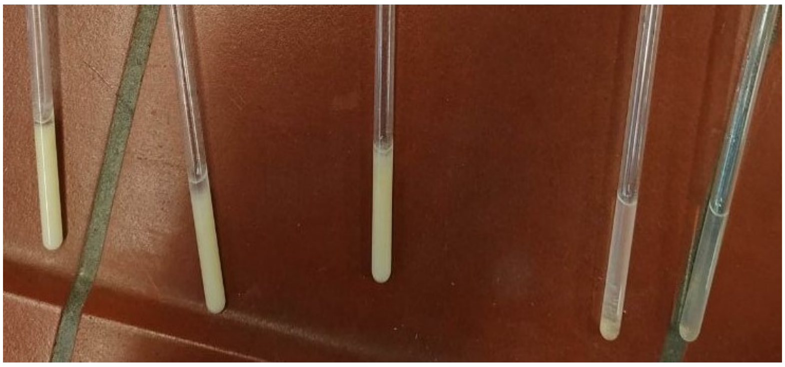


Figure 4.10.: Samples taken from reaction carried out in the flask after time steps 1h to 7h (left to right).

The spectra recording was strongly influenced by the inhomogeneity of the samples which caused a poor signal resolution. The comparison of the signals in the 1D proton spectra of both approaches showed no significant difference between the reaction process in a flask or the probe head.

For both approaches, no full conversion was achieved after the described reaction time of 8h. Especially the aromatic protons of 4-aminobenzoic acid and the most significant signals H3b of compound **2** at ~5.1 ppm are still visible, as can be seen in the HSQCed spectrum 4.11.

A closer look further reveals that multiple CH_2 groups were formed in the reaction. This is a first indicator that the reaction is not as clean as described. In the range of 0.5–2.0 ppm of the aliphatic region no signal which could belong to SDS was obtained. Thus, it is concluded that the formed precipitate is aggregated SDS.

In Figure 4.12 the TOCSY spectra of substrate **2** (purple) superimposed to the crude product (black) are shown.

In all recorded spectra too many signal were observed, so that it can be deduced that a mixture was formed. An exact assignment of the aromatic product signals cannot be achieved.

The enlarged representation shows the signals that can be assigned to the product. A second, similar spin system is identified in the TOCSY correlation (marked with grey dashed lines).

2D DOSY experiments were also recorded but could not be clearly evaluated due to the large number of signals.

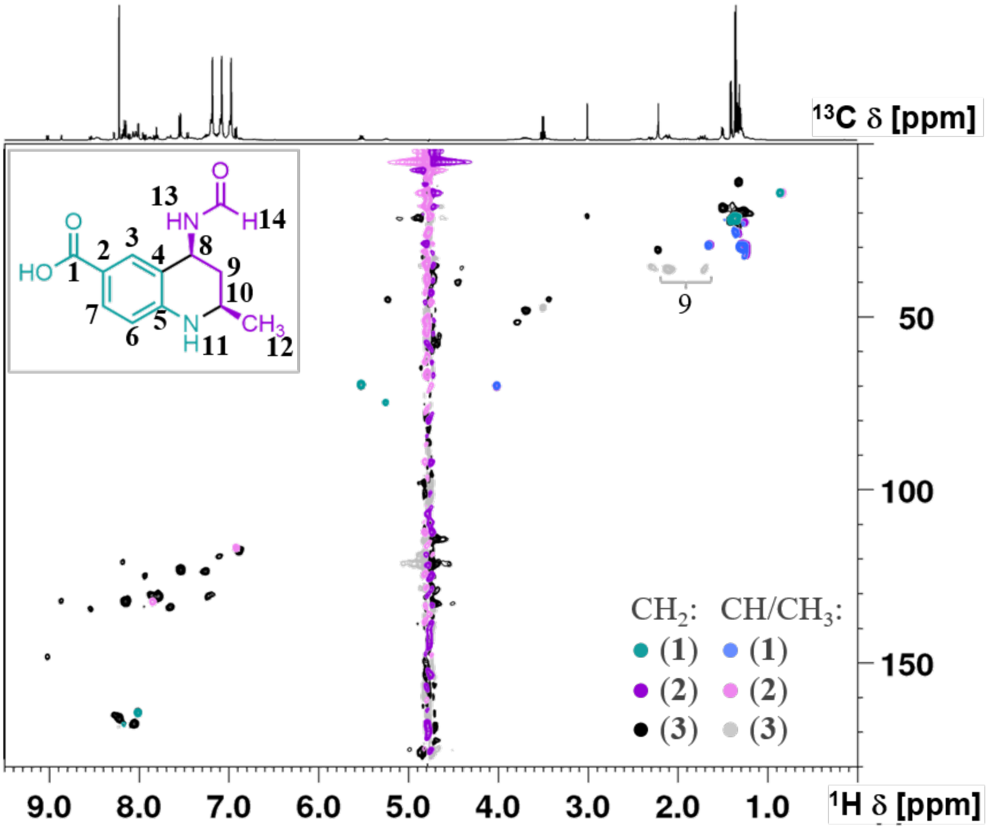


Figure 4.11.: 2d TOCSY spectra of both starting materials **1** (green/blue) and **2** (purple/violet) as well as the crude reaction product (black/grey). The phase colours are described in the legend (H_2O/D_2O 9:1, pH 1, 600 MHz, 298 K).

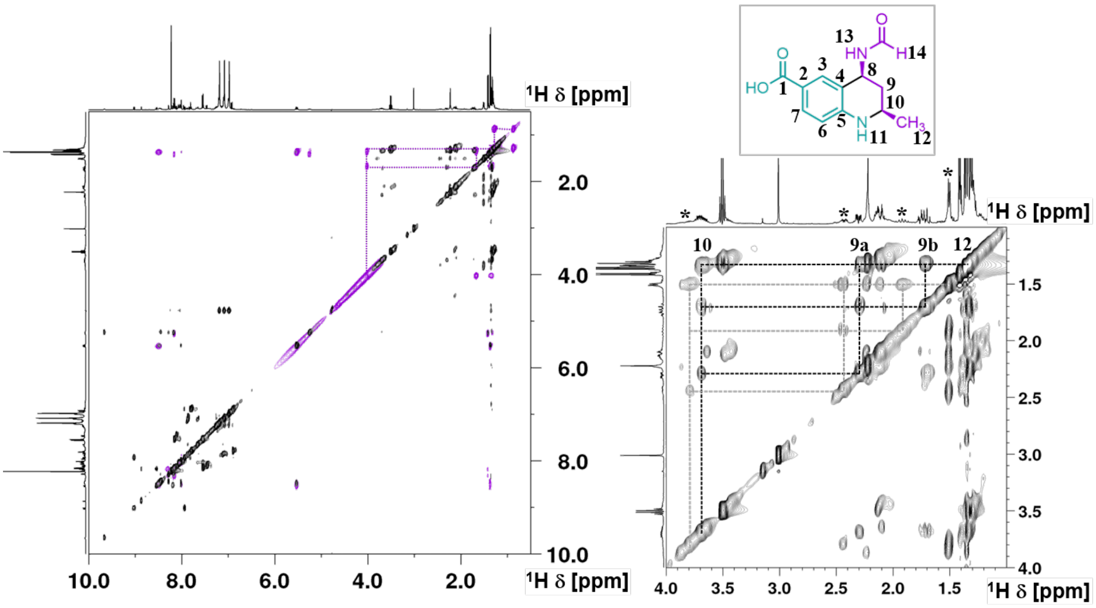


Figure 4.12.: Superimposed phase sensitive HSQC spectra of both starting materials **1** (green/blue) and **2** (purple/violet) as well as the crude reaction product (black/grey) (H_2O/D_2O 9:1, pH 1, 600 MHz, 298 K).

5. Summary and Outlook

The positioning of small organic molecules in or on a micelle was investigated using NMR spectroscopic methods. Four molecules were selected that were similar in size but had very opposite polarity.

The influence of the small molecules on the diffusion properties of the micelle-small molecule complex is summarised in Table 4.1

The diffusion properties of SDS change whenever they interact with one of the small molecules. These effective interactions have already been recognised in the 1D proton spectra. Interestingly, micelles tend to appear smaller when interacting with small organic molecules.

Only the interaction of phenol with micelles causes a decrease in the diffusion velocity of the micelle. This was attributed to the fact that the dissociated protons of the hydroxy group reduce the charge density of the micelle.

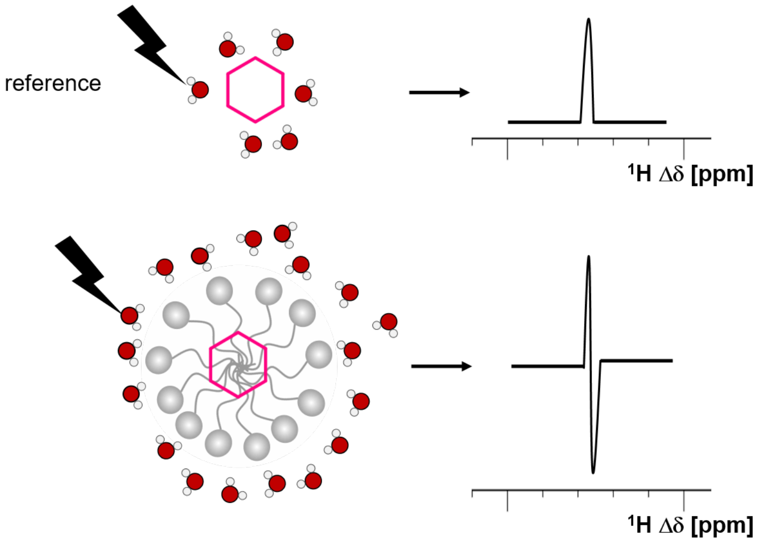


Figure 5.1.: The waterLOGSY experiment could be used in the context of this work to test the exposure to water of small components in the presence of micelles.

The positioning studies using inter-molecular NOE acquisitions showed that all four organic molecules are located in the micelle. The NOE acquisition of SDS upon implementation of dioxane further showed that the SDS signals near the head group were more strongly developed upon selective irradiation of dioxane than the remaining SDS signals. The irradiation of SDS signal B also showed that dioxane is located in spatial proximity. The other molecules seem to be anchored quite deeply inside the micelle. The assumption of deep anchoring of toluene, cyclohexane and phenol in the micelle interior was also supported by waterLOGSY experiments.

It was found that the waterLOGSY experiment cannot be applied here in a conventional way. Instead, it is more a way of making a conclusion about the general water exposure of a molecule, as drawn schematically in Figure 5.1.

The monitoring of a micelle catalysed reaction as an application in the field of green chemistry worked moderately. For a further investigation on micelle promoted reactions it would be important to optimise the reaction conditions.

Part VI.

APPENDICES

A. Bibliography

[1] P. Berche, *Clinical Microbiology and Infection* **2012**, *18*, 1–6.

[2] H. D. Flack, *Acta Crystallographica Section A* **2009**, *65*, 371–389.

[3] D. Marion, *Mol. Cell. Proteomics* **2013**, *12*, 3006–3025.

[4] M. Hesse, H. Meier, *Spektroskopische Methoden in der organischen Chemie, 8. überarb. Auflage 2011*, Georg Thieme Verlag, Stuttgart, **2014**.

[5] P. C. Driscoll, A. M. Gronenborn, L. Beress, G. M. Clore, *Biochemistry* **1989**, *28*, 2188–2198.

[6] F. Bloch, W. W. Hansen, M. Packard, *Phys. Rev.* **1946**, *70*.

[7] E. M. Purcell, H. C. Torrey, R. V. Pound, *Phys. Rev.* **1946**, *69*, 37–38.

[8] R. R. Ernst, W. A. Anderson, *Rev. Sci. Instrum.* **1966**, *37*.

[9] M. P. Williamson, T. F. Havel, K. Wüthrich, *Journal of Molecular Biology* **1985**, *182*, 295–315.

[10] F. Castiglione, M. Ferro, A. Mele, *Microbiologia Medica* **2018**, *32*.

[11] A. Yee, X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, T. Ramelot, G. M. Lee, S. Bhattacharyya, P. Gutierrez, A. Denisov, C.-H. Lee, J. R. Cort, G. Kozlov, J. Liao, G. Finak, L. Chen, D. Wishart, W. Lee, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. M. Edwards, C. H. Arrowsmith, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1825–30.

[12] J. M. Berg, L. Stryer, J. L. Tymoczko, *Stryer Biochemie*, Springer-Verlag, Berlin-Heidelberg-New York, **2015**.

[13] N. Sewald, H.-D. Jakubke, *Peptides: chemistry and biology*, John Wiley & Sons, New Jersey, **2015**.

[14] I. Bertini, K. S. McGreevy, G. Parigi, *NMR of Biomolecules: Towards Mechanistic Systems Biology*, **2012**.

[15] K. Wüthrich, *Europhys. News* **1986**, *17*, 11–13.

[16] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95–99.

[17] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531–1534.

[18] H. Kessler, M. Gehrke, C. Griesinger, *Angew. Chem. Int. Ed.* **1988**, *27*, 490–536.

[19] P. J. Hore, *Nuclear magnetic resonance*, Oxford University Press, United Kingdom, **2015**.

[20] J. Keeler, *Understanding NMR Spectroscopy* **2011**, 526.

[21] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, *Protein NMR spectroscopy: principles and practice*, Elsevier Academic Press, Burlington-San Diego-London, **1995**.

[22] P. W. Atkins, J. De Paula, *Physical Chemistry*, John Wiley & Sons, New Jersey, **2014**.

[23] T. D. W. Claridge, *High-resolution NMR techniques in organic chemistry*, *Vol. 27*, Elsevier, **2016**.

[24] J. Shin, W. Lee, W. Lee, *Expert Rev. Proteomics* **2008**, *5*, 589–601.

[25] P. Güntert, *Q. Rev. Biophys.* **1998**, *31*, 145–237.

[26] K. Wüthrich, *NMR of proteins and nucleic acids*, Wiley, New Jersey, **1986**.

[27] L. Y. Lian, G. Roberts, *Protein NMR Spectroscopy: Practical Techniques and Applications*, **2011**.

[28] T. Cierpicki, J. Otlewski, *J. Biomol. NMR* **2001**, *21*, 249–61.

[29] N. J. Baxter, M. P. Williamson, *J. Biomol. NMR* **1997**, *9*, 359–369.

[30] C. F. Geraldes, *Encyclopedia of Inorganic and Bioinorganic Chemistry* **2012**, 1–20.

[31] T. Sugiki, K. Furuita, T. Fujiwara, C. Kojima, *Molecules* **2018**, *23*.

[32] A. Bundi, K. Wüthrich, *Biopolymers* **1979**, *18*, 285–297.

[33] K. Wüthrich, G. Wider, G. Wagner, W. Braun, *J. Mol. Biol.* **1982**, *155*, 311–319.

[34] D. Wishart in *Encyclopedia of Biophysics*, (Ed.: G. C. K. Roberts), Springer Berlin Heidelberg, Berlin, Heidelberg, **2013**, pp. 279–280.

[35] D. S. Wishart, B. D. Sykes, F. M. Richards, *Biochemistry* **1992**, *31*, 1647–1651.

[36] M. P. Williamson in *Mod. Magn. Reson.* **2018**.

[37] F. L. Nobrega, E. J. Cabrita, *J. Chem. Educ.* **2011**, 990–994.

[38] S. Walpole, S. Monaco, R. Nepravishta, J. Angulo, *STD NMR as a Technique for Ligand Screening and Structural Studies*, *Vol. 615*, 1st ed., Elsevier Inc., **2019**, pp. 423–451.

[39] V. V. Krishnan, *Curr. Anal. Chem.* **2005**, *1*, 307–320.

[40] A. Antanasijevic, B. Ramirez, M. Caffrey, *J. Biomol. NMR* **2014**, *60*.

[41] C. Raingeval, O. Cala, B. Brion, M. Le Borgne, R. E. Hubbard, I. Krimm, *J. Enzyme Inhib. Med. Chem.* **2019**, *34*, 1218–1225.

[42] C. Dalvit, G. P. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349–359.

[43] R. Huang, A. Bonnichon, T. D. W. Claridge, I. K. H. Leung, *Nature Publishing Group* **2017**, 1–6.

[44] R. Huang, I. K. Leung, *Protein–Small Molecule Interactions by WaterLOGSY*, *Vol. 615*, 1st ed., Elsevier Inc., **2019**, pp. 477–500.

[45] Y. Cohen, L. Avram, L. Frish, *Angew. Chem. Int. Ed.* **2005**, *44*, 520–554.

[46] R. Kerssebaum, *TopSpin User Manual*, Bruker BioSpin GmbH, Rheinstetten, **2016**.

[47] P. Güntert, *Methods in molecular biology (Clifton N.J.)* **2004**, *278*, 353–378.

[48] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*.

[49] P. Güntert, L. Buchner, *J. Biomol. NMR* **2015**, *62*, 453–471.

[50] M. Moradi, V. Babin, C. Roland, T. A. Darden, C. Sagui, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 20746–20751.

[51] D. Pal, P. Chakrabarti, *J. Mol. Biol.* **1999**, *294*, 271–288.

[52] W. J. Wedemeyer, E. Welker, H. A. Scheraga, *Biochemistry* **2002**, *41*, 14637–14644.

[53] K. K. Mehta, E. E. Paskaleva, X. Wu, N. Grover, R. V. Mundra, K. Chen, Y. Zhang, Z. Yang, H. Feng, J. S. Dordick, R. S. Kane, *Biotechnology and Bioengineering* **2016**, *113*.

[54] L. Kyne, M. B. Hamel, R. Polavaram, C. P. Kelly, *Clin. Infect. Dis.* **2002**, *34*, 346–353.

[55] C. Lübbert, C. Johann, A. S. Kekulé, D. Worlitzsch, S. Weis, J. Mössner, S. Behl, T. Seufferlein, *Zeitschrift fur Gastroenterologie* **2013**, *51*.

[56] S. Dial, J. A. C. Delaney, A. N. Barkun, S. Suissa, *JAMA* **2005**, *294*, 2989–2995.

[57] D. E. Voth, J. D. Ballard, Clostridium difficile toxins: Mechanism of action and role in disease, **2005**.

[58] J. R. O'Connor, S. Johnson, D. N. Gerding, *Gastroenterology* **2009**, *136*, 1913–1924.

[59] N. M. Chumbler, M. A. Farrow, L. A. Lapierre, J. L. Franklin, D. Haslam, J. R. Goldenring, D. B. Lacy, *PLoS Pathogens* **2012**, *8*.

[60] V. Arato, G. Gasperini, F. Giusti, I. Ferlenghi, M. Scarselli, R. Leuzzi, *Scientific Reports* **2019**, *9*.

[61] J. Nesper, I. Hug, S. Kato, C. S. Hee, J. M. Habazettl, P. Manfredi, S. Grzesiek, T. Schirmer, T. Emonet, U. Jenal, *eLife* **2017**, *6*.

[62] P. J. Hensbergen, O. I. Klychnikov, D. Bakker, I. Dragan, M. L. Kelly, N. P. Minton, J. Corver, E. J. Kuijper, J. W. Drijfhout, H. C. van Leeuwen, *FEBS Lett.* **2015**, *589*, 3952–3958.

[63] J. Corver, V. Cordo', H. C. van Leeuwen, O. I. Klychnikov, P. J. Hensbergen, *Mol. Microbiol.* **2017**, *105*, 663–673.

[64] R. Tamayo, *PLoS Pathogens* **2019**, *15*.

[65] U. Jenal, A. Reinders, C. Lori, *Nat. Rev. Microbiol.* **2017**, *15*, 271–284.

[66] E. B. Purcell, R. W. McKee, S. M. McBride, C. M. Waters, R. Tamayo, *J. Bacteriol.* **2012**, *194*.

[67] C. Pichlo, L. Juetten, F. Wojtalla, M. Schacherl, D. Diaz, U. Baumann, *J. Biol. Chem.* **2019**, *294*.

[68] M. Schacherl, C. Pichlo, I. Neundorf, U. Baumann, *Structure* **2015**, *23*, 1632–1642.

[69] P. J. Hensbergen, O. I. Klychnikov, D. Bakker, V. J. C. van Winden, N. Ras, A. C. Kemp, R. A. Cordfunke, I. Dragan, A. M. Deelder, E. J. Kuijper, J. Corver, J. W. Drijfhout, H. C. van Leeuwen, *Mol. Cell. Proteomics* **2014**, *13*, 1231–1244.

[70] R. L. J. Keller, *The Computer Aided Resonance Assignment Tutorial*, Cantina, Goldau, **2004**.

[71] Schrödinger LLC, *The PyMOL Molecular Graphics System, Version 1.8*, **2015**.

[72] O. Trott, A. J. Olson, *J. Comp. Chem.* **2009**.

[73] M. Schacherl, C. Pichlo, I. Neundorf, U. Baumann, *Structure* **2015**, *23*, 1632–1642.

[74] T. F. Vieira, S. F. Sousa, *Appl. Sci.* **2019**, *9*.

[75] R. Rentzsch, B. Y. Renard, *Briefings in Bioinformatics* **2015**, *16*, 1045–1056.

[76] H. Derakhshankhah, S. Jafari, *Biomed. Pharmacother.* **2018**, *108*, 1090–1096.

[77] S. G. Patel, E. J. Sayers, L. He, R. Narayan, T. L. Williams, E. M. Mills, R. K. Allemann, L. Y. Luk, A. T. Jones, Y. H. Tsai, *Scientific Reports* **2019**, *9*, 1–9.

[78] A. Prochiantz, *Curr. Opin. in Cell Biol.* **2000**, *12*, 400–406.

[79] K. Wagstaff, D. Jans, *Curr. Med. Chem.* **2006**, *13*.

[80] D. Derossi, S. Calvet, A. Trembleau, A. Brunissen, G. rard Chassaing, A. Prochiantz, *J. Biol. Chem.* **1996**, *271*, 18188–18193.

[81] A. Gräslund, F. Madani, S. Lindberg, Ü. Langel, S. Futaki, *J. Biophys.* **2011**, 414729.

[82] H. Myrberg, L. Zhang, M. Mäe, Ü. Langel, *Bioconjugate Chem.* **2008**, *19*.

[83] S. Aroui, D. Mili, S. Brahim, M. D. Waard, A. Kenani, *Biochem. Biophys. Res. Commun.* **2010**, *396*.

[84] J. Xie, Y. Bi, H. Zhang, S. Dong, L. Teng, R. J. Lee, Z. Yang, *Frontiers in Pharmacology* **2020**, *11*, 1–23.

[85] J. L. Lau, M. K. Dunn, *Bioorg. Med. Chem.* **2018**, *26*.

[86] S. Sachdeva, *Int. J. Pept. Res. Ther.* **2017**, *23*, 49–60.

[87] D. Kalafatovic, E. Giralt, *Molecules* **2017**, *22*.

[88] M. Horn, F. Reichart, S. Natividad-Tietz, D. Diaz, I. Neundorf, *Chem. Comm.* **2016**, *52*, 2261–2264.

[89] I. Nagaoka, S. Hirota, F. Niyonsaba, M. Hirata, Y. Adachi, H. Tamura, D. Heumann, *J. Immunol.* **2001**, *167*, 3329–3338.

[90] J. Hoyer, U. Schatzschneider, M. Schulz-Siegmund, I. Neundorf, *Beilstein J. Org. Chem.* **2012**, *8*, 1788–1797.

[91] C. Chen, R. Brock, F. Luh, P. J. Chou, J. W. Larrick, R. F. Huang, T. huang Huang, *FEBS Letters* **1995**, *370*, 46–52.

[92] J. Hoyer, U. Schatzschneider, M. Schulz-Siegmund, I. Neundorf, *Beilstein J. Org. Chem* **2012**, *8*, 1788–1797.

[93] I. Neundorf, R. Rennert, J. Hoyer, F. Schramm, K. Löbner, I. Kitanovic, S. Wölfl, *Pharmaceuticals* **2009**, *2*, 49–65.

[94] M. Horn, I. Neundorf, *Scientific Reports* **2018**, *8*, 1–12.

[95] M. B. Martins, I. Carvalho, *Tetrahedron* **2007**, *63*, 9923–9932.

[96] A. D. Borthwick, *Chem. Rev.* **2012**, *112*, 3641–3716.

[97] A. S. M. Ressurreição, R. Delatouche, C. Gennari, U. Piarulli, *European J. Org. Chem.* **2011**, 217–228.

[98] I. Guzzetti, M. Civera, F. Vasile, D. Arosio, C. Tringali, U. Piarulli, C. Gennari, L. Pignataro, L. Belvisi, D. Potenza, *ChemistryOpen* **2017**, *6*, 128–136.

[99] A. S. M. Ressurreição, A. Bordessa, M. Civera, L. Belvisi, C. Gennari, U. Piarulli, *J. Org. Chem.* **2008**, *73*, 652–660.

[100] M. Hällbrink, J. Oehlke, G. Papsdorf, M. Bienert, *Biochim. Biophys. Acta - Biomembranes* **2004**, *1667*.

[101] E. Eiríksdóttir, K. Konate, Ü. Langel, G. Divita, S. Deshayes, *Biochim. Biophys. Acta - Biomembranes* **2010**, *1798*.

[102] M. Di Pisa, G. Chassaing, J. M. Swiecicki, *Biochemistry* **2015**, *54*, 194–207.

[103] F. Duchardt, I. R. Ruttekolk, W. P. Verdurmen, H. Lortat-Jacob, J. Bürck, H. Hufnagel, R. Fischer, M. van den Heuvel, D. W. Löwik, G. W. Vuister, A. Ulrich, M. de Waard, R. Brock, *Journal of Biological Chemistry* **2009**, *284*, 36099–36108.

[104] P. Bayer, M. Kraft, A. Ejchart, M. Westendorp, R. Frank, P. Rösch, *J. Mol. Biol.* **1995**, *247*.

[105] L. Feni, L. Jütten, S. Parente, U. Piarulli, I. Neundorf, D. Diaz, *Chem. Commun.* **2020**, *56*, 5685–5688.

[106] *TopSpin3.5*, Bruker BioSpin GmbH, Karlsruhe, **2016**.

[107] Schrödinger LLC, *Schroedinger Release 2019-2: Maestro*, New York, NY, **2019**.

[108] Schrödinger LLC, *Schroedinger Release 2019-2: MacroModel 9.7*, New York, NY, **2019**.

[109] E. Polak, G. Ribiere, *Revue française d'informatique et de recherche opérationnelle. Série rouge* **1969**, *3*.

[110] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, R. M. Levy, *J. Comput. Chem.* **2005**, *26*.

[111] F. Guarnieri, W. C. Still, *J. Comput. Chem.* **1994**, *15*.

[112] N. H. Andersen, J. W. Neidigh, S. M. Harris, G. M. Lee, Z. Liu, H. Tong, *J. Am. Chem. Soc.* **1997**, *119*, 8547–8561.

[113] T. Cierpicki, J. Otlewski, *J. Biomol. NMR* **2001**, 249–261.

[114] I. Guzzetti, M. Civera, F. Vasile, D. Arosio, C. Tringali, U. Piarulli, C. Gennari, L. Pignataro, L. Belvisi, D. Potenza, *ChemistryOpen* **2016**, *6*.

[115] A. S. M. da Ressurreição, A. Vidu, M. Civera, L. Belvisi, D. Potenza, L. Manzoni, S. Ongeri, C. Gennari, U. Piarulli, *Chem. Eur. J.* **2009**, *15*, 12184–12188.

[116] N. E. Hafsa, D. Arndt, D. S. Wishart, *Nucleic Acids Research* **2015**, *43*, W370–W377.

[117] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. Ulrich Hartl, *Molecular Chaperone Functions in Protein Folding and Proteostasis*, *Vol. 82*, PMID: 23746257, **2013**, pp. 323–355.

[118] C. B. Anfinsen, *Science* **1973**, *181*, 223–230.

[119] G. Zhang, Z. Ignatova, *Curr. Opin. Struct. Biol.* **2011**, *21*, 25–31.

[120] D. Cui, S. Ou, S. Patel, *Proteins: Struct. Funct. Bioinf.* **2014**, *82*.

[121] S. Mojtabavi, N. Samadi, M. A. Faramarzi, *Iran. J. Pharm. Res.* **2019**, *18*, 13–30.

[122] C. M. Dobson, A. Šali, M. Karplus, *Angew. Chem. Int. Ed.* **1998**, *37*, 868–893.

[123] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*.

[124] T. Eichner, S. E. Radford, *Mol. Cell* **2011**, *43*.

[125] A. C. Muntau, J. Leandro, M. Staudigl, F. Mayer, S. W. Gersting, *J. Inherit. Metab. Dis.* **2014**, *37*, 505–523.

[126] D. Whitley, S. P. Goldberg, W. D. Jordan, *J. Vasc. Surg.* **1999**, *29*.

[127] A. E. Rydeen, E. M. Brustad, G. J. Pielak, *Journal of the American Chemical Society* **2018**, *140*, 7441–7444.

[128] P. H. Yancey, *J. Exp. Biol.* **2005**, *208*, 2819–2830.

[129] S. H. Khan, N. Ahmad, F. Ahmad, R. Kumar, *IUBMB Life* **2010**, *62*.

[130] X. Fei, D. Yang, N. LaRonde-LeBlanc, G. H. Lorimer, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*.

[131] D. Balchin, M. Hayer-Hartl, F. U. Hartl, *Science* **2016**, *353*.

[132] E. P. Melo, N. Estrela, C. Lopes, A. C. Matias, E. Tavares, V. Ochoa-Mendes, *Curr. Protein Pept. Sci.* **2011**, *11*.

