Structure and function of the autotransporter protein YadA

A solid-state MAS NMR study

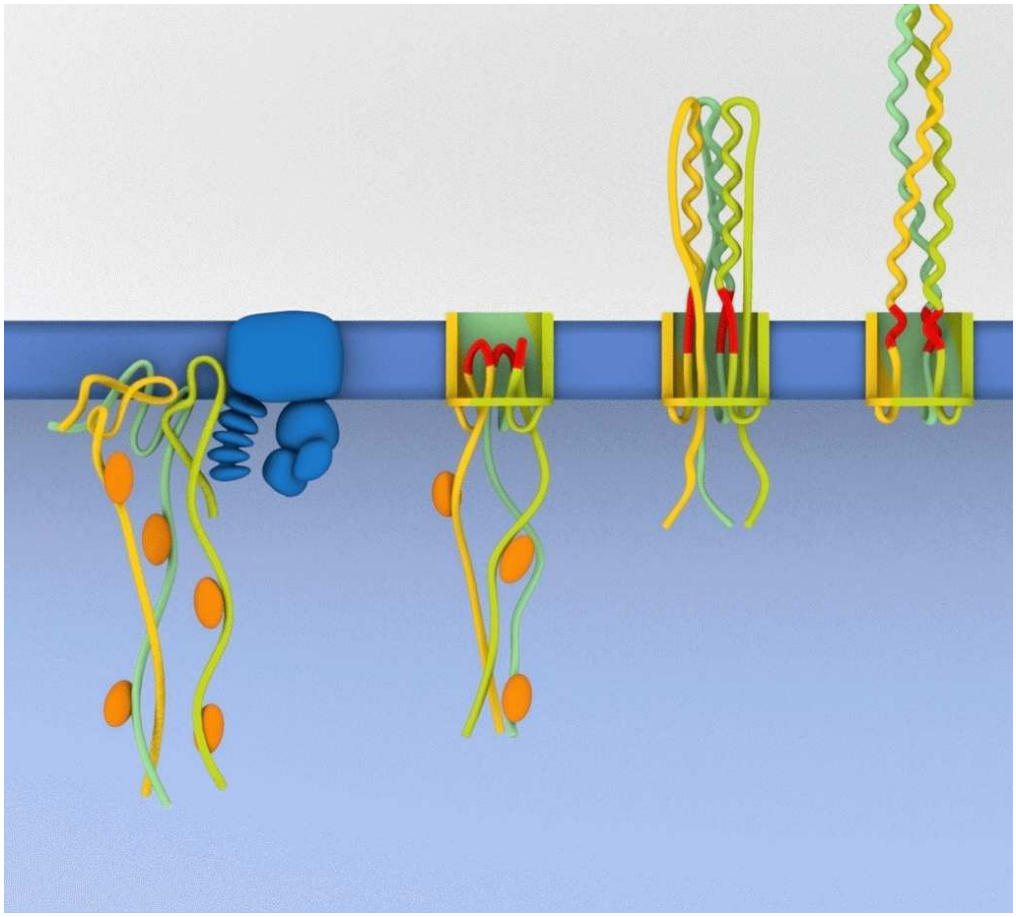
Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin



By

Shakeel Ahmad Shahid

Muridke, Pakistan

March 2012

This work was accomplished in the period from October 2008 to March 2012 at Leibniz-Institut für Molekulare Pharmakologie (FMP) Berlin, and Max Planck Institute for Developmental Biology Tübingen, in supervision of Dr. Barth-Jan van Rossum, PD. Dr. Dirk Linke and Prof. Dr. Hartmut Oschkinat

1st Reviewer: Prof. Dr. Hartmut Oschkinat

2nd Reviewer: PD. Dr. Dirk Linke

date of defence: 16 November 2012

Acknowledgements

This doctoral work might not have been accomplished without the help of several people. I would like to avail the opportunity of acknowledging them.

Firstly, I would thank my immediate supervisor Dr. Barth-Jan van Rossum. He has given me all the moral and scientific support, encouragement and motivation to groom from the very beginning of this project; overlooked many of my mistakes and helped me to be a confident science student. His good instructions, friendly company, skills to explain complicated ideas in simple ways and irreplaceable advices at every step have helped me to excel in this project. It would not have been possible to write this thesis without his guidance.

I would also acknowledge my second supervisor, PD Dr. Dirk Linke from Max Planck Institute for Developmental Biology, Tübingen. I am thankful to him for his trust on me and for giving me the opportunity to work on such a fascinating project. His kind assistance, inspirational teaching, thought-provoking discussions and brainstorming about the project have helped me a lot to understand and complete this research project.

I would like to express my deepest gratitude to my Professor, Prof. Dr. Hartmut Oschkinat for his cooperation and for the opportunity to complete my dissertation in his lab. I found him very sympathetic, caring, and supportive and always ready to help throughout my stay at FMP. Indeed, I have profited myself to a great deal in his research group, being a part of his dedicated team and working in this excellent scientific environment.

Similarly, my thanks would go to Prof. Dr. Andrei Lupas from Max Planck Institute for Developmental Biology, Tübingen for his special interest and useful suggestions on this project. I am thankful to him for his invaluable moral and scientific advices during my stays in Tübingen. I found him very soft-hearted and eager to see the students from third-world countries progressing in the field of science.

It is my pleasure to thank Dr. Benjamin Bardiaux and Dr. Michael Habeck for the structure calculation of YadA-M. It was a perfect collaboration and I learned a lot from both of them. They were ever willing to help me and have offered their support from the day one of this

project. I am thankful to them for partly reading the manuscript of the thesis and for giving their useful comments to improve it.

I have many friends to say thanks in the institute. From my officemates, I would acknowledge Stefan Markovic and Trent Franks who not only had helped me recording NMR spectra but also offered a great company of friendship. I am thankful to Trent for the useful discussions on NMR and for his precious suggestions during my thesis writing. I am grateful to Monica Santos de Freitas, Anne Wartenberg, Britta Kunert and Marco Röben for thoughtful discussions in the office.

I would acknowledge my colleagues, Janet Zapke, Anup Chowdhary, Joren Retel, Andrew J. Nieuwkoop, Anja Voreck, Ümit Akbey, Arne Linden, Sascha Lange and Mathias Hiller for providing a delightful atmosphere in the institute. I am indebted to Iftikhar Ahmad, Arne Linden and Sascha Lange for the German translation of my thesis' summary. I am also thankful to Victoria A. Higman for her support in partly checking the chemical shift assignments of YadA-M. I am thankful to Ahsanullah for the great time which we had at the campus. Many thanks to the friendliest Andrea Steuer for her help in the official work and to Gabriele Schumacher for handling the bureaucracy.

Indeed, this work is indebted to the financial support from Baden Württemberg Stiftung. I am highly thankful to them for financing this project.

I owe a great deal to the great country, Germany and its wonderful cities Berlin and Tübingen for providing all the necessary support and opportunities for a foreign student to achieve his dream.

In the end, I would thank my family members, especially my parents, whose prayers and best wishes were always with me throughout my studies.

I thank all of you.

Shakeel Ahmad Shahid

Berlin, March 2012

Dedication

To my father, **Muhammad Yousaf Shahid**:

Whose keen interest in my education, encouragement, and wish to see me become a doctor has inspired me to achieve this milestone.

Contents

1 GENERAL INTRODUCTION 1

1.1 SECRETION SYSTEMS IN GRAM-NEGATIVE BACTERIA 2
1.2 TRIMERIC AUTOTRANSPORTER ADHESINS (TAAS) 4
1.2.1 MOLECULAR STRUCTURE OF TAAS 7
1.3 MODELS OF AUTOTRANSPORT IN TAAS 7
1.3.1 MULTIMERIC MODEL 7
1.3.2 SINGLE-CHAIN MODEL 8
1.3.3 ASSISTED MODEL 9
1.4 YERSINIA ADHESIN A (YADA) 10
1.4.1 YERSINIOSIS 14
1.5 PROTEIN STRUCTURE DETERMINATION 14
1.5.1 INTRODUCTION 14
1.5.2 FUNDAMENTALS OF NMR 15
1.5.3 SOLID-STATE NMR 18
1.5.4 CROSS-POLARIZATION MAGIC-ANGLE SPINNING NMR 20
1.6 INFERENTIAL STRUCTURE DETERMINATION (ISD) 22
1.6.1 INTRODUCTION 22
1.6.2 STRUCTURE DETERMINATION FROM CONVENTIONAL METHODS 23
1.6.3 PRINCIPLES OF ISD 25
1.6.4 EXAMPLE 28
1.7 AIM OF THE PROJECT 29

2 MATERIALS AND METHODS 31

2.1 YADA-M BIOSYNTHESIS, PURIFICATION AND CRYSTALLIZATION 31
2.2 SOLID-STATE MAS NMR 32
2.2.1 HOMONUCLEAR ^{13}C-^{13}C CORRELATION SPECTRA 33
2.2.2 HETERONUCLEAR ^{15}N, ^{13}C CORRELATION SPECTRA 34
2.2.3 METHYL-FILTERED SPECTRA 35
2.2.4 CHHC AND NHHC 36
2.3 DATA PROCESSING AND ANALYSIS 36

3 CHEMICAL SHIFT ASSIGNMENT OF YADA-M 37

3.1 INTRODUCTION 37
3.2 YADA-M, AN IDEAL MEMBRANE PROTEIN FOR SOLID-STATE MAS NMR 38
3.3 RESIDUE SPECIFIC ASSIGNMENT OF YADA-M 39
3.3.1 STARTING POINTS FOR SEQUENCE SPECIFIC ASSIGNMENT 42
3.4 SEQUENCE-SPECIFIC ASSIGNMENT OF YADA-M 43
3.4.1 NCACX AND NCOCX 44
3.4.2 SELECTIVE TRANSFERS USING DREAM MIXING 48
3.4.3 CONSISTENCY CHECK OF THE SEQUENCE-SPECIFIC ASSIGNMENT 50
3.4.4 ASSIGNMENTS FROM 15N – 13C TEDOR 53
3.4.5 METHYL-FILTERED POST-C7 55
3.5 SUMMARY 56

4 THE STRUCTURE OF YADA-M 59

4.1 INTRODUCTION 59
4.2 SECONDARY STRUCTURE OF YADA-M 60
4.3 TERTIARY STRUCTURE 62
4.3.1 STRUCTURE-DEFINING NMR RESTRAINTS 62
4.3.2 THE SHEAR NUMBER (S) FOR β-BARREL OF YADA-M 65
4.3.3 FILTERED RESTRAINTS 70
4.4 *DE NOVO* STRUCTURE DETERMINATION OF YADA-M 74
4.4.1 DISTANCE RESTRAINTS USED BY ISD 74
4.4.2 STRUCTURE OF THE MONOMER (SIMULATION 1-3) 76
4.4.3 ASSEMBLY OF THE TRIMER (SIMULATION 4-5) 80
4.4.4 FINAL REFINEMENT (SIMULATION 6) 82
4.5 STRUCTURE OF YADA-M BY ARIA 86
4.6 DESCRIPTION OF THE STRUCTURE 87
4.7 CONCLUSION 87

5 DISCUSSION 88

5.1 DIFFERENT INTERACTIONS THAT STABILIZE YADA-M 88
5.1.1 INTERACTIONS INVOLVING AROMATIC RESIDUES 88
5.1.2 INTERACTIONS INVOLVING CHARGED AND POLAR RESIDUES 89
5.1.3 INTERACTIONS INVOLVING HYDROPHOBIC RESIDUES 90
5.2 CRYSTAL CONTACTS OBSERVED IN SOLID-STATE MAS NMR SPECTRA 90

5.3 ROLE OF GLYCINES IN YADA-M 93
5.3.1 SIGNIFICANCE OF HIGHLY CONSERVED G72 94
5.4 ASSA REGION 96
5.4.1 EVOLUTIONARY ANALYSIS OF YADA-M 99
5.4.2 INTERACTIONS OF ASSA RESIDUES WITH BARREL INTERIOR 101
5.4.3 NON-COVALENT S···OINTERACTIONS 102
5.4.4 THREE SERINE PAIRS IN YADA-M 103
5.5 PROPOSED MECHANISM OF AUTOTRANSPORT IN YADA AND TAAS 104
5.5.1 EVIDENCE OF THE SUGGESTED AUTOTRANSPORT MODEL 106
5.5.2 POSSIBLE EXPLANATION OF THE AUTOTRANSPORT MECHANISM 108

6 BIBLIOGRAPHY 109

7 APPENDICES 116

Abbreviations

MPs	**M**embrane **P**roteins
TAAs	**T**rimeric **A**utotransporter **A**dhesins
YadA	*Y**ersinia*** **A**dhesin **A**
YadA-M	Membrane anchor domain of *Yersinia* Adhesin A
OM	**O**uter **M**embrane
NMR	**N**uclear **M**agnetic **R**esonance
1D, 2D, 3D	**O**ne-**d**imensional, **T**wo-**d**imensional, **T**hree-**d**imensional
MAS	**M**agic-**A**ngle **S**pinning
CP	**C**ross **P**olarization
DARR	**D**ipolar **A**ssisted **R**otational **R**esonance
PDSD	**P**roton **D**riven **S**pin **D**iffusion
PAR	**P**roton **A**ssisted **R**ecoupling
REDOR	**R**otational **E**cho **D**ouble **R**esonance
TEDOR	**T**ransferred **E**cho **D**ouble **R**esonance
POST-C7	**P**ermutationally **O**ffset **S**tabilized **C7**
CPPI	**C**ross **P**olarization **P**olarization **I**nversion
DD	**D**ipolar **D**ephasing
ISD	**I**nferential **S**tructure **D**etermination
ARIA	**A**mbiguous **R**estraints for **I**terative **A**ssignment
ppm	**P**arts **P**er **M**illion
γ	Gyromagnetic ratio
rf	**R**adio **F**requency
FID	**F**ree **I**nduction **D**ecay
CSA	**C**hemical **S**hift **A**nisotropy
kHz, MHz	**k**ilo **H**ertz, **m**ega **H**ertz
ASSA region	Region from residues A37-A41 in YadA-M
RMC	**R**eplica-exchange **M**onte **C**arlo
RMSD	**R**oot **M**ean **S**quare **D**eviation
MDSA	**M**olecular **D**ynamics-based **S**imulated **A**nnealing

Summary

Bacterial infection starts with colonization of the host organism. For this purpose, pathogenic bacteria assemble several molecular complexes on their surfaces such as pili, fimbriae, membrane-anchored fibres etc., collectively called adhesins. *Yersinia* adhesin A (Yad A) is one of the virulence factors of *Yersinia Enterocolitica* and *Yersinia pseudotuberculosis* and is the prototype of non-fimbrial, non-pilus Trimeric Autotransporter Adhesins (TAAs). The C-terminal transmembrane anchor domain of YadA forms a highly stable trimeric β-barrel in the outer membrane. The general feature of all TAAs is that they export their own coiled-coil stalk, with their head at the N-terminus (stalk and head domains are collectively called passenger domain) into the extracellular milieu without aid of ATP. Once outside the cell, the sticky head domain can bind to extracellular matrix proteins of the host such as, fibronectin, collagen, and laminin. Invasion of the host can lead to various enteric foodborne diseases, e.g., enterocolitis, acute enteritis, diarrhoea, mesenteric lymphadenitis, septicaemia, and reactive arthritis to autoimmune disorders. Although several attempts have been made to understand the autotransport mechanism, detailed insight in the mechanism based on experimental evidence is missing. Knowledge of the structure and dynamics of domains involved in autotransport can contribute to understanding the autotransport mechanism and can shed light on the process of bacterial adhesion.