[133] I. Lidbury, J. C. Murrell, Y. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*.

[134] T. D. Linley, M. E. Gerringer, P. H. Yancey, J. C. Drazen, C. L. Weinstock, A. J. Jamieson, *Deep Sea Res. Part I: Oceanographic Research Papers* **2016**, *114*.

[135] M. E. Gerringer, T. D. Linley, A. J. Jamieson, E. Goetze, J. C. Drazen, *Zootaxa* **2017**, *4358*.

[136] P. H. Yancey, J. F. Siebenaller, *J. Exp. Biol.* **1999**, *202*.

[137] T. Y. Lin, S. N. Timasheff, *Biochemistry* **1994**, *33*.

[138] G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623–16633.

[139] M. Ufnal, A. Zadlo, R. Ostaszewski, *Nutrition* **2015**, *31*, 1317–1323.

[140] L. R. Singh, N. K. Poddar, T. A. Dar, S. Rahman, R. Kumar, F. Ahmad, *J. Iran. Chem. Soc.* **2011**, *8*.

[141] M. Auton, D. W. Bolen, J. Rösgen, *Proteins: Struct. Funct. Genet.* **2008**, *73*.

[142] D. R. Canchi, P. Jayasimha, D. C. Rau, G. I. Makhatadze, A. E. Garcia, *J. Phys. Chem. B* **2012**, *116*.

[143] L. Hua, R. Zhou, D. Thirumalai C, B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*.

[144] H. Kokubo, C. Y. Hu, B. M. Pettitt, *Journal of the American Chemical Society* **2011**, *133*.

[145] Y. T. Liao, A. C. Manson, M. R. DeLyser, W. G. Noid, P. S. Cremer, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*.

[146] J. Hunger, N. Ottosson, K. Mazur, M. Bonn, H. J. Bakker, *Phys. Chemistry Chemical Physics* **2015**, *17*.

[147] E. A. Oprzeska-Zingrebe, J. Smiatek, *J. Phys. Chem. B* **2019**, *123*.

[148] J. Ma, I. M. Pazos, F. Gai, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*.

[149] Y. Yang, Y. Mu, W. Li, *Phys. Chem. Chem. Phys.* **2016**, *18*.

[150] J. Dyson, P. Wright, *Chem. rev.* **2004**, *104*, 3607–22.

[151] J. W. Neidigh, R. M. Fesinmeyer, K. S. Prickett, N. H. Andersen, *Biochemistry* **2001**, *40*.

[152] B. Barua, N. Andersen in, *Vol. 8*, **2001**, pp. 221–226.

[153] J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen, *Nature* **2002**, *9*, 425–430.

[154] K. G. Hovagimyan, J. T. Gerig, *J. Phys. Chem. B* **2005**, *109*, 24142–24151.

[155] A. J. Pearson, Y. Yamamoto, *Encyclopedia of Reagents for Organic Synthesis* **2010**.

[156] T. Dwars, E. Paetzold, G. Oehme, *Angewandte Chemie - International Edition* **2005**, *44*.

[157] B. Fabry, *Chemie in unserer Zeit* **1991**, *25*.

[158] P. Anastas, N. Eghbali, *Chem. Soc. Rev.* **2010**, *39.*

[159] J. Engberts, S. Otto, *Diels-Alder Reactions in Micellar Media*, New York, NY: Marcel Dekker, **2001**.

[160] K. S. Mineev, K. D. Nadezhdin, *Nanotechnol. Rev.* **2017**, *6*, 15–32.

[161] D. E. Warschawski, A. A. Arnold, M. Beaugrand, A. Gravel, É. Chartrand, I. Marcotte, *Biochim. Biophys. Acta - Biomembranes* **2011**, *1808*, 1957–1974.

[162] P. Mukerjee, K.J.Mysels, *Critical micelle concentrations of aqueous surfactant systems*, Natl. Stand. Ref. Data. Ser., Washington, DC: US. Government Printing, **1971**.

[163] D. Blokhin, E. A. Filippova, V. V. Klochkov, *Appl. Magn. Reson.* **2014**, *45*, 715–721.

[164] X. Cui, S. Mao, M. Liu, H. Yuan, Y. Du, *Langmuir* **2008**, *24*, 10771–10775.

[165] J. N. Israelachvili, *Interactions of Biological Membranes and Structures*, **2011**, pp. 577–616.

[166] X. Cui, S. Mao, M. Liu, H. Yuan, Y. Du, *Langmuir* **2008**, *24*, 10771–10775.

[167] J. W. Steed, J. L. Atwood, *Supramolecular Chemistry: Second Edition*, **2009**, pp. 1–970.

[168] G. Clayden, W. Warren, N. Greeves, P. Wothers, *Organic Chemistry*, Oxford University Press, **2001**.

[169] I. Muthukrishnan, V. Sridharan, J. C. Menéndez, *Chem. Rev.* **2019**, *119*, 5057–5191.

[170] J. S. B. Forero, J. Jones Junior, F. M. da Silva, *Curr. Org. Syn.* **2016**, *13*, 157–175.

[171] D. R. Merchán Arenas, C. A. Martínez Bonilla, V. V. Kouznetsov, *Org. Biomol. Chem.* **2013**, *11*, 3655–3663.

[172] L. S. Povarov, B. Mikhaĭlov, *Russ. Chem. Bull.* **1963**, *12*, 871–871.

[173] L. S. Povarov, V. I. Grigos, B. M. Mikhailov, *Bulletin of the Academy of Sciences USSR Division of Chemical Science* **1963**, *12.*

[174] K. N. Houk, R. W. Strozier, *J. Am. Chem. Soc.* **1973**, *95.*

[175] V. V. Kouznetsov, *Tetrahedron* **2009**, *65*.

[176] J. A. Platts, K. Gkionis, *Physical Chemistry Chemical Physics* **2009**, *11*, 10331–10339.

[177] U. Pyell in *Capillary Electromigration Separation Methods*, **2018**.

[178] R. Consonni, L. Santomo, R. Tenni, R. Longhi, L. Zetta, *FEBS Letters* **1999**, *458*, 81.

List of Figures

1.1. 3D solution structure which was solved by NMR spectroscopy and hybrid distance geometry-dynamical simulated annealing (pdb-code: 2bds).[5] . . 4

2.1. Formation of a peptide bond between two amino acids through a condensation reaction.[12] 8

2.2. Schematic drawing of the definition of dihedral/torsion angles of a peptide backbone (left) with the definition of *cis*- (ω=0°) and *trans*-conformation (ω=180°) (R_{sc}: side chain of any amino acid but proline).[12] 8

2.3. Two frequently observed secondary structure elements: an α-helix (left) and β-sheet (right).(NMR derived structures are taken from the BMRB: α-helix: penetratin, DOI: doi:10.13018/BMR5290, pdb-code: 1kz2, pdb-code: 1KZ2 , β-sheet: part of the sequence doi:10.13018/BMR17375, pdb-code: 1axc) 9

2.4. NOE pattern for different secondary structure elements. Thickness of lines shows how likely this spacial contact is.[15] 10

2.5. Representation of a general ramachandran plot. 11

2.6. Illustration of the spatial quantisation for spin-half (left side of the sphere) and spin-one (right side of the sphere) nuclei. The values $-I$ to $+I$ for three states are given in integer steps for spin-one nuclei.[19] 13

2.7. Flow chart of structure elucidation process by NMR spectroscopic methods.[25] 14

2.8. Effekct of pH on exchanging protons with $\log K_{\text{inter}}$ is plotted as function of the pH value. The NH protons of the backbone (bb) show the minimum at pH 3.0.[26] 16

2.9. Illustrative representation of typical ^{1}H chemical shift ranges where protons from peptides or proteins resonate in a 1D spectrum.[26] 18

2.10. Strategy for spin system assignment through J-coupled nuclei. The arrow shows amide proton as a starting point and goes to adjacent α- and β-protons (R: side chain). On the right side TOCSY signals are displayed.[3] 19

2.11. Energy-level diagram of a two spin-system. ω with concomitant number (0-2) indicates transition probabilities along the arrows. Equilibrium state of population is shown as grey balls. 20

2.12. Left: NOE enhancement η plotted as function of dimensionless molecular tumbling $\omega_0\tau_c$ rate. Emphasised are the size and NOE intensities which are expected for molecular sizes. The box highlights the zero crossing of the NOE intensity for which another approach is necessary, shown on the right: ROE of homonuclear two-spin system as function of molecular tumbling in comparison to transient NOE curve.[23] 21

2.13. Heteronuclear enhancement η as function of the molecular tumbling rate τ_c.[15] 22

2.14. Connection of the spin systems is mainly perfored through the identification of NOE signals of adjacent H_i^N and H_{i-1}^α protons (R: side chain). On the right side the exemplary NOSY signals are displayed.[3] 23

2.15. Bar graph of chemical shift index of H^α-protons. Arrows and loops are indications for the kind of fold, β-turn and α-helices, respectively. y-axis: CSI value, x-axis: residue number.[35] 24

2.16. Schematic representation of saturation transfer difference (STD) NMR experiment. Colour gradient indicates saturation state of ligand from black: highly saturated to white: not saturated. Non-binding ligands are not detected in 1D proton spectrum as shown on right hand side. In the off-resonance experiment, selective saturation is set far from all signals (around 30.000 ppm) to detect blank spectrum with all involved components as shown right.[31] 25

2.17. Schematic representation of waterLOGSY experiment with 1D proton trace on right hand side. After selective saturation on water signal, magnetisation is transferred from bulk water to protein to ligand. Non-binders (circles) show opposite signal intensities to binders (triangles).[31] 26

2.18. waterLOGSY spectrum of BSA:Trp:Suc model sample (600 MHz, mixing time: 0.8 s, 298 K, pH 6.08, H_2O/D_2O 9:1, BSA:Trp:Suc 1:100:100). 27

2.19. Effect of the absence (a) and presence (b) of diffusion on the phase shift and signal intensity.[45] 28

2.20. Standard simulating annealing protocol of NMR restrained structure calculation software.[14] 29

1.1. *Cis*/*trans*-conformers of Pro–Xaa bonds with labeles on relevant dihedral angles. 33

1.2. Schematic representation of *C. difficiles* modes of action: adhesion (top) and motility (bottom). High c-di-GMP level causes an higher expression of CD2831 and CD3246 surface anchors by interaction with type II riboswitches. Low levels of bacterium messenger leads to an upregulation of toxins Tcd A and B as well as enhanced flagella biosynthesis for its motility. Missing interaction of c-di-GMP further causes higher release of *P*ro-*P*ro *e*ndo*p*eptidase-1 (abbr. PPEP-1), which has a unique specificity for Pro-Pro bonds.[62–65] 35

1.3. Representation of PPEP-1 with short hepta-peptide **1** (Ace–GVNPPVP–NH_2) bound to active side. All residues involved in recognition of the substrate are represented in cartoon for backbone atoms and side chains with sticks. Only domains important for protein-ligand complexion are emphasized. The enlargement (right) has all those amino acids labeled. The peptide is drawn in stick representation, colored by atom type (N: blue, O: red, C: green, H: white)(PDB-code: 5a0x).[68] 37

4.1. Structural formula of 2-aminobenzoyl (Abz) and 2,4-dinitrophenyl (Dnp). 44

4.2. ^{1}H NMR spectrum of peptides **4-6**. The regions where additional signal sets are detected are enlarged and marked with asterics (**a, b**: 600 MHz/**c**: 500 MHz, PBS (H_2O/D_2O, 9:1), 283 K. 45

4.3. 2D TOCSY details of the fingerprint region of peptide **4**. Identified spin-systems are emphasised with vertical dashed lines (600 MHz, 100 ms mixing time, 283 K). 46

4.4. 2D NOESY excerpts (black) of the fingerprint region of peptide **5** superimposed with related TOCSY spectra (red) excerpts. NOE cross-peaks of $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ are emphasised and labelled (500/600 MHz, 200-300 ms mixing time, 283 K). 47

4.5. Expected NOE contacts in Xaa-Pro *cis* (left) and *trans* (right) configured peptide bonds. 48

4.6. 2D NOESY excerpt of Pro–Pro bonds. Indicated are spacial close contacts, which clearly reveal an all-*trans* conformation for the ω-angles is present (500/600 MHz, 200-300 ms mixing time, 283 K). 49

4.7. 2D NOESY excerpt of amide-amide region of peptide **4** (600 MHz, 200-300 ms mixing time, 283 K). 50

4.8. Superimposition of 10 energy minimised 3D solution structures along backbone atoms (N, C_α, C, O) from residue 2 to 14 (terminal caps Abz and NH_2 groups were excluded in alignment process). 51

4.9. STD amplification factor as function of applied saturation times. In black are all build-up curve with an amplification factor larger than 1, in grey all other curves belonging to proline moieties. Amide signals and all other protons of the heptapeptide are left out for the sake of clarity (600 MHz, 298 K). 52

4.10. Epitope mapping graphic of the relevant part which is located within the binding pocket during interaction. 52

4.11. Results of the docking studies displayed with the Ligplot$^+$ application (green dashed lines: hydrogen bonds, red beams with semicircle: residues involved in hydrophobic contacts that do not belong to the ligand – black spheres of ligand with beams: atoms belonging to ligand which are the corresponding to hydrophobic contacts) 53

1.1. Three considered mechanisms of CPP translocation. Through formation of transitory structure (left), direct penetration (middle) and endocytosis pathway (right).[76] 62

4.1. 1D ^{1}H NMR spectra of the peptides (a) linear DKP3-sC18* and (b) cyclic DKP3-sC18*. The enlargements shows the region of amide and aromatic protons. (peptide concentration ca. 1.3 mM, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). 71

4.2. Excerpts of the fingerprint region of 2D TOCSY spectra of (a) cyclic DKP1-sC18* and (b) cyclic DKP3-sC18*. Diastereotopicity of the methylene protons of glycine and DKP3 are emphasised in a box (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 80-120 ms, 600 MHz). 72

4.3. Sequential assignment of linear DKP3-sC18* peptide. For the sake of clarity, labelling is only shown for NOE signals that do not overlap. Assignment of signals in severe overcrowded area are indicated (red: 2D NOESY, black: 2D TOCSY, 50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz). 73

4.4. Top: Expected intra-residual NOE (dashed lines) in the scaffolds, that is drawn in Haworth projection. Bottom: intra-residual NOE cross-signals of both DKP scaffolds within the aliphatic region of 2D NOESY spectra (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz). 74

4.5. Excerpt of 2D NOESY spectra of both cyclic peptides cyc DKP1-sC18* (left) and cyc DKP3-sC18* (right). Inter-residual NOE signals are labelled (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 200 ms, 600 MHz). 75

4.6. Superimposition of 20 energy minimised structures of cyclic **DKP3**-sC18* peptide calculated with NMR experimental data evaluated from spectra acquired at 283 K (residue colours: violett: charged, green: hydrophobic, yellow: non-polar, red: polar uncharged, blue: DKP scaffold). 77

4.7. Hydrogen bond interactions (yellow dashes) derived from NMR experimental data collected at 283 K. 77

4.8. (a) Backbone trace of energy lowest structure calculated based on NMR experimental data at 283 K. (b) Preferred intra-molecular hydrogen bond pattern (dotted lines) obtained through MC/SD simulations. Colour code of the dotted lines is indicated in the legend. 78

4.9. 1D ^1H NMR spectra of linear and cyclic peptides in presence of SDS micelles (a) lin DKP1-sC18*, (b) cyc DKP1-sC18* (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 500 MHz cryoprobe or 600 MHz). 79

4.10. DOSY NMR spectra of cyclic DKP1-sC18* in aqueous solution (black trace) in comparison to their presence in SDS micelles (gray trace). TSP is superimposed and taken as internal reference. 1D projection shows peptide together with micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). 79

4.11. 2D TOCSY spectrum of an excerpt of the finger print region of linear DKP3 peptidomimetic in presence of micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). 80

4.12. 2D NOESY of the fingerprint region of linear DKP3-sC18* peptide (middle). Two exemplary excerpts of inter-residual contacts in an α-helix are shown (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). 81

4.13. 2D NOESY amide-amide region of DKP1-sC18* peptide (middle). Two exemplary excerpts of inter-residual contacts in an α-helix are shown (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). 82

4.14. The helical wheel (left) of the sC18* peptide sequence together with a cartoon presentation of the α-helix (right) that was calculated derived from NMR spectroscopic restraints. The CSI (bottom) was generated based in H^N and H^α chemical shift of the DKP1-sC18* peptide.[116] 83

4.15. Backbone superimposition of 10 energy-minimised structures derived from experimental data (a) and an illustration of location of charged side chains of arginine and lysine residues (violet)(b). 83

5.1. Charged side chains of cyclic DKP-sC18* peptidomimetics are mostly arranged on one side of the peptide providing a highly efficient interaction with the anionic surface of a micelle. 85

1.1. Energy landscape of protein folding into its native structure (left) and into misfolded structural elements (right). Also shown are kinetically favourable states in which the protein can be trapped on its folding pathway. Chaperones can assist to overcome free energy barriers so that a transition to more energetically favourable conformations can take place.[117] 90

1.2. Schematic drawing of the protein biosynthesis pathway with possibly involved chemical or molecular chaperones (exemplary shown with GroEL complex).[130,131] 93

1.3. Protein folding and unfolding induced by TMAO and urea, respectively. TMAO is excluded from the hydration layer (left) and keeps the protein in its native form, whereas urea interacts more tightly with the protein backbone to cause an unfold of the protein.[128,139] 94

4.1. ^{1}H NMR spectrum of the Trp-cage mini-protein (peptide concentration ca. 1.3 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz) 101

4.2. 2D TOCSY NMR spectrum of the Trp-cage mini-protein of the aliphatic region (peptide concentration ca. 1.3 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz) 102

4.3. 2D NOESY NMR excerpts of the Trp-cage with a) correlation from aromatic $\epsilon 1$ protons and b) correlation of aromatic $\eta 2$ and $\zeta 2$ protons to proline residues of the hydrophobic pocket. Nomenclature of the Trp residue is adopted from the CYANA software standard library (peptide concentration ca. 1.0 mM, PBS, pH 7.0, H_2O/D_2O 9:1, 280 K, 499 MHz) 103

4.4. Ramachandran diagram of the Trp-cage mini-protein, derived from NMR experimental data. 103

4.5. 20 energy minimised structures calculated based on NMR experimental restraints. Red: Trp side chain, green: residues that form the hydrophobic pocket. 104

4.6. 1D NOE spectra at different mixing times of TMAO and cGV. Irradiation was applied by a selective pulse on the proton resonance of TMAO. The 1D proton spectrum shows the dipeptide in water (H_2O/D_2O 9:1, 298 K, 400 MHz). 105

4.7. ^{1}H^{19}F 1D HOE NMR spectra of cGV with HIFP together with the full 1D proton spectrum of cGV. Irradiation was set on the flourine signal of HIPF (^{19}F: 76.33 ppm). The mixing time for each 1D spectrum is indicated in the respective spectrum (H_2O/D_2O 9:1, 298 K, 400 MHz). 106

4.8. ^{1}H^{19}F 2D HOESY NMR spectrum of cGV with HIFP (H_2O/D_2O 9:1, 298 K, 400 MHz). 106

4.9. ^{15}N^{1}H 2D HOESY NMR spectrum of labelled urea in presence of methanol (DMSO-d_6, 298 K, 400 MHz). 107

5.1. Visual representation of the hetero-nuclear H–N NOE between the Trp-cage mini-protein and isotopically labelled TMAO (arrow: envisioned HOE, O: red, N: blue, C: grey, H, white). 109

1.1. Membrane mimetic types that are used in BioNMR spectroscopy.[160] . . . 114

1.2. Simplified illustration of a SDS micelle with emphasise on double layer, consisting of a positive charged surface and the contrary charged *Stern* layer. The diffusive layer constitutes the area where the charge decreases from high (double layer) to low (bulk solvent).[165,166] 115

1.3. Molecule orbital presentation of the normal electron demand Diels-Alder reaction (left side) and the reverse version (right side). How good orbitals of HOMO and LUMO overlap is determined through the size of the energy gap, which is larger (gray dashed arrow) for the reverse reaction in opposite to the normal one (black dashed arrow). The symmetry of the orbitals during a reaction is emphasised on the outside for both cases.[168] 116

4.1. 1D NMR trace of SDS at different concentrations: below (3 mM), at (8.2 mM) and above cmc (30 mM). Reference is an external capillary with acetone in water (H_2O/D_2O 9:1, 298 K, 600 MHz). 123

4.2. 2D DOSY spectra of three peptide concentrations. Lowest (3 mM) and highest (30 mM) are shown in black traces. Diffusion coefficients are written on related dotted dashed lines. 8.2 mM diffusion dimension in displayed in grey. All spectra are referenced with an external acetone/water capillary, which was set to a diffusion coefficient of –8.1 $\log(m^2/s)$ (H_2O/D_2O 9:1, 600 MHz, 298 K, NS 32, 64k (f1) 32 (F2), linear ramp (2–98%), d1 1.5 s, d20 70–120 ms). 125

4.3. Solubility of the four compounds under investigation. 126

4.4. The change in the chemical shift of the SDS micelle signals among addition of a small organic molecule (30 mM SDS, 50 mM additive, H_2O/D_2O 9:1, 600 MHz, 298 K. All related ^{1}H NMR spectra are included in the appendix).126

4.5. 1D NOESY spectra of dioxane in presence of SDS micelles. The selective irradiated signal in indicated by a flash H_2O/D_2O 9:1, 500 MHz, mixing time: 0.9 sec, 298 K). 129

4.6. waterLOGSY (bottom) and 1D proton (top) spectrum of the sample including cyclohexane and micelles (H_2O/D_2O 9:1, 500 MHz, mixing time: 0.8 sec, 298 K). 130

4.7. 1D proton spectra of the $MnCl_2$ titration experiment of the sample including dioxane and SDS micelles. $MnCl_2$ concentration is indicated for every spectrum (H_2O/D_2O 9:1, 500 MHz, 298 K). 131

4.8. 1D proton spectra of substrate **2** at pH 1 and 7. At pH 7 (bottom trace), the amide form (left) is predominant. At pH 1, the amide form begins to rearrange to the imine form (right). At t=0h (middle trace), H3a signals of both forms are detected, whereas at t=6h (top trace) conversion to imine form is passed almost completely (H_2O/D_2O 9:1, 298 K, 600 MHz). 133

4.9. Proposed reaction mechanism by Kouznetsov *et al.*.[171] 134

4.10. Samples taken from reaction carried out in the flask after time steps 1h to 7h (left to right). 135

4.11. 2d TOCSY spectra of both starting materials **1** (green/blue) and **2** (purple/violet) as well as the crude reaction product (black/grey). The phase colours are described in the legend (H_2O/D_2O 9:1, pH 1, 600 MHz, 298 K). 136

4.12. Superimposed phase sensitive HSQC spectra of both starting materials **1** (green/blue) and **2** (purple/violet) as well as the crude reaction product (black/grey) (H_2O/D_2O 9:1, pH 1, 600 MHz, 298 K). 136

5.1. The waterLOGSY experiment could be used in the context of this work to test the exposure to water of small components in the presence of micelles. 137

C.1. Amino acids occurring in this thesis. xxxix

D.1. Superimposition of 10 energy minimized 3D solution structures along backbone atoms (N, Cα, C, O) from residue 2 to 14 (proline: red, terminal caps Abz and NH_2 groups were excluded in alignment process). left: Abz-PP-pep, middle: Abz-AP-pep, right: Abz-PA-pep xliv

D.2. 2D TOCSY details of the fingerprint region of all three peptides **4-6** (600 MHz, 100 ms mixing time, 283 K). xlv

D.3. 2D NOESY excerpts (black) of the fingerprint region superimposed with related TOCSY spectra (gray) excerpts. NOE cross-peaks of $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ are emphasised and labelled (500/600 MHz, 200-300 ms mixing time, 283 K).xlvi

D.4. 2D NOESY excerpts of Xaa–Pro and Pro–Pro bonds. Indicated are spacial close contacts, which clearly reveal an all-*trans* conformation for all three ω-angles (500/600 MHz, 200-300 ms mixing time, 283 K). xlvii

E.1. 3D solution structures of DKP1-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K. lxi

E.2. 3D solution structures of DKP3-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K. lxii

E.3. 3D solution structures of cyclic DKP1-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K. lxiii

E.4. 3D solution structures of cyclic DKP3-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K. lxiv

E.5. a. 3D solution structure of linear DKP1-sC18* peptide in presence of SDS-micelles; b. Statistics for 20 energy minimised structures computed with CYANA software; c. sequencial plot of short-, medium- and long-range NOE connectivities used in 3D structure calculation. lxv

E.6. a. 3D solution structure of linear DKP3-sC18* peptide in presence of SDS-micelles; b. Statistics for 20 energy minimised structures computed with CYANA software; c. sequencial plot of short-, medium- and long-range NOE connectivities used in 3D structure calculation. lxvi

E.7. (a) Backbone trace of energy lowest structure calculated based on NMR experimental data at 283 K. (b) Preferred intramolecular hydrogen bond pattern gained through MC/SD simulations. Color code in legend are related to dotted lines in wedge formula and give the propensity of the occurrence of hydrogen bonds. lxvii

E.8. Representative structures of the mostly populated clusters as obtained by MC/SD calculations after energy minimisation of cyclic **DKP1**-sC18* and **DKP3**-sC18* peptides. dark grey: O, N, C backbone atoms, white: H atoms; H atoms bound to carbon as well as side chains are omitted for sake of clarity. lxvii

E.9. 1D ^{1}H NMR spectra of the peptides (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18*, (c) cyclic **DKP1**-sC18* and (d) cyclic **DKP3**-sC18* in solution. On the right hand side, a enlargements of labeled protons are shown, partially with coupling constants which allow an estimation about dihedral angles (peptide concentration ca. 1.3 mM, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). lxix

E.10. Excerpts of fingerprint region (amide-aliphatic region) of 2D TOCSY spectra of (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18*, (c) cyclic **DKP1**-sC18* and (d) cyclic **DKP3**-sC18* peptide. The box in (d) shows the diastereotopic behaviour of methylene protons belonging to glycine and the scaffold itself. For sake of clarity, only H^N-H^α chemical shifts are indicated (F: phenylalanine, G: glycine, K: lysine, L: leucine, N: asparagine, R: arginine) (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 80-120 ms, 600 MHz). lxx

E.11. Amide-amide region of the 2D NOESY spectrum of linear DKP3-sC18*. Inter-residual NOE signals are labelled (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz). lxxi

E.12. 1D ^{1}H NMR spectra of linear and cyclic peptides in presence of SDS micelles (a) lin **DKP1**-sC18*, (b) lin **DKP3**-sC18*, (c) cyc **DKP1**-sC18* and (d) cyc **DKP3**-sC18* (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 500 MHz cryoprobe or 600 MHz). . . lxxi

E.13. DOSY NMR spectra of cyclic DKP3-sC18* free in solution (black trace) in comparison to their presence in SDS micelles (gray trace). TSP is superimposed and taken as internal reference. 1D projection shows peptide together with micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz). lxxii

E.14. Fingerprint excerpt of 2D TOCSY spectra of (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18* and (c) cyclic **DKP1**-sC18* peptides in presence of SDS micelles. (peptide to micelle concentration: around 1:1, PBS, pH 6.08 H_2O/D_2O, 298 K, mixing time 80-120 ms, 600 MHz). lxxiii

E.15. Amide-amide regions of 2D NOESY spectra of both linear peptides (a) and (b) and (c) cyclic DKP1-sC18* peptide in presence of SDS micelles. All spacial close backbone amide signals are labelled (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, mixing time 200 ms, 600 MHz). lxxiv

F.1. H-N HMQC spectrum of synthesised isotopic labelled TMAO with TMA as external reference (H_2O/D_2O 9:1, 298 K, 400 MHz). lxxx

F.2. ESI spectrum of ^{15}N-labelled TMAO. lxxx

G.1. 1D proton spectra of SDS below cmc with and without dioxane as additive (top box) and SDS above cmc with and without dioxane (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxii

G.2. 1D proton spectra of SDS below cmc with an without phenol as additive (top box) and SDS above cmc with and without phenol (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxiii

G.3. 1D proton spectra of SDS underneath cmc with an without toluene as additive (top box) and SDS above cmc with and without phenol (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxiv

G.4. 1D proton spectra of SDS below cmc in the absence of SDS (top) and in presence of 3 mM SDS immediately after addition of toluene (middle trace) and after 1 one day (bottom trace) (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxv

G.5. 1D proton spectra of cyclohexane in water acquired immediately after addition (top trace), after 1 hour (middle trace) and after 24 hours (bottom trace) (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxvi

G.6. 1D proton spectra of SDS below (left) and above (right) cmc with cyclohexane. The signals of cyclohexane are emphasised in a box (H_2O/D_2O 9:1, 600 MHz, 298 K). lxxxvi

G.7. Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with dioxane (H_2O/D_2O 9:1, 600 MHz, 298 K, d20: 150–200 ms). lxxxvii

G.8. Superimposition of 2D DOSY spectra of SDS below cmc with toluene. Measurements were performed immediately after addition of toluene (t=0) and after one day (t=1d) (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms). lxxxvii

G.9. Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with toluene (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms). lxxxviii

G.10. Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with phenol (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms). lxxxviii

G.11. Superimposition of 2D DOSY spectra of SDS under (grey trace) and above (black trace) cmc with cyclohexane (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 70-120 ms). lxxxix

G.12. 1D NOESY spectra of dioxane in presence of SDS micelles at various mixing times. Mixing times are indicated for 1D traces (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)). lxxxix

G.13. 1D NOESY spectra of phenol in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation points (sorted descending): aromatic protons of phenol, SDS signal *B* and and *D* (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)). xc

G.14.1D NOESY spectra of toluene in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation points (sorted descending): methyl group of phenol and SDS signal *D* (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)). xci

G.15.1D NOESY spectrum of cyclohexane in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation was performed on the SDS signal *D* (H_2O/D_2O 9:1, 600 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)).xcii

G.16.WaterLOGSY spectra of dioxane in water (middle) and dioxane in micelle solution (bottom) (600 MHz, 298 K, mixing time: 0.8 sec). xcii

G.17.WaterLOGSY spectra of phenole in water (middle) and in micelle solution (bottom) (600 MHz, 298 K, mixing time: 1.2 sec). xciii

G.18.WaterLOGSY spectra of toluene in water (middle) and in micelle solution (bottom) (600 MHz, 298 K, mixing time: 0.8 sec). xciii

G.19.1D ^{1}H NMR spectrum of NVF in water at pH 1 (H_2O/D_2O 9:1, 298 K, 600 MHz). xciv

G.20.Phase sensitive HSQC experiment of substrate 2 at pH 1 and pH 7 in stacked layout. Crossed-out signals belong to micelles. CH/CH_3 show opposite sign to CH_2. Substrate 2 at pH 1 is plotted in darker grey for CH/CH_3 and black for CH_2; for pH 7 CH/CH_3 is light grey and CH_2 in black (H_2O/D_2O 9:1, 298 K, 600 MHz). xciv

G.21.2D DOSY spectra of NVF in pH 1 (black) and pH 7 (grey) solution in presence of 12 mM micelles. All spectra are referenced with an external acetone/water capillary, which was set to a diffusion coefficient of –8.1 $\log(m^2/s)$ (H_2O/D_2O 9:1, 600 MHz, 298 K, NS 32, 8k (f1) 32 (F2), linear ramp (2–98%), d1 1.5 s, d20 70 ms). xcv

List of Tables

2.1. Properties selected nuclei.[23] 22

2.1. Investigated peptide sequences. Differing residues are outlined in a box. All sequences were synthesized and kindly provided by the *Baumann*-group (Biochemical Insitute of the University of Cologne) 40

4.1. Peptide sequences that were extensively investigated previously. They were kindly provided by the *Baumann*-group (Biochemical Insitute of the University of Cologne) 43

4.2. Investigated peptide sequences. The motif of interest is outlined in a box. 44

2.1. Peptide sequences including amino acid numbering studied in the scope of this work. 65

4.1. Temperature coefficient $\frac{\Delta\delta}{\Delta T}\left[\frac{ppb}{K}\right]$ of amide protons of all four peptide sequences determined for an experimental temperature range of 283 K–298 K. Values which might indicate an involvement in an hydrogen bond are emphasised in boxes (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 600 MHz). 76

4.1. Experimentally determined diffusion coefficients. Values are relative to the external standard acetone/water which was set to -8.1 $log(^{m^2}/_{s})$. Coefficients belong always to bold emphasized compounds. D values are all in $[\cdot 10^{-9}\ ^{m^2}/_{s}]$ (ΔD: relative deviation in reference to pure component, CMC: critical micellar concentration, Dxn: dioxane, Tln: toluene, cyH: cyclohexane). 128

D.1. ^1H chemical shifts for long proline-rich **peptides 4-6** in 50 mM phosphate buffer. Dinitrophenyl group attached to Lys^{12} side chain is not included except of bridging H^N proton (PP/AP: 600 MHz, PA: 500 MHz, 283 K, pH 7.0, H_2O/D_2O, d_4-TSP ^1H $\delta = 0$ ppm). xlii

D.2. Statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at 283 K. xliii

D.3. Hydrogen bonds in 20 energy minimised structures calculated based on NMR experimental data with CYANA-software. Cut-off for hydrogen bond length was chosen as 1Å. xliii

E.1. Overview of calculated amounts and ratios of peptide vs micelle in solution. An average of 60 SDS per micelle was chosen. xlix

E.2. ^{1}H chemical shifts of *linear* ***DKP1***-*sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). li

E.3. ^{1}H chemical shifts of *linear* ***DKP1***-*sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lii

E.4. ^{1}H chemical shifts for *linear* ***DKP3***-*sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). liii

E.5. ^{1}H chemical shifts of *linear* ***DKP3***-*sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). liv

E.6. ^{1}H chemical shifts of *cyclic* ***DKP1***-*sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lv

E.7. ^{1}H chemical shifts of *cyclic* ***DKP1***-*sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lvi

E.8. ^{1}H chemical shifts of *cyclic* ***DKP3***-*sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lvii

E.9. ^{1}H chemical shifts of *cyclic* ***DKP3***-*sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lviii

E.10. ^{1}H chemical shifts of *linear* ***DKP1***-*sC18** (left), *linear* ***DKP3***-*sC18** (middle) and *cyclic* ***DKP1***-*sC18** (right) in presence of *SDS micelles* in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lx

F.1. ^{1}H chemical shifts of Trp-cage mini-protein in 50 mM phosphate buffer at *280 K* (600 MHz, pH 7.0, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lxxv

F.2. ^{1}H chemical shifts of Trp-cage mini-protein in 50 mM phosphate buffer at *298 K* (600 MHz, pH 7.0, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm). lxxvii

F.3. Statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at 280 K. lxxix
F.4. Hydrogen bonds in 20 energy minimised structures calculated based on NMR experimental data with CYANA-software. Cut-off for hydrogen bond length was chosen as 1Å. lxxix

G.1. Overview of the chemical shift change of the SDS signals under the addition of small organic molecules at an SDS concentration of 3 mM(arrows: shift direction, X: no shift). lxxxi
G.2. Overview of the change in diffusion rate of single SDS molecules at a concentration of 3 mM induced by the interaction with small molecules. The change in diffusion rate of the additive itself, compared to diffusion through water, is also shown (arrow up: increased coefficient, arrow down: decreased coefficient). lxxxi

B. Non-standard residue library files

Library file of the non-standard residue *Abz* for 3D structure calculations using CYANA software.

```
RESIDUE   abz      1   17    3   16
   1 OMEGA    0    0    0.0000   16   15   14    5    0
   1 C    C_BYL   0    0.0000    2.393     0.234     0.284     2    3    0    0    0
   2 O    O_BYL   0    0.0000    2.918     1.274     0.660     1    0    0    0    0
   3 N    N_AMI   0    0.0000    1.113     0.195    -0.190     1    4    5    0    0
   4 H    H_AMI   0    0.0000    0.705     1.115    -0.282     3    0    0    0    0
   5 C2   C_VIN   0    0.0000    0.467    -0.903    -0.798     5    6   14    0    0
   6 CE1  C_ARO   0    0.0000   -0.446    -0.546    -1.807     5    8    7    0    0
   7 HE1  H_ARO   0    0.0000   -0.655     0.501    -2.017     6    0    0    0    0
   8 CD1  C_ARO   0    0.0000   -1.052    -1.514    -2.606     6   10    9    0    0
   9 HD1  H_ARO   0    0.0000   -1.726    -1.214    -3.406     8    0    0    0    0
  10 CD2  C_ARO   0    0.0000   -0.785    -2.858    -2.386     8   12   11    0    0
  11 HD2  H_ARO   0    0.0000   -1.255    -3.615    -3.010    10    0    0    0    0
  12 CE2  C_ARO   0    0.0000    0.053    -3.237    -1.335    10   14   13    0    0
  13 HE2  H_ARO   0    0.0000    0.160    -4.300    -1.140    12    0    0    0    0
  14 C3   C_VIN   0    0.0000    0.684    -2.272    -0.524    12    5   15    0    0
  15 C    C_BYL   0    0.0000    1.406    -2.738     0.706    14   16   17    0    0
  16 O    O_BYL   0    0.0000    1.437    -2.033     1.708    15    0    0    0    0
  17 N    N_AMI   0    0.0000    2.011    -3.977     0.663    15    0    0    0    0
```

Library file of the non-standard residue *Lys-Dnp* for 3D structure calculations using CYANA software.

```
RESIDUE   LYK     11   42    3   41
   1 OMEGA    0    0    0.0000    2    1    3    4    0
   2 PHI      0    0    0.0000    1    3    5   28    0
   3 CHI1     0    0    0.0000    3    5    7   11   34
   4 CHI2     0    0    0.0000    5    7   11   15   34
   5 CHI3     0    0    0.0000    7   11   15   19   34
   6 CHI4     0    0    0.0000   11   15   19   23   34
   7 CHI5     0    0    0.0000   15   19   23   25   34
   8 CHI6     0    0    0.0000   19   23   25   26   34
   9 CHI7     0    0    0.0000   23   25   26   27   28
  10 CHI8     0    0    0.0000   30   32   33   34   35
  11 PSI      0    0    0.0000    3    5   40   42    0
   1 C    C_BYL   0    0.0000    3.251    -0.866    -0.491     2    3    0    0    0
   2 O    O_BYL   0   0.0000   4.456   -0.771   -0.670    1   0   0   0   0
   3 N    N_AMI   0   0.0000   2.469    0.174   -0.092    1   4   5   0   0
   4 H    H_AMI   0   0.0000   2.931    1.078   -0.015    3   0   0   0   0
   5 CA   C_ALI   0   0.0000   1.019    0.123    0.041    3   6   7  40   0
   6 HA   H_ALI   0   0.0000   0.653   -0.677   -0.613    5   0   0   0   0
   7 CB   C_ALI   0   0.0000   0.583   -0.116    1.492    5   8   9  11   0
   8 HB2  H_ALI   0   0.0000  -0.511   -0.047    1.552    7   0   0   0  10
   9 HB3  H_ALI   0   0.0000   0.978    0.680    2.137    7   0   0   0  10
  10 QB   PSEUD   0   0.0000   0.2335   0.3165   1.8445   0   0   0   0   0
  11 CG   C_ALI   0   0.0000   1.030   -1.476    2.037    7  12  13  15   0
  12 HG2  H_ALI   0   0.0000   2.123   -1.503    2.117   11   0   0   0  14
  13 HG3  H_ALI   0   0.0000   0.733   -2.268    1.340   11   0   0   0  14
  14 QG   PSEUD   0   0.0000   1.428   -1.8855   1.7285   0   0   0   0   0
  15 CD   C_ALI   0   0.0000   0.418   -1.756    3.411   11  16  17  19   0
  16 HD2  H_ALI   0   0.0000  -0.676   -1.763    3.327   15   0   0   0  18
  17 HD3  H_ALI   0   0.0000   0.676   -0.951    4.110   15   0   0   0  18
  18 QD   PSEUD   0   0.0000   0.000   -1.357    3.7185   0   0   0   0   0
  19 CE   C_ALI   0   0.0000   0.904   -3.094    3.971   15  20  21  23   0
  20 HE2  H_ALI   0   0.0000   1.983   -3.039    4.159   19   0   0   0  22
  21 HE3  H_ALI   0   0.0000   0.726   -3.889    3.236   19   0   0   0  22
  22 QE   PSEUD   0   0.0000   1.3545  -3.464    3.6975   0   0   0   0   0
  23 NH   N_AMO   0   0.0000   0.177   -3.391    5.211   19  24  25   0   0
  24 HN   H_AMI   0   0.0000  -0.837   -3.370    5.037   23   0   0   0   0
  25 C1   C_ARO   0   0.0000   0.542   -4.634    5.780   23  26  38   0   0
  26 C2   C_ARO   0   0.0000  -0.181   -5.836    5.613   25  27  30   0   0
  27 N3   N_AMO   0   0.0000  -1.426   -5.866    4.837   26  28  29   0   0
  28 O31  O_XXX   0   0.0000  -1.741   -6.941    4.307   27   0   0   0   0
  29 O32  O_XXX   0   0.0000  -2.082   -4.818    4.743   27   0   0   0   0
  30 C3   C_ARO   0   0.0000   0.276   -7.048    6.160   26  31  32   0   0
  31 H3   H_ALI   0   0.0000  -0.296   -7.965    6.024   30   0   0   0   0
  32 C4   C_ARO   0   0.0000   1.479   -7.062    6.875   30  33  36   0   0
  33 N4   N_AMO   0   0.0000   1.970   -8.321    7.452   32  34  35   0   0
  34 O41  O_XXX   0   0.0000   3.063   -8.301    8.033   33   0   0   0   0
  35 O42  O_XXX   0   0.0000   1.259   -9.327    7.320   33   0   0   0   0
  36 C5   C_ARO   0   0.0000   2.222   -5.889    7.041   32  37  38   0   0
  37 H5   H_ALI   0   0.0000   3.160   -5.883    7.593   36   0   0   0   0
  38 C6   C_ARO   0   0.0000   1.755   -4.692    6.489   36  39  25   0   0
  39 H6   H_ALI   0   0.0000   2.338   -3.782    6.628   38   0   0   0   0
  40 C    C_BYL   0   0.0000   0.439    1.457   -0.451    5  41  42   0   0
  41 O    O_BYL   0   0.0000   1.060    2.511   -0.461   40   0   0   0   0
  42 N    N_AMI   0   0.0000  -0.853    1.380   -0.848   40   0   0   0   0
```

Library file of the non-standard residue *DKP-1* for 3D structure calculations using CYANA software.

```
RESIDUE   DKP       4   44    3   43
    1 OMEGA     0    0    0.0000    2    1    3    4    0
    2 PHI       0    0    0.0000    1    3    5    9    0
    3 CHI1      0    0    0.0000    3    5    9   19   17
    4 CHI2      0    0    0.0000    5    9   19   17   15
    1 C    C_BYL    0    0.0000   -4.4958   -2.7346    1.3698    2    3    0    0    0
    2 O    O_BYL    0    0.0000   -5.5151   -3.0916    1.9460    1    0    0    0    0
    3 N    N_AMI    0    0.0000   -3.7260   -1.6693    1.7358    1    4    5    0    0
    4 H    H_AMI    0    0.0000   -3.0402   -1.2878    1.0819    3    0    0    0    0
    5 CB2  C_ALI    0    0.0000   -4.2404   -0.7948    2.7217    3    7    8    9    0
    6 QB2  PSEUD    0    0.0000   -4.8542   -0.7390    2.7919    0    0    0    0    0
    7 HB21 H_ALI    0    0.0000   -4.8729   -0.0693    2.1969    5    0    0    0    6
    8 HB22 H_ALI    0    0.0000   -4.8355   -1.4087    3.3869    5    0    0    0    6
    9 CA2  C_ALI    0    0.0000   -3.1942   -0.1223    3.5700    5   10   11   19    0
   10 HA2  H_ALI    0    0.0000   -2.5725   -0.8514    4.1020    9    0    0    0    0
   11 CO2  C_BYL    0    0.0000   -2.2095    0.5548    2.6210    9   12   13    0    0
   12 O2   O_BYL    0    0.0000   -1.8591    0.1179    1.5296   11    0    0    0    0
   13 N1   N_AMI    0    0.0000   -1.5864    1.6288    3.2086   11   14   15    0    0
   14 HN1  H_AMI    0    0.0000   -0.7742    1.9545    2.7016   13    0    0    0    0
   15 CA1  C_ALI    0    0.0000   -2.4832    2.5736    3.8592   13   16   17   20    0
   16 HA1  H_ALI    0    0.0000   -1.8739    3.4094    4.2130   15    0    0    0    0
   17 CO1  C_BYL    0    0.0000   -3.1677    1.8959    5.0239   15   18   19    0    0
   18 O1   O_BYL    0    0.0000   -3.3856    2.4940    6.0715   17    0    0    0    0
   19 N2   N_AMI    0    0.0000   -3.8248    0.7456    4.5958    9   17   24    0    0
   20 CB1  C_ALI    0    0.0000   -3.6878    3.0223    2.9933   15   22   23   42    0
   21 QB1  PSEUD    0    0.0000   -4.3172    2.9147    3.1163    0    0    0    0    0
   22 HB11 H_ALI    0    0.0000   -4.2823    2.1603    2.6587   20    0    0    0   21
   23 HB12 H_ALI    0    0.0000   -4.3521    3.6691    3.5740   20    0    0    0   21
   24 CG   C_ALI    0    0.0000   -5.0114    0.2854    5.3145   19   26   27   28    0
   25 QG   PSEUD    0    0.0000   -4.9739    0.0877    5.9254    0    0    0    0    0
   26 HG1  H_ALI    0    0.0000   -5.1708    0.9048    6.1990   24    0    0    0   25
   27 HG2  H_ALI    0    0.0000   -4.7770   -0.7295    5.6517   24    0    0    0   25
   28 CD   C_VIN    0    0.0000   -6.3040    0.3044    4.4959   24   29   31    0    0
   29 CE1  C_ARO    0    0.0000   -6.7612    1.4403    3.8086   28   30   34    0    0
   30 HE1  H_ARO    0    0.0000   -6.2119    2.3736    3.8429   29    0    0    0   33
   31 CE2  C_ARO    0    0.0000   -7.0715   -0.8707    4.3850   28   32   36    0    0
   32 HE2  H_ARO    0    0.0000   -6.7317   -1.7969    4.8457   31    0    0    0   33
   33 QE   PSEUD    0    0.0000   -6.4718    0.2884    4.3443    0    0    0    0   41
   34 CZ1  C_ARO    0    0.0000   -7.9315    1.4004    3.0368   28   35   39    0    0
```

```
  35 HZ1  H_ARO   0    0.0000   -8.2594    2.2847    2.4938   34    0    0    0   38
  36 CZ2  C_ARO   0    0.0000   -8.2423   -0.9093    3.6208   31   37   39    0    0
  37 HZ2  H_ARO   0    0.0000   -8.8000   -1.8386    3.5314   36    0    0    0   38
  38 QZ   PSEUD   0    0.0000   -8.5297    0.2231    3.0127    0    0    0    0   41
  39 CH   C_ARO   0    0.0000   -8.6803    0.2303    2.9564   34   36   40    0    0
  40 HH   H_ARO   0    0.0000   -9.5972    0.2025    2.3725   39    0    0    0    0
  41 QH   PSEUD   0    0.0000   -7.5008    0.2557    3.6785    0    0    0    0    0
  42 C    C_BYL   0    0.0000   -3.3850    3.7374    1.7345   20   43   44    0    0
  43 O    O_BYL   0    0.0000   -4.2378    3.8739    0.8602   42    0    0    0    0
  44 N    N_AMI   0    0.0000   -2.1092    4.2100    1.6499   42    0    0    0    0
```

Library file of the non-standard residue *DKP-3* for 3D structure calculations using CYANA software.

```
RESIDUE   DKP      4   44    3   43
   1 OMEGA    0    0    0.0000    2    1    3    4    0
   2 PHI      0    0    0.0000    1    3    5    9    0
   3 CHI1     0    0    0.0000    3    5    9   19   17
   4 CHI2     0    0    0.0000    5    9   19   17   15
   1 C    C_BYL   0    0.0000   -4.2166   -2.8466    0.9403    2    3    0    0    0
   2 O    O_BYL   0    0.0000   -5.4099   -2.8311    0.6738    1    0    0    0    0
   3 N    N_AMI   0    0.0000   -3.6071   -2.1882    1.9751    1    4    5    0    0
   4 H    H_AMI   0    0.0000   -2.5937   -2.0975    1.9318    3    0    0    0    0
   5 CB2  C_ALI   0    0.0000   -4.3293   -1.2192    2.7827    3    7    8    9    0
   6 QB2  PSEUD   0    0.0000   -4.9644   -1.3008    2.8744    0    0    0    0    0
   7 HB21 H_ALI   0    0.0000   -5.1334   -0.7876    2.1750    5    0    0    0    6
   8 HB22 H_ALI   0    0.0000   -4.7954   -1.8140    3.5738    5    0    0    0    6
   9 CA2  C_ALI   0    0.0000   -3.4779   -0.0801    3.4113    5   10   11   19    0
  10 HA2  H_ALI   0    0.0000   -2.7858   -0.5238    4.1350    9    0    0    0    0
  11 CO2  C_BYL   0    0.0000   -2.6207    0.5389    2.2804    9   12   13    0    0
  12 O2   O_BYL   0    0.0000   -1.9727   -0.1306    1.4716   11    0    0    0    0
  13 N1   N_AMI   0    0.0000   -2.6156    1.8984    2.1696   11   14   15    0    0
  14 HN1  H_AMI   0    0.0000   -1.9502    2.2316    1.4861   13    0    0    0    0
  15 CA1  C_ALI   0    0.0000   -2.8610    2.7442    3.3239   13   16   17   20    0
  16 HA1  H_ALI   0    0.0000   -3.3365    3.6239    2.8662   15    0    0    0    0
  17 CO1  C_BYL   0    0.0000   -3.9239    2.2250    4.3045   15   18   19    0    0
  18 O1   O_BYL  0    0.0000   -4.4354    2.9874    5.1302   17    0    0    0    0
  19 N2   N_AMI   0    0.0000   -4.3262    0.8965    4.1550    9   17   24    0    0
  20 CB1  C_ALI   0    0.0000   -1.6074    3.2668    4.0236   15   22   23   42    0
  21 QB1  PSEUD   0    0.0000   -1.4216    3.8840    3.9965    0    0    0    0    0
  22 HB11 H_ALI   0    0.0000   -1.9037    4.0610    4.7219   20    0    0    0   21
  23 HB12 H_ALI   0    0.0000   -0.9394    3.7070    3.2710   20    0    0    0   21
```

```
24 CG    C_ALI   0    0.0000   -5.2850    0.3847    5.1346   19   26   27   28    0
25 QG    PSEUD   0    0.0000   -5.1337    0.2403    5.7651    0    0    0    0    0
26 HG1   H_ALI   0    0.0000   -5.3585    1.0488    6.0041   24    0    0    0   25
27 HG2   H_ALI   0    0.0000   -4.9089   -0.5681    5.5261   24    0    0    0   25
28 CD    C_VIN   0    0.0000   -6.6731    0.2339    4.5488   24   29   31    0    0
29 CE1   C_ARO   0    0.0000   -7.3334    1.3392    3.9913   28   30   34    0    0
30 HE1   H_ARO   0    0.0000   -6.8552    2.3172    3.9963   29    0    0    0   33
31 CE2   C_ARO   0    0.0000   -7.3249   -1.0078    4.5511   28   32   36    0    0
32 HE2   H_ARO   0    0.0000   -6.8439   -1.8817    4.9842   31    0    0    0   33
33 QE    PSEUD   0    0.0000   -6.8496    0.2178    4.4903    0    0    0    0   41
34 CZ1   C_ARO   0    0.0000   -8.5965    1.1989    3.4155   28   35   39    0    0
35 HZ1   H_ARO   0    0.0000   -9.0861    2.0618    2.9733   34    0    0    0   38
36 CZ2   C_ARO   0    0.0000   -8.5862   -1.1486    3.9692   31   37   39    0    0
37 HZ2   H_ARO   0    0.0000   -9.0676   -2.1225    3.9505   36    0    0    0   38
38 QZ    PSEUD   0    0.0000   -9.0769   -0.0304    3.4619    0    0    0    0   41
39 CH    C_ARO   0    0.0000   -9.2193   -0.0468    3.3978   34   36   40    0    0
40 HH    H_ARO   0    0.0000  -10.1952   -0.1598    2.9349   39    0    0    0    0
41 QH    PSEUD   0    0.0000   -5.5704    2.0049    3.7826    0    0    0    0    0
42 C     C_BYL   0    0.0000   -0.8234    2.2250    4.8027   20   43   44    0    0
43 O     O_BYL   0    0.0000   -1.1363    1.0436    4.8293   42    0    0    0    0
44 N     N_AMI   0    0.0000    0.2268    2.7476    5.4275   42    0    0    0    0
```

C. Structure chart of selected amino acids

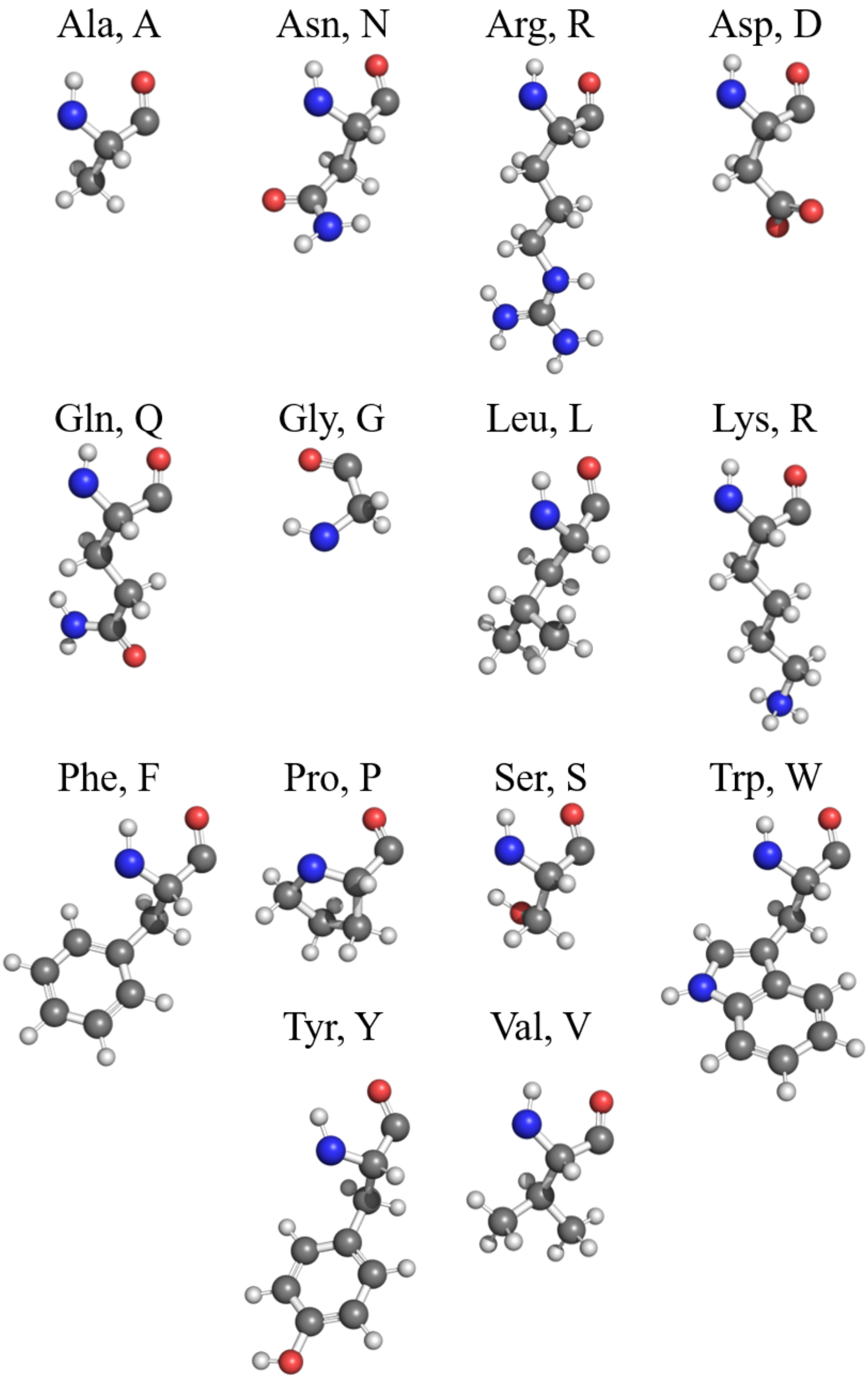

Ala, A
Asn, N
Arg, R
Asp, D
Gln, Q
Gly, G
Leu, L
Lys, R
Phe, F
Pro, P
Ser, S
Trp, W
Tyr, Y
Val, V

Figure C.1.: Amino acids occurring in this thesis.

D. Appendices Part I

D.1. Proton assignment of peptides 4-6

Table D.1.: ^{1}H chemical shifts for long proline-rich **peptides 4-6** in 50 mM phosphate buffer. Dinitrophenyl group attached to Lys^{12} side chain is not included except of bridging H^N proton (PP/AP: 600 MHz, PA: 500 MHz, 283 K, pH 7.0, H_2O/D_2O, d_4-TSP ^{1}H $\delta = 0$ ppm).