Membrane proteins are heavily underrepresented in the Protein Data Bank (PDB) despite being involved in numerous cellular processes of phenomenal importance and being prime focus of pharmaceutical industry. The large molecular size, slow tumbling motion, low solubility, intrinsic heterogeneity, and the presence of flexible loops are inherent characteristics of membrane proteins which make them difficult to study by most high resolution techniques. Solid-state NMR (ssNMR), on the other hand, does not rely on the availability of high quality crystals, rapid molecular tumbling, or the size of the complex; therefore, ssNMR serves as a promising method for structural studies of quasi-immobilized solid or solid-like macromolecular complexes including membrane proteins and amyloid fibrils. In addition, ssNMR is also able to give information about the dynamics of pharmaceutically important protein-protein and protein-ligand binding sites.

The first step towards the structure determination by magic-angle spinning (MAS) ssNMR as reported in this thesis was the sequence-specific chemical shift assignment of virtually all spin systems in the symmetric YadA-M trimer (11.3 kDa, 105 amino-acid residues per monomer). Several two-dimensional (2D) and three-dimensional (3D) homo- and heteronuclear correlation spectra were recorded on a single, uniformly ^{13}C, ^{15}N labelled micro-crystallized YadA-M (membrane anchor domain of YadA) preparation. 2D homonuclear ^{13}C correlation spectra served to identify different spin systems upon their characteristic fingerprint cross-peak pattern. Sequential links among the spin systems were established by selective transfer between the backbone ^{15}N amide to either the Cα or C` in 3D NCACX and 3D NCOCX experiments. The chemical shift values (^{13}C and ^{15}N) were used to predict the secondary structure of YadA-M which consists of four anti-parallel β-strands and one α-helix for each monomer. In the next step, various experiments under specific and non-specific recoupling conditions at multiple mixing times were recorded to achieve a sufficient number of medium and long-range structural restraints. The definition and registry of intra- and inter-monomer β-strands was determined by careful observation of 2D ChhC type spectra which are rich in long range cross-strand contacts. In addition, 2D PAR, DARR, TEDOR, NhhC and PDSD spectra at different mixing times provided complementary distance information. Single- and double-methyl filtered 2D DARR spectra with extended mixing times were a useful addition to the data set and provided distance restraints that originated specifically from side-chain methyl groups. Long-range contacts between strands and helices were identified that assisted in defining the trimeric structure of the membrane anchor domain. Analysis of the NMR data set resulted in a substantial amount of structure-defining restraints that served as input for multiple computational tools, such as ISD and ARIA, to derive the tertiary and quaternary (trimeric) structure of YadA-M. Thus, a *de novo* structure of the β-barrel membrane protein was obtained on the basis of MAS solid-state NMR data.

It was found that a quartet of helical residues (named the “ASSA” region) displays random-coil-like chemical shifts, low order-parameters, reduced helix propensity, and a drop in signal intensity, pointing towards a relatively increased flexibility. Evolutionary studies revealed that this region is highly conserved and shows high preference for small side-chain residues. From coiled-coil analysis, a switch from a classical seven-residue repeat

to a straighter, eleven-residue repeat was observed, which ends at the ASSA region. Based on combined structural and evolutionary studies, the work reported in this thesis is in strong favour of the 'hairpin model' as mechanism of the autotransport; the ASSA region was identified as flexible linker that functions as the hairpin at the onset of the autotransport and plays an important role in the C- to N-terminal passage of the polypeptide through the β-barrel. Residues in the ASSA region show interactions with the small side-chain residues of the interior wall, i.e., A37-A68, A41-G61, and L45-G72, which necessarily stabilize the ASSA region after the transport is accomplished. In addition, an uncommon, non-covalent S···O interaction was reported which was observed between the side-chain oxygen atoms of S38-S39 and the sulphur atom of M96. Intermolecular peaks (crystal contacts) were also observed which gave a tilted up-and-down organization for the neighbouring trimers. Amino-acid residues involved in crystalline contacts exhibit heterogeneously broadened NMR signals.

This thesis demonstrates, for the first time, that solid-state MAS NMR is able to solve the structure of membrane proteins from a single, uniformly ^{13}C, ^{15}N labelled sample. The structure of YadA-M provides important insights into the structural and functional aspects of the autotransporter domain of YadA; moreover, the results are in good agreement with the hairpin model, which is most likely conserved for all trimeric autotransporters.

Zusammenfassung

Bakterielle Infektionen beginnen mit dem Eindringen in die Wirtszelle. Zu diesem Zweck exprimieren pathogene Bakterien verschiedene molekulare Komplexe an ihrer Oberfläche, wie z.B. Pili, Fimbrien, Membran-verankerte Fäden usw. Diese werden zusammenfassend als Adhäsine bezeichnet. *Yersinia* Adhäsin A (YadA) ist ein Virulenzfaktor von *Yersinia enterocolitia* und *Yersinia pseudotuberculosis* sowie ein typisches Beispiel eines trimären Autotransporter-Adhäsins (TAA), welche weder Fimbrien noch Pili sind. Die C-terminale Ankerdomäne von YadA formt ein höchst stabiles trimäres β-Fass in der äußeren Membran. Alle TAAs können ohne Hilfe von ATP ihren eigenen Doppelhelix-Fortsatz in die extrazelluläre Matrix exportieren. Am N-Terminus dieses Fortsatzes befindet sich eine Kopfdomäne und ihre Gesamtheit wird als „passenger" Domäne bezeichent. Außerhalb der Zelle kann sich die klebrige Kopfdomäne an verschiedene Zelltypen und an die extrazellulären Matrixproteine des Wirts (z.B. Fibronectin, Kollagen und Laminin) binden. Die Infektion des Wirts kann zu verschiedenen Lebensmittelvergiftungen und nachfolgenden Krankheitsbildern führen, wie z.B. zu Enterocolitis, akute Enteritis, Diarrhoe, mesenterialer Lymphadenitis, Blutvergiftung, bis hin zu Autoimmunstörungen durch reaktive Arthritis. Obwohl verschiedene Versuche unternommen wurden den Autotransport-Mechanismus zu verstehen, fehlt es an experimentellen Daten für ein detailliertes Verständnis. Erkenntnisse bezüglich der Struktur und Dynamik der am Autotransport beteiligten Domänen können zum Verständnis des Autotransport-Mechanismuses beitragen sowie den Prozess der bakteriellen Adhäsion erhellen.

Obwohl Membranproteine in viele zelluläre Prozesse maßgeblich involviert sind und daher im Fokus der pharmazeutischen Industrie stehen, sind sie in der Protein Datenbank (PDB) stark unterrepräsentiert. Membranproteine sind aufgrund ihrer molekularen Größe, der damit verbundenen langsamen Taumelbewegung, sowie ihrer schlechten Löslichkeit, ihrer strukturellen Heterogenität und ihren flexiblen Schleifen schlecht mit hochauflösenden Techniken zu studieren. Festkörper-NMR (solid-state NMR; ssNMR) hingegen ist nicht abhängig von der Qualität der Kristalle, einer schnellen molekularen Bewegung oder der Größe des Komplexes. Daher ist ssNMR eine vielversprechende Methode für strukturelle Studien an immobilisierten, makromolekularen Komplexen von Membranproteinen oder

Amyloidfibrillen. Zudem ist ssNMR in der Lage, Informationen über die Dynamik pharmakologisch wichtiger Strukturen zu geben, wie z.B. die Bereiche von Protein-Protein Interaktionen oder Liganden-Bindungen.

Der erste Schritt zur Strukturbestimmung durch ssNMR unter Drehung im magischen Winkel (magic-angle spinning; MAS) ist, wie in dieser Arbeit beschrieben, die sequenzspezifische Zuordnung der chemischen Verschiebungen nahezu aller Spinsysteme im symetrischem YadA-M Trimär (11,3 kDa, 105 Aminosäuren pro Monomer). Verschiedene 2D und 3D homo- und hetronukleare Korrelationsspektren wurden von uniform ^{13}C, ^{15}N markiertem, mikrokristallinem YadA-M aufgenommen. 2D homonukleare ^{13}C Korrelationsspektren wurden benutzt, um verschiedene Spinsysteme durch ihre charakteristischen Muster aus Kreuz-Signalen zu identifizieren. Sequenziell verknüpft wurden die Spinsysteme durch einen selektiven Transfer ausgehend von den ^{15}N Amiden der Hauptkette zu entweder Cα oder C` mit 3D NCACX und 3D NCOCX Experimenten. Die chemischen Verschiebungen (^{13}C und ^{15}N) wurden verwandt um die sekundäre Struktur des YadA-M zu bestimmen, wobei jedes Monomer aus vier anti-parallelen ß-Strängen und einer α-Helix besteht. Anschließend wurden verschiedene Experimente unter spezifischen und nicht-spezifischen „recoupling“ Bedingungen und mit unterschiedlichen Mischzeiten aufgenommen. Damit konnte eine hinreichende Anzahl von Atomabständen mittlerer und langer Distanz bestimmt werden. Die intra- und intermonomäre Anordung der β-Stränge wurde durch 2D ChhC Spektren ermittelt. Diese zeigen viele Kontakte über lange Entfernungen und zwischen benachbarten Strängen. Weiterhin lieferten 2D PAR, DARR, TEDOR, NhhC und PDSD Spektren mit unterschiedlichen Mischzeiten ergänzende Abstands-Informationen. Einfach und doppelt Methyl-gefilterte 2D DARR Spektren mit verlängerten Mischzeiten ergänzten die Abstands-Informationen, ausgehend von den Methylgruppen der Seitenketten. Zusätzlich halfen Kontakte über lange Abstände zwischen den einzelnen Strängen und Helices bei der Bestimmung der trimären Struktur der Ankerdomäne. Insgesamt resultierte die Auswertung der NMR Daten in einer Vielzahl von struktur-bestimmenden Abstands-Einschränkungen. Diese bildeten die Grundlage für verschiedene computergestütze Methoden (z.B. ISD und ARIA), um die tertiäre und quartäre (trimäre) Struktur des YadA-M abzuleiten. So wurde de novo die Struktur des Membranproteins als ß-Fass bestimmt.

Es konnte gezeigt werden, dass ein Quartett von Seitenketten in einer Helix (ASSA Region) chemische Verschiebungen aufweist, die typisch für eine ungeordnete Sekundärstruktur, jedoch untypisch für eine Helix sind. Des Weiteren deutet eine reduzierte Signalintensität auf eine relativ hohe Flexibilität dieser Seitenketten hin. Evolutionsstudien haben gezeigt, dass diese Region stark konserviert ist und meistens aus Aminosäuren mit kleinen Seitenketten beteht. Eine Analyse der Doppelhelix ergab, dass statt des klassischen Modells (sieben Aminosäuren pro Windung) eine geradere Form mit elf Aminosäuren pro Windung vorliegt. Diese Helix endet an der ASSA Region. Basierend auf einer Kombination der strukturellen Daten und evolutionärer Studien, folgt der Mechanismus des Autotransports dem sogenannten „hairpin"-Modell. Die ASSA Region wurde dabei als das flexible Bindeglied identifiziert, welches beim Beginn des Autotransports den „hairpin" formt. Sie spielt damit eine entscheidende Rolle bei der Passage des Polypeptids vom C- zum N-Terminus durch das ß-Fass. Die Seitenketten der ASSA Region interagieren mit kleinen Seitenketten der Innenwand, z.B. A37-A68, A41-G61 und L45-G72, und stabilisieren zwangsläufig die ASSA Region nach erfolgtem Transport. Zudem wurde eine ungewöhnliche, nicht-kovalente S-O Interaktion zwischen den Sauerstoff Seitenketten-Atomen von S38, S39 und dem Schwefelatom von M96 beobachtet. Intermolekulare Signale (Kristall-Kontakte) wurden gemessen, welche auf eine alternierende Anordung (oben-unten) von benachbarten Trimären hinweisen. Resonanzen von Aminosäuren, die an Kristallkontakten beteiligt sind, sind heterogen verbreitert.

Die vorliegende Dissertation zeigt erstmalig, dass ss-MAS-NMR geeignet ist, die Struktur von Membranproteinen unter Verwendung einer einzigen uniform ^{13}C, ^{15}N markierten Probe aufzuklären. Die Struktur von YadA-M ermöglicht einen wichtigen Einblick in die strukturellen und funktionellen Aspekte der Autotransport-Domäne von YadA. Des Weiteren stützen die Resultate das „hairpin"-Modell, welches höchst wahrscheinlich für alle trimären Autotransporter konserviert ist.