	$\boxed{P^6P^7}$ Peptide				$\boxed{A^6P^7}$ Peptide				$\boxed{P^6A^7}$ Peptide			
Res.	H^N	H^α	**other**		H^N	H^α	**other**		H^N	H^α	**other**	
	$^3J_{H^\alpha H^N}$ [Hz]				$^3J_{H^\alpha H^N}$ [Hz]				$^3J_{H^\alpha H^N}$ [Hz]			
Abz			H^{ar}	7.472, 7.342, 6.953, 6.877			H^{ar}	7.468, 7.342, 6.888, 6.847			H^{ar}	7.466, 7.338, 6.880, 6.851
Asp^1	8.633 (d, 6.60)	4.803	β	2.778, 2.660	8.637	4.793	β	2.804, 2.678	8.639 (d, 6.66)	4.779	β	2.775, 2.654
Val^2	8.331	4.168	β	2.073	8.348	4.166	β	2.087	8.352	4.154	β	2.080
			γ	0.951	(d, 8.31)		γ	0.955	(d, 5.57)		γ	0.940
Val^3	8.429	4.079	β	2.023	8.422	4.081	β	2.051	8.451	4.053	β	2.026
			γ	0.932			γ	0.941			γ	0.916
Asn^4	8.704 (d, 7.48)	5.004	β	2.791, 2.653	8.645	4.706	β	2.797, 2.685	8.741 (d, 7.37)	4.994	β	2.854, 2.682
			δ	7.742, 6.992			δ	7.753, 6.956			δ	7.783, 7.045
Pro^5 or Ala^5		4.677	β	2.333, 2.041	8.323	4.576	β	1.351		4.396	β	2.282, 2.033
			γ	1.924							γ	1.961
			δ	3.814, 3.703							δ	3.828, 3.756
Pro^6 or Ala^6		4.436	β	2.363, 2.018		4.444	β	2.284, 2.021	8.389	4.290	β	1.366
			γ	1.881			γ	1.859				
			δ	3.900, 3.650			δ	3.786, 3.631				
Val^7	8.352	4.374	β	2.033	8.423	4.364	β	2.042	8.158	4.387	β	2.032
			γ	0.961			γ	0.973	(d, 8.35)		γ	0.946
Pro^8		4.677	β	2.312, 2.020		4.693	β	2.356, 2.035		4.662	β	2.342, 2.021
			γ	1.898			γ	1.884			γ	1.880
			δ	3.902, 3.644			δ	3.908, 3.649			δ	3.880, 3.648
Pro^9		4.455	β	2.288		4.683	β	2.295, 2.023		4.456	β	2.312, 2.001
			γ	2.007			γ	1.892			γ	1.921
			δ	3.792, 3.641			δ	3.809, 3.657			δ	3.789, 3.641
Ser^{10}	8.499 (d, 6.67)	4.423	β	3.873	8.494 (d, 6.72)	4.442	β	3.872	8.498 (d, 6.53)	4.421	β	3.857
Lys^{11}	8.579	4.348	β	1.900	8.574	4.352	β	1.891	8.582	4.348	β	1.889
(dnp)	(d, 6.88)		γ	1.492	(d, 7.07)		γ	1.485	(d, 7.00)		γ	1.498
			δ	1.792			δ	1.770			δ	1.774
			ϵ	3.537			ϵ	3.534			ϵ	3.520
			H^N	8.871 (t, 6.09)			H^N	8.872 (t, 5.85)			H^N	8.873 (t, 5.95)
Asp^{12}	8.411	4.588	β	2.729, 2.592	8.413	4.589	β	2.722, 2.558	8.424	4.579	β	2.725, 2.582
Asp^{13}	8.305	4.529	β	2.670	8.307	4.538	β	2.674	8.316	4.525	β	2.650

D.2. Summary tables of NMR-derived structure calculations of proline-rich peptides 4-6

Table D.2.: Statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at 283 K.

	$\boxed{\mathbf{P^6P^7}}$	$\boxed{\mathbf{A^6P^7}}$	$\boxed{\mathbf{P^6A^7}}$
NOE upper distance limits			
Short-range $\lvert i-j\rvert \leq 1$	61	47	61
Medium range $\lvert i-j\rvert \leq 5$	0	6	1
Long-range $\lvert i-j\rvert \geq 5$	0	0	0
Violations > 0.2Å	0	0	0
CYANA target function (Å^2)	0.52	0.20	1.42
RMSD to mean coordinates (Å)			
Backbone N, Cα, CO	2.59 +/− 0.70	2.24 +/− 0.40	2.46 +/− 0.31
All heavy atoms of residue	3.56 +/− 0.69	3.18 +/− 0.52	3.71 +/− 0.59
Ramachandran statistics (%)			
Most favored	55.0	17.0	36.0
Additionally allowed	45.0	67.0	61.0
Generously allowed	0.0	6.0	0.0
Disallowed	0.0	1.0	3.0

*50 mM phosphate buffer, pH 6.08, H_2O/D_2O 9:1, 283 K

Table D.3.: Hydrogen bonds in 20 energy minimised structures calculated based on NMR experimental data with CYANA-software. Cut-off for hydrogen bond length was chosen as 1Å.

sequence No.		H ··· O		# of hydrogen bonds
4	$\boxed{\mathbf{P^5P^6}}$	N^4	V^2	3
		D^{10}	K^{11}	8
5	$\boxed{\mathbf{A^5P^6}}$	N^4	V^2	1
		A^5	V^3	3
6	$\boxed{\mathbf{P^5A^6}}$	V^7	P^5	1

D.3. NMR derived 3D structure ensembles

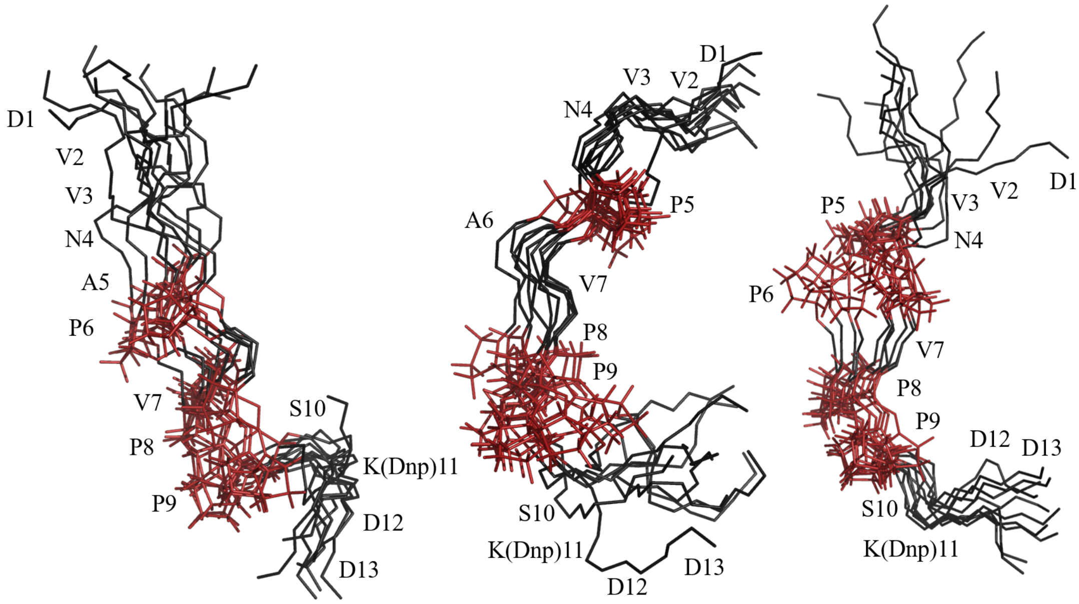

D1
V2
V3
N4
A5
P6
V7
P8
P9
S10
K(Dnp)11
D12
D13
D1
V3
V2
N4
P5
A6
V7
P8
P9
S10
K(Dnp)11
D12
D13
D1
V2
V3
P5
N4
P6
V7
P8
P9
D12
D13
S10
K(Dnp)11

Figure D.1.: Superimposition of 10 energy minimized 3D solution structures along backbone atoms (N, Cα, C, O) from residue 2 to 14 (proline: red, terminal caps Abz and NH_2 groups were excluded in alignment process). left: Abz-PP-pep, middle: Abz-AP-pep, right: Abz-PA-pep

D.4. Selected spectra of Proline-rich peptides 4-6

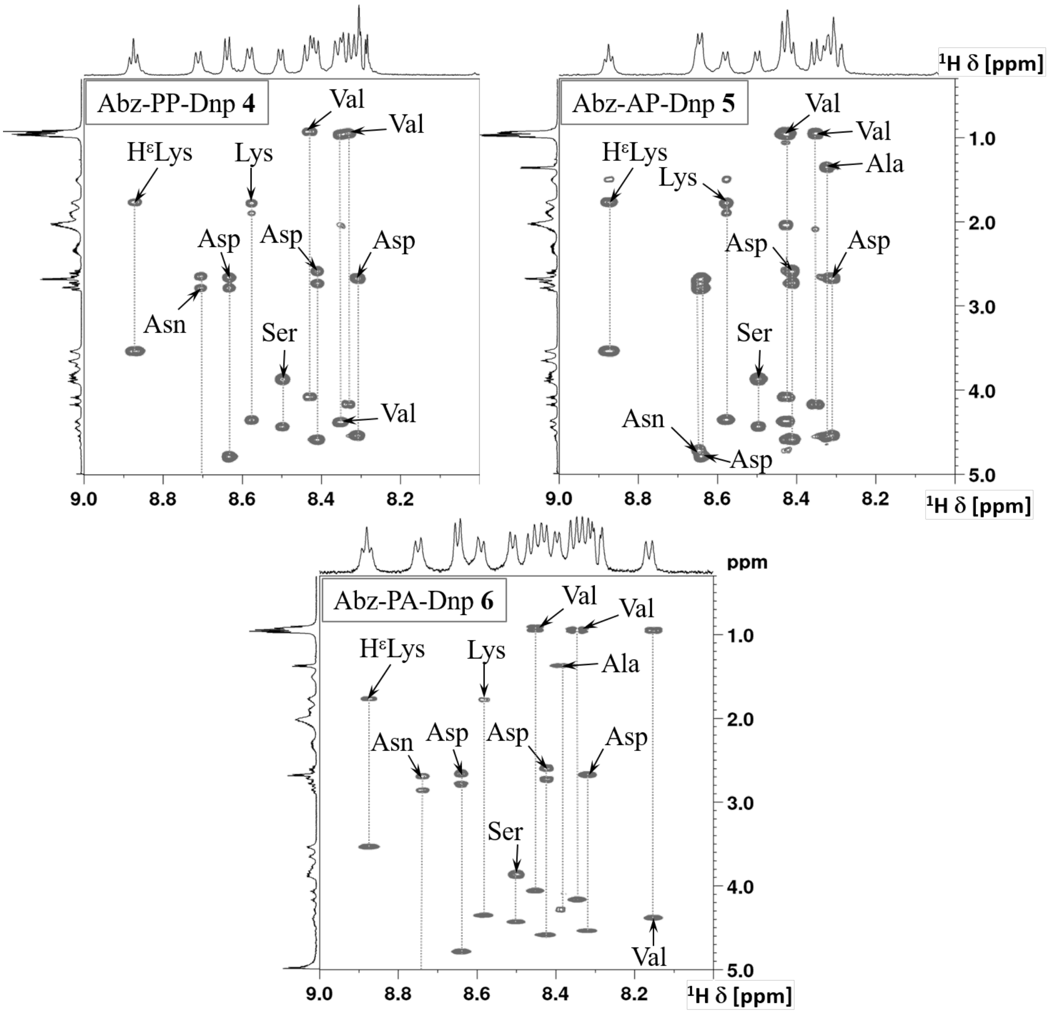


Figure D.2.: 2D TOCSY details of the fingerprint region of all three peptides **4-6** (600 MHz, 100 ms mixing time, 283 K).

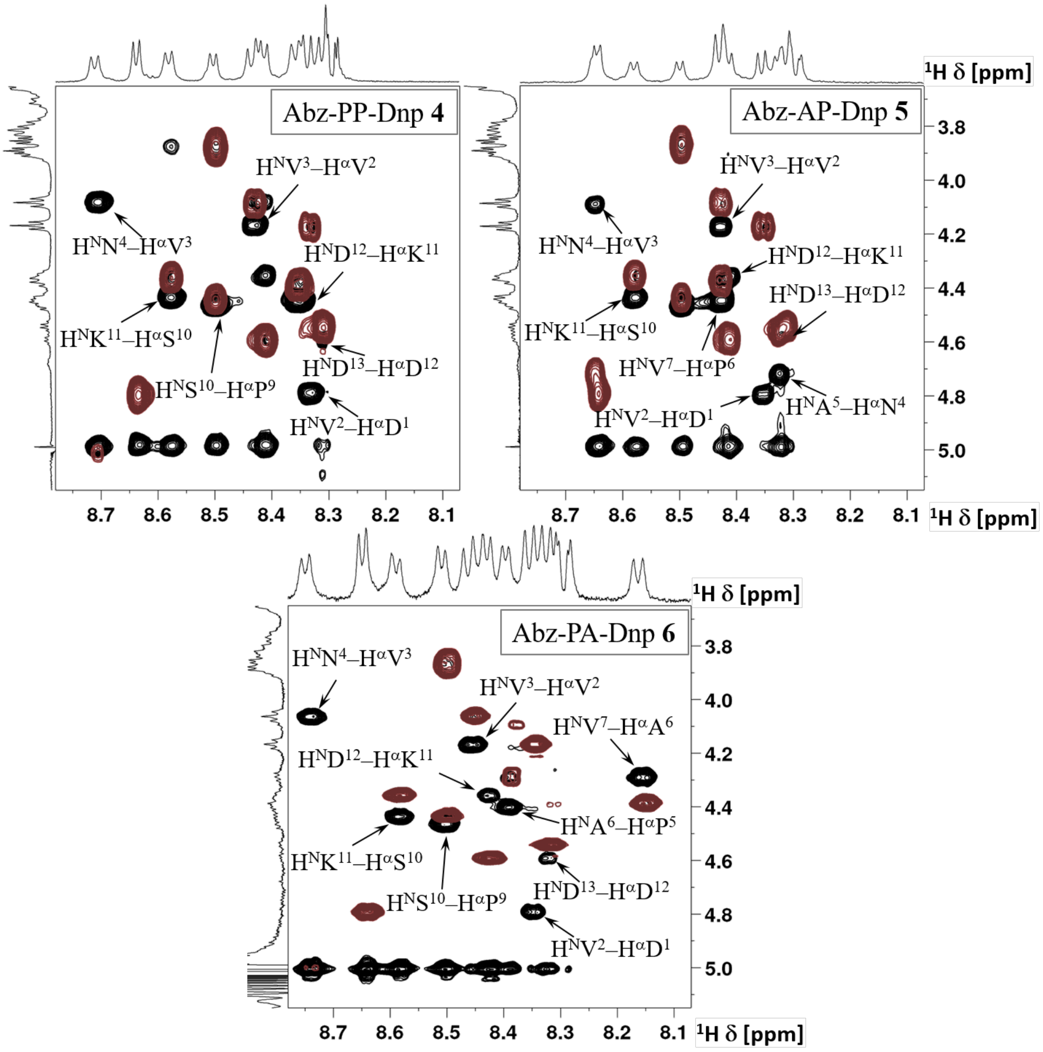


Figure D.3.: 2D NOESY excerpts (black) of the fingerprint region superimposed with related TOCSY spectra (gray) excerpts. NOE cross-peaks of $H^{\alpha}_{(i)}$–$H^{N}_{(i+1)}$ are emphasised and labelled (500/600 MHz, 200-300 ms mixing time, 283 K).

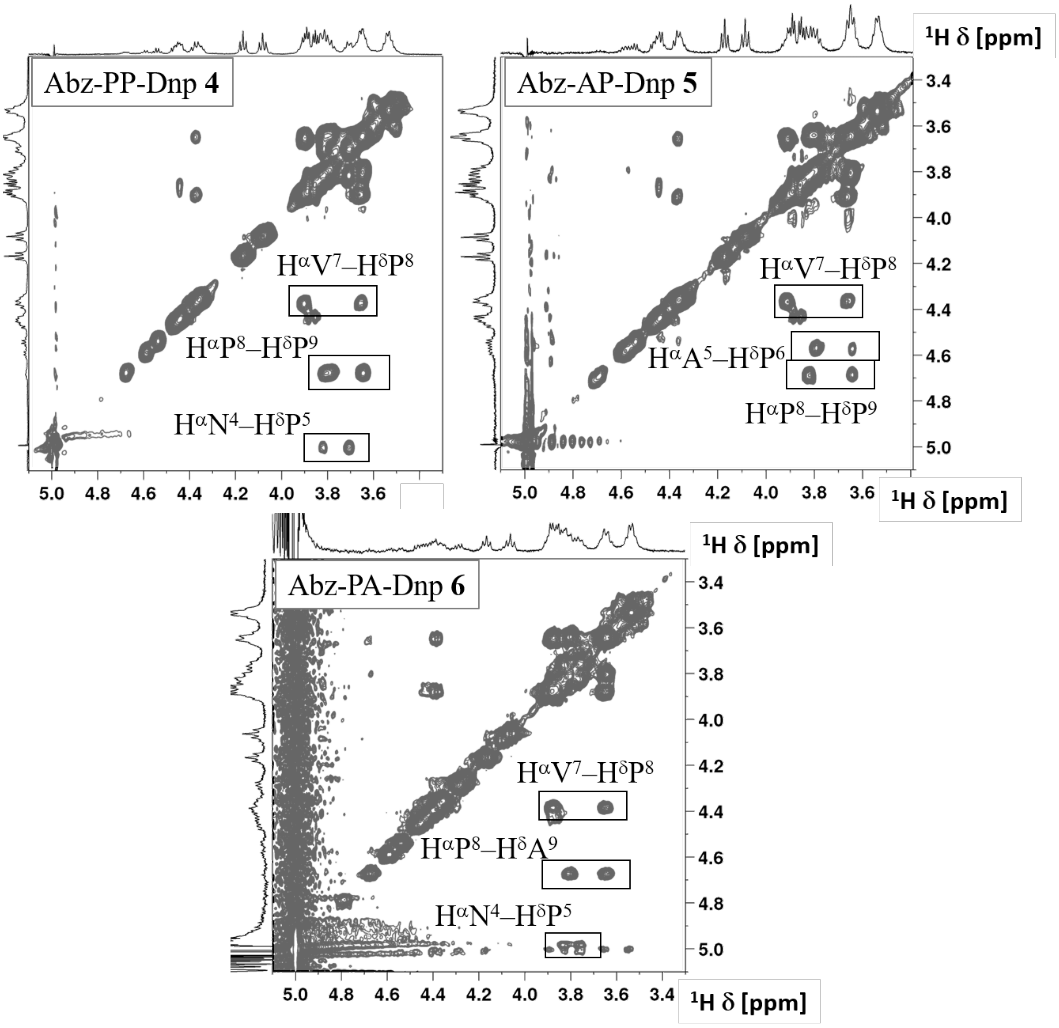


Figure D.4.: 2D NOESY excerpts of Xaa–Pro and Pro–Pro bonds. Indicated are spacial close contacts, which clearly reveal an all-*trans* conformation for all three ω-angles (500/600 MHz, 200-300 ms mixing time, 283 K).

E.1. Sample preparation of DKP peptidomimetics

Samples were always prepared with a freshly made SDS stock solution of 200 mM. During the sample preparation the cyclic peptides always precipitated immediately when getting in contact with SDS micelles. In contrast, samples of SDS with linear peptides showed no precipitate. Various sample preparation attempts and sample conditions were carried out. The variation of pH values, the ion pressure and an exchange of the phospharte buffer were taken into account. The final solution was an optimal ratio of around 1–3 peptide molecules per micelle. However, a slight precipitate was still observed, which lead to sample inhomogenities. Those inhomogeniteies hampered the spactra acquisition In Table E.1, all samples which were finally used for all investigations in terms of micelle interaction are displayed:

Table E.1.: Overview of calculated amounts and ratios of peptide vs micelle in solution. An average of 60 SDS per micelle was chosen.

	DKP1-sC18*		**DKP3-sC18***	
	linear	cylic	linear	cyclic
pH	6.07	6.01	6.07	6.01
Sample Volume [μl]	200	500	200	500
amount [mg]	1.1	1.4	1.6	1.4
concentration [mM]				
peptides	1.99	1.07	2.90	1.07
SDS	100	21.1	100	21.1
amount of molecules [$\cdot$ E(17) mol^{-1}]				
peptides	2.40	3.21	3.49	3.21
micelles	2.00	1.06	2.00	1.06
ratio peptide:micelle	1.2:1.0	3.0:1.0	0.9:1.0	3.0:1.0

All suitable samples show a peptide to micelle ratio of 1:1 to 3:1.

The initially prepared samples that have shown precipitated .

E.2. Proton assignment of DKP peptides

Table E.2.: ^{1}H chemical shifts of *linear* ***DKP1**-sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

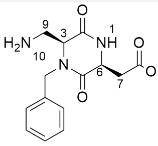
Residue	**^{1}H δ (ppm) 283 K**				
	H^{N}	**$^{3}J_{H^{\alpha}H^{N}}$** [Hz]	**H$^{\alpha}$**		**Others**
Gly^{1}	8.727	t, 6.15	4.020, 3.973		
Leu^{2}	8.331	d, 6.54	4.298	β	1.621
				γ	1.609
				δ	0.895, 0.858
Arg^{3}	8.483		4.286	β	1.832, 1.743
				γ	1.642
				δ	3.185
				ϵ	7.361
Lys^{4}	8.410	d, 6.80	4.253	β	1.781
				γ	1.447
				δ	1.701
				ϵ	2.980
Arg^{5}	8.505	d, 7.41	4.309	β	1.831, 1.782
				γ	1.646
				δ	3.188
				ϵ	7.377
Leu^{6}	8.467		4.353	β	1.657
				γ	1.586
				δ	0.948, 0.889
Arg^{7}	8.488		4.299	β	1.745
				γ	1.648
				δ	3.177
				ϵ	7.341
Lys^{8}	8.454		4.253	β	1.750, 1.689
				γ	1.283, 1.336
				δ	1.570
				ϵ	2.967
Phe^{9}	8.439		4.626	β	3.032, 3.132
				o 7.270, ***m*** 7.355, ***p*** 7.328	
Arg^{10}	8.371	d, 7.49	4.313	β	1.725, 1.824
				γ	1.576
				δ	3.182
				ϵ	7.320
Asn^{11}	8.558	d, 7.07	4.647	β	2.872, 2.774
				δ	7.766, 7.039
Lys^{12}	7.996	d, 7.73	4.150	β	1.725, 1.818
				γ	1.396
				δ	1.678
				ϵ	2.975
DKP1		 9 3 1 O H_2N 10 NH N 6 O OH 7 O	**H3** 4.445	**H7**	3.121, 2.949
			H6 4.570	**H9**	3.484, 3.439
				CH_2Ph	4.692, 4.855
				CH_2**Ph**	7.378, 7.433, 7.404

Table E.3.: ^{1}H chemical shifts of *linear* ***DKP1****-sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

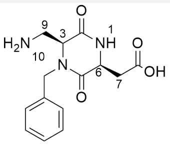
Residue	**^{1}H 298 K δ (ppm)**				
	H^N	**$^3J_{H^\alpha H^N}$** [Hz]	H^α	**Others**	
Gly^1	8.633		4.019, 3.955		
Leu^2	8.221	d, 6.73	4.298	β	1.606
				δ	0.865
Arg^3	8.321		4.286	β	1.817, 1.766
				γ	1.612
				δ	3.175
				ϵ	7.286
Lys^4	8.257		4.253	β	1.777
				γ	1.420
				δ	1.673
				ϵ	2.979
Arg^5	8.365		4.309	β	1.807, 1.735
				γ	1.602
				δ	3.182
				ϵ	7.296
Leu^6	8.312		4.353	β	1.623
				δ	0.922
Arg^7	8.342		4.299	β	1.735
				γ	1.587
				δ	3.175
				ϵ	7.278
Lys^8	8.309		4.253	β	1.660
				γ	1.319
				ϵ	2.961
Phe^9	8.299		4.626	β	3.132, 3.023
				o 7.270, ***m*** 7.365, ***p*** 7.313	
Arg^{10}	8.275		4.313	β	1.832, 1.735
				γ	1.586
				δ	3.183
				ϵ	7.274
Asn^{11}	8.459		4.647	β	2.843, 2.765
				δ	7.659, 6.947
Lys^{12}	7.877	d, 7.98	4.150	β	1.810
				γ	1.395
				δ	1.721
				ϵ	2.972
DKP1			**H3** 4.423	**H7**	3.092, 2.951
			H6 4.588	**H9**	3.471, 3.428
				CH_2Ph	7.359, 7.440, 7.409

Table E.4.: ^{1}H chemical shifts for *linear* ***DKP3**-sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

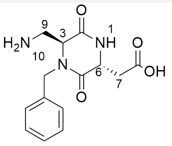
Residue	**^{1}H 283 K δ (ppm)**				
	H^N	$^3J_{H^\alpha H^N}$ [Hz]	**H^α**		**Others**
Gly^1	8.557		3.959		
Leu^2	8.296	d, 7.32	4.339	β	1.614
				δ	0.896
Arg^3	8.515		4.291	β	1.747
				γ	1.556
				δ	3.170
				ϵ	7.252
Lys^4	8.433		4.257	β	1.785
				γ	1.418
				δ	1.720
				ϵ	2.980
Arg^5	8.481		4.360	β	1.664
				γ	1.581
				δ	3.194
				ϵ	7.252
Leu^6	8.463		4.270	β	1.645
				γ	1.569
				δ	0.932, 0.851
Arg^7	8.519		4.294	β	1.801, 1.741
				γ	1.614
				δ	3.171
				ϵ	7.252
Lys^8	8.468		4.264	γ	1.349
				δ	1.660
				ϵ	3.188
Phe^9	8.444		4.628	β	3.126, 3.030
				o 7.270, ***m*** 7.355, ***p*** 7.328	
Arg^{10}	8.378		4.318	β	1.836, 1.731
				γ	1.574
				δ	3.199
				ϵ	7.255
Asn^{11}	8.564		4.658	β	2.868, 2.778
				δ	7.768, 7.044
Lys^{12}	8.007	d, 6.51	4.140	β	1.833, 1.727
				γ	1.400
				δ	1.673
				ϵ	2.861
DKP3		 O 9 3 1 H_2N NH O 10 N 6 OH 7 O	**H3** 4.276	**H7**	3.136, 2.959
			H6 4.683	**H9**	3.506
				CH_2Ph	5.153, 4.391

Table E.5.: ^{1}H chemical shifts of *linear* ***DKP3**-sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

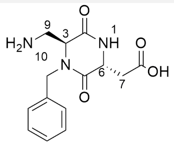
Residue	**^{1}H 298 K δ (ppm)**				
	H^N	**$^{3}J_{H^{\alpha}H^{N}}$ [Hz]**	**H^{α}**		**Others**
Gly^{1}	8.432		3.946		
Leu^{2}	8.184	d, 6.28	4.324	β	1.600
				δ	0.886
Arg^{3}	8.389		4.279	β	1.806, 1.731
				γ	1.597
				δ	3.149
				ϵ	7.252
Lys^{4}	8.277		4.365	β	1.813
				γ	1.418
				δ	1.732
				ϵ	2.983
Arg^{5}	8.382		4.299	β	1.801, 1.741
				γ	1.597
				δ	3.169
				ϵ	7.525
Leu^{6}	8.313		4.350	β	1.640
				δ	0.906
Arg^{7}	8.336		4.371	β	1.747
				γ	1.556
				δ	3.170
				ϵ	7.525
Lys^{8}	8.307		4.247	β	1.660
				γ	1.299
				ϵ	2.957
Phe^{9}	8.301		4.635	β	3.131, 3.025
				o 7.262, ***m*** 7.366, ***p*** 7.316	
Arg^{10}	8.288		4.314	β	1.819, 1.731
				γ	1.574
				δ	3.180
				ϵ	7.255
Asn^{11}	8.457		4.663	β	2.846, 2.767
				δ	7.661, 6.930
Lys^{12}	7.881	d, 7.43	4.144	β	1.816, 1.727
				γ	1.383
				δ	1.673
				ϵ	2.975
DKP3		 O 9 3 1 H_2N NH O 10 N 6 OH 7 O	**H3** 4.275	**H7**	3.094, 2.948
			H6 4.682	**H9**	3.479
				CH_2Ph	5.130, 4.391

Table E.6.: ^{1}H chemical shifts of *cyclic* ***DKP1**-sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

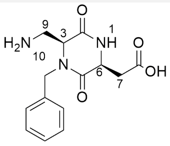
Residue	**^{1}H 283 K δ (ppm)**				
	H^N	**$^{3}J_{H^\alpha H^N}$ [Hz]**	**H^α**		**Others**
Gly^{1}	8.643		3.975		
Leu^{2}	8.444		4.349	β	1.609
				γ	1.696
				δ	0.851
Arg^{3}	8.445		4.308	β	1.850, 1.748
				γ	1.607
				δ	3.168
				ϵ	7.391
Lys^{4}	8.349		4.232	β	1.660, 1.630
				γ	1.380
				δ	1.781
				ϵ	2.967
Arg^{5}	8.380		4.223	β	1.845, 1.763
				γ	1.606
				δ	3.187
				ϵ	7.396
Leu^{6}	8.360		4.378	β	1.714
				γ	1.626
				δ	0.891, 0.947
Arg^{7}	8.427		4.231	β	1.834, 1.749
				γ	1.560
				δ	3.173
				ϵ	7.343
Lys^{8}	8.293		4.169	β	1.820
				γ	1.146
				δ	1.610
				ϵ	2.895
Phe^{9}	8.134	d, 6.78	4.689	β	3.185, 3.014
				o 7.254, ***m*** 7.371, ***p*** 7.323	
Arg^{10}	8.372		4.231	β	1.625, 1.602
				γ	1.807
				δ	3.171
				ϵ	7.401
Asn^{11}	8.504		4.660	β	2.809, 2.852
				δ	7.742, 7.029
Lys^{12}	8.322		4.367	β	1.672
				γ	1.381
				δ	1.809
				ϵ	2.955
DKP1	**H10** 8.418	 O 9 3 1 H_2N NH O 10 N 6 OH 7 O	**H3** 4.168	**H7**	2.918
			H6 4.587	**H9**	3.775
				CH_2Ph	4.366, 5.092

Table E.7.: ^{1}H chemical shifts of *cyclic* ***DKP1****-sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