1 General Introduction

Membrane proteins lie at the heart of most cellular processes, acting as channels, pumps, enzymes and receptors. The fact that mutations in the membrane proteins (MPs) directly or indirectly lead to diseases like epilepsy, retinitis pigmentosa, cystic fibrosis and nephrogenic diabetes insipidus (Hubner and Jentsch 2002; Schoneberg, Schulz et al. 2004), makes them important drug targets. In fact, 90% of all drug targets are MPs that constitute roughly 30% of the proteome. Knowledge about the structure of MPs is the most important pre-requisite to understand their biological functions. However, structure determination of MPs is generally not a trivial job and is accompanied by various experimental complexities. This is the main reason why, despite being involved in numerous cellular processes of high biological importance, the number of high-resolution structures of MPs deposited in the Protein Data Bank (PDB) makes up less than 1% of all entries. This low representation of the MPs is due to several reasons: first, the purification of MPs in itself is a wearisomely difficult task. This is because MPs dislike to be removed from their native lipid environment, and finding the right conditions and detergent is highly tedious. Second, MPs are infamous for being crystallized; any degree of flexibility in the crystal lattice induced by presence of detergent micelles is detrimental for growing high quality 3D crystals, hence this renders them difficult to be studied by X-ray crystallography. Finally, their relatively large molecular size, heterogeneity, presence of flexible loops and slow molecular tumbling motion which is further reduced in the native lipid bilayer, render them difficult to be studied with solution NMR spectroscopy. Solid-state NMR, on the other hand, does not rely on availability of high-quality crystals or rapid molecular tumbling, hence serves as a promising method for structural studies of MPs (McDermott 2009). It is an emerging biophysical technique which has successfully been applied to quasi-immobilized solids or solid-like macromolecular complexes, including MPs and amyloid systems (Judge and Watts 2011). Solid-state NMR is a unique tool to achieve structure of biomolecules under near-physiological conditions and to probe their dynamics which is directly related to their biological functioning. In short, solid-state NMR is not only eligible to fill the gap left by other high-resolution techniques but also able to give information about structure and dynamics of pharmaceutically important protein-protein and protein-ligand binding sites.

This thesis presents the full structure determination of a membrane protein, the transmembrane anchor domain (~45 kDa) of Yersinia Adhesin A (YadA), solved exclusively with solid-state magic-angle spinning (MAS) NMR data, using a single, uniformly $^{13}C/^{15}N$ labelled microcrystalline sample. This was a collaborative project in which sample preparation (Wollmann, Zeth et al. 2006), and structure calculation by ISD (Rieping, Habeck et al. 2005) were carried out in close collaboration with the Max Planck Institute for Developmental Biology, Tübingen. The work described here is entirely based on a single isotope-labelled, micro-crystalline YadA-M sample provided by Dirk Linke (MPI, Tübingen).

The experimental details of the solid-state MAS NMR spectra and the YadA-M sample preparation (Wollmann, Zeth et al. 2006) are given in **Chapter 2**. The most fundamental protein structure determination step by solid-state MAS NMR is the chemical shift assignment of individual nuclei. The strategies used to solve this giant puzzle are described in **Chapter 3**. The determined chemical shift values are directly associated with the electronic environment of the nuclei. Secondary-structure information can be predicted from these chemical shifts using programmes like TALOS+ (Shen, Delaglio et al. 2009) or RCI (Berjanskii and Wishart 2005). The secondary structure information helps to guide evaluation of the spectra. The assignment of multiple-range distance restraints provides tertiary structure information. These distance restraints can be converted into geometrical restraints from which a protein structure can be calculated using automated simulation programmes like ISD (Rieping, Nilges et al. 2008) and ARIA (Bardiaux, Bernard et al. 2009) (**Chapter 4**). An accurate YadA-M structure at atomic resolution, where a distinction between relatively rigid and dynamic domains can be made, is able to reveal some functional aspects of the protein (**Chapter 5**). The cited references are listed in **Chapter 6**. Additional information, for instance the interactions involved in stability of YadA-M, table of chemical shifts, etc. are listed in Appendices, **Chapter 7**.

1.1 Secretion systems in Gram-negative bacteria

Gram-negative bacteria face the challenge of transporting their surface proteins across two membranes, i.e., the inner membrane and the outer membrane. To overcome this barrier, Gram-negative bacteria have developed different protein secretion systems both in inner- and outer membranes. General translocation systems in the inner membrane are the Sec (secretion) system and twin arginine translocation (Tat) system. The former is responsible for

transporting unfolded proteins and the latter for folded proteins across the inner membrane. Both systems perform their duty by recognizing a cleavable N-terminal signal peptide on transport substrates (Leo, Grin et al. 2011). In Gram-negative bacteria, at least eight different secretion systems for delivering proteins to the extracellular environment have been reported. Each type of system is composed of large protein complexes and is labelled with Roman numerals from type I-VIII. **Type I secretion system** transports substrates of various kinds from ions to proteins across both membranes altogether in one step, bypassing the periplasmic compartment (Delepelaire 2004). For this purpose, it makes a continuous channel which is extended from the cytoplasmic to the outer membrane. This system is composed of an inner-membrane ABC transporter and an outer-membrane pore which are coupled through a membrane fusion protein in the periplasm. **Type II secretion system** is a complex apparatus and transports fully folded proteins and enzymes from the periplasm across the outer membrane (Sandkvist 2001). It is a two partner secretion system (TPS) where unfolded proteins, in the first step, are transported across the cytoplasmic membrane into the periplasm by the Sec complex. The target proteins are recognized by the Sec machinery by their N-terminal signal peptide which is cleaved after a successful export into the periplasm. Once in the periplasmic compartment, the proteins are folded into a transportable conformation and are ready to be secreted into the extracellular milieu. T**ype III secretion system** is like a molecular syringe; the pathogens use it as a weapon for injecting toxins into the host cell. This, like type I secretion system, is independent of Sec machinery and exports unfolded proteins directly from cytoplasmic membrane into the host cell in a single step (Cornelis and Van Gijsegem 2000). The injected proteins destroy the functioning and communication system of the cell under attack, making the survival of invading bacteria more favourable. **Type IV secretion system** is composed of up to 12 proteins, and injects the DNA or protein-DNA complex directly from the bacterial cytoplasm into the eukaryotic host cells. This system of transport is used by Gram-negative bacteria to achieve two objectives, i.e., genetic exchange, and delivery of effector molecules to eukaryotic cells. Genetic exchange helps pathogens to adapt the changing environment during host invasion, whereas interaction of the host and effector molecules not only suppresses the host defence systems, but also assists in the bacterial colonization (Cascales and Christie 2003). The molecular apparatus in both type III and IV secretion systems spans both the bacterial inner- and outer membranes to reach the host cell membranes. Both these systems are able to secrete proteins

either in the extracellular milieu or into the eukaryotic cytoplasm. **Type V secretion system** represents the largest group of secreted proteins in Gram-negative bacteria. This system has been studied extensively and is classified into five subclasses (Leo, Grin et al. 2011). Different autotransporters in this secretion system are: monomeric or classical autotransporters (Va), two-partner secretion systems (Vb), trimeric autotransporters (Vc), fused two-partner secretion systems (Vd), and inverted classical autotransporters (Ve). A comprehensive review about this class of autotransporters has been written by (Leo, Grin et al. 2011). In general, the C-terminus of the proteins (referred to as translocator domain, β-domain, C-domain, transport domain, helper domain) in type V(a-e) secretion systems, forms a β-barrel in the bacterial outer membrane and facilitates the N-terminus (referred to as passenger domain, α-domain, N-domain, extracellular domain) to reach the extracellular environment. In all the cases, Sec machinery is responsible for transporting the proteins across the cytoplasmic membrane into the periplasm. The N-terminal passenger domains, in some cases are cleaved after the autotransport, whereas in others, both C-terminal β-barrel and the N-terminal passenger domain remain intact before and after the autotransport. Despite many known details about this secretion system, there remain some open questions, such as which chaperons keep these periplasmic proteins in the unfolded state; what is the exact mechanism of the autotransport. YadA belongs to the type Vc secretion system and will be discussed here in detail. **Type VI secretion system** is recently discovered and can be resembled to a large cell puncturing device (Bingle, Bailey et al. 2008). It is also Sec independent and spans both the inner- and the outer membranes to secrete proteins outside the bacterial cell envelope. **Type VII secretion systems** do not belong to Gram-negative bacteria, but are limited to myobacteria, actinobactera and Gram-positive bacteria (Simeone, Bottai et al. 2009). **Type VIII secretion systems** are dedicated for transporting the bacterial curli which belong to amyloids and are engaged in biofilm formation, adhesion to the surfaces and cell aggregation (Barnhart and Chapman 2006).

1.2 Trimeric Autotransporter Adhesins (TAAs)

Pathogenic bacteria need to adhere to the cell surfaces for colonizing and infecting their hosts. Pathogenicity of the Gram-negative bacteria is therefore entirely dependent on this first step of adhesion. For this purpose, bacteria assemble several protein complexes on their

surfaces, e.g., pili, fimbriae, membrane-anchored fibres etc. (Linke, Riess et al. 2006). These surface complexes are called adhesins and have diverse modes of attachment, domain architecture and composition. Interestingly enough, these secreted proteins bring themselves in the outside world without aid of ATP. The C-terminal part forms a β-barrel in the lipid bilayer that helps to export the N-terminus through the membrane. These autonomous proteins are called autotransporter adhesins. In many Gram-negative pathogens, these secreted virulence factors are symmetrical trimers and therefore are called Trimeric Autotransporter Adhesins (TAAs). The TAAs generally have high content of α-helical coiled-coils and β-helical or β-trefoil structures (Linke, Riess et al. 2006). TAAs belong to type Vc secretion system; members of this family are important pathogenicity factors in Gram-negative bacteria. Almost every family member of TAAs functions as adhesin and is a homotrimer; therefore, the term trimeric autotransporter adhesin is given to them (Henderson, Navarro-Garcia et al. 2004). Protein secretion occurs in two steps. The first step follows the Sec-dependent pathway in which the N-terminal signal peptide on the target protein is recognized by Sec machinery, and is cleaved after the protein is translocated into the periplasmic compartment. In the periplasm, it is kept into a transport-competent (unfolded) conformation by unknown chaperons. In the second step, C-terminal part of the protein forms a β-barrel in the outer membrane and directs the N-terminus towards the extracellular milieu (Grosskinsky, Schutz et al. 2007). Although models on the export of N-terminus, which would be discussed later on in this chapter, have been presented, yet the information about the actual mechanism of autotransport is limited. It was previously thought that no auxiliary proteins are needed for this transport, however, recently Bam (beta-barrel assembly machinery) complex has been reported to assist β-barrel formation in TAAs (Lehr, Schutz et al. 2010). Nonetheless, the transport in TAAs does not require ATP or proton gradient, hence they are deservedly called autotransporters. All members of this family are important virulence factors and cause different diseases in mammals and plants by mediating adhesion to host cells and tissues. Most frequent diseases in human are diarrhoea, urninary tract infections and airway infections. The colonization of host is accomplished by the interaction of these proteins with the host tissues. Moreover, TAAs bind not only to a wide range of host molecules but also with each other, causing autoagglutination which is directly correlated with the bacterial virulence (Laird and Cavanaugh 1980).

Table 1.1 – Trimeric Autotransporter Adhesins (TAAs)

Protein	Full name	Pathogen	Host	Diseases	Functions
YadA	*Yersinia* adhesin A	*Yersinia enterocolitica*, *Y.pseudotuber-culosis*	Mammals (human, rodents, pigs)	Enteritis, mesenteric lymphadenitis, reactive arthritis	Binds to ECM, epithelial cells and neutrophils, Serum, phagocytosis resi-istance,
YadB, C	*Yersinia* adhesin B, C	*Yersinia pestis*	Mammals (human, animals)	Plague	Invades epithelial cells
BadA	*Bartonella* adhesin A	*Bartonella henselae*	Mammals (human, cats)	Cat scratch disease, bacillary angiomatosis, peliosis hepatis	Binds to epithelial cells and collagen phagocytosis resistance, proangiogenic factor
Hia	*Haemophilus influenzae* adhesin	*H. influenzae*	Mammals (human)	Conjunctivitis menigitis, pneumonia, sinusitis	Binds to epethelial cells
Hsf	*Haemophilus* surface fibrils	*H. influenzae*	Mammals (human)	Conjunctivitis menigitis, pneumonia, sinusitis	Binds to epethelial cells
NadA	*Neisseria* adhesin A	*Neisseria meningitidis*	Mammals (human)	Meningitis, sepsis,	Adhesion and invasion of epithelial cells
UspA1, UspA2	*Ubiquitous* surface protein 1 and 2	*Moraxella catarrhalis*	Mammals (human)	Otitis media, bronchitis, sinusitis, laryngitis	Binds to ECM and epithelial cells, serum and phagocytosis resistance
Vomp-ABCD	Variable outer-membrane protein	*Bartonella quintana*	Mammals (human)	Trench fever, bacillary angiomatosis, peliosis hepatis	Collagen binding, autoagglutination
XadA	*Xanthomonas* adhesin A	*Xanthomonas oryzae*	Plants (rice)	Rice blight	Leaf attachment and entry

Table 1.1 – Some important members of TAAs with their pathogenicity factors (proteins), biological functions, diseases and hosts. All TAAs follow the same head-stalk-anchor architecture and are defined by their C-terminal β-barrel domain in the outer membrane. The Table was adapted from references (Linke, Riess et al. 2006; Lyskowski, Leo et al. 2011).

Proteins of extracellular matrix (ECM) (e.g, fibronectin, collagen, laminin) and cell surface receptors are among the components to which TAAs bind. A few family members of TAAs are listed in Table 1.1 along with some of the information regarding their reservoirs, biological functions and diseases in which they are involved.