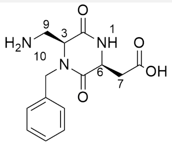
Residue	**^{1}H 298 K δ (ppm)**				
	H^N	**$^3J_{H^\alpha H^N}$ [Hz]**	**H^α**	**Others**	
Gly^1	8.556		3.963		
Leu^2	8.316		4.334		
				γ	1.619
				δ	0.877
Arg^3	8.357		4.309	β	1.845, 1.763
				γ	1.608
				δ	3.187
				ϵ	7.304
Lys^4	8.217		4.369	β	1.824
				γ	1.385
				δ	1.720
				ϵ	2.963
Arg^5	8.324		4.242	β	1.841, 1.760
				γ	1.605
				δ	3.169
				ϵ	7.324
Leu^6	8.190		4.377	β	1.696
				γ	1.621
				δ	0.908
Arg^7	8.315		4.337	β	1.843
				γ	1.607
				δ	3.189
				ϵ	7.307
				β	1.712
Lys^8	8.172		4.179	γ	1.164
				δ	1.623
				ϵ	2.901
Phe^9	8.039	d, 7.51	4.690	β	3.200, 3.019
				o 7.252, ***m*** 7.366, ***p*** 7.319	
Arg^{10}	8.284		4.240	β	1.834, 1.749
				γ	1.560
				δ	3.173
Asn^{11}	8.387		4.664	β	2.947, 2.809
				δ	7.643, 6.943
Lys^{12}	8.210		4.234	β	1.827
				γ	1.377
				δ	1.724
				ϵ	2.972
DKP1	**H10** 8.312		**H3** 4.592	**H7**	2.933
			H6 4.592	**H9**	3.765
				CH_2**Ph**	7.310, 7.411, 7.390

Table E.8.: ^{1}H chemical shifts of *cyclic* ***DKP3****-sC18** in 50 mM phosphate buffer at *283 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

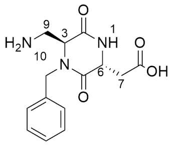
Residue	**^{1}H 283 K δ (ppm)**				
	H^N	**$^3J_{H^\alpha H^N}$** [Hz]	**H^α**		**Others**
Gly^1	8.474		3.848, 4.074		
Leu^2	8.360		4.352	β	1,711
				γ	1.608
				δ	0.879, 0.927
Arg^3	8.466		4.250	β	1.734, 1.838
				γ	1.592
				δ	3.104
				ϵ	7.363
Lys^4	8.334		4.148	β	1.755, 1.812
				γ	1.391
				δ	1.653
				ϵ	2.962
Arg^5	8.416		4.246	β	1.809
				γ	1.598
				δ	3.164
				ϵ	7.356
Leu^6	8.347		4.382	β	1.734
				γ	1.630
				δ	0.896, 0.965
Arg^7	8.374		4.219	β	1.792
				γ	1.573
				δ	3.173
Lys^8	8.347		4.171	γ	1.126
				δ	1.571
				ϵ	2.883
Phe^9	8.132	d, 7.81	4.706	β	3.026, 3.228
					o 7.269, ***m*** 7.372
Arg^{10}	8.361		4.226	β	1.805
				γ	1.582
				δ	3.162
				ϵ	7.415
Asn^{11}	8.570	d, 5.82	4.645	β	2.830, 2.881
				δ	7.046, 7.789
Lys^{12}	8.246	d, 8.10	4.334	β	1.758, 1.848
				γ	1.378, 1.454
				δ	1.663
				ϵ	2.965
DKP3	**H1** 8.445		**H3** 4.030	**H7**	2.891, 3.026
	H10 8.531	t, 6.32	**H6** 4.607	**H9**	3.738, 3.877
				**CH_2**Ph	4.162, 5.280
				CH_2**Ph**	7.330

Table E.9.: ^{1}H chemical shifts of *cyclic* ***DKP3**-sC18** in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

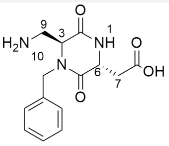
Residue	**^{1}H 298 K δ (ppm)**				
	H^N	**$^3J_{H^\alpha H^N}$ [Hz]**	**H$^\alpha$**		**Others**
Gly^1	8.370		3.849, 4.067		
Leu^2	8.232		4.351	β	1.696
				γ	1.600
				δ	0.911, 0.872
Arg^3	8.360		4.254	β	1.850, 1.744
				γ	1.599
				δ	3.122
Lys^4	8.208		4.172	β	1,701
				γ	1.147
				δ	1.600
				ϵ	2.886
Arg^5	8.310		4.240	β	1.799
				γ	1.585
				δ	3.165
Leu^6	8.188		4.394	β	1,799
				δ	0.944, 0.886
Arg^7	8.288		4.239	β	1.790
				γ	1.610
				δ	3.170
Lys^8	8.187		4.166	γ	1.378
				δ	1.628
				ϵ	2.963
Phe^9	8.030	d, 7.92	4.706	β	3.218, 3.016
				o 7.266, ***m*** 7.365, ***p*** 7.333	
Arg^{10}	8.281		4.235	β	1.845, 1.778
				γ	1.571
				δ	3.175
				ϵ	7.358
Asn^{11}	8.449		4.647	β	2.876, 2.814
				δ	7.685, 6.951
Lys^{12}	8.138		4.338	β	1.850, 1.749
				γ	1.416
				δ	1.653
				ϵ	2.968
DKP3	**H1** 8.404	 H2N 9 3 1 NH O O 10 N 6 7 OH O	**H3** 4.029	**H7**	3.011, 2.904
	H10 8.309		**H6** 4.034	**H9**	3.857, 3.730
				CH$_2$Ph	5.258, 4.163
				CH$_2$**Ph**	7.415, 7.372, 7.318

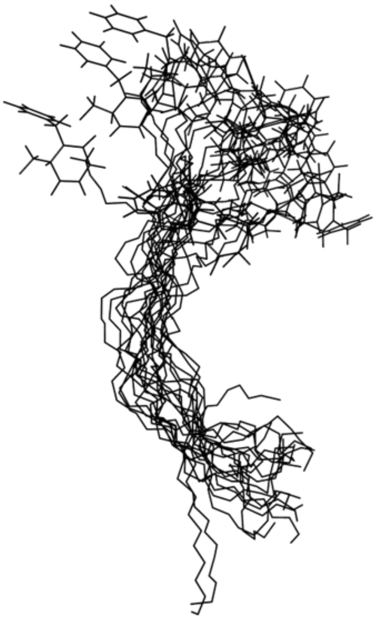
Table E.10.: ^{1}H chemical shifts of *linear **DKP1**-sC18** (left), *linear **DKP3**-sC18** (middle) and *cyclic **DKP1**-sC18** (right) in presence of *SDS micelles* in 50 mM phosphate buffer at *298 K* (600 MHz, pH 6.08, H$_2$O/D$_2$O 9:1, d$_4$-TSP ^{1}H δ=0 ppm).

Residue	***linear DKP1-sC18****			***linear DKP3-sC18****			***cyclic DKP1-sC18****		
	H^N	**H$^\alpha$**	**Others**	**H^N**	**H$^\alpha$**	**Others**	**H^N**	**H$^\alpha$**	**Others**
Gly^1	8.755	3.937		8.326	3.963		8.418	3.978	
Leu^2	7.799	4.053	β 1.823, 1.697 γ 1.611 δ 1.000	8.155	4.071	β 1.830 γ 1.644 δ 0.908, 0.748	8.392	4.303	β 1.833 γ 1.660 δ 0.854
Arg^3	8.207	3.837	β 1.918, 1.691 γ 1.612 δ 3.175 ϵ 7.286	8.277	3.911	β 1.938, 1.823 γ 1.685 δ 3.188 ϵ 7.238	7.670	4.404	β 1.906, 1.763 γ 1.608 δ 3.270 ϵ 7.266
Lys^4	7.694	4.001	β 1.919, 1.756 γ 1.394 δ 1.669 ϵ 3.183	7.690	4.068	β 1.874, 1.676 γ 1.419 δ 1.491 ϵ 3.003	8.306	4.304	β 1.824 γ 1.459 δ 1.710 ϵ 3.026
Arg^5	7.845	4.147	β 1.872 γ 1.670 δ 3.176 ϵ 7.153	7.873	4.189	β 1.951, 1.777 γ 1.601 ϵ 7.163	7.862	4.410	β 1.901, 1.760 γ 1.605 δ 3.263 ϵ 7.086
Leu^6	8.052	4.164	β 1.954, 1.954, γ 1.668 δ 0.911	7.962	4.202	δ 0.912	7.922	4.239	β 1.696 γ 1.621 δ 0.895
Arg^7	7.952	4.064	β 1.953, 1.827 γ 1.698 δ 3.181 ϵ 7.305	7.853	4.108	β 1.898 γ 1.676 δ 3.186 ϵ 7.160	7.771	4.268	β 1.782 γ 1.579 δ 3.160 ϵ 7.057
Lys^8	7.644	4.104	β 1.969, 1.828 γ 1.341 δ 1.664 ϵ 2.945	7.686	4.110	β 1.811 γ 1.356 δ 1.640 ϵ 2.977	7.886	4.334	β 1.712 γ 1.266 δ 1.623 ϵ 2.992
Phe^9	7.892	4.345	β 3.165 ***o*** 7.298, ***m*** 7.235	7.904	4.421	β 3.236, 3.178	7.983	4.552	β 3.202, 3.019 ***o*** 7.325, ***m*** 7.210
Arg^{10}	7.301	4.162	β 1.965, 1.759 γ 1.630 δ 2.939 ϵ 7.302	7.883	4.216	β 1.805 1.805 γ 1.667 δ 3.186	7.980	4.429	β 1.979, 1.692 γ 1.479 δ 3.214 ϵ 7.169
Asn^{11}	7.955	4.688	β 2.877, 2.688 δ 7.558, 6.850	8.024	4.718	β 2.889, 2.698 δ 7.562, 6.867	8.338	4.518	β 3.265, 2.983 δ 7.596, 6.931
Lys^{12}	7.650	4.105	β 1.892, 1.810 γ 1.351 δ 1.679 ϵ 3.173	7.680	4.140	β 1.802 γ 1.334 δ 1.631 ϵ 2.951	7.695	4.655	β 1.827 γ 1.437 δ 1.716 ϵ 3.017
DKP1/3		**H3** 4.469 **H6** 4.553	**H7** 2.942 **H9** 3.526 3.526 **CH$_2$**Ph 4.962, 4.327 CH$_2$**Ph** 7.359, 7.440, 7.409	**H1** 8.330	**H3** 4.327 **H6** 4.817	**H7** 2.952 **H9** 3.447, 3.374 **CH$_2$**Ph 4.960, 4.551 CH$_2$**Ph** 7.409, 7.343	**H1** 8.419 **H10** 8.135	**H3** 4.216 **H6** 4.602	**H7** 2.904, 2.530 **H9** 3.808, **H9** 3.477 **CH$_2$**Ph 5.563, 4.058 CH$_2$**Ph** 7.278 7.278

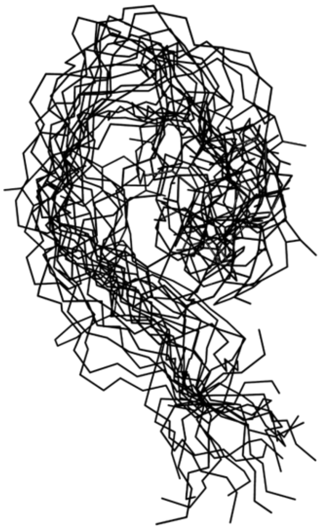
E.3. Summary tables and structure ensembles of NMR-derived structure calculations of DKP-linked peptides

E.3.1. Linear Peptides in aqueous medium

Linear DKP1-sC18*



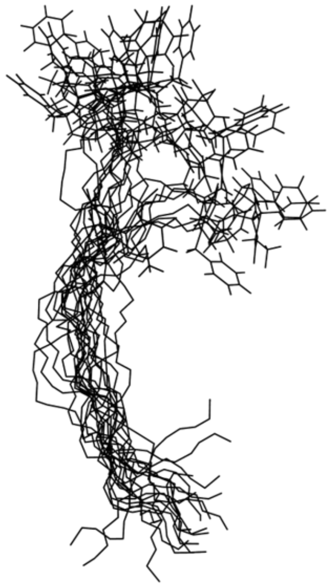
NOE cross-peaks	333
Assigned	329
NOE upper distance limits	98
Short-range $\|i-j\| \leq 1$	98
Medium range $\|i-j\| \leq 5$	98
Long-range $\|i-j\| \geq 5$	0
Violations > 0.2Å	0
CYANA target function (Å^2)	0.98
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	2.93
All heavy atoms of residue	4.44
Ramachandran statistics (%)	
Most favored	70.5
Additionally allowed	29.5
Generously allowed	0.0
Disallowed	0.0



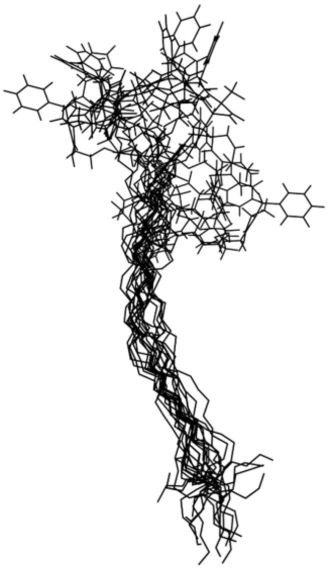
NOE cross-peaks	290
Assigned	282
NOE upper distance limits	109
Short-range $\|i-j\| \leq 1$	102
Medium range $\|i-j\| \leq 5$	0
Long-range $\|i-j\| \geq 5$	7
Violations > 0.2Å	0
CYANA target function (Å^2)	0.98
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	2.53
All heavy atoms of residue	4.17
Ramachandran statistics (%)	
Most favored	60.5
Additionally allowed	39.5
Generously allowed	0.0
Disallowed	0.0

Figure E.1.: 3D solution structures of DKP1-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K.

Linear DKP3-sC18*



NOE cross-peaks	393
Assigned	386
NOE upper distance limits	122
Short-range $\|i-j\| \leq 1$	122
Medium range $\|i-j\| \leq 5$	0
Long-range $\|i-j\| \geq 5$	0
Violations > 0.2Å	0
CYANA target function (Å^2)	0.58
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	2.61
All heavy atoms of residue	4.19
Ramachandran statistics (%)	
Most favored	57.0
Additionally allowed	43.0
Generously allowed	0.0
Disallowed	0.0

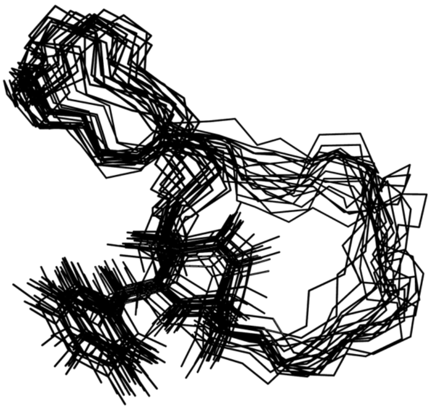


NOE cross-peaks	369
Assigned	361
NOE upper distance limits	105
Short-range $\|i-j\| \leq 1$	104
Medium range $\|i-j\| \leq 5$	1
Long-range $\|i-j\| \geq 5$	0
Violations > 0.2Å	0
CYANA target function (Å^2)	0.58
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	2.48
All heavy atoms of residue	4.14
Ramachandran statistics (%)	
Most favored	67.5
Additionally allowed	32.5
Generously allowed	0.0
Disallowed	0.0

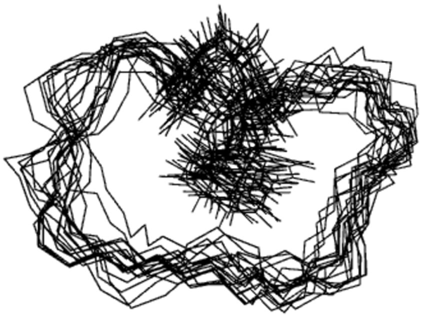
Figure E.2.: 3D solution structures of DKP3-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K.

E.3.2. Cyclic Peptides in aqueous medium

Cyclic DKP1-sC18*



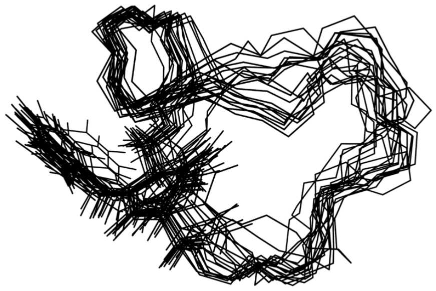
NOE cross-peaks	442
Assigned	428
NOE upper distance limits	169
Short-range $\lvert i-j\rvert \leq 1$	145
Medium range $\lvert i-j\rvert \leq 5$	8
Long-range $\lvert i-j\rvert \geq 5$	16
Violations > 0.2Å	0
CYANA target function (Å^2)	0.69
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	0.89
All heavy atoms of residue	2.06
Ramachandran statistics (%)	
Most favored	51.0
Additionally allowed	48.5
Generously allowed	0.5
Disallowed	0.0



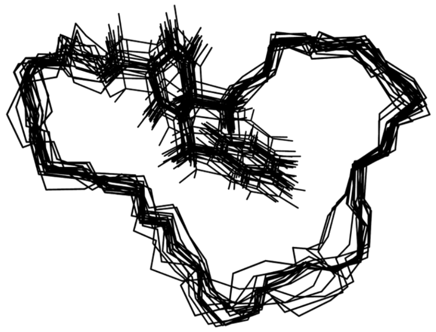
NOE cross-peaks	405
Assigned	396
NOE upper distance limits	98
Short-range $\lvert i-j\rvert \leq 1$	80
Medium range $\lvert i-j\rvert \leq 5$	10
Long-range $\lvert i-j\rvert \geq 5$	8
Violations > 0.2Å	0
CYANA target function (Å^2)	0.69
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	1.09
All heavy atoms of residue	2.44
Ramachandran statistics (%)	
Most favored	33.5
Additionally allowed	65.0
Generously allowed	1.5
Disallowed	0.0

Figure E.3.: 3D solution structures of cyclic DKP1-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K.

Cyclic DKP3-sC18*



NOE cross-peaks	658
Assigned	632
NOE upper distance limits	183
Short-range $\lvert i-j \rvert \leq 1$	157
Medium range $\lvert i-j \rvert \leq 5$	16
Long-range $\lvert i-j \rvert \geq 5$	10
Violations > 0.2Å	0
CYANA target function (Å^2)	0.59
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	1.00
All heavy atoms of residue	2.29
Ramachandran statistics (%)	
Most favored	48.0
Additionally allowed	51.5
Generously allowed	0.5
Disallowed	0.0



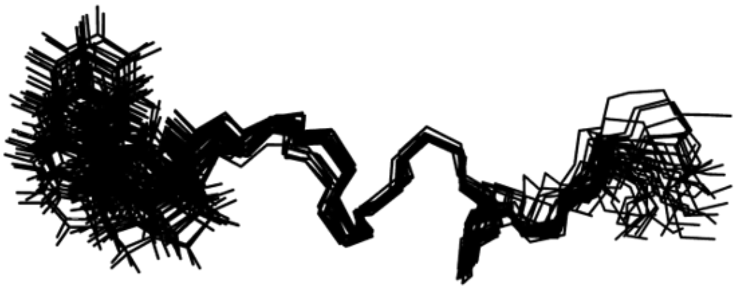
NOE cross-peaks	547
Assigned	541
NOE upper distance limits	150
Short-range $\lvert i-j \rvert \leq 1$	113
Medium range $\lvert i-j \rvert \leq 5$	17
Long-range $\lvert i-j \rvert \geq 5$	20
Violations > 0.2Å	0
CYANA target function (Å^2)	0.59
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	0.72
All heavy atoms of residue	1.95
Ramachandran statistics (%)	
Most favored	46.0
Additionally allowed	52.0
Generously allowed	2.0
Disallowed	0.0

Figure E.4.: 3D solution structures of cyclic DKP3-equipped peptide with statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at top: 283 K, bottom: 298 K.

E.3.3. Linear Peptides in presence of SDS

Linear DKP1-sC18*

a.



b.

NOE cross-peaks	621
Assigned	588
NOE upper distance limits	193
Short-range $\|i-j\| \leq 1$	157
Medium range $\|i-j\| \leq 5$	36
Long-range $\|i-j\| \geq 5$	0
Violations > 0.2Å	0
CYANA target function (Å^2)	0.98
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	0.85
All heavy atoms of residue	1.82
Ramachandran statistics (%)	
Most favored	58.9
Additionally allowed	40.8
Generously allowed	0.3
Disallowed	0.0

c.

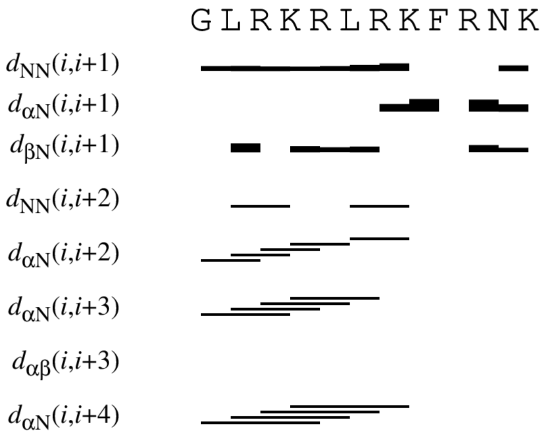
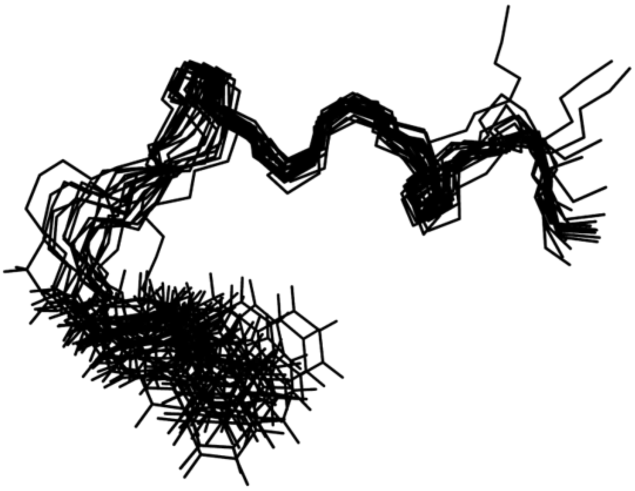


Figure E.5.: a. 3D solution structure of linear DKP1-sC18* peptide in presence of SDS-micelles; b. Statistics for 20 energy minimised structures computed with CYANA software; c. sequencial plot of short-, medium- and long-range NOE connectivities used in 3D structure calculation.

Linear DKP3-sC18*

a.



b.

NOE cross-peaks	651
Assigned	626
NOE upper distance limits	169
Short-range $\lvert i-j \rvert \leq 1$	139
Medium range $\lvert i-j \rvert \leq 5$	27
Long-range $\lvert i-j \rvert \geq 5$	3
Violations > 0.2Å	0
CYANA target function (Å^2)	0.58
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	1.06
All heavy atoms of residue	1.90
Ramachandran statistics (%)	
Most favored	72.4
Additionally allowed	27.6
Generously allowed	0.0
Disallowed	0.0

c.

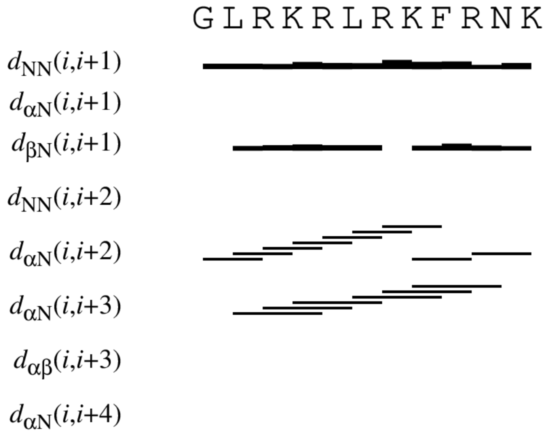

G L R K R L R K F R N K
$d_{NN}(i,i+1)$
$d_{\alpha N}(i,i+1)$
$d_{\beta N}(i,i+1)$
$d_{NN}(i,i+2)$
$d_{\alpha N}(i,i+2)$
$d_{\alpha N}(i,i+3)$
$d_{\alpha\beta}(i,i+3)$
$d_{\alpha N}(i,i+4)$

Figure E.6.: a. 3D solution structure of linear DKP3-sC18* peptide in presence of SDS-micelles; b. Statistics for 20 energy minimised structures computed with CYANA software; c. sequencial plot of short-, medium- and long-range NOE connectivities used in 3D structure calculation.

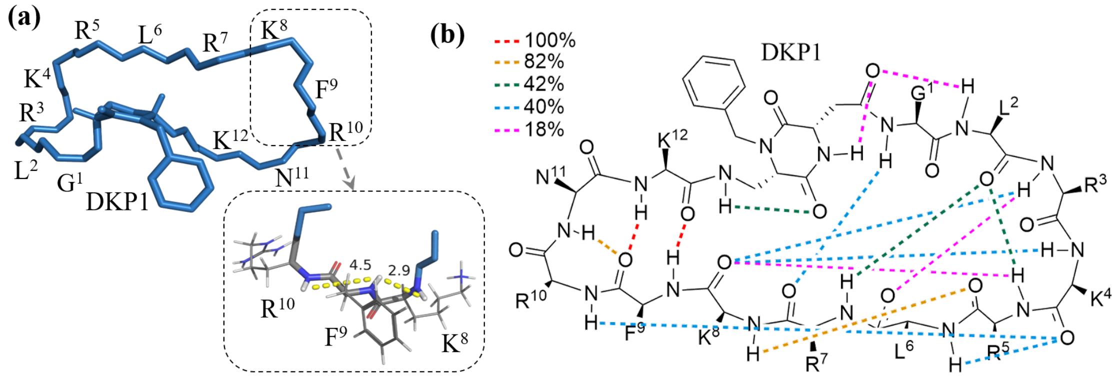

(a)
R5
L6
R7
K8
K4
R3
F9
R10
K12
L2
G1
DKP1
N11
4.5
2.9
R10
F9
K8
(b)
100%
82%
42%
40%
18%
DKP1
G1
L2
R3
K4
R5
L6
R7
K8
F9
R10
N11
K12

Figure E.7.: (a) Backbone trace of energy lowest structure calculated based on NMR experimental data at 283 K. (b) Preferred intramolecular hydrogen bond pattern gained through MC/SD simulations. Color code in legend are related to dotted lines in wedge formula and give the propensity of the occurrence of hydrogen bonds.

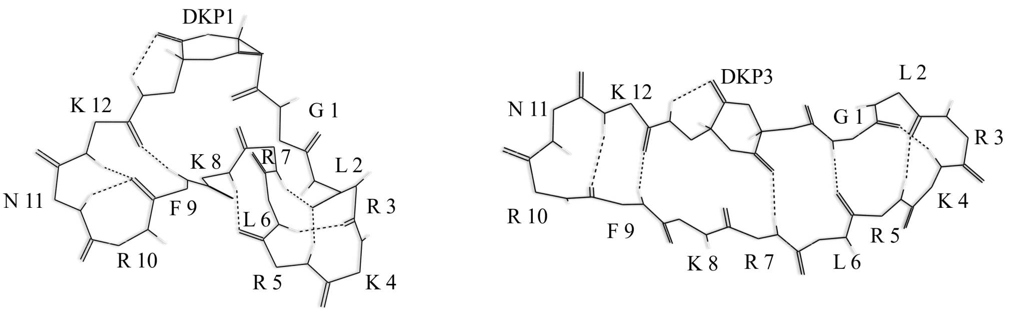

DKP1
K 12
G 1
N 11
K 8
R 7
L 2
F 9
L 6
R 3
R 10
R 5
K 4
DKP3
L 2
N 11
K 12
G 1
R 3
R 10
K 4
F 9
R 5
K 8
R 7
L 6

Figure E.8.: Representative structures of the mostly populated clusters as obtained by MC/SD calculations after energy minimisation of cyclic **DKP1**-sC18* and **DKP3**-sC18* peptides. dark grey: O, N, C backbone atoms, white: H atoms; H atoms bound to carbon as well as side chains are omitted for sake of clarity.