1.2.1 Molecular structure of TAAs

The structural organization of TAAs is simple; each molecular fibre consists of three major domains, i.e., the head, stalk and anchor domains. The head domain is a bulky part of the protein and performs the duty of adhering to epithelial cells and the ECM components. The stalk is variable in size and confers protection against host defence mechanism. It is fibrous in nature, extremely rich in coiled-coils and functions as a spacer between the bacterial cell envelope and eukaryotic cells (Linke, Riess et al. 2006). The anchor domain serves two functions. Firstly, it anchors the protein in the OM, and secondly, it translocates head and stalk (collectively called passenger domain) through its pore, into the extracellular environment. Structures of head domain of YadA and anchor domain of *Haemophilus* (Hia) have been determined to atomic resolution (Nummelin, Merckel et al. 2004; Yeo, Cotter et al. 2004). Head of YadA is a trimer of single-stranded, left-handed β-helices (cf. Figure 1.2).

1.3 Models of Autotransport in TAAs

Different models have been proposed for the transport of N-terminal passenger domain through the OM. These models can be divided into three categories, i.e., multimeric model, single chain model and assisted model. A brief introduction about these existing models is presented here along with their drawbacks.

1.3.1 Multimeric model

According to this model, at least six molecules of the autotransporter proteins assemble in the bacterial outer membrane and form a hydrophilic mega pore, through which the N-terminal passenger domains of the participating monomers are exported out (Veiga, Sugawara et al. 2002). This is illustrated in Figure 1.1A; transporter domains are shown as cylinders and passenger domains as ovals. Translocation is achieved by a combined effort of several translocation domains which form a sufficiently wide pore to allow the passage of even folded proteins. Autotransport by this model is dependent upon the presence of at least

six polypeptide chains in close vicinity inside the outer membrane. Keeping in mind that the passenger helical domains are lying inside their respective β-barrels and are not only interwound but also make contacts with the barrel inner walls, this model faces a twofold problem. Firstly, the helical domain must unwind and secondly, the barrel should open at different places after the autotransport, to let helices get inside the barrels. The fact of the matter is that barrel is the most stable part of TAAs and such opening seems highly improbable (Lyskowski, Leo et al. 2011).

1.3.2 Single-chain model

This model states that C-terminal part of the polypeptide forms a β-barrel into the OM before bringing its passenger domain into the extracellular environment. Thus, all essential elements for the autotransport are contained within a single polypeptide chain and the transport adopts an independent route. The passenger domain passes through the β-barrel in an unfolded state and folds only after it faces the extracellular milieu. This model has two variants which agree in the formation of β-barrel but differ in the transport of helical domain.

According to the **threading model** of autotransport, the N-terminal head part is the first to start travelling into the newly formed pore, dragging the rest half along with it (Figure 1.1B). Three unfolded chains shall make a loop to insert their N-termini into the pore in a concerted manner. The threading model is inconsistent with the fact that deletions and modifications in the N-termini of autotransporters did not inhibit translocation (Bernstein 2007). This suggests that N-terminal part lacks the needed information to direct it into the pore lumen.

The **Hairpin model** proposes that after the formation of the β-pore, a hairpin is formed at the C-terminus of the passenger domain. This hairpin inserts into the pore lumen and pulls the unfolded polypeptide segments through the pore lumen with C-to-N-terminal polarity (Figure 1.1B) (Lyskowski, Leo et al. 2011). Thus, the hairpin region of the polypeptide is actively involved in the initiation of autotransport. Any translocation should fail by removing or modifying this region. This is corroborated by studies which have shown that at least 76 and 70 C-terminal residues are required for the β-barrel formation and passenger-domain secretion in Hia and YadA, respectively (Roggenkamp, Ackermann et al. 2003; Surana, Cutter et al. 2004).

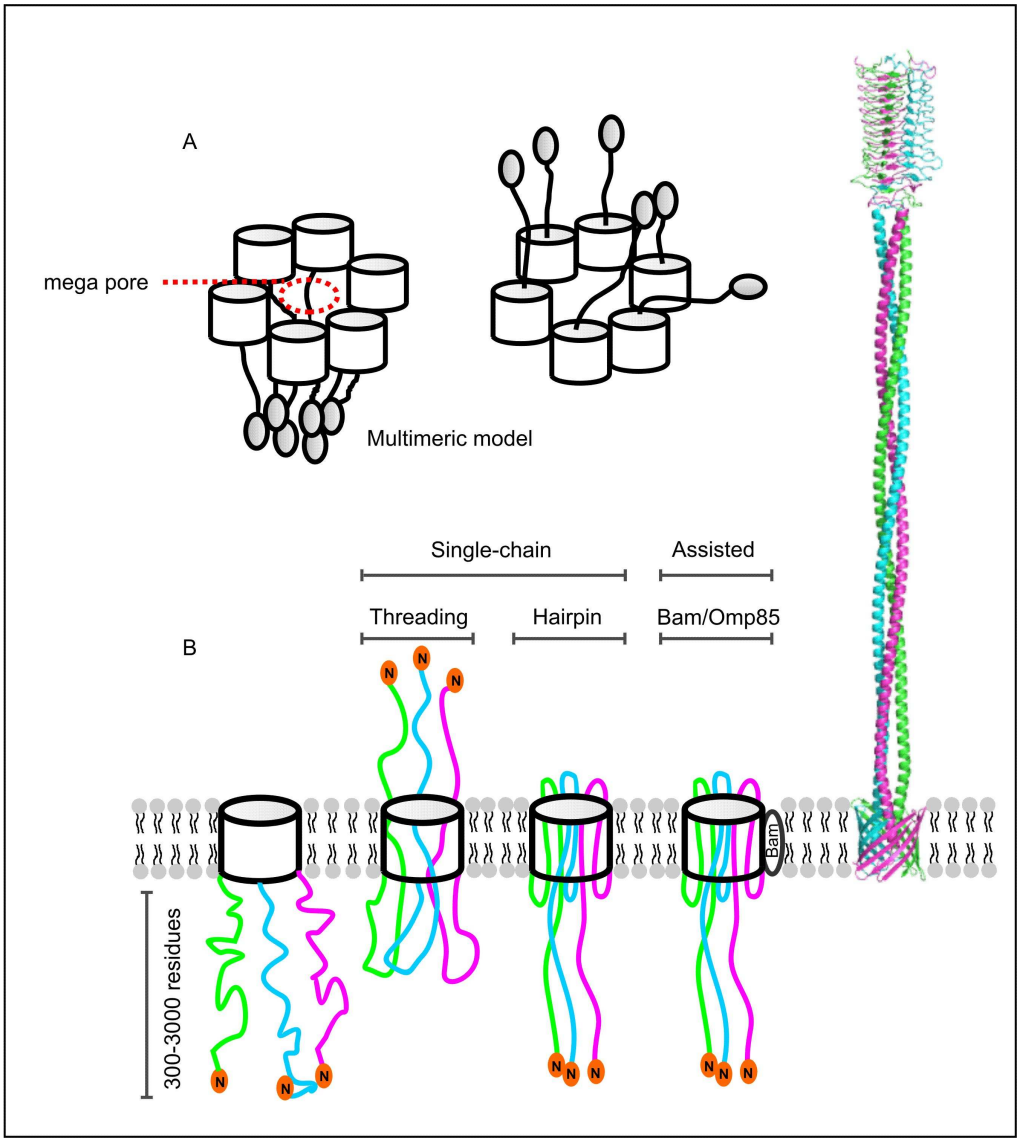


Figure 1.1 – Different models of the autotransport mechanism in TAAs. **(A)** Multimeric model is depicted in Figure 1.1A, where an oligomeric pore is formed by insertion of at least six molecules into the outer membrane. The C-terminal domains are shown as cylinders and the N-terminal passenger domains are shown as ovals. **(B)** The single-chain model, which is subdivided into two categories, i.e., threading and the hairpin model. Both have consensus in formation of the C-terminal β-pore in the outer membrane. Export of the passenger domain may start either from the N-terminus (threading model) or the C-terminus (hairpin model). Assisted model needs other OM proteins for β-pore biogenesis and translocation of passenger domain. The Figures A and B are adapted from references (Veiga, Sugawara et al. 2002) and (Lyskowski, Leo et al. 2011), respectively.

1.3.3 Assisted model

According to this model some helper outer-membrane proteins are involved in the biogenesis of β-barrel, and secretion of α-domain in TAAs. The fact that the Bam complex is involved in the formation of several outer-membrane β-barrel proteins makes this model highly credible. This is further strengthened by a recent study which shows that Bam

complex interacts with the translocation domain during transport in classical (monomeric) autoransporters (Lyskowski, Leo et al. 2011). It can therefore be assumed that auxiliary OM proteins like Bam complex are also engaged both in β-barrel biogenesis and translocation of passenger domain in TAAs (Figure 1.1B). Although the above mentioned three models have been proposed for autotransport in TAAs, the experimental evidence which can prove a single model, by suggesting the exact autotransport mechanism, is missing.

1.4 Yersinia Adhesin A (YadA)

Yersinia Adhesin A (YadA) is the prototype and a distinctive member of non-fimbrial, non-pilus trimeric autotransporter adhesins (TAAs). It is a virulence factor of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* (Bolin, Norlander et al. 1982). Its trans-membrane anchor domain (hereafter called YadA-M) forms a highly stable trimeric β-barrel in the outer membrane and transports its own N-terminal domain through the membrane pore (type Vc secretion system). The three N-termini build a coiled-coil stalk with a sticky head at the end that adheres to the ECM components on the host cell surface. This adherence is one of the first steps in the infection pathways. The multifaceted pathogenic factor is involved in several food-borne diseases such as diarrhoea, septicaemia, reactive arthritis, mesenteric lymphadenitis, iritis (Cover and Aber 1989). Like other members of TAA family, it forms trimeric complexes composed of three major structural domains, i.e., the head, stalk and the anchor domains. A structural model of YadA was presented by Koretke et al. which shows a lollipop-like organization (Koretke, Szczesny et al. 2006) (Figure 1.2A). A single protein molecule is a symmetrical trimer where each protomer consists of 422-455 amino-acid residues with 200-240 kDa molecular mass in different *Yersinia* strains (El Tahir and Skurnik 2001). The **YadA head domain** binds to the host cell surfaces and the ECM proteins, such as collagen, fibronectin, laminin. The preference of ligand for YadA is different in different species; YadA from *Y.enterocolitica* binds favourably to collagen while that of *Y. pseudotuberculosis* binds preferably to fibronectin. The crystal structure of YadA head domain is available which shows a left-handed β-helical arrangement (Figure 1.2D, left) (Nummelin, Merckel et al. 2004). The YadA-collagen interaction is exceptionally strong and renders YadA resistant against extreme temperatures (20 minutes at 80 °C), pH (5-10) and denaturing conditions (< 3 mol/L of urea) (Leo and Skurnik 2011).

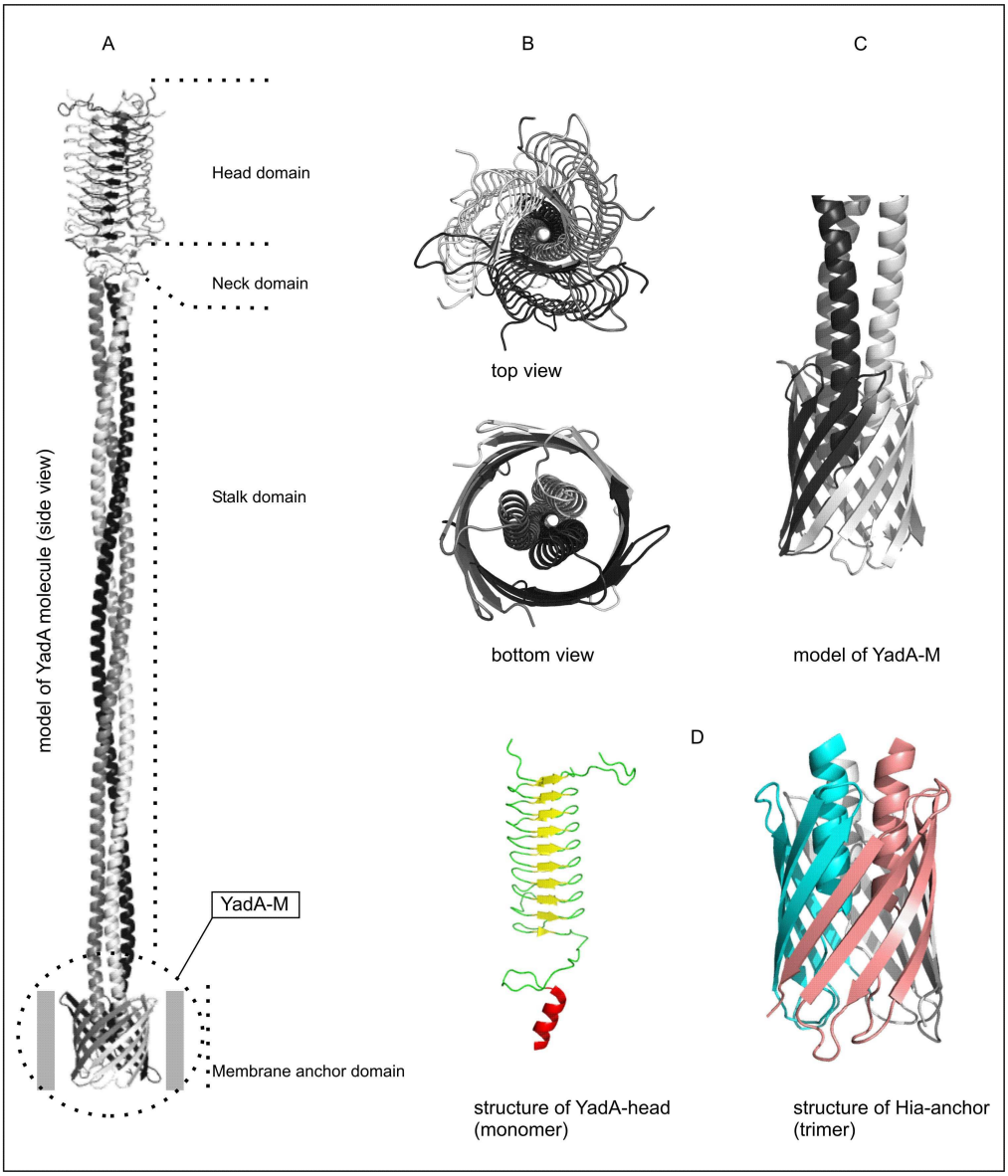


Figure 1.2 – Molecular architecture of YadA. **(A)** Hypothetical model of the full YadA fibre where three monomers are depicted in different colours. Head, stalk and anchor domains collectively constitute a lollipop-like YadA molecule. The head is connected to the stalk with a short neck (red-coloured helix in Figure D). Autotransport (C-terminal) domain is buried in the lipid bilayer, shown as grey bars. YadA-M refers to the construct which was used for the solid-state NMR experiments. **(B)** The same molecular model is shown as seen through the head- (top) and the anchor domains (bottom). **(C)** A model of YadA membrane anchor domain. **(D)** Crystal structures of YadA head (PDB:1P9H) and Hia membrane anchor domain (PDB: 2GR8) (left and right side, respectively), solved by X-ray crystallography. The molecular model of the full protein fibre was built by combining the crystal structure of the head and theoretical models of the stalk and anchor domains (Koretke, Szczesny et al. 2006).