E.4. NMR spectra of studied DKP-sC18* peptidomimetics

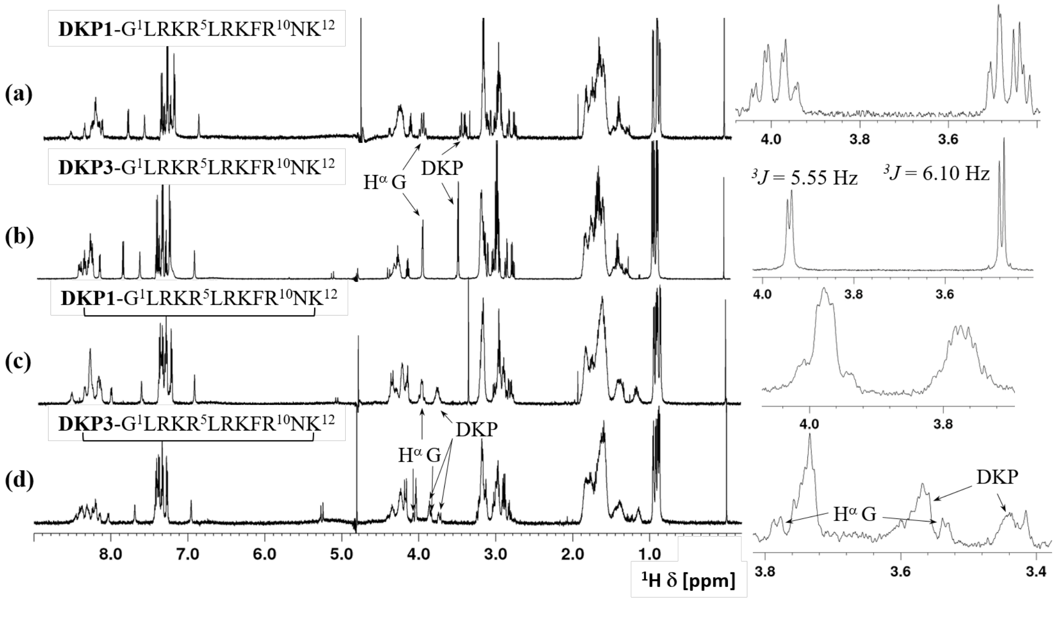


Figure E.9.: 1D ^{1}H NMR spectra of the peptides (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18*, (c) cyclic **DKP1**-sC18* and (d) cyclic **DKP3**-sC18* in solution. On the right hand side, a enlargements of labeled protons are shown, partially with coupling constants which allow an estimation about dihedral angles (peptide concentration ca. 1.3 mM, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

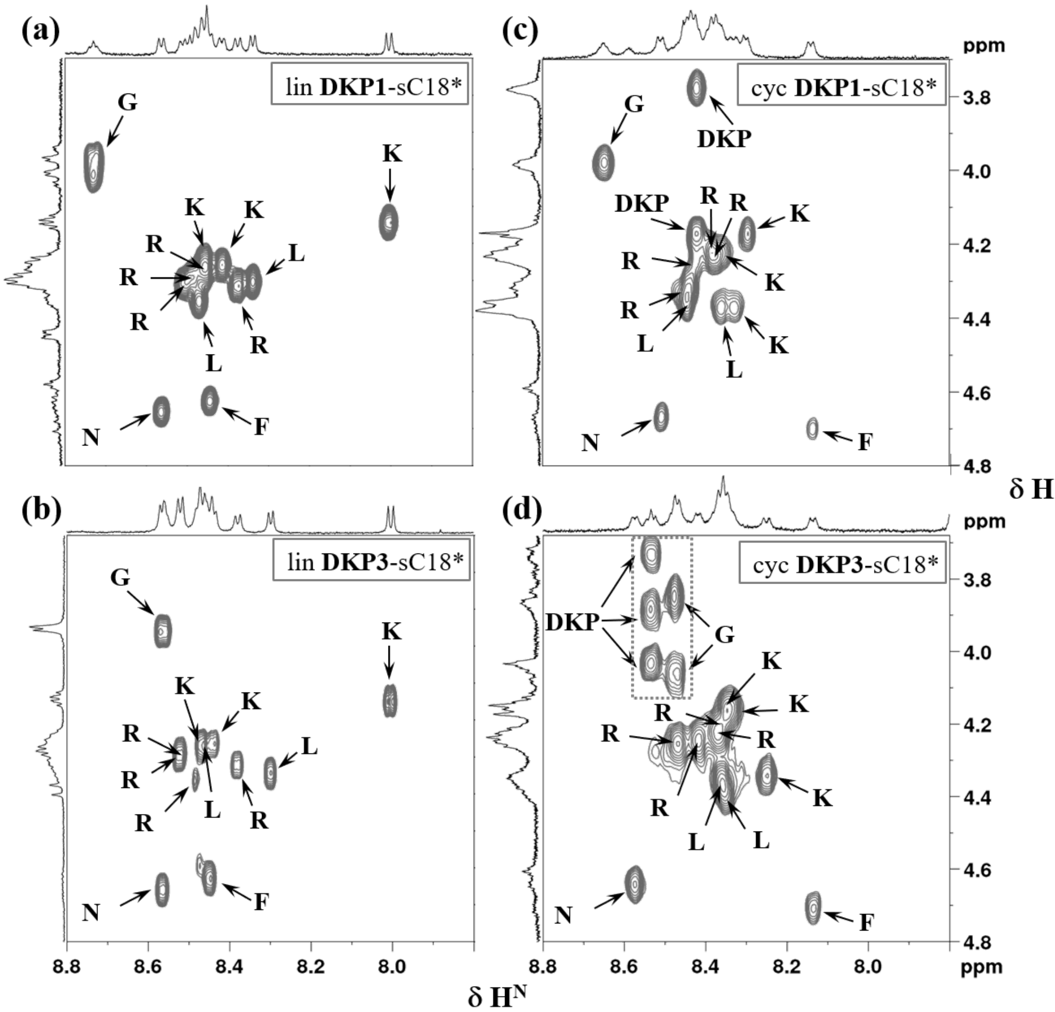


Figure E.10.: Excerpts of fingerprint region (amide-aliphatic region) of 2D TOCSY spectra of (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18*, (c) cyclic **DKP1**-sC18* and (d) cyclic **DKP3**-sC18* peptide. The box in (d) shows the diastereotopic behaviour of methylene protons belonging to glycine and the scaffold itself. For sake of clarity, only H^N-H^α chemical shifts are indicated (F: phenylalanine, G: glycine, K: lysine, L: leucine, N: asparagine, R: arginine) (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 80-120 ms, 600 MHz).

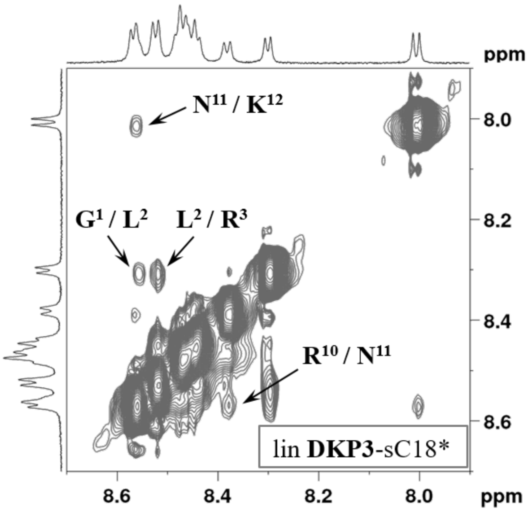


Figure E.11.: Amide-amide region of the 2D NOESY spectrum of linear DKP3-sC18*. Inter-residual NOE signals are labelled (50 mM phosphate buffer, pH 6.08, H_2O/D_2O, 283 K, mixing time 300 ms, 600 MHz).

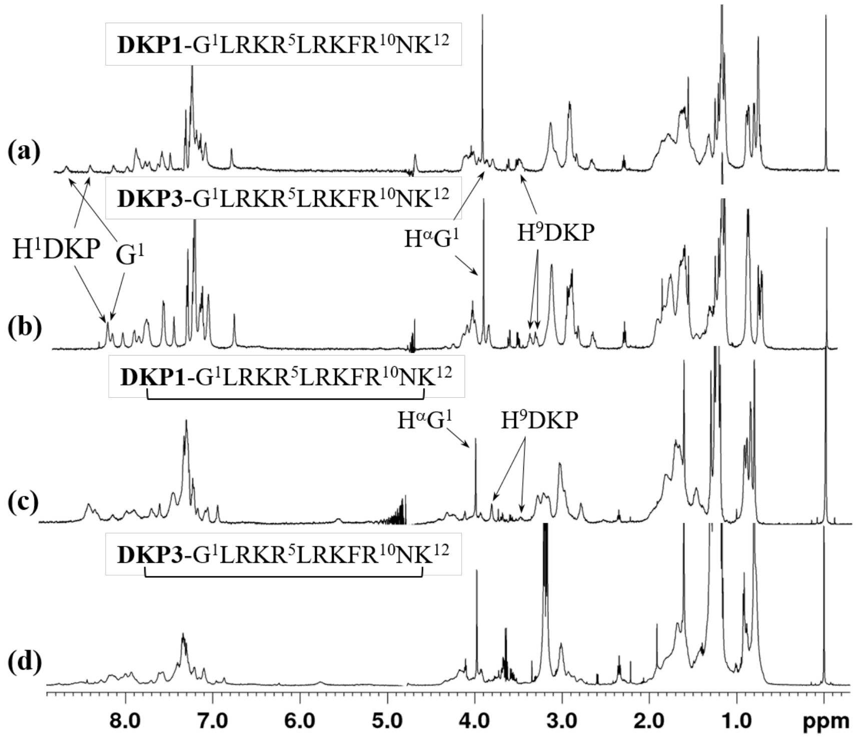


Figure E.12.: 1D ^{1}H NMR spectra of linear and cyclic peptides in presence of SDS micelles (a) lin **DKP1**-sC18*, (b) lin **DKP3**-sC18*, (c) cyc **DKP1**-sC18* and (d) cyc **DKP3**-sC18* (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 500 MHz cryoprobe or 600 MHz).

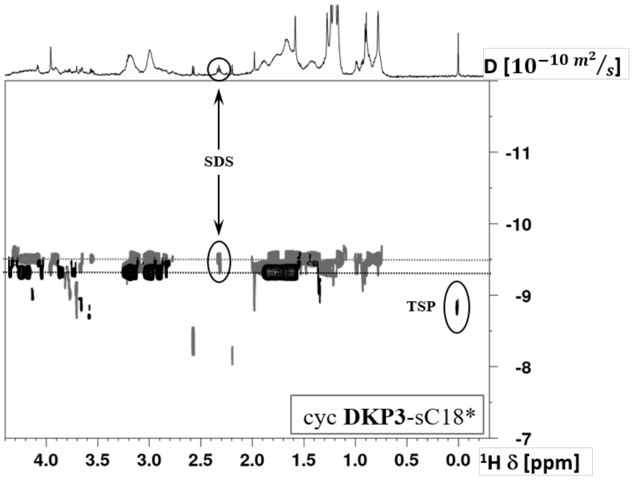


Figure E.13.: DOSY NMR spectra of cyclic DKP3-sC18* free in solution (black trace) in comparison to their presence in SDS micelles (gray trace). TSP is superimposed and taken as internal reference. 1D projection shows peptide together with micelles (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, 600 MHz).

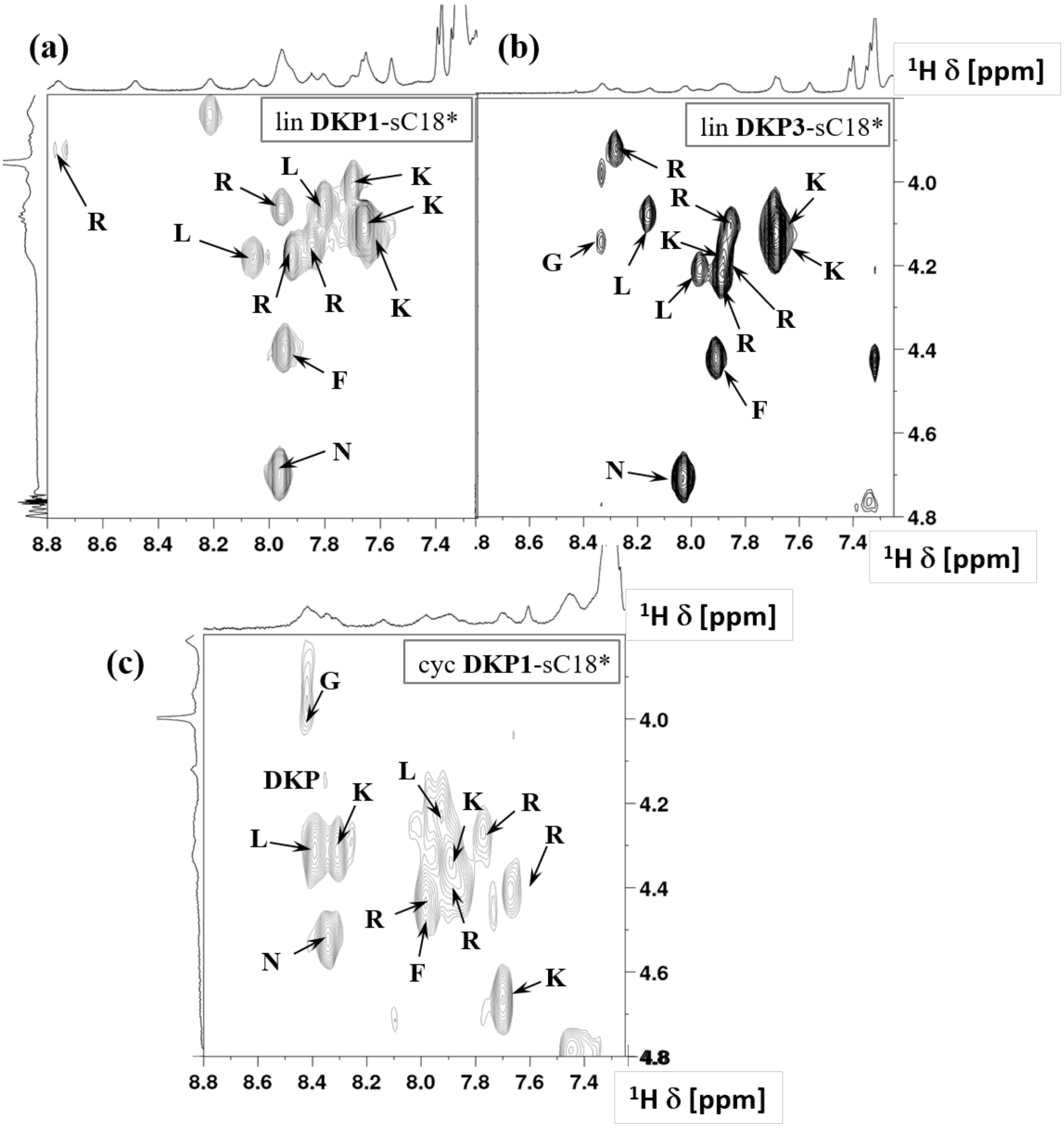


Figure E.14.: Fingerprint excerpt of 2D TOCSY spectra of (a) linear **DKP1**-sC18*, (b) linear **DKP3**-sC18* and (c) cyclic **DKP1**-sC18* peptides in presence of SDS micelles. (peptide to micelle concentration: around 1:1, PBS, pH 6.08 H_2O/D_2O, 298 K, mixing time 80-120 ms, 600 MHz).

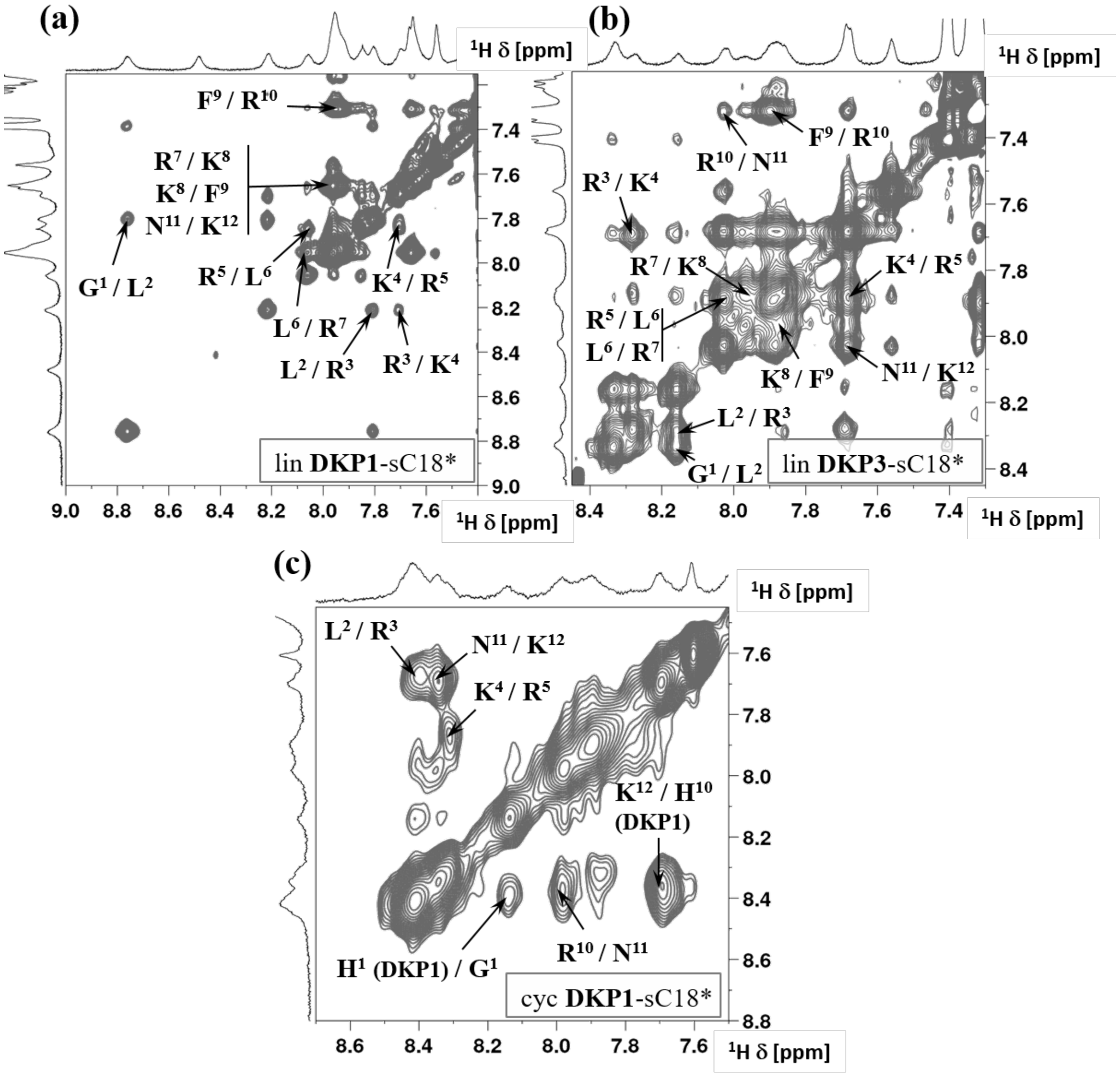


Figure E.15.: Amide-amide regions of 2D NOESY spectra of both linear peptides (a) and (b) and (c) cyclic DKP1-sC18* peptide in presence of SDS micelles. All spacial close backbone amide signals are labelled (peptide to micelle concentration: around 1:1, PBS, pH 6.08, H_2O/D_2O 9:1, 298 K, mixing time 200 ms, 600 MHz).

F. Appendices Part III

F.1. Proton and carbon assignment of the Trp-cage mini-protein

Table F.1.: ^{1}H chemical shifts of Trp-cage mini-protein in 50 mM phosphate buffer at *280 K* (600 MHz, pH 7.0, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

Residue						
	H^N	**H^α**		**Others**	C	
Asn^1		3.930	β	2.909	C^α	53.937
			δ	7.948, 7.212		
Leu^2		4.188	β	1.827	C^α	57.116
			γ	1.472	C^δ	25.232
			δ	0.994, 0.937		
Tyr^3	8.678	4.160	β	3.164, 3.123	C^α	57.146
			δ	7.052		
			ϵ	6.848		
Ile^4	8.114	3.767	β	1.964	C^α	63.272
			γ	1.719, 1.406	C^γ	25.735
			δ	0.954	C^δ	12.159
Gln^5	7.983	3.932	β	2.202, 2.098	C^α	57.993
			γ	2.418	C^γ	33.657
			ϵ	7.874, 7.053		
Trp^6	8.073	4.298	β	3.583, 3.186	C^α	62.672
			HD 1	7.034		
			HE 1	9.777		
			HE 3	7.132		
			HH 2	7.236		
			HZ 2	7.232		
			HZ 3	7.158		
Leu^7	8.321	3.476	β	1.932, 1.607	C^α	57.439
			γ	1.360	C^δ 1	23.204

			δ	0.954, 0.860		
Lys^{8}	8.330	3.952	β	1.933	C$^{\alpha}$	58.695
			γ	1.444	C$^{\epsilon}$	42.079
			δ	1.631, 1.567		
			ϵ	2.950		
Asp^{9}	8.031	4.530	β	2.909, 2.744	C$^{\beta}$	41.379
Gly^{10}	7.627	4.158, 3.504			C$^{\alpha}$	44.615
Gly^{11}	8.488	3.218, 0.971			C$^{\alpha}$	45.563
Pro^{12}		4.636	β	2.520, 2.168	C$^{\beta}$	27.260
			γ	2.074		
			δ	3.821, 3.490		
Ser^{13}	7.822	4.484	β	3.936	C$^{\alpha}$	59.300
Ser^{14}	8.274	4.173	β	3.870, 3.523	C$^{\alpha}$	59.865
Gly^{15}	8.034	4.278, 3.823			C$^{\alpha}$	45.336
Arg^{16}	8.189	5.039	β	1.901, 1.822		
			γ	1.669		
			δ	3.330, 3.229		
			ϵ	7.681		
Pro^{17}		4.769	β	2.352, 2.015	C$^{\beta}$	27.104
			γ	1.796		
			δ	3.879, 3.679		
Pro^{18}		2.668	β	1.354, 0.477		
			γ	1.731		
			δ	3.558, 3.531		
Pro^{19}		4.372	β	2.239	C$^{\beta}$	27.182
			γ	1.949, 1.848		
			δ	3.210, 3.002		
Ser^{20}	8.343	4.303	β	3.827		

Table F.2.: ^{1}H chemical shifts of Trp-cage mini-protein in 50 mM phosphate buffer at *298 K* (600 MHz, pH 7.0, H_2O/D_2O 9:1, d_4-TSP ^{1}H δ=0 ppm).

Residue							
	HN	**H$^\alpha$**		**Others**	C_δ		
Asn^1		3.930	β	2.843	C^α	53.937
			δ	7.584, 7.080		
Leu^2		4.195	β	1.751	C^α	57.116
			γ	1.473	C^δ	25.232
			δ	0.969, 0.912		
Tyr^3	8.505	4.228	β	3.146, 3.092	C^α	57.146
			δ	7.045		
			ϵ	6.830		
Ile^4	7.965	3.816	β	1.927	C^α	63.272
			γ	1.643, 1.349	C^γ	25.735
			δ	0.916	C^δ 1	12.159
Gln^5	7.962	3.974	β	2.148, 2.087	C^α	57.993
			γ	2.381	C^γ	33.657
			ϵ	7.710, 6.903		
Trp^6	7.997	4.372	β	3.525, 3.202	C^α	62.672
			HD 1	7.075		
			HE 1	9.806		
			HE 3	7.276		
			HH 2	7.145		
			HZ 2	7.229		
			HZ 3	7.176		
Leu^7	8.215	3.603	β	1.850, 1.593	C^α	57.439
			γ	1.417	C^δ 1	23.204
			δ	0.925, 0.845		

Lys^{8}	82.778	3.987	β	1.891	C$^{\alpha}$	58.695
			γ	1.428	C$^{\epsilon}$	42.079
			δ	1.648, 1.542		
			ϵ	2.955		
Asp^{9}	7.990	4.574	β	2.847, 2.718	C$^{\beta}$	41.379
Gly^{10}	7.704	4.110, 3.589			C$^{\alpha}$	44.615
Gly^{11}	8.352	3.310, 1.633			C$^{\alpha}$	45.563
Pro^{12}		4.561	β	2.463, 2.123	C$^{\beta}$	27.260
			γ	2.046		
			δ	3.747, 3.433		
Ser^{13}	7.889	4.474	β	3.926	C$^{\alpha}$	59.300
Ser^{14}	8.202	4.235	β	3.870, 3.613	C$^{\alpha}$	59.865
Gly^{15}	8.021	4.198, 3.851			C$^{\alpha}$	45.336
Arg^{16}	8.109	4.928	β	1.845, 1.760		
			γ	1.677		
			δ	3.293, 3.229		
			ϵ	7.657		
Pro^{17}		4.722	β	2.328, 2.015	C$^{\beta}$	27.104
			γ	1.814		
			δ	3.851, 3.642		
Pro^{18}		3.088	β	1.461, 0.882		
			γ	1.760		
			δ	3.596, 3.531		
Pro^{19}		4.372	β	2.235	C$^{\beta}$	27.182
			γ	1.949, 1.879		
			δ	3.295, 3.105		
Ser^{20}	8.188	4.323	β	3.827		

F.2. Summary tables of NMR-derived structure calculations of Trp-cage miniprotein

Table F.3.: Statistics and average values over 20 energy minimised CYANA conformers. Data were collected and calculated at 280 K.

NOE upper distance limits	
Short-range $\|i-j\| \leq 1$	151
Medium range $\|i-j\| \leq 5$	36
Long-range $\|i-j\| \geq 5$	31
Violations > 0.2Å	0
CYANA target function (Å^2)	0.17
RMSD to mean coordinates (Å)	
Backbone N, Cα, CO	0.58 +/− 0.13
All heavy atoms of residue	0.91 +/− 0.16
Ramachandran statistics (%)	
Most favored	67.3
Additionally allowed	32.7
Generously allowed	0.0
Disallowed	0.0

Table F.4.: Hydrogen bonds in 20 energy minimised structures calculated based on NMR experimental data with CYANA-software. Cut-off for hydrogen bond length was chosen as 1Å.

H $\cdots$	O	# of hydrogen bonds
L^7	I^4	5
K^8	I^4	12
D^9	Q^5	2
D^9	W^6	10
G^{10}	L^7	8
G^{11}	L^7	2
S^{14}	G^{11}	3
S^{14}	P^{12}	3
$H\epsilon$ W^6	R^{16}	12

F.3. Selected analytics of ^{15}N TMAO

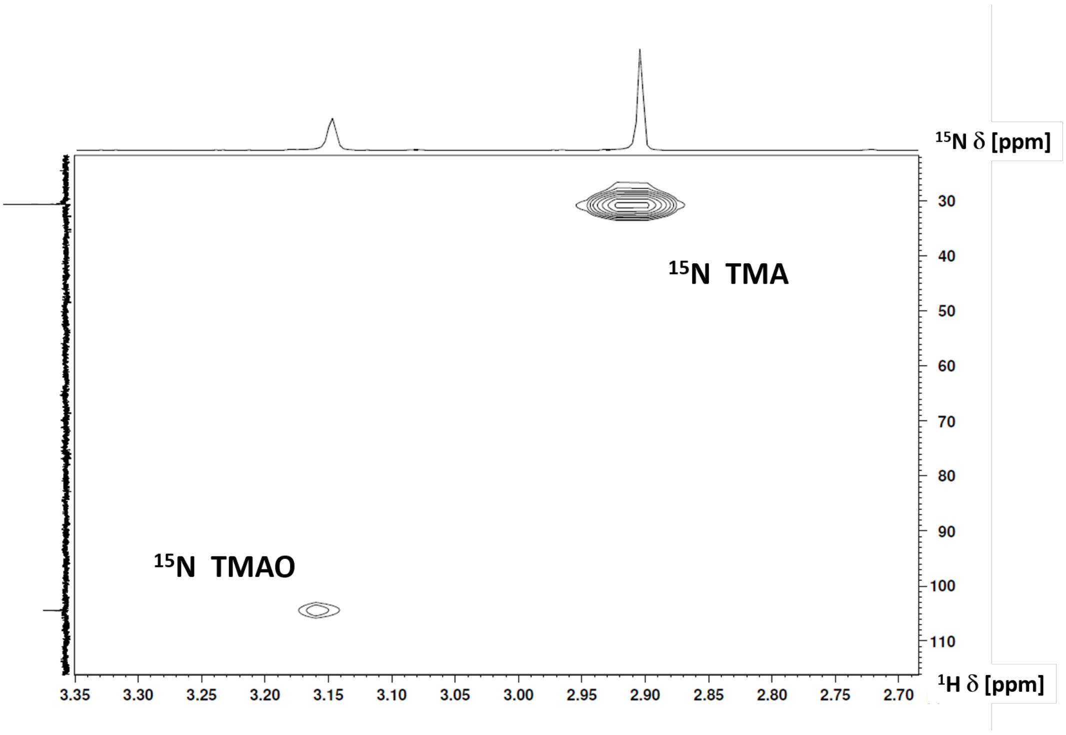


Figure F.1.: H-N HMQC spectrum of synthesised isotopic labelled TMAO with TMA as external reference (H_2O/D_2O 9:1, 298 K, 400 MHz).

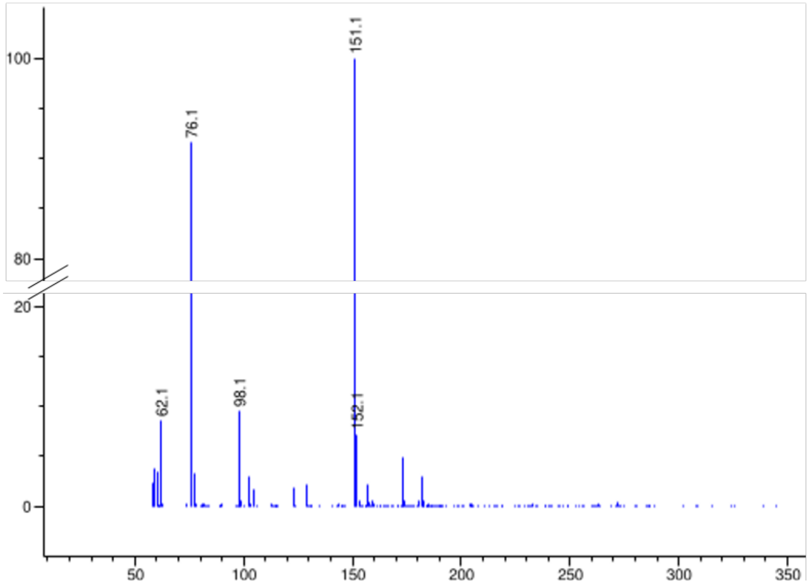


Figure F.2.: ESI spectrum of ^{15}N-labelled TMAO.