The single-stranded, left-handed β-helical head domain is approximately 5 nm long and its affinity for collagen binding is necessary for the virulence of YadA. Another biological role of N-terminal head is the autoagglutination which is characterised by the aggregation of the head regions with each other in a zipper-like fashion, forming a biofilm and displaying an

excellent mechanism against host defence system (Skurnik, Bolin et al. 1984; Hoiczyk, Roggenkamp et al. 2000). Autoagglutination is essential for virulence in *Yersiniae*. In fact, the strains which fail to agglutinate were found avirulent (Laird and Cavanaugh 1980). The N-terminus of the head also plays a critical role in hemagglutination and neutrophil binding (Kapperud and Lassen 1983; Roggenkamp, Ruckdeschel et al. 1996). A hypothetical model of the anchor and the full fibre of YadA with different orientations are shown in Figure 1.2A, B and C. The models were made on the basis of available crystal structures of YadA-head and Hia-anchor domains (Figure 1.2D) (Koretke, Szczesny et al. 2006). The head is connected to the stalk with a short sequence, collectively called the **neck domain**. This domain functions as an adapter to link wide left-handed β-roll (head) to the relatively narrow coiled-coil stalk (Nummelin, Merckel et al. 2004). **The Stalk** of YadA is an 18 nm long, fibrous structure and functions as a spacer to project the head away from the bacterial membrane to host cell surfaces. There is a transition from right- to left-handed supercoiling in the region close to the membrane anchor (Alvarez, Gruber et al. 2010). Its involvement in serum resistance makes it an important part of the protein (Foster and Hook 1998). YadA defies the bactericidal activity of host serum complement presumably by direct or indirect binding of the stalk with the complement factor H (China, Sory et al. 1993; Roggenkamp, Ruckdeschel et al. 1996). The stalk in TAAs is extremely variable in length. BadA which is also called a giant TAA has a height of 300 nm as compared to 23 nm long YadA; the difference is due to lengths of stalks in both members.

The C-terminal membrane anchor domain or autotransport domain of YadA is embedded in the outer membrane and forms a β-pore which is occluded after a successful transport of head and stalk domains. According to the theoretical model (Koretke, Szczesny et al. 2006), each monomer contributes four β-strands that assemble in the membrane to form a twelve-stranded β-barrel. The diameter of the pore is dependent upon the tilt of β-strands with respect to the perpendicular of outer membrane. This tilt is called the shear number, represented by "S" and is very crucial for determining the exact dimensions of the barrel. It is defined as the change of residue numbers on a β-strand when a point moves in the left hydrogen bond direction around the barrel, back to the same strand (McLachlan 1979; Chou, Carlacci et al. 1990; Murzin, Lesk et al. 1994). The shear number is directly proportional to the width and inversely proportional to the height of the β-barrel. A hypothetical model of YadA-M based on the X-ray structure of Hia is shown in Figure 1.2C. An atomic resolution

X-ray structure of the membrane anchor domain of Hia is already available and is shown in Figure 1.2D (right side) (Meng, Surana et al. 2006). The choice of the shear number plays an important role in the modelling (a shear number of 16 was initially used to model the barrel) (Koretke, Szczesny et al. 2006). A barrel with shear number 12 would also be sufficient to accommodate three helices if the lumen-facing residues, on average, are equal to the size of an alanine (Linke, Riess et al. 2006). The role of the membrane anchor domain is evident from its name; firstly, it anchors the protein in the bacterial outer membrane and secondly, it exports the polypeptide chain into the extracellular milieu. In contrast to the head and stalk domains, anchor domain is homologous in TAA family. In-frame deletion mutagenesis studies on YadA have shown that membrane anchor domain is responsible for autoransport in all TAAs (Grosskinsky, Schutz et al. 2007). It was demonstrated that deletions/modifications on the head and stalk domains do not stop barrel trimerization, and autotransport of the N-terminal polypeptide chain. In reality, C-terminal 70 amino-acid residues are believed to be sufficient to make a barrel and translocate the N-terminus across the outer membrane (Roggenkamp, Ackermann et al. 2003).

A glycine residue in the second β-strand, facing the pore lumen, is highly conserved throughout the family of TAAs and is believed to be involved in the autotransport process. Mutation studies of this glycine with long side-chain residues indeed affected the expression levels of the protein (Grosskinsky, Schutz et al. 2007). Sequence alignment of the membrane anchor of some members of TAAs is shown in Figure 1.3.

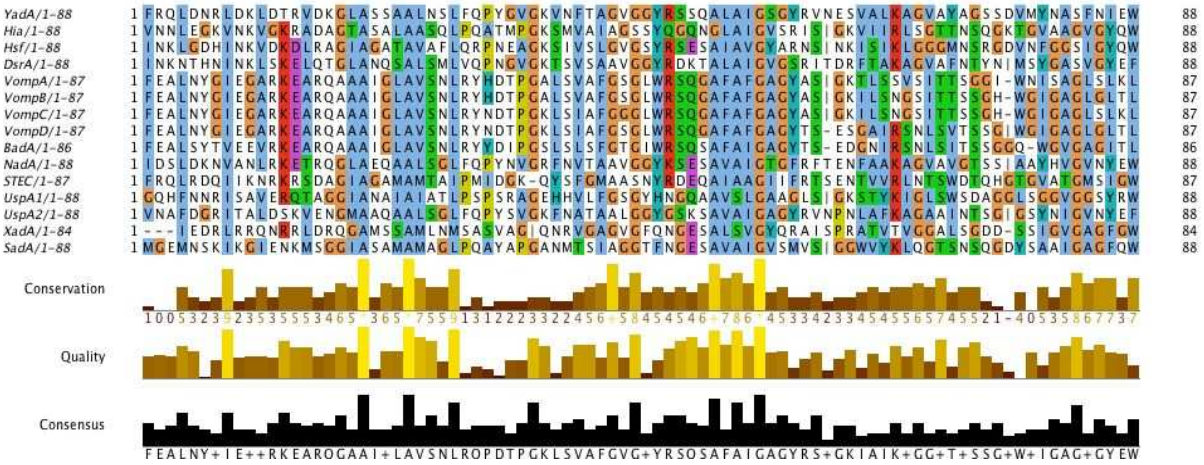


Figure 1.3 – Sequence alignment of membrane anchor domains of some TAAs. The colour-coding is according to the residue type. β-barrel G72 (highlighted in orange) and α-helical A37 (highlighted in blue) are completely conserved throughout the alignment. The degree of conservation, its quality and consensus are also given at the bottom of alignment results. The alignment was edited in Jalview using the sequence alignment in reference (Grosskinsky, Schutz et al. 2007).

1.4.1 Yersiniosis

Bacteria from *Yersinia* are involved in several infectious diseases called yersiniosis. *Y.enterocolitica* is the most common species which affects human and is involved in several enteric diseases. Children are most likely to be affected by *Y.enterocolitica*. Common symptoms of infection are fever, abdominal pain and diarrhoea (often with blood). Pigs are found to be the major reservoir of *Y.enterocolitica*. Infection occurs by eating raw or undercooked pork products, especially raw pork intestines (chitterlings). Further sources of infection are contaminated, unpasteurized milk or unhygienic water. Other *Yersinia* species are found in domesticated mammals as well, such as rabbits, cattle, horses, cats, dogs and sheep (www.cdc.gov).

1.5 Protein structure determination

1.5.1 Introduction

Molecular structure forms the fundamental basis of understanding the biological function of proteins. Several techniques are used in structural biology to solve a protein's structure. The strength of these techniques in terms of atomic-scale resolution is important; the derived biological information is dependant upon this resolution. For instance, a light microscope can give an overall picture of living cells (e.g., yeast or bacteria), which are moving and bumping on the microscope slide (Goodsell 1996). From such an image, information about the cellular shape, and sometimes about cellular compartments can be observed, however, the molecules which form these compartments are thousands of times smaller, and a conventional light microscope cannot resolve them. An electron microscope is more powerful and focuses a beam of electrons on the specimen to give images of the individual molecules. Biological molecules are composed of light atoms, hence unable to stop most electrons. Heavy atoms (e.g., gold and uranium) are sometimes added in the biological samples so as to make them adequately opaque to create an image. Electron microscopes can give at best, the external shape and size of the molecules which provide valuable information. However, to understand most biological functions, one has to zoom further into the molecules to see the arrangement of building blocks, the atoms. X-ray crystallography, in this respect is a highly successful tool to discern structures at atomic resolution. Instead of

focussing a light or electron beam on the sample (there is no way to focus X-rays, therefore no X-ray microscope is available), it bombards X-rays on the specimen. The scattered rays make a distinct pattern of the electronic cloud on a sheet or film. Highly dedicated computers then calculate the structure of the molecule to atomic detail on the basis of this diffraction pattern. Since a single molecule cannot scatter enough X-rays, quadrillions of molecules are stacked together in a regular geometry to form a three-dimensional (3D) single crystal. The process of crystallization is time-consuming, besides that, many biological samples thus far could not be crystallized. Another problem is that the biological samples are highly sensitive and are destroyed after the X-ray bombardment. Solution-state Nuclear Magnetic Resonance (NMR) spectroscopy is an alternative high-resolution atomic structure determination technique which does not rely on the availability of crystals; neither does it destroy the sample after measuring it. However, solids or solid-like materials of high biological importance, such as MPs, amyloid fibrils, which are involved in many diseases, cannot easily be accessed with solution NMR. As discussed earlier, solid-state NMR is highly suitable for measuring samples that are inaccessible to other structure determination tools. This chapter describes very briefly the basic principles of solid-state NMR. The readers interested in details are directed to the books, 'Spin dynamics: Basics of Nuclear Magnetic Resonance' by Malcolm H. Levitt, and 'Introduction to Solid-state NMR Spectroscopy' by Melinda J. Duer.

1.5.2 Fundamentals of NMR

Nuclear Magnetic Resonance (NMR) spectroscopy is the technique in which interactions of nuclear magnetic moments with an external magnetic field are studied. The nuclear spin is one of the four fundamental physical properties of a nucleus, the other three are: mass, electric charge, and magnetism. A nucleus possesses an angular momentum (i.e., spin angular momentum donated by 'I') that is a vector quantity and indicates the axis of the spin. The isotopes which have a zero spin angular momentum ($I = 0$) are NMR silent; only those with $I > 0$ are inherently magnetic and give rise to an NMR signal. The ^{1}H, ^{13}C and ^{15}N spins ($I = 1/2$) are the most important spin systems in the field of structural biology. Nuclei with $I > 1/2$ give rise to quadrupolar interactions. In this thesis, only spins $I = 1/2$ are studied. As a consequence of magnetic moment and angular momentum, nuclear spins when placed in a static magnetic field (B_0), start precessing around the B_0 axis with a certain frequency, like a

spinning top (cf. Figure 1.4B, red arrow). This angular velocity is called nuclear Larmor frequency (ω_0) and is proportional to the applied B_0 and the gyromagnetic ratio γ of the nucleus under observation, i.e.,

$$\omega_0 = -\gamma B_0$$

Therefore, each nucleus shows a different Larmor frequency in the magnetic field (e.g., at B_0=14.1 T, ^{1}H, ^{13}C and ^{15}N have a ω_0 of 600, 150.9, and 60.8 MHz, respectively). NMR is the measurement of this Larmor frequency and probes the interaction of induced magnetic moments with the external magnetic field. The fact that a single kind of nucleus gives different signals in an NMR spectrum comes from the shielding effect of the electrons. The electronic cloud that surrounds the nuclei perturbs (or shields) the net magnetic field felt by a nucleus, thus resulting into different Larmor frequency depending on the chemical environment. These differences are called chemical shifts. In a field B_0, the spins are polarized and precess in a cone with a constant angle; the angle of the cone for each spin depends upon its initial position in the absence of B_0. The magnetic moment of spin 1/2 nuclei is quantized and shows two spin states in the presence of B_0, i.e., +1/2 (aligned along the B_0, also called spin-up or α-spin, low energy state) and -1/2 (aligned against the B_0, also called spin-down or β-spin, high energy state). The splitting in spin-up and spin-down states is called the Zeeman effect and is proportional to the gyromagnetic ratio of the nuclei, and the strength of B_0 (cf. Figure 1.4A). Outside the B_0, the energy difference between two states is zero. On a macroscopic scale the α-spins slightly outnumber the β-spins; the difference of population is given by the Boltzmann distribution:

$$\frac{pi}{pj} = \exp\left(\frac{\Delta E}{k_b T}\right)$$

Where ΔE is the difference in energy between the two spin states and is given by:

$$\Delta E = \left(\frac{\gamma \hbar B_0}{2\pi}\right)$$

For a ^{1}H nuclear spin at 20 T, the difference in energy between two spin states is ΔE= 5.6 x 10^{-25} J. The available thermal energy ($k_b T$) at room temperature (298 K) is 4.1 x 10^{-21} J and is higher than ΔE. Therefore, the lower energy state is slightly more populated, though the difference is extremely small, ~1/10^6 in a routinely used NMR magnet. Due to very large

number of spins, there arises a macroscopic nuclear net magnetic moment (M_0) which is aligned along the B_0 axis (Figure 1.4B, red arrow).

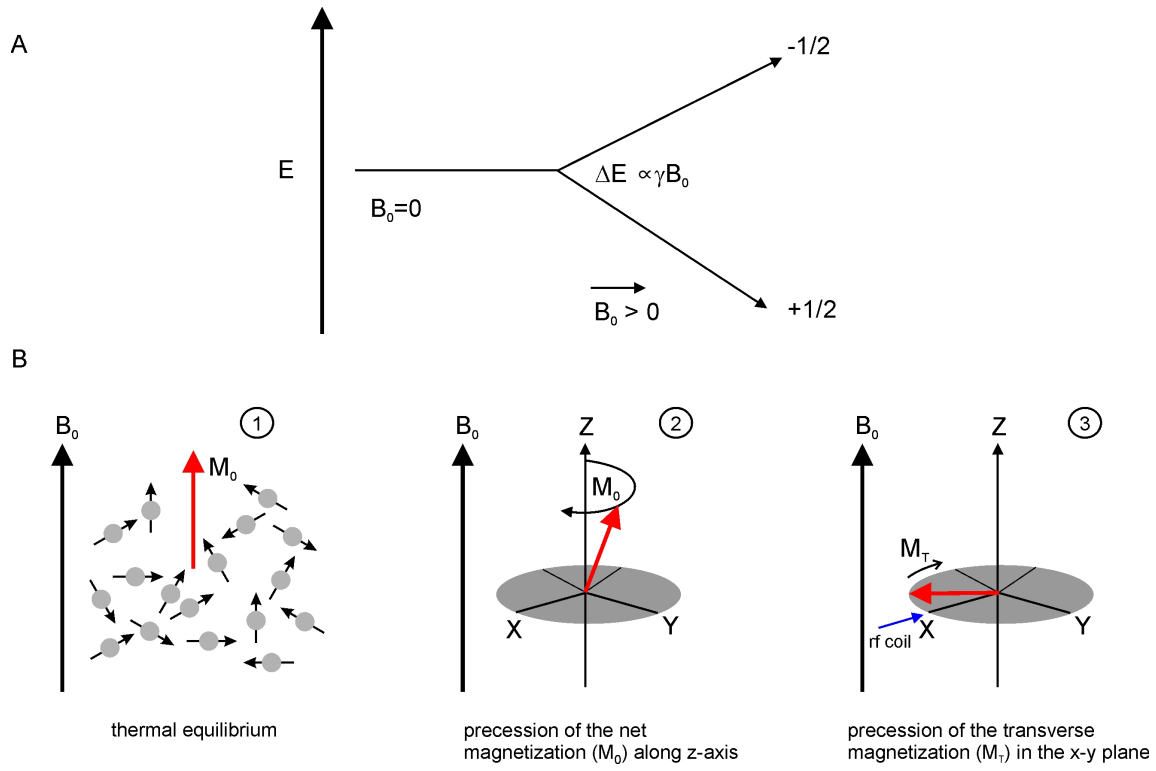


Figure 1.4 – Basics of NMR. (A) Zeeman splitting in the presence of B_0. The energy difference (and hence the population difference) between α- and β-spin states depends upon the strength of B_0 and γ of nuclei. Since γ for nuclei is constant, stronger NMR magnets are being developed to increase the signal intensity. By the time of writing, 23.5 T (which causes ^{1}H to precess with a frequency of 1 GHz) magnets are available for NMR studies. (B) On the left side, the net magnetic moment (M_0) is shown with a red arrow while the spins are indicated by grey circles with arrow heads representing the direction of spin angular momentum for individual nuclei. It is easier to understand the behaviour of M_0 in a Cartesian coordinate system (shown in the middle of the Figure). The magnetization in the transverse plane is denoted by M_T and rotates in the *xy*-plane. The NMR signal is recorded by the rf-coil as shown on the right side of the Figure.

In a Cartesian coordinate system, the orientation of B_0, by convention is along the *z*-axis. M_0 is much weaker (measured in Hz) as compared to the B_0 (measured in MHz), hence cannot be measured along the *z*-axis. A radio frequency pulse (rf-pulse) with an oscillating magnetic field (B_1 field) near resonance with the spins under study (i.e., the frequency of the rf-pulse is very close to the ω_0 of the spins, $\omega_{rf} \sim \omega_0$) is applied to rotate the net magnetization into the *xy*-plane. Viewed from the rotating frame of reference, the B_1 field is static and by adjusting its phase, it can be applied along the *x*-axis. M_0 will precess around B_1, and can be rotated to the *xy*-plane for detection. This magnetization, which now is perpendicular to B_0, is called

transverse magnetization (M_T). When the rf-pulse is switched off, the transverse magnetization M_T of different spins will precess in the *xy*-plane at different frequencies due to the chemical shift differences. As different nuclei are subjected to dissimilar interactions, M_T gets out of phase and relaxes away (T_2 or transverse relaxation) and slowly relaxes back to its equilibrium state along the *z*-axis (T_1 or longitudinal relaxation). Due to electromagnetic induction, the rotating magnetic field induces an electric current in a coil in the *xy*-plane. The same rf-coil used for rotating M_T serves the purpose of detecting this oscillating current, which is processed to give an NMR signal in the form of *free-induction decay* (FID). The FID is associated with X- and Y-magnetization and gives the amplitude as well as the phase of the signal. The FID is Fourier transformed to achieve the NMR spectrum.

1.5.3 Solid-state NMR

In solid-state NMR, as opposed to solution-state NMR, the anisotropic interactions (chemical shift anisotropy, dipolar and quadrupolar couplings) largely dominate the spectrum, resulting in broader signals. Although, the broad featureless spectral peaks in solids are overloaded with structural and dynamic information, the difficulty in interpreting them renders them useless for further studies. The most commonly observed spins in solid-state NMR experiments are spin 1/2 nuclei (e.g., ^{1}H, ^{13}C, ^{15}N). ^{1}H is the most abundant spin (99.99%) with high gyromagnetic ratio, however, it is not routinely used in solid-state NMR because of its strong homonuclear dipolar coupling effect. Among low gamma, low abundance nuclei, ^{13}C and ^{15}N are the most often observed spin systems for structural studies by biologic solid-state MAS NMR.

Let us take an example of ^{1}H and ^{13}C nuclei which are close in space. By convention, the naturally abundant spins are represented by *I* (e.g., ^{1}H) and rare spins by *S* (e.g., ^{13}C). Both nuclei produce tiny magnetic fields which affect the neighbouring nuclei (within a few Å) in such a way that the net external magnetic field (B_0) felt by them is either increased or decreased. This through-space interaction is called heteronuclear dipolar coupling and forms the basis of structural studies via distance restraints in solid-state NMR. The strength of heteronuclear dipolar coupling is given by the following Hamiltonian:

$$H_{IS} = -d(3\cos^2\theta - 1)I_Z S_Z \qquad (1)$$

Where *d* is the heteronuclear dipolar coupling constant and is given by:

$$d = \left(\frac{\mu_0}{4\pi}\right)\frac{\hbar\gamma_I\gamma_S}{r^3{}_{IS}}$$

The above equation implies that heteronuclear dipolar coupling is inversely proportional to the cube of the distance between two spins; $\frac{1}{r^3}$ is the distance dependence between interacting spins and is exploited to observe distance restraints in NMR. The angle between the B_0 axis and the internuclear vector is given by θ and represents the orientational dependence of the dipolar coupling relative to B_0. The angle θ is distributed randomly in solids leading to broad, featureless lineshapes. Two nuclei with same distance but different orientations would result in different dipolar interactions. In a powder sample, there are all possible orientations of the crystallites which result in countless signals at different chemical shift values; all of them merge together to form a very broad peak. The second feature which causes line broadening in solid-state NMR is the chemical shift anisotropy (CSA). The chemical shift of a nucleus strongly depends upon the orientation of its asymmetric electron distribution with respect to B_0 axis. The external magnetic field is shielded by the surrounding electrons, and for certain orientations the shielding effect is higher than others. For instance, the resonance frequency of a carbonyl moiety can differ to 120 ppm depending upon its orientation with regard to the static magnetic field.

In contrast, solution-state NMR gives well-resolved, narrow spectral lines. The reason lies in the freedom of molecules to move and reorient in all possible directions in very short time in a solution. The internuclear vector samples all possible values of θ in little time, resulting into a time-averaged narrow NMR signal. The rapid molecular tumbling motion in liquids is so fast on the NMR time scale that a molecule reorients before the dipolar coupling and CSA have time to evolve. Therefore both these interactions, i.e., dipolar coupling and chemical shift anisotropy which depend upon the term $(3\cos^2\theta - 1)$ are averaged in solutions, resulting in isotropic spectral lines. The lack of this rapid tumbling motion in solids, results in the dominance of the chemical shift anisotropy (CSA) and strong dipolar couplings. In solid-state NMR, two general approaches are used to overcome these anisotropic interactions and to achieve high-resolution spectra. The first approach relies on orienting individual molecules in the strong magnetic field, and thereby retrieving structural-rich orientational restraints. In fact, the first structure of a membrane protein deposited in the PDB (trans-membrane region of the M2 ion channel protein; PDB: 1MP6)

was resolved from orientational constraints (Wang, Kim et al. 2001). The second approach to obtain well-resolved signals in solid-state NMR is magic-angle-spinning (MAS) (illustrated in Figure 1.5, right). In MAS NMR, the sample rotor is aligned at the "magic angle" (i.e., 54.7°) relative to B_0 and mechanically rotated with high frequencies (typically 10 - 20 kHz) to get rid of CSA effects. With MAS, the *average* angle θ for each pair of interacting spins is at the magic angle. The term for dipolar coupling in equation 1, as well as the CSA, becomes zero if θ_m =54.7°. Solid-state MAS NMR exploits this fact to achieve solution-like narrow spectral lines.

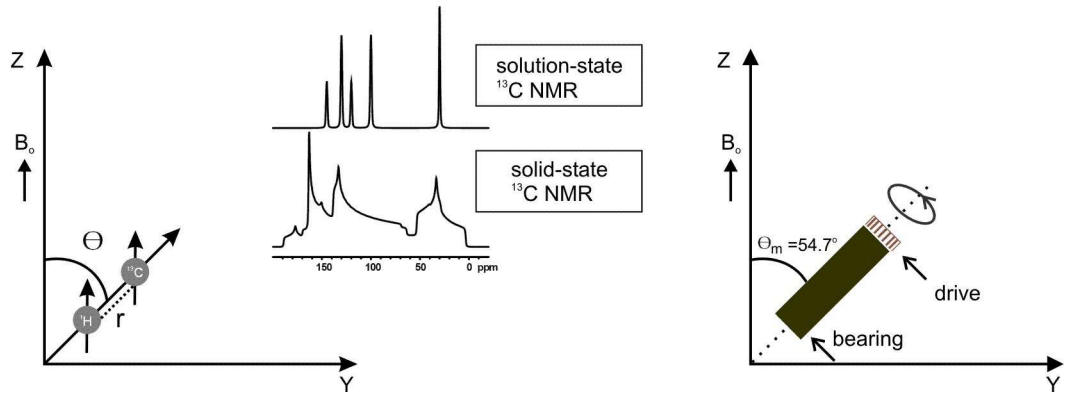


Figure 1.5 – Solid-state magic-angle spinning (MAS) NMR. On the left side of the Figure, the orientational dependence of dipolar coupling is shown. A hypothetical solution- and solid-state NMR spectrum is shown in the middle of the Figure. On the right side of the Figure the solid-state MAS NMR technique is illustrated. The bearing air pressure lifts the sample rotor while drive air is applied on its cap for spinning the rotor at high frequencies.

1.5.4 Cross-polarization magic-angle spinning NMR

As already mentioned in this chapter, protons are the most sensitive NMR nuclei because of their high natural abundance and gyromagnetic ratio. Besides, the relaxation times for ^{1}H are very short which minimize the acquisition times since successive experiments can be repeated more rapidly ('scanning'). In solid-state NMR, however, for non-deuterated samples, detection of ^{1}H is not frequently done. This is due to the strong ^{1}H-^{1}H dipolar couplings, which create many technical difficulties. Therefore, in solid-state NMR of biological samples, ^{13}C and ^{15}N are commonly used for signal detection. The signal intensity for these low-γ nuclei is very weak due to their low isotopic natural abundance. In solid-state NMR, the weak signals are enhanced by isotopic labelling of the sample, and by using a standard method called cross polarization (CP). In this technique, a polarization-exchange

contact is established between *I* and *S* spins, in which the magnetization flows from highly polarized *I* (^{1}H) to the less polarized *S* (^{13}C/^{15}N) spins. The pulse sequence for a standard CP experiment is shown in Figure 1.6. The experiment starts with a 90° proton excitation pulse (highlighted with solid, black bar in Figure 1.6A), which brings the ^{1}H magnetization in the *xy*-plane. Next, simultaneous pulses (B_1 fields) are applied on both ^{1}H and ^{13}C/^{15}N nuclei to lock them along the rotating frame *x*-axis. Like the Zeeman interaction, the spin-lock pulses also result into a so-called spin splitting, but this time in the rotating frame. Whereas the external magnetic field causes a huge spin splitting (i.e., $\omega_0^H/2\pi$ ~ 600 MHz, $\omega_0^C/2\pi$ ~ 150 MHz in a 17.4 T magnet), the spin splitting caused by these rf-pulses is in the range of kHz (i.e., usually $\omega_1^H/2\pi$ ~ 50 kHz). The splitting in this case depends entirely upon the amplitude of spin-lock fields for $B_1^{C/N}$ and B_1^H. Another way of realizing CP is that the rf-pulses cause the transverse magnetization to nutate with a frequency, called nutation frequency which is dependent on the strength of applied pulses. The strength of both fields is carefully selected to satisfy the so-called 'Hartmann-Hahn matching' condition (i.e., $\omega_1^H=\omega_1^C$) in which the nutation frequencies of both *I* and *S* spins is almost equal.