G. Spectra and tables of ^{1}H NMR studies of interaction on small molecules and micelles

Table G.1.: Overview of the chemical shift change of the SDS signals under the addition of small organic molecules at an SDS concentration of 3 mM(arrows: shift direction, X: no shift).

Signal **Additive**	*A*	*B*	*C*	*D*	Additive
Dioxane	←	←	←	←	→
Phenol	←	←	←	←	×
Toluene	←	←	←	←	←
Cyclohexane	←	←	←	←	←

Table G.2.: Overview of the change in diffusion rate of single SDS molecules at a concentration of 3 mM induced by the interaction with small molecules. The change in diffusion rate of the additive itself, compared to diffusion through water, is also shown (arrow up: increased coefficient, arrow down: decreased coefficient).

Signal **Additive**	SDS	Additive
Dioxane	↑	↓
Phenol	↓	↓
Toluene	↓	↓
Cyclohexane	↓	↑

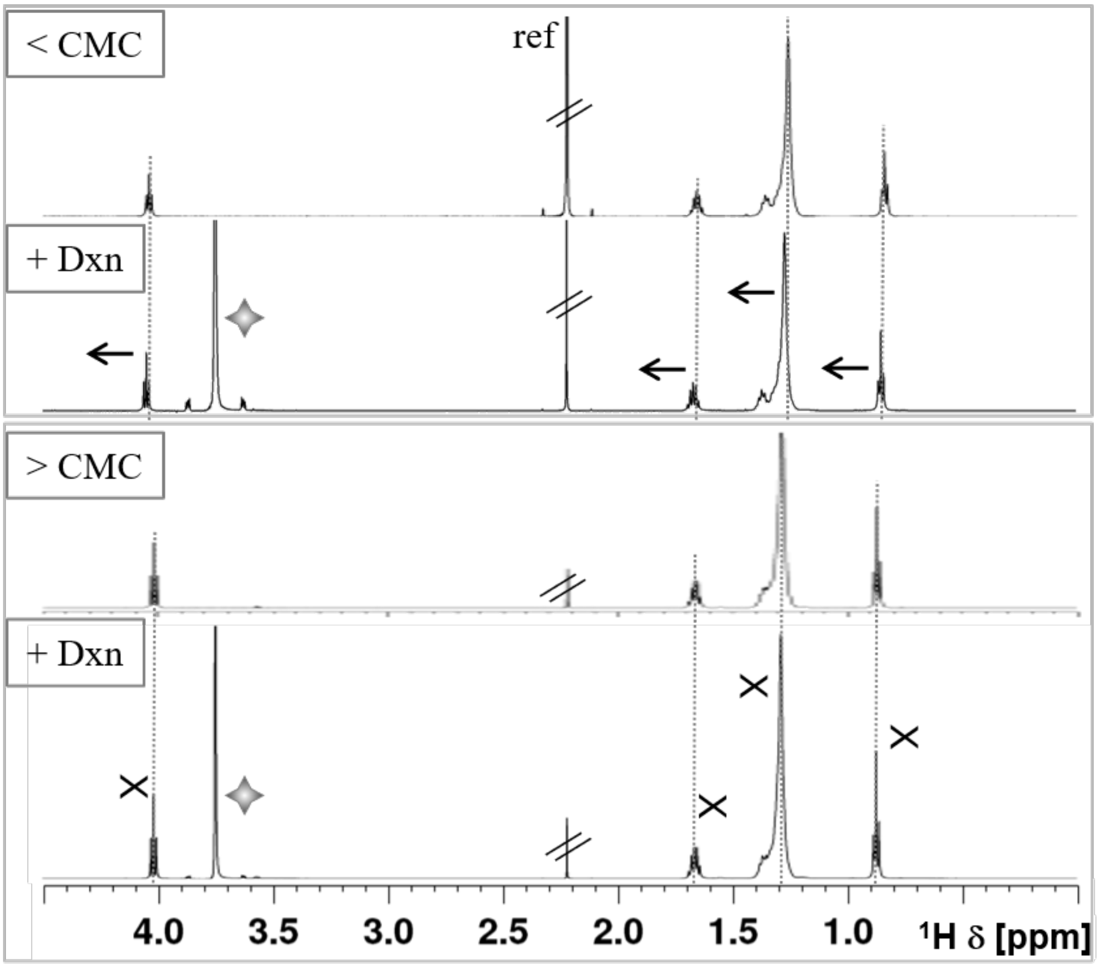


Figure G.1.: 1D proton spectra of SDS below cmc with and without dioxane as additive (top box) and SDS above cmc with and without dioxane (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K).

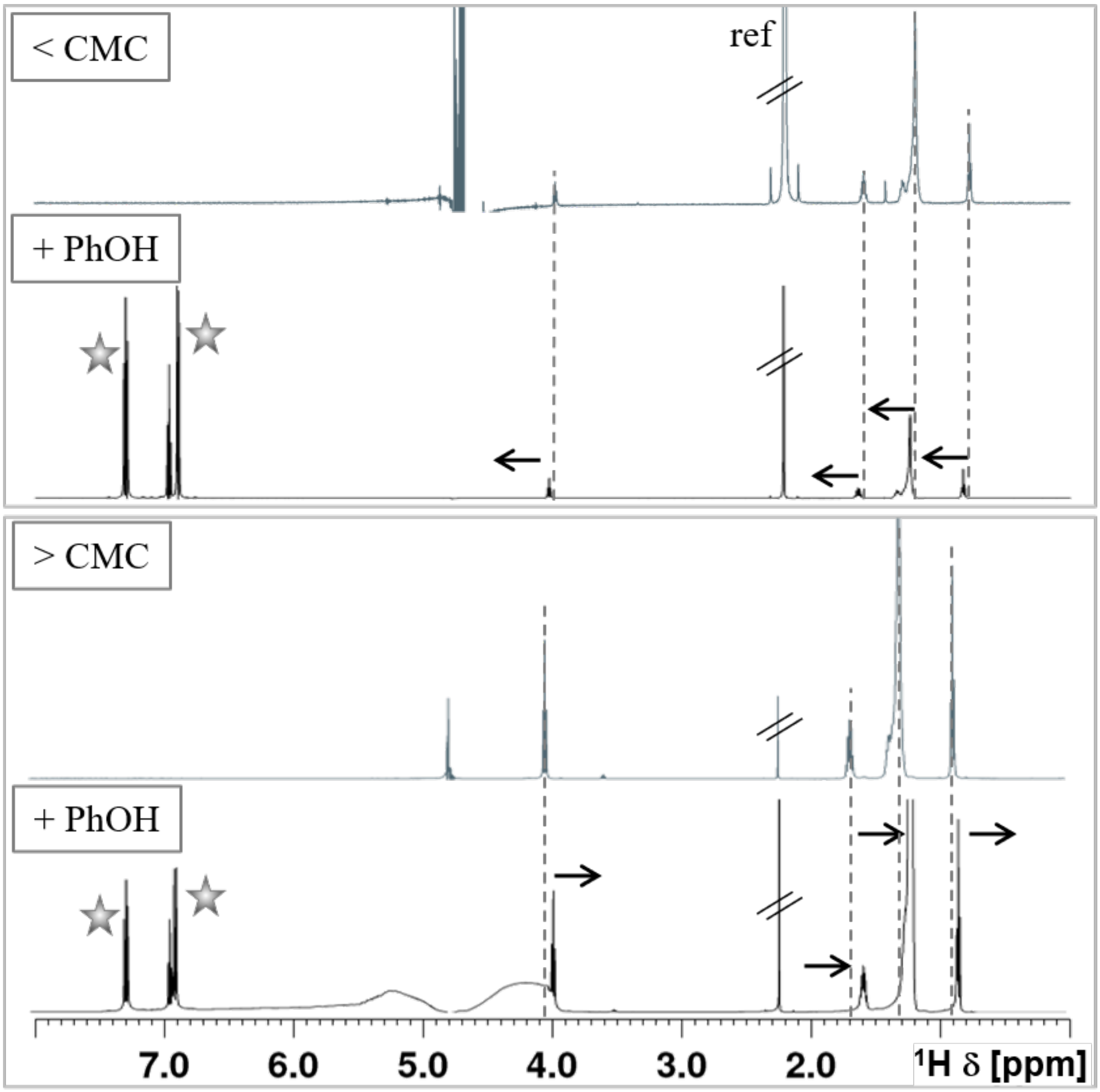


Figure G.2.: 1D proton spectra of SDS below cmc with an without phenol as additive (top box) and SDS above cmc with and without phenol (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K).

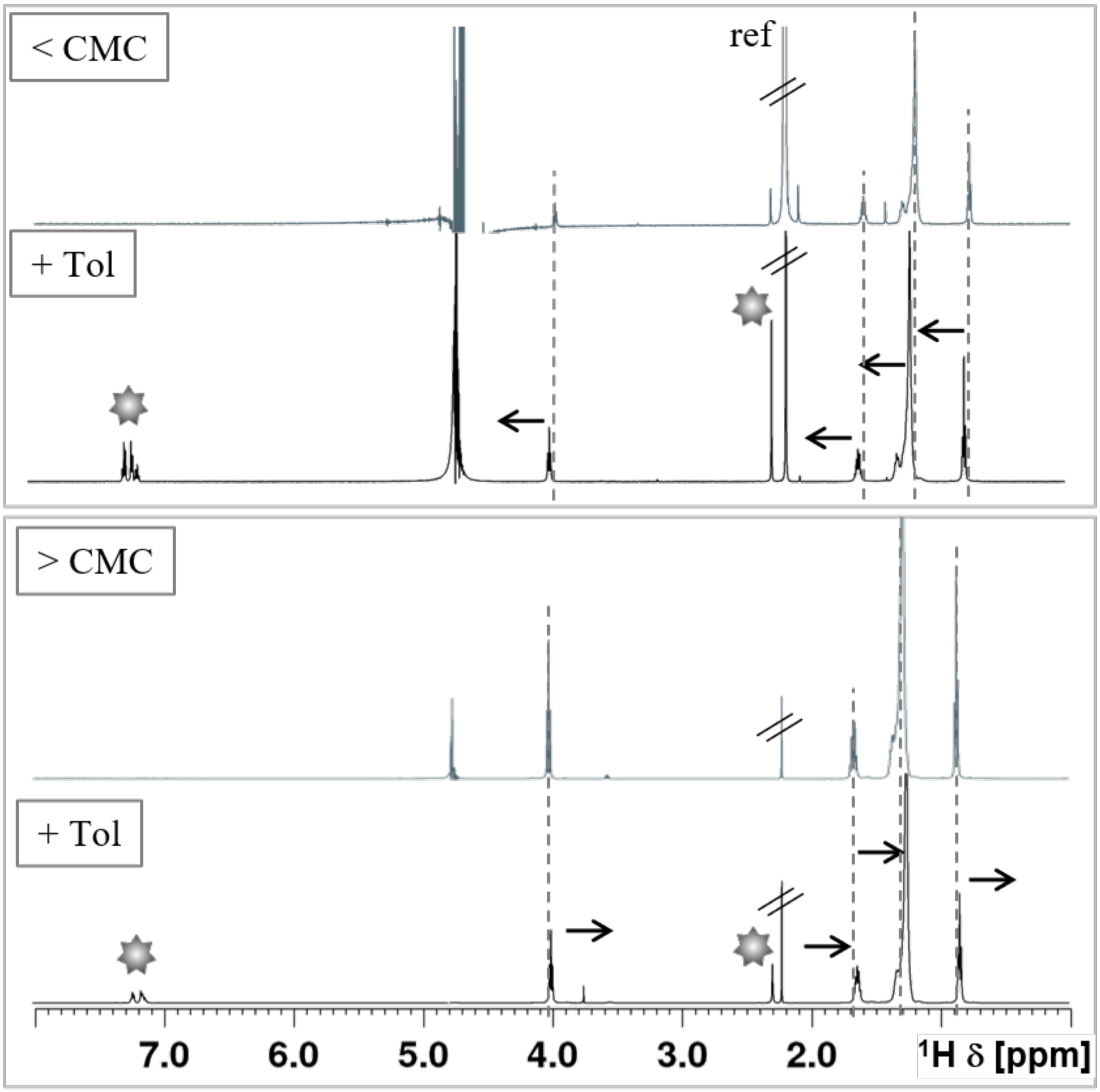


Figure G.3.: 1D proton spectra of SDS underneath cmc with an without toluene as additive (top box) and SDS above cmc with and without phenol (box in the bottom). Acetone/water external capillary is the reference (H_2O/D_2O 9:1, 600 MHz, 298 K).

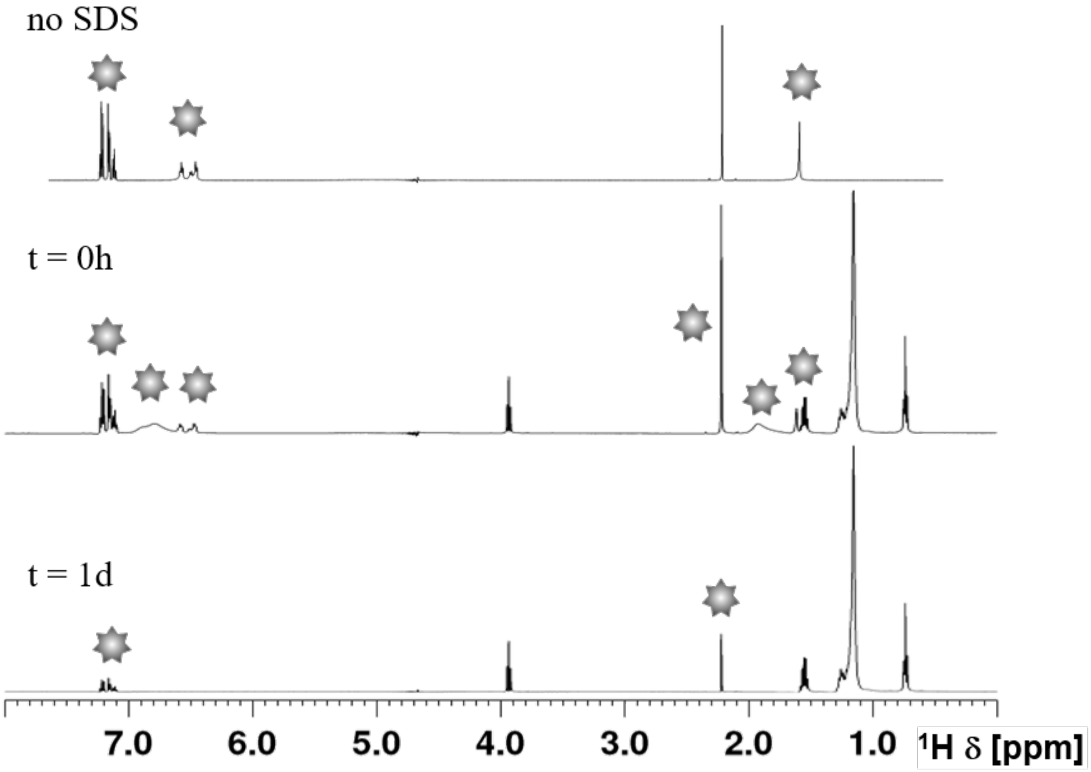


Figure G.4.: 1D proton spectra of SDS below cmc in the absence of SDS (top) and in presence of 3 mM SDS immediately after addition of toluene (middle trace) and after 1 one day (bottom trace) (H_2O/D_2O 9:1, 600 MHz, 298 K).

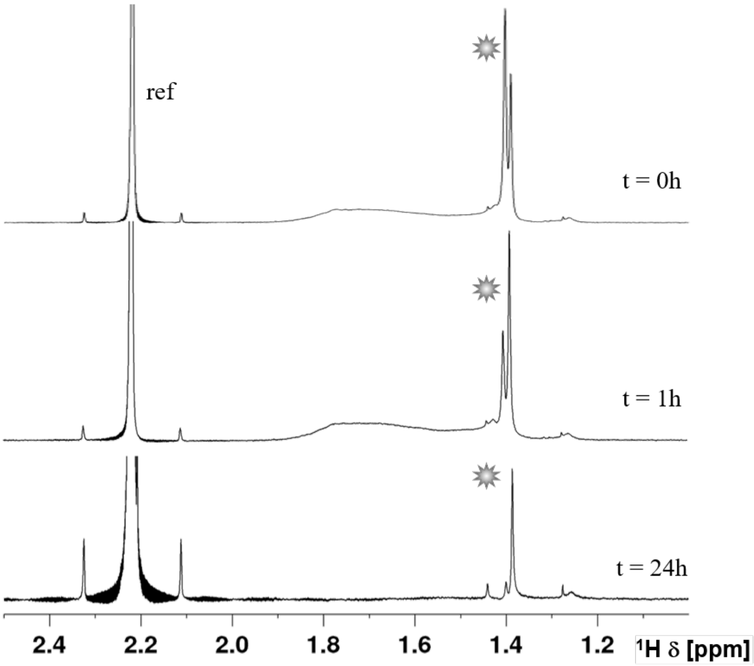


Figure G.5.: 1D proton spectra of cyclohexane in water acquired immediately after addition (top trace), after 1 hour (middle trace) and after 24 hours (bottom trace) (H_2O/D_2O 9:1, 600 MHz, 298 K).

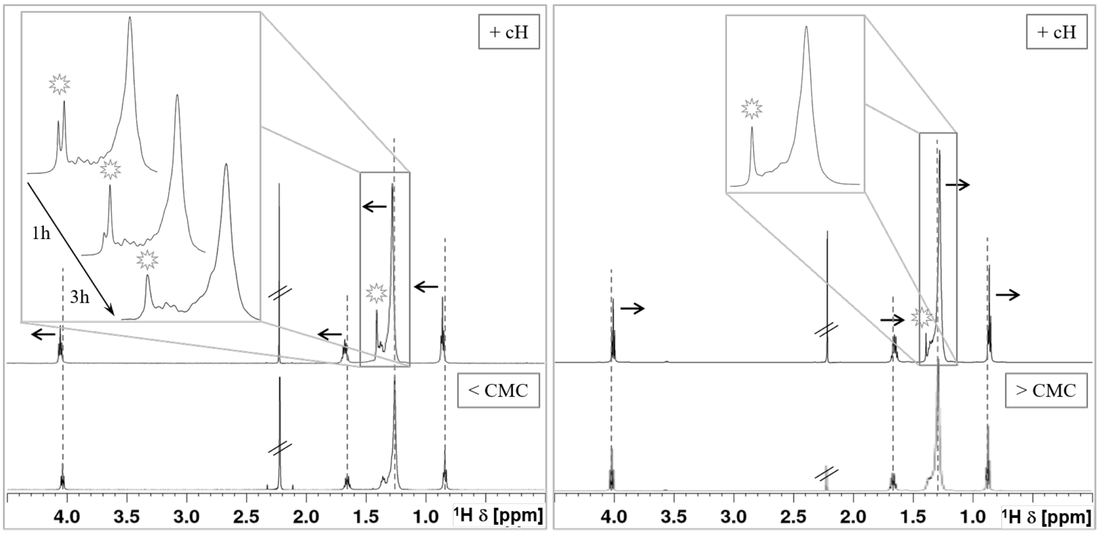


Figure G.6.: 1D proton spectra of SDS below (left) and above (right) cmc with cyclohexane. The signals of cyclohexane are emphasised in a box (H_2O/D_2O 9:1, 600 MHz, 298 K).

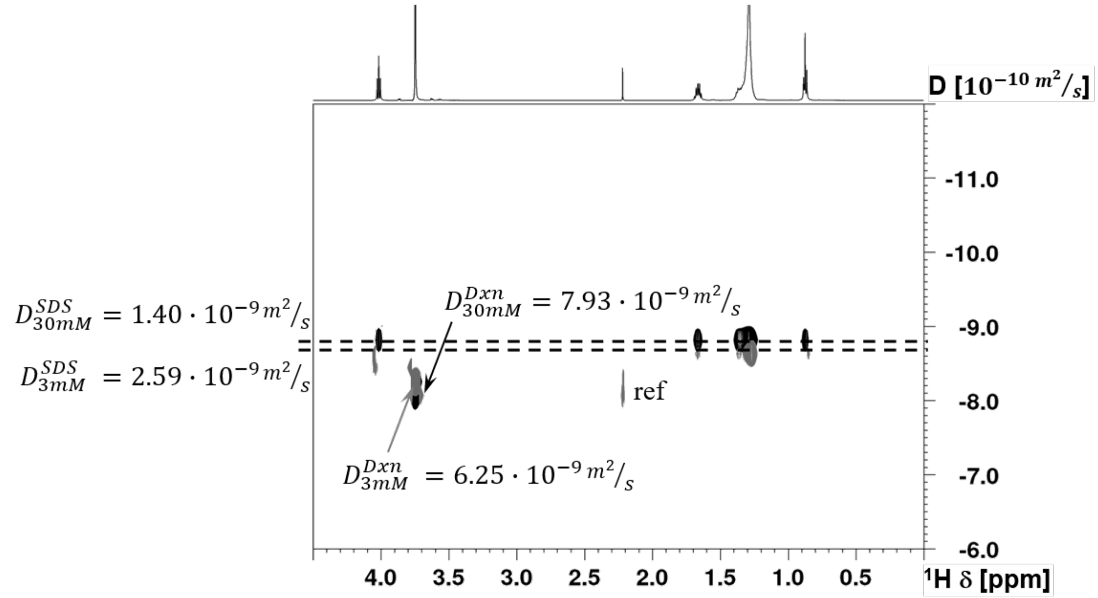


Figure G.7.: Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with dioxane (H_2O/D_2O 9:1, 600 MHz, 298 K, d20: 150–200 ms).

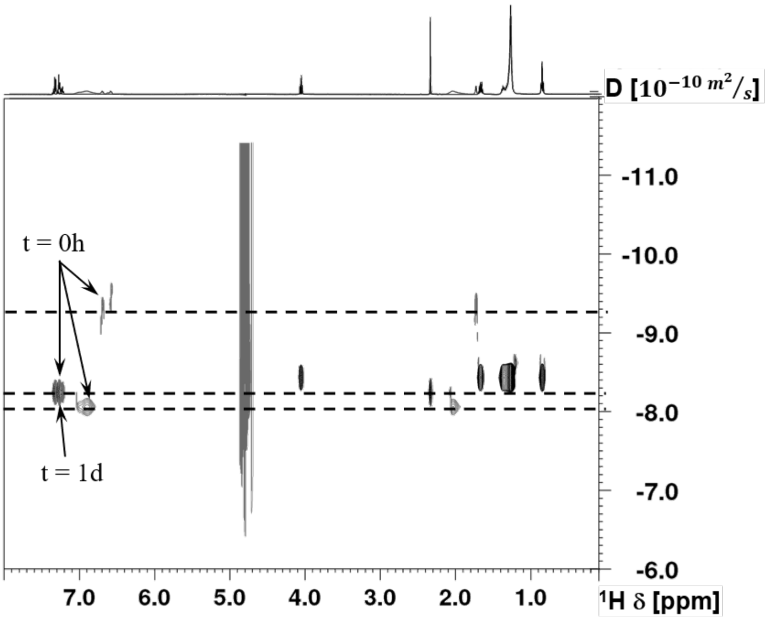


Figure G.8.: Superimposition of 2D DOSY spectra of SDS below cmc with toluene. Measurements were performed immediately after addition of toluene (t=0) and after one day (t=1d) (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms).

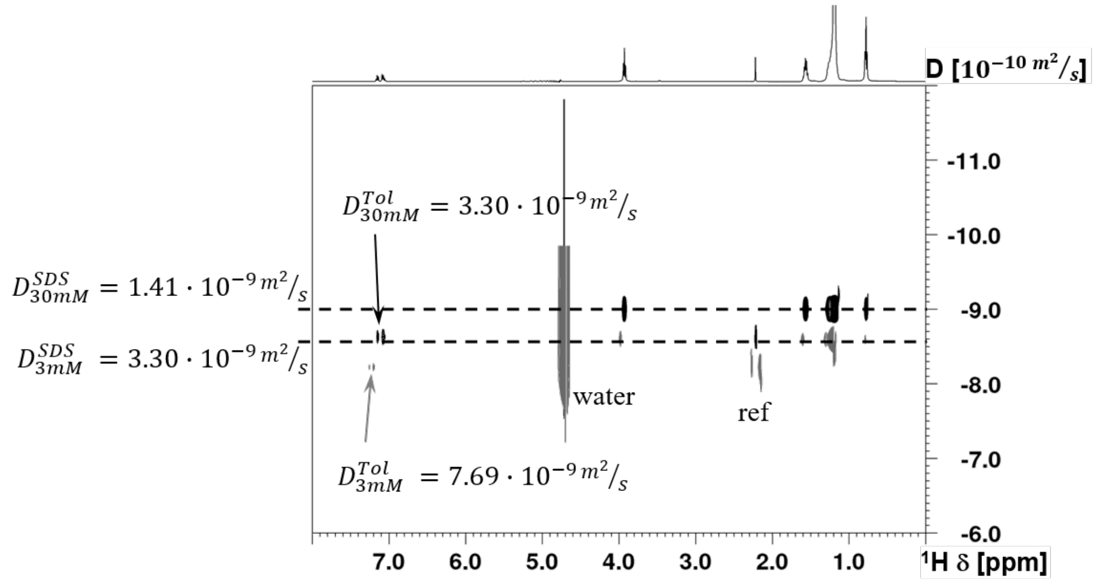


Figure G.9.: Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with toluene (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms).

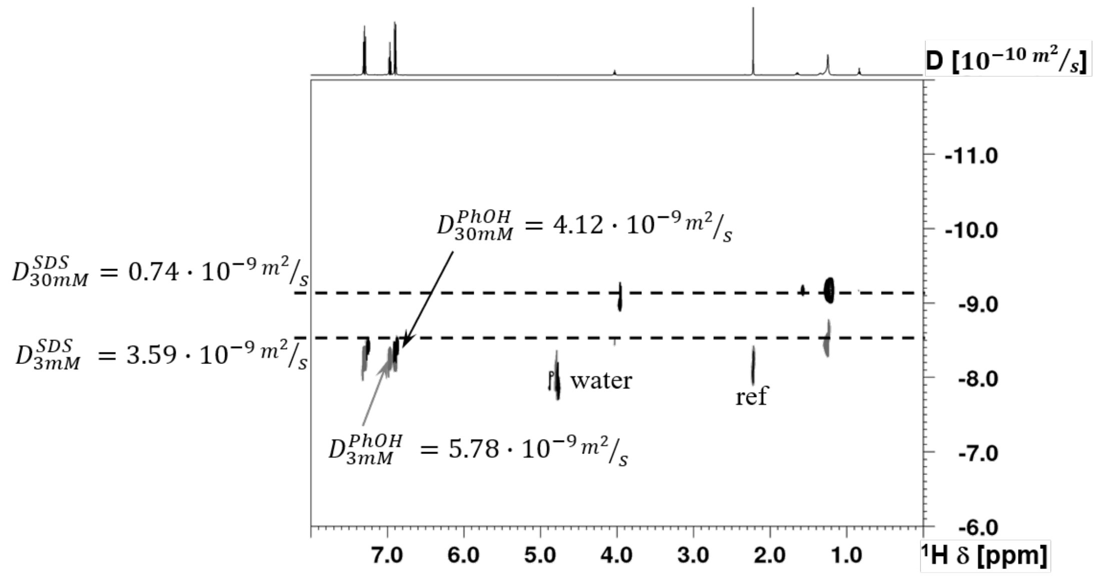


Figure G.10.: Superimposition of 2D DOSY spectra of SDS below (grey trace) and above (black trace) cmc with phenol (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 30-80 ms).

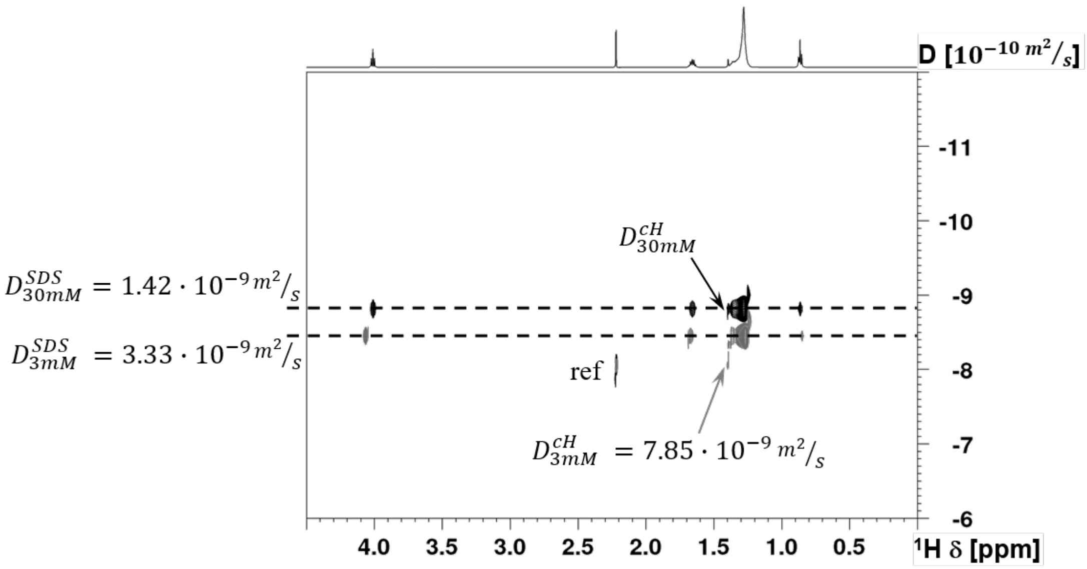


Figure G.11.: Superimposition of 2D DOSY spectra of SDS under (grey trace) and above (black trace) cmc with cyclohexane (H_2O/D_2O 9:1, 600 MHz, 298 K, d20 70-120 ms).