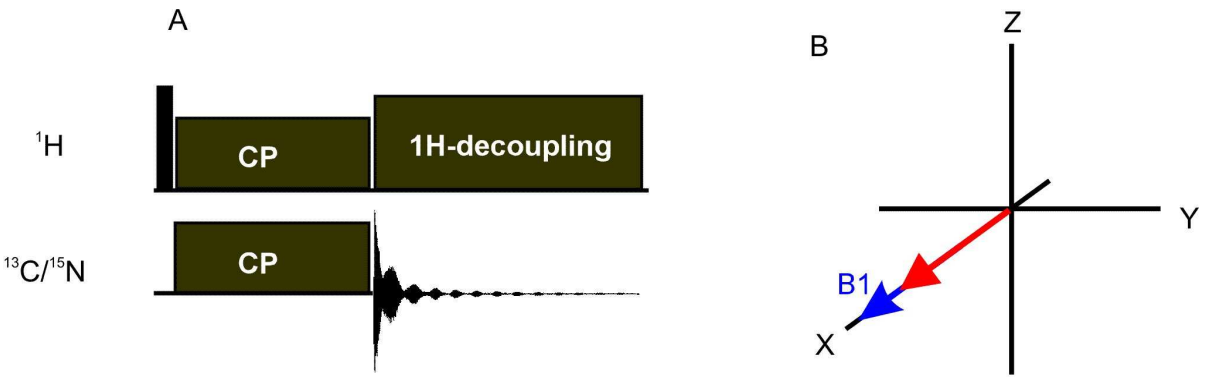


Figure 1.6 – Standard cross polarization (CP) experiment in solid-state magic-angle spinning NMR. Pulse sequence for a simple one-dimensional CP experiment is shown in Figure A. The CP contact times are shown as CP blocks which trap or 'spin-lock' *I* and *S* spins in the transverse plane. The FID on *S* spins is acquired under high-power decoupling of *I* spins. Red and blue arrows in Figure B depict the trapped magnetization and applied B_1 field, respectively.

The CP experiments are usually accompanied with magic-angle spinning (MAS), for which Hartmann-Hahn matching condition is given by: $\omega_1^H - \omega_1^{C/N} = \pm n\ \omega_r$ (with $n = \pm 1, \pm 2$). For instance, for a spinning sample with MAS frequency $\omega_r/2\pi$ = 10 kHz, a Hartmann-Hahn matching condition would be ω_1^H = 60 kHz, and ω_1^C = 50 or 70 kHz. The magnetization flows from highly polarized ^{1}H to the less polarized ^{13}C/^{15}N spins during the established CP

contact. Finally, the signal is acquired on $^{13}C/^{15}N$ spins while applying a high-power decoupling on ^{1}H spins. Proton decoupling is a process of applying continuous rf-pulses that rotate the ^{1}H spins between spin-up and spin-down states, thus partially averaging the proton dipolar coupling.

As most of the dipolar couplings, which are rich in distance information, are lost during MAS, several recoupling schemes are used in solid-state MAS NMR to bring them back. A train of pulses that are synchronized with the MAS is applied which results in an incomplete averaging of the dipolar couplings. This enables through-space magnetization transfer between the coupled spins. Homonuclear recoupling is achieved by, e.g., radio frequency-driven recoupling (RFDR), while heteronuclear recoupling is accomplished by rotational echo double resonance (REDOR/TEDOR) pulse schemes. DARR and PDSD schemes are used to allow spin diffusion, which gives correlation signals between coupled nuclei, allowing the chemical shift assignment of individual amino-acid residues and detection of distance restraints. Comprehensive application and details of these recoupling schemes to attain chemical shift assignment of amino-acid residues and to achieve distance restraint information are given in Chapter 3 and 4, respectively.

1.6 Inferential Structure Determination (ISD)

1.6.1 Introduction

High-resolution NMR spectroscopy has emerged as a unique bio-molecular method for three-dimensional (3D) structure determination of macromolecules. Despite its many-sided advantages, NMR structure determination is often considered to be a subjective method. Contrary to X-ray crystallography, NMR is rather an indirect method and numerous human interventions in data handling are required to reach the final structure. NMR data in its raw state cannot be used for structure calculation. Several pre-processing steps, such as resonance assignment, peak picking and assignment of medium and long range distance restraints are needed before the data are ready for structure calculation (a schematic summary of different steps to achieve a 3D protein structure is given in Chapter 4, Figure 4.1). There is always a likelihood of error or bias during these steps; personal beliefs or choices may lead to a number of wrong resonance assignments and distance restraints. The subjective choices in data treatment and manual parameter settings result in a structure that not exclusively

represents the experimental data. Wrongly assigned restraints or inappropriate parameter settings, in the worst-case scenario, may lead to a precise (the structures in the final ensemble are consistent and reproducible relative to each other) however, totally inaccurate solution. Thus, there is no principle way to objectively judge the accuracy and reliability of an NMR structure. One way to judge the quality of an NMR structure in traditional techniques is to analyse distance restraint violations and ensemble root-mean-square deviations (RMSDs), both of which are derived in the aforementioned steps of spectral analysis. There are software packages that can establish the validity and quality of a structure independently by evaluating the fit of the structure to the experimental restraints and the geometric properties of the structure, e.g., WHAT IF (Vriend 1990), PROCHECK (Laskowski, Macarthur et al. 1993), PROCHECK-NMR (Laskowski, Rullmann et al. 1996), QUEEN (Nabuurs, Spronk et al. 2003), just to name a few. For a detailed review see reference (Spronk, Nabuurs et al. 2004; Nabuurs, Spronk et al. 2006). However, the structure calculation programmes should ideally be able to simultaneously calculate the structure and evaluate its reliability, alone from the experimental data. Inferential structure determination (ISD) is a tool which uses Bayesian inference to obtain a probability distribution from which both the unknown structure and its uncertainty can be objectively determined. In this chapter, a brief introduction about the general principles of ISD and its differences from conventional methods are presented. The readers interested in the details are directed to the references (Habeck, Nilges et al. 2005; Rieping, Habeck et al. 2005; Habeck 2012)

1.6.2 Structure determination from conventional methods

The goal of every structure determination method is to deduce the spatial positions of individual atoms (atomic coordinates) of the molecule from experimental data. Traditional methods use energy minimization to calculate molecular structures from experimental data (Brunger and Nilges 1993). Let us take ARIA (Nilges 1995; Rieping, Habeck et al. 2007) as an example to illustrate how a protein structure is calculated. In the first step, ARIA uses experimental data (in our case NMR data), imposes a force field, introduces additional parameters and runs a molecular dynamics-based simulated annealing protocol (MDSA) (Nilges, Clore et al. 1988; Guntert, Mumenthaler et al. 1997) to get an ensemble of conformations. The force field (data-independent knowledge about the protein potential energy) imposes physical constraints on the structure to compensate a lack of data. The non-

linear optimization problem is solved iteratively to achieve a final ensemble of lowest hybrid energy structures which is in accord with the experimental data. The hybrid energy is given by:

$$E_{\text{hybrid}}(X) = w_{\text{data}}\, E_{\text{data}}(X) + E_{\text{phys}}(X) \qquad (1)$$

where:
E_{hybrid}: Hybrid energy or total cost function
E_{data}: Cost function that evaluates the goodness of the fit (of the data) for a specific structure X
E_{phys}: Physical energy of the system, determined by molecular force field
w_{data}: Weighting factor

In the next step, theoretical data is back calculated from the ensemble and compared with the observed data to determine its consistency, using a heuristic or cross validation. Non-satisfied restraints are either removed or reassigned and the refined NMR data are used in the next iteration to obtain another set of lowest energy structures. This process is iterated until the best fit between the calculated structure and the experimental data are achieved (Figure 1.7A). The value of the weighting factor in equation 1 is manually provided to the software. This value instructs the programme about how much should it rely on the experimental data by controlling its contribution with respect to the force field. If wrongly assigned restraints are present in the data, a higher value would result into an inaccurate structure (over fitting). A lower value means that the programme may discard the correct peaks and would end up again into a wrong structure (under fitting). Therefore, selection of the weighting factor plays a crucial role in the whole structure determination. Likewise, other auxiliary parameters (e.g., NMR calibration scales, alignment tensors, etc.) which though they are not of immediate interest, but have to be introduced in order to describe the problem adequately, are required to run a calculation. These additional parameters, in Bayesian theory, are called "nuisance parameters" (Nilges, Habeck et al. 2008). In the above-mentioned case, the weighting factor is one of such nuisance parameters. In traditional methods, values of nuisance parameters are given either by a rule of thumb or determined heuristically; again, personal beliefs and manual intervention come into play. If the experimental data are good and complete in terms of information, traditional methods are able to get a correct structure even with incorrect nuisance parameters.

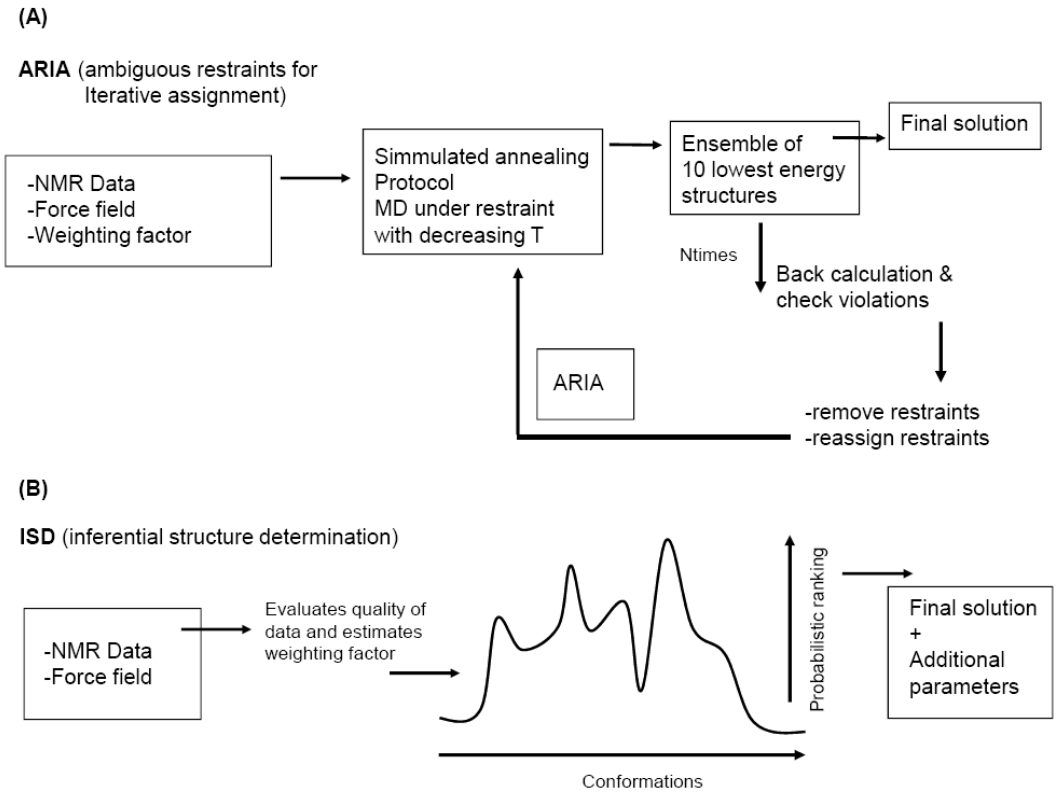
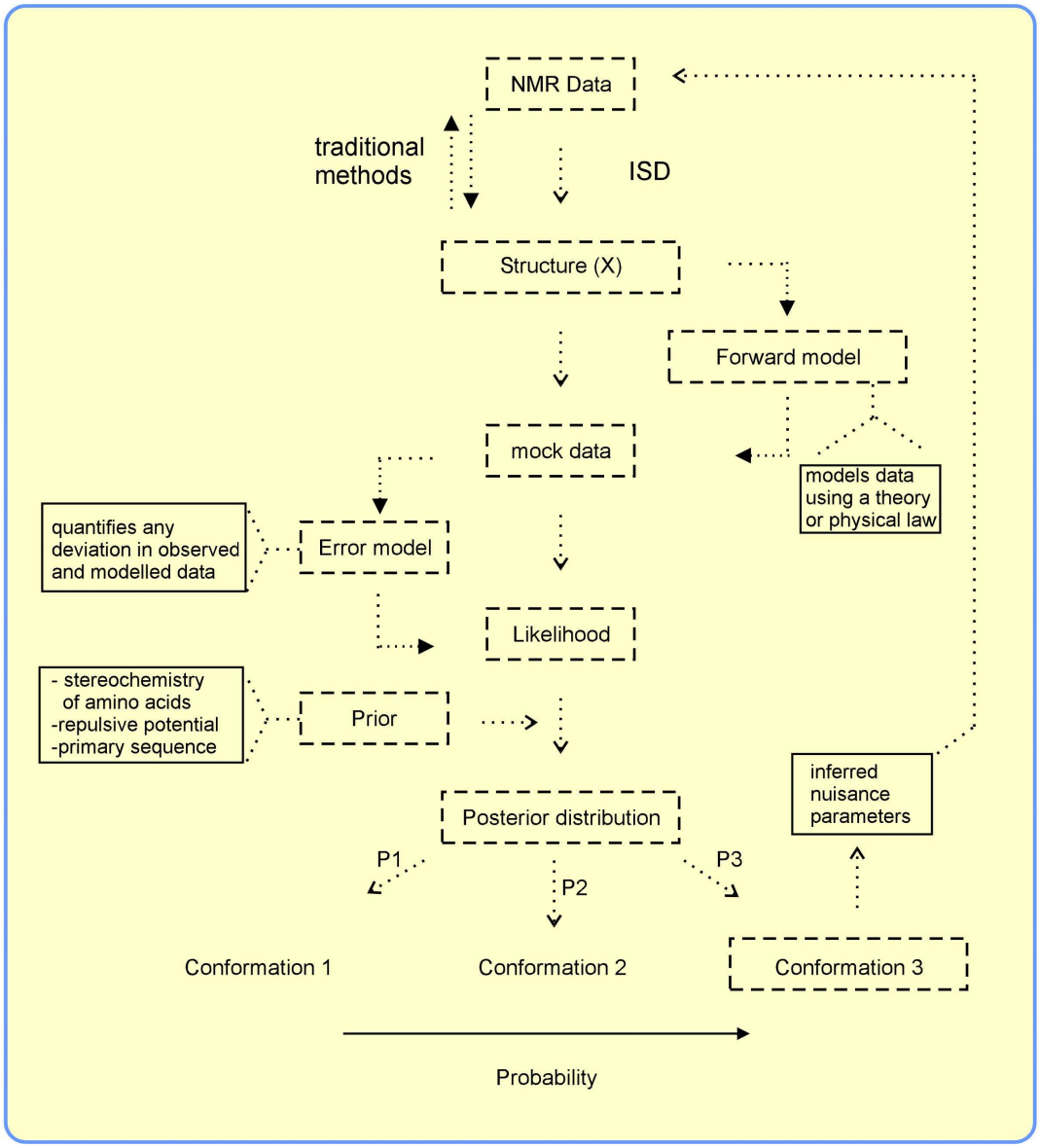