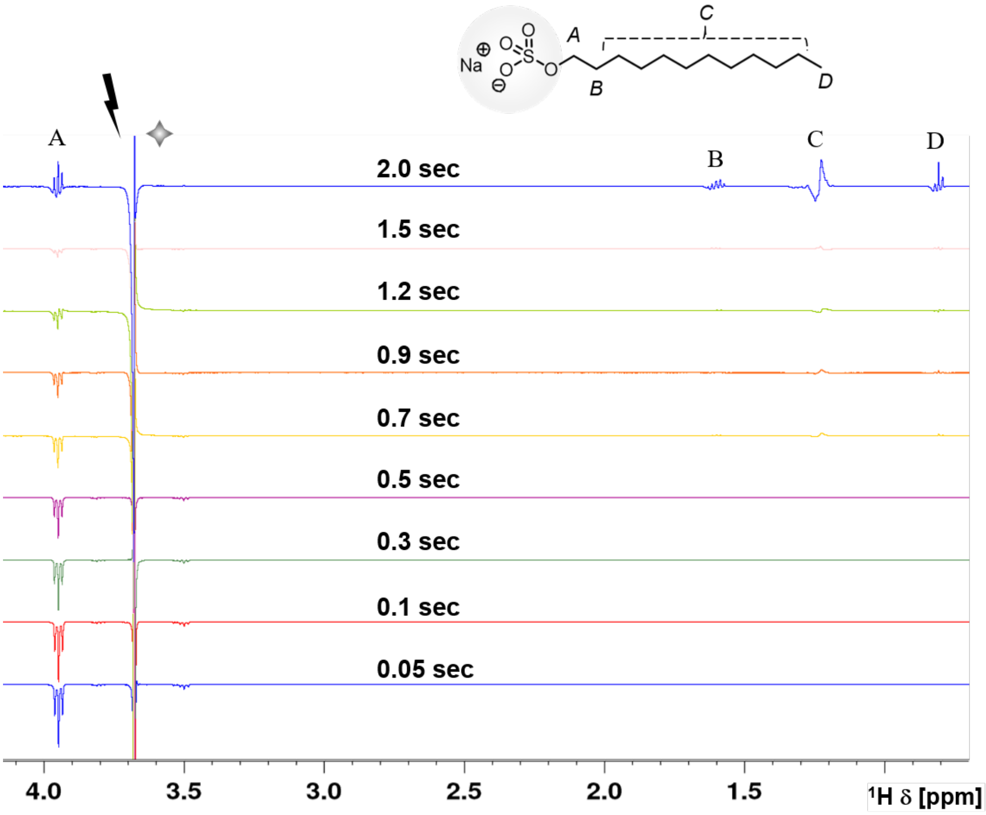


Figure G.12.: 1D NOESY spectra of dioxane in presence of SDS micelles at various mixing times. Mixing times are indicated for 1D traces (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)).

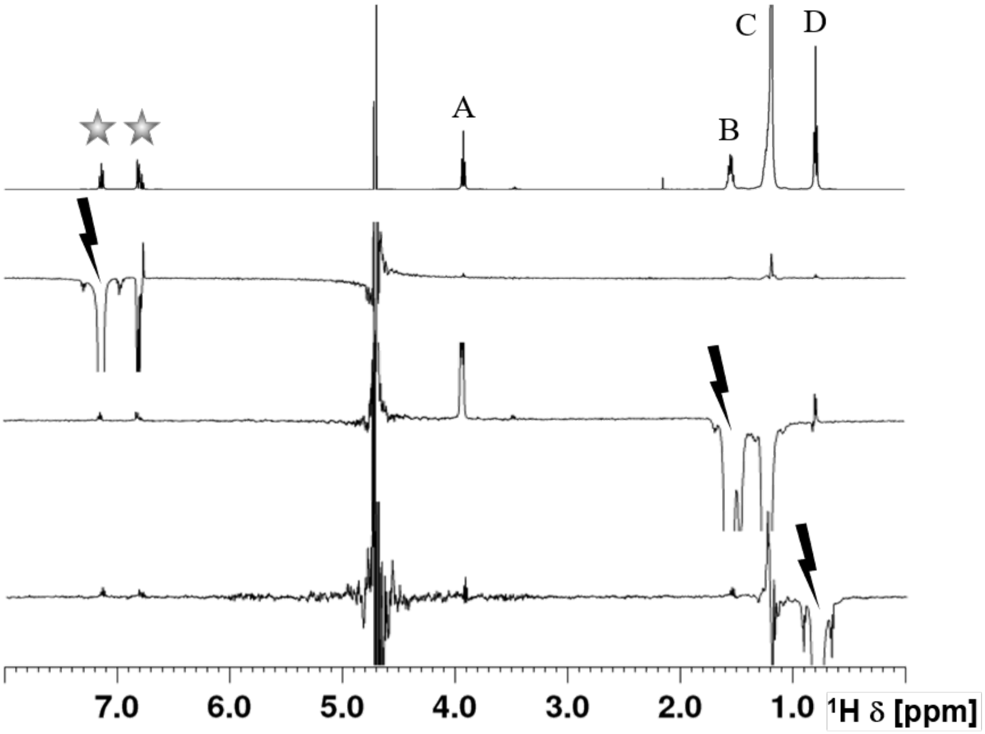


Figure G.13.: 1D NOESY spectra of phenol in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation points (sorted descending): aromatic protons of phenol, SDS signal *B* and and *D* (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)).

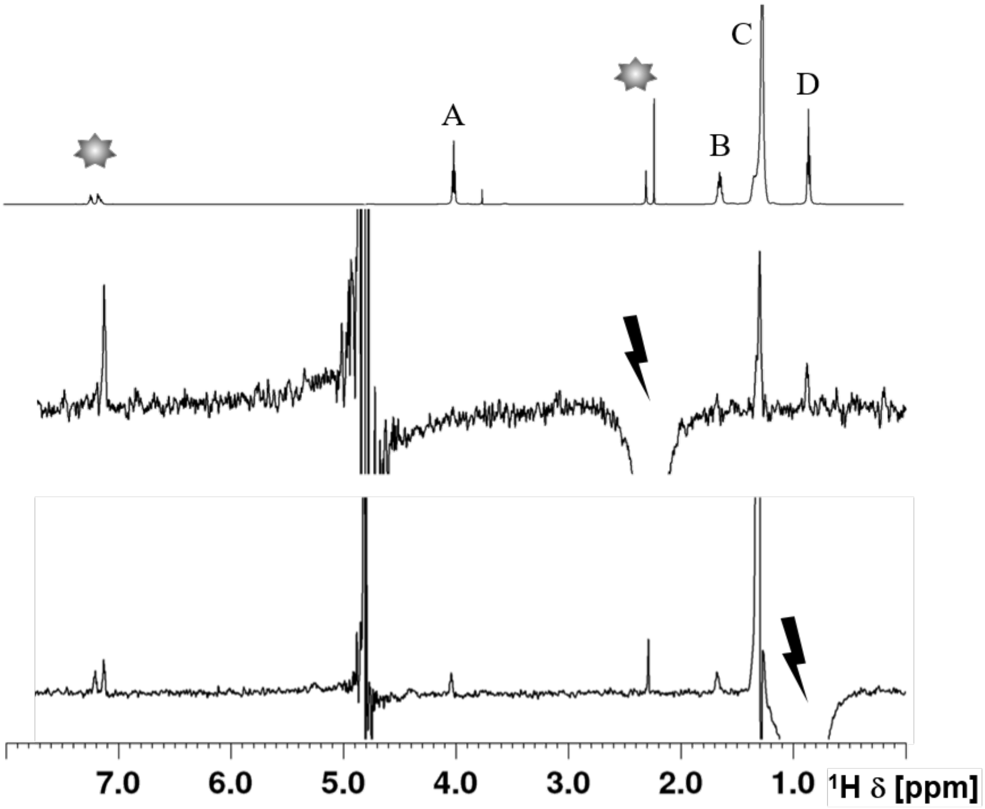


Figure G.14.: 1D NOESY spectra of toluene in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation points (sorted descending): methyl group of phenol and SDS signal *D* (H_2O/D_2O 9:1, 500 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)).

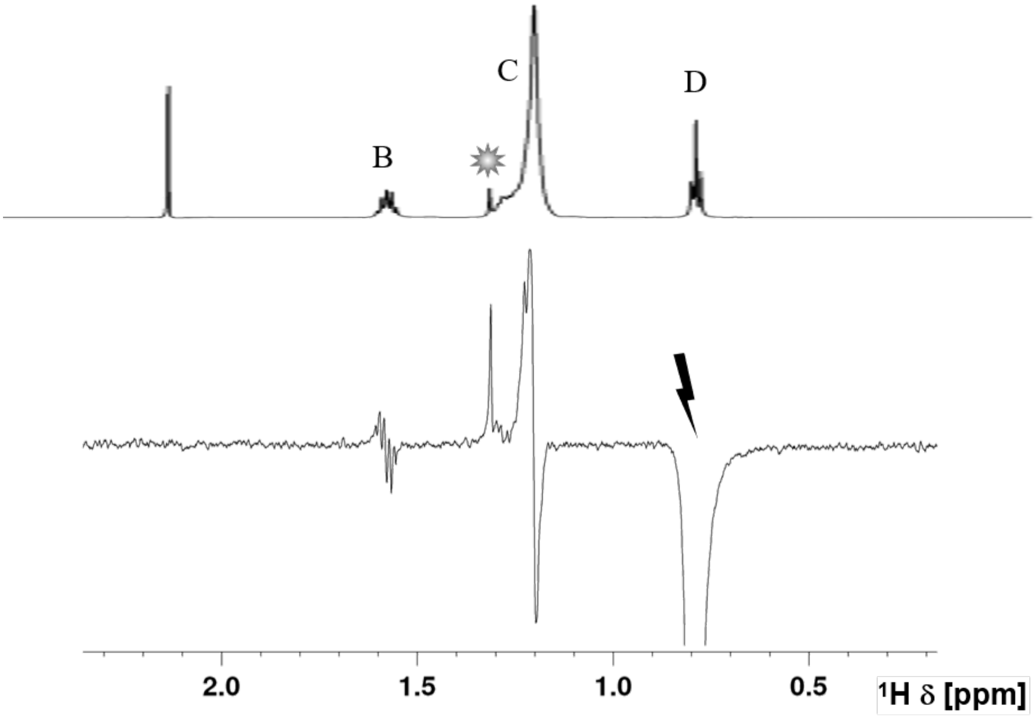


Figure G.15.: 1D NOESY spectrum of cyclohexane in presence of SDS micelles. 1D proton reference spectrum is shown in the top trace. Irradiation was performed on the SDS signal *D* (H_2O/D_2O 9:1, 600 MHz, 298 K, d1: 15 s, mixing time: 1 s, NS 256, bandwidth excitation: 33 Hz (Gaus1_180r.1000)).

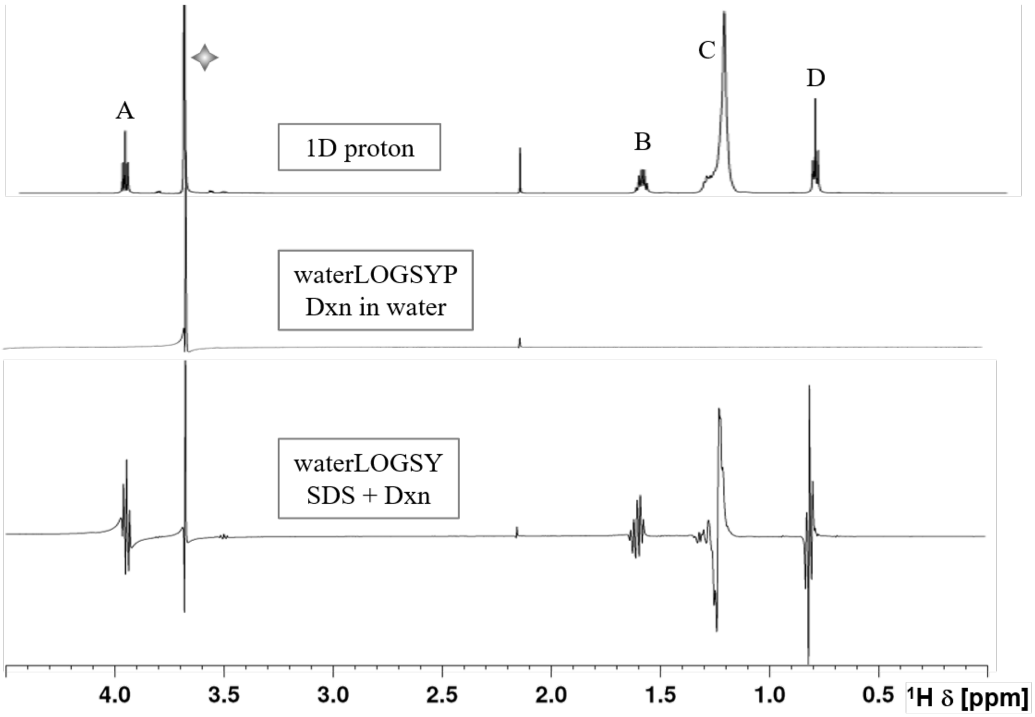


Figure G.16.: WaterLOGSY spectra of dioxane in water (middle) and dioxane in micelle solution (bottom) (600 MHz, 298 K, mixing time: 0.8 sec).

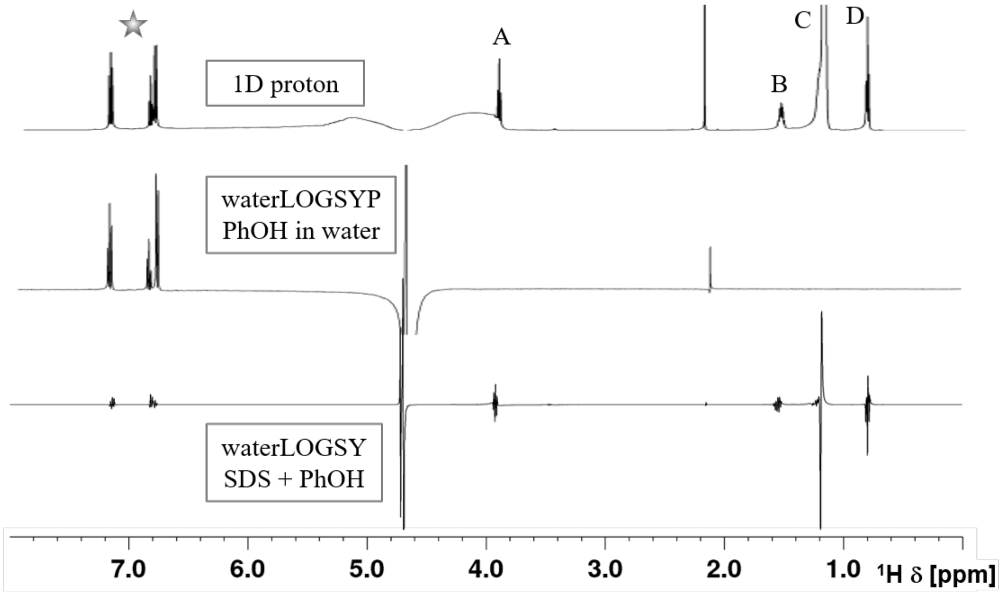


Figure G.17.: WaterLOGSY spectra of phenole in water (middle) and in micelle solution (bottom) (600 MHz, 298 K, mixing time: 1.2 sec).

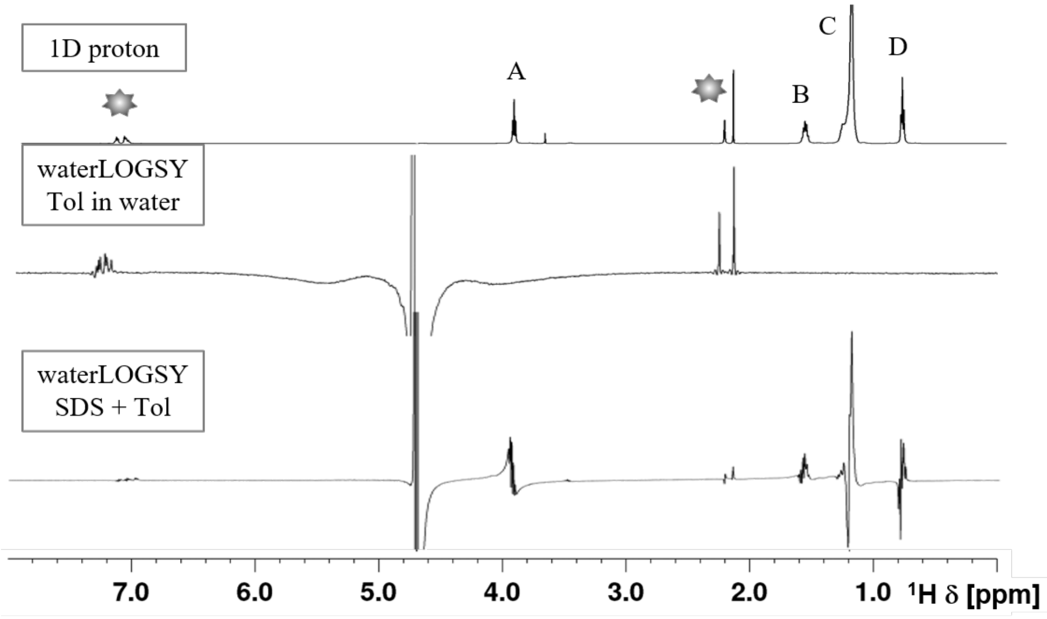


Figure G.18.: WaterLOGSY spectra of toluene in water (middle) and in micelle solution (bottom) (600 MHz, 298 K, mixing time: 0.8 sec).

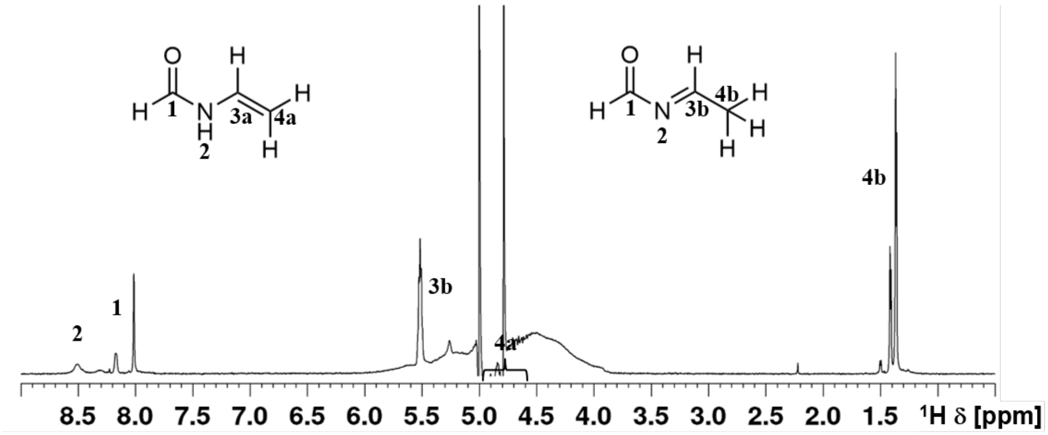


Figure G.19.: 1D ^{1}H NMR spectrum of NVF in water at pH 1 (H_2O/D_2O 9:1, 298 K, 600 MHz).

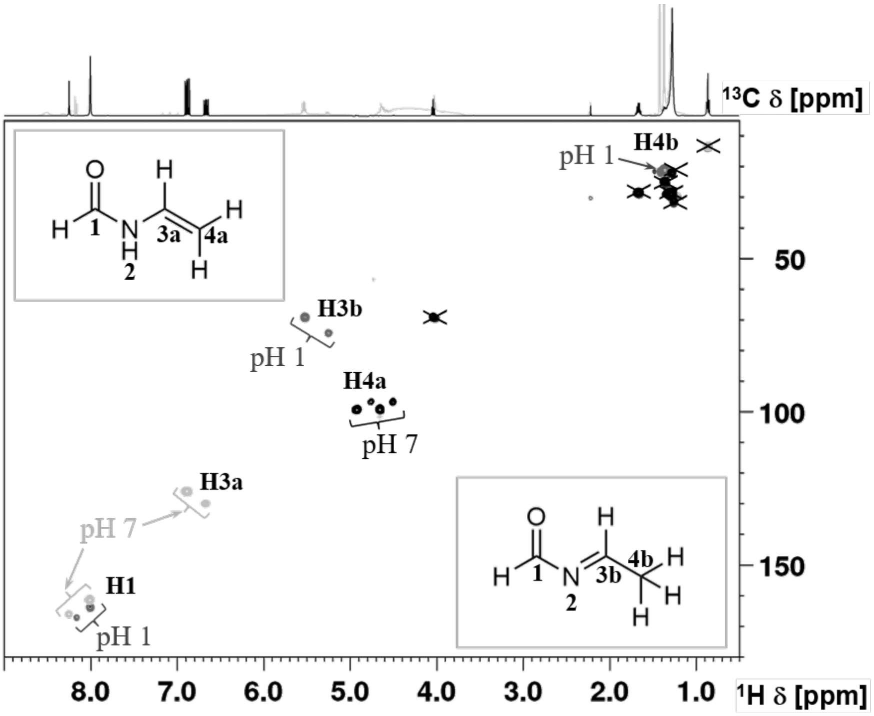


Figure G.20.: Phase sensitive HSQC experiment of substrate 2 at pH 1 and pH 7 in stacked layout. Crossed-out signals belong to micelles. CH/CH_3 show opposite sign to CH_2. Substrate 2 at pH 1 is plotted in darker grey for CH/CH_3 and black for CH_2; for pH 7 CH/CH_3 is light grey and CH_2 in black (H_2O/D_2O 9:1, 298 K, 600 MHz).

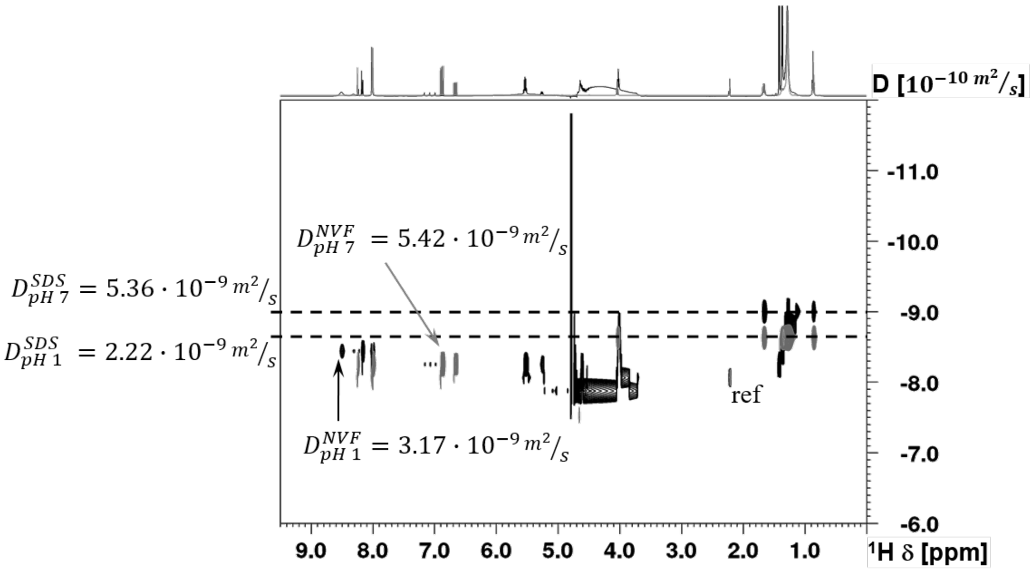


Figure G.21.: 2D DOSY spectra of NVF in pH 1 (black) and pH 7 (grey) solution in presence of 12 mM micelles. All spectra are referenced with an external acetone/water capillary, which was set to a diffusion coefficient of –8.1 log(m^2/s) (H_2O/D_2O 9:1, 600 MHz, 298 K, NS 32, 8k (f1) 32 (F2), linear ramp (2–98%), d1 1.5 s, d20 70 ms).

H. Contribution from others

Cell penetrating peptides were designed and synthesised in the working group of Prof *Ines Neundorf* (Institute of Biochemistry, University of Cologne, Zülpicher Straße 47a, 50674 Köln) by *Lucia Feni* and *Sara Parente*. Prof *Umberto Piarulli* (DISAT, University of Insubria, Como, Italy) designed the 2,5-diketopiperazines.
The Trp-cage mini-protein sequence was also synthesised in Prof *Ines Neundorf* group and kindly provided for NMR spectroscopic investigations.
Proline-rich peptide sequences were designed, synthesised and crystallised in the working group of Prof *Ulrich Baumann* (Institute of Biochemistry, University of Cologne, Zülpicher Straße 47a, 50674 Köln) by *Christian Pichlo*.
^{15}N labelled TMA·HCl was purchased and kindly provided by Prof. Albrecht Berkessel (Institute of Chemistry, University of Cologne, Greinstraße 4, 50939 Köln).

Erklärung zur Dissertation

gemäß der Promotionsordnung vom 12. März 2020

Diese Erklärung muss in der Dissertation enthalten sein.
(*This version must be included in the doctoral thesis*)

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht."

Teilpublikationen:

L. Feni, L. Jütten, S. Parente, U. Piarulli, I. Neundorf, D. Diaz, Cell-penetrating peptides containing 2,5-diketopiperazine (DKP) scaffolds as shuttles for anti-cancer drugs: conformational studies and biological activity, Chem. Commun. 2020, 56, 5685-5688.

C. Pichlo, L. Jütten, F. Wojtalla, M. Schacherl, D. Diaz, U. Baumann, Molecular determinants of the mechanism and substrate specificity of Clostridium difficile proline-proline endopeptidase-1, J. Biol. Chem. 2019, 294 (30), 11525–11535.

Datum, Name und Unterschrift

06.04.2021, Linda Jütten

Curriculum Vitae

Linda Jütten
University of Cologne, Department of Chemistry
Institute of Organic Chemistry – NMR Department, Cologne, Germany
☎ +49 176 23625857, ✉ LJuetten@smail.uni-koeln.de

Education

04/2017-today	**PhD Program in Chemistry**, University of Cologne
04/2015-03/2017	**MSc in Chemistry** , University of Cologne
	Thesis title: "NMR based studies of permethylated monosaccharides and proline-rich peptides", *Advisor:* Dr. Dolores Diaz
10/2011-06/2015	**BSc in Chemistry**, University of Cologne
	Thesis title: "Synthesis and transformation of ProM-2", *Advisor:* Prof. Hans-Günther Schmalz
02/2011-09/2011	**Study preparatory intership**, Juelich Research Centre
2006-2010	**Abitur**, Gymnasium Haus Overbach
	Jülich-Barmen, Germany
2000-2006	**Fachoberschulreife**, Mädchengymnasium Jülich
	Jülich, Germany
1997-2000	Katholische Grundschule Haaren-Waldfreucht, Germany
1996-1997	Gemeinschaftsgrundschule Beckrath, Germany

Posters

- L. Jütten, D. Diaz, NMR based studies on the role of osmolytes on protein folding, *bio*-N^{3}MR Annual Meeting, Forschungszentrum Jülich, Jülich, Germany.
- L. Jütten, L. Feni, S. Parente, I. Neundorf, D. Diaz, How to Derive the 3D Structure of Non-Labeled Peptides: Protocol for NMR-Based Structural Elucidation **2018**, 1st PhD Symposium, University of Cologne, Cologne, Germany.
- L. Jütten, L. Feni, S. Parente, U. Piarulli, I. Neundorf, and D. Diaz, The NMR structure of 2,5-Diketopiperazine containing peptides in solution and in membrane mimetic medium (SDS), 40th FGMR Annual Discussion Meeting **2018**, Leipzig, Germany.
- L. Jütten, C. Pichlo, U. Baumann, D. Diaz, The NMR solution structure of a proline-rich short peptideand the interaction with its receptor, 39th FGMR Annual Disussion Meeting **2017**, Bayreuth, Germany.

Publications

- L. Feni, L. Jütten, S. Parente, U. Piarulli, I. Neundorf, D. Diaz, Cell-penetrating peptides containing 2,5-diketopiperazine (DKP) scaffolds as shuttles for anti-cancer drugs: conformational studies and biological activity, *Chem. Commun.* **2020**, *56*, 5685-5688.
- C. Pichlo, L. Jütten, F. Wojtalla, M. Schacherl, D. Diaz, U. Baumann, Molecular determinants of the mechanism and substrate specificity of Clostridium difficile proline-proline endopeptidase-1, *J. Biol. Chem.* **2019**, *294*(30), 11525–11535.
- L. Jütten, K. Ramirez-Gualito, A. Weilhard, B. Albrecht, G. Cuevas, M. Fernandez-Alonso, J. Jimeenez-Barbero, N. Schloerer, D. Diaz, Diffusion NMR studies on the role of solvation of sugars and their interaction with aromatic moieties, *ACS Omega*, **2018**, *3*, 536-543

Köln, der 06.04.2021