Figure 1.7 – Schematic representation of ARIA and ISD. (A) Different steps involved in ARIA calculation are summarized. An ensemble of preliminary structures is derived from the data using a hybrid energy function (energy minimization approach). In cross validation, consistently violating restraints are either removed or reassigned and the refined data are used in the subsequent iterations. The process is repeated several times until the best fit between the experimental and modelled data are achieved. **(B)** ISD converts experimental data into different conformations which are ranked according to their frequency of appearance. This is shown by a hypothetical graph which is plotted against the conformational space and the probability. The conformation with highest probability gives the final solution.

1.6.3 Principles of ISD

ISD views structure calculation as an inference problem. The 3D structure of a protein is calculated using an entirely probabilistic approach. Instead of converting data into geometrical constraints (like traditional methods do), ISD uses the experimental data to rank all possible conformations of a molecule in proportion to their probability. Next, the distribution of these probabilities is explored and the conformations with highest probability (which are more supported by the data) are selected for further analysis (Habeck 2012). A big advantage of ISD is that in addition to the structure, it objectively determines the missing nuisance parameters. Unlike traditional methods, it does not impose a fixed value of the

weighting factor w_{data}. ISD has the power to evaluate the quality of data and to estimate the weighting factor accordingly; if the quality of data is good, ISD will estimate a higher w_{data} value and vice versa. Individual probabilities (*P*i) are evaluated for every conformation (*X*i) using Bayes' theorem. A schematic summary of structure calculation by ISD is shown in Figure 1.7B. The NMR data and the force field which in this case is the prior knowledge about the system, are used to derive a probability distribution of all possible conformations. The quality of determined structures depends upon the spread of the distribution: if only one conformation has a non-zero probability, ISD would uniquely determine a single structure from the data. On the other hand, if every conformation has equal probability, the experimental data would be uninformative (Nilges, Habeck et al. 2008). However, real cases lie in between these two extremes. The probabilities are determined in an objective manner because they rely solely on the experimental data and the prior information. Prior knowledge is the already established information about the system under study. For example, we already know the primary sequence of the protein, how atoms in different amino-acid residues should occupy volume in space, what are the typical bond lengths, the hydrogen bonding pattern in β-sheet and α-helix, etc. The Scheme 1.1 illustrates different steps in ISD structure calculation. Structure calculation in traditional methods follows the following route: the interactions between atoms (i.e., sequential, short, medium and long-range assignments) are translated into conformational restraints, such as distance- and torsion angle restraints. The preliminary set of structures calculated from these conformational restraints is directly reverted to modelled data to check restraint violations in the experimental data. ISD, on the contrary, uses a probabilistic model and delegates the inversion task to the Bayesian inference machinery. The data is modelled in two steps. In the first step, idealized "mock data" are calculated using a forward model which may be a theory or physical law (Habeck 2012). The "mock data" in general, shows deviations from the experimental data (reasons for deviations may be systematic errors, experimental noise, theoretical shortcomings). For this reason, an error model is used to quantify any such deviation between the observed and modelled data. This results in a data probability, also called likelihood function. The likelihood is the probability of data with respect to a given structure "X". In the next step, the probability of every conformation has to be estimated so as to rank all of them accordingly.



Scheme 1.1 – Schematic representation of inferential structure determination (ISD); see text for details.

ISD uses a Monte Carlo based sampling algorithm (Gibbs sampler) for this purpose. There is yet a serious problem with Gibbs sampling: the Markov chain sometimes gets trapped in a single mode and cannot sample the entire posterior distribution. The replica-exchange Monte Carlo (RMC) or parallel temperature method (Swendsen and Wang 1986) circumvents this problem by using a temperature-like parameter which flattens the probability density. Molecular dynamics simulated annealing (MDSA) also uses the same principle by heating and slowly cooling the system, thereby releasing the trapped conformations. A drawback of MDSA however, is that if the system is not cooled slowly enough, it can still get trapped in

local minima. In contrast, RMC uses multiple heat-baths at different temperatures; each bath is called a replica. These replicas are sampled independently without interacting with each other. Nevertheless, they are allowed to exchange the sampled states. Thus, RMC maintains heat-baths at all temperatures throughout the simulation, allowing continuous escape of trapped states. The final step is the distribution of different conformations with respect to their probabilities. This distribution (called posterior distribution) is the product of likelihood function and prior, and encodes the final solution ensemble (Scheme 1.1). The equation to calculate the posterior distribution is:

$$\Pr(X \mid D,I) = \Pr(X \mid I)\ \Pr(D \mid X,I) \qquad (2)$$

$\Pr(X \mid D,I) \propto$ Posterior distribution

$\Pr(X \mid I)$ = Prior

$\Pr(D \mid X,I)$ = Likelihood

1.6.4 Example

Let us consider an example of inferring the torsion angle (φ) value for proteins with ISD. We already know that in most cases, the value of φ for proteins is negative. Therefore, this is our prior knowledge [$\Pr(X \mid I)$] which describes the probability to find a particular torsion angle even before we have acquired any experimental data (black dashed line in Figure 1.9). The likelihood [$\Pr(D \mid X,I)$] of the data is inferred (two curves with broken line, highlighted in orange) using the Bayesian inference machinery. This likelihood adds more information to our knowledge about the torsion angle. The posterior distribution [$\Pr(X \mid D,I)$] (red curve) is the product of prior probability and likelihood of data and is the total knowledge about the torsion angle (cf. equation 2). To rank each conformation according to its probability, Bayes' theorem is used which is particularly suitable to solve complex data analysis problems (Habeck 2012).

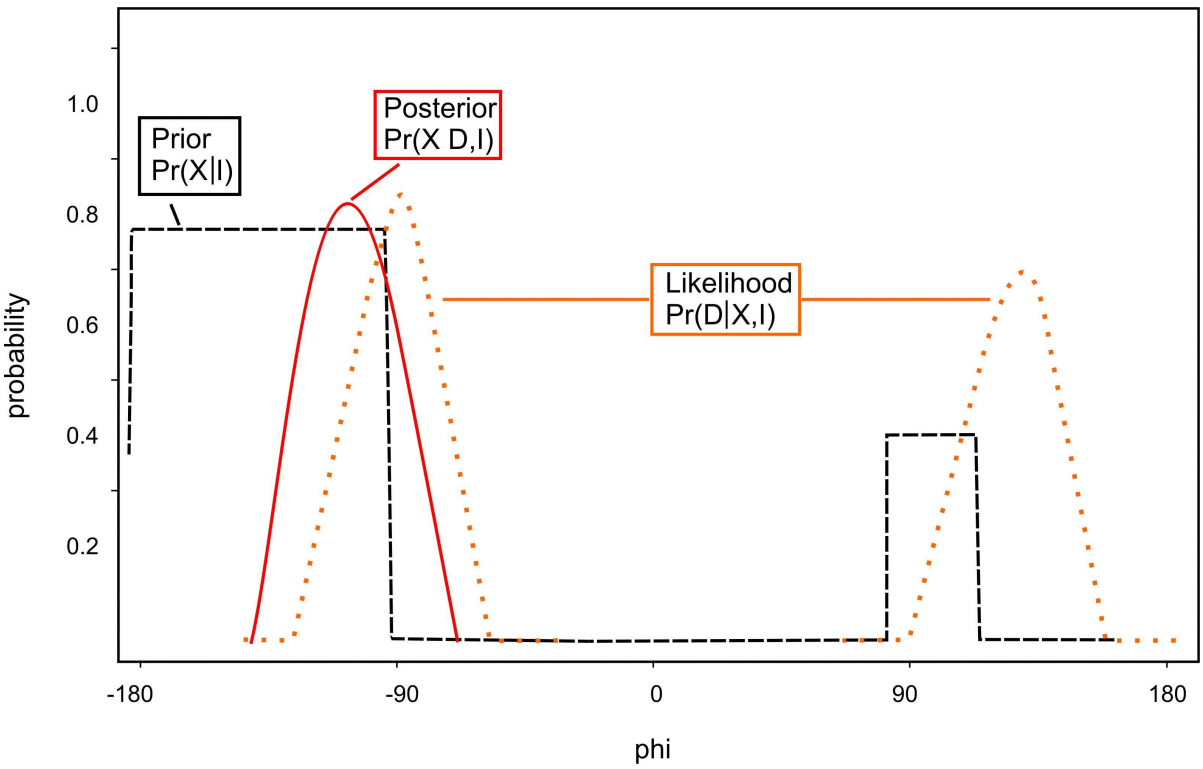


Figure 1.9 – Estimation of posterior distribution for torsion angle phi (φ). Black dashed line represents the prior knowledge about "φ". Except for glycine, the value of "φ" for all other residues is negative (i.e., lies in the left region of the Ramachandran plot). The likelihood of the data is shown by orange broken curves. Red curve shows the product of prior and likelihood that is called posterior distribution. The example and the Figure is adapted from reference (Habeck 2012).

1.7 Aim of the project

The autotransport mechanism in TAAs, despite being studied on YadA from the enteropathogen *Yersinia enterocolitica* (Grosskinsky, Schutz et al. 2007), at the moment is poorly understood. Knowledge about the structure and dynamics of the domains involved in the autotransport will contribute to a better understanding of this process and will shed light on the pathogenicity of TAAs. As membrane anchor domains are the definitive and conserved elements of TAA family, a high-resolution structure of the β-barrel domain of any member, may help in understanding the mechanistic details of the autotransport process, which should be applicable to all family members. The aim of this project was to determine the high-resolution structure of the prototypical YadA membrane anchor domain by solid-state MAS NMR. The membrane anchor domain of YadA (YadA-M) is an important target for structure determination because of its intermediate size, high stability and its availability in the form of microcrystals (Wollmann, Zeth et al. 2006). A general model of various steps from YadA translation to secretion is given in Figure 1.4 (Grosskinsky, Schutz et al. 2007).

After being synthesized, YadA polypeptide is delivered to the periplasm through Sec machinery where it is kept unfolded by some unknown chaperones. The C-termini are the first to insert in the outer membrane and fold to form the barrel. The latter stages in which the N-termini are exported out, remain unclear.

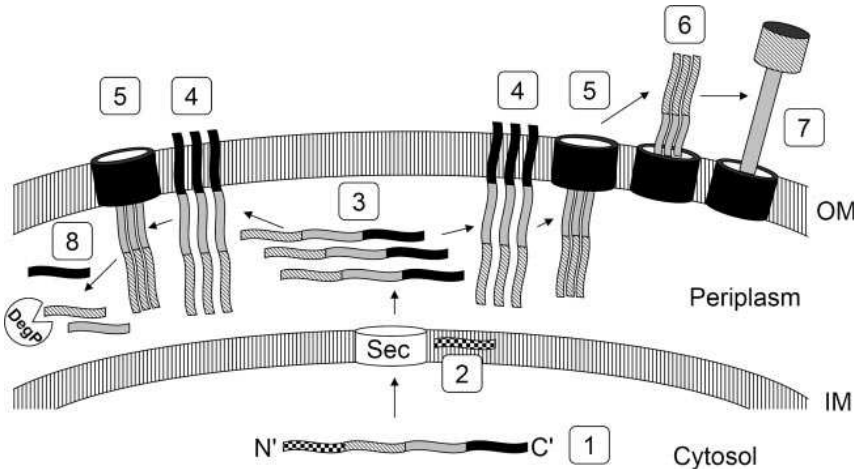


Figure 1.4 – Overall mechanism of autotransport in YadA, illustrated in different steps. N-terminal signal peptide is cleaved and the polypeptide is transported into the periplasm through Sec machinery. In the next step, the C-termini of the proteins get inserted in the outer membrane and oligomerize to form a β-barrel. Any misfolding at this stage leads to degradation of the protein by DegP. A successful transport is dependent upon the correct folding of the barrel domain. BamA is involved in the biogenesis of YadA β-barrel (Lehr, Schutz et al. 2010). The Figure was reproduced with permission from reference (Grosskinsky, Schutz et al. 2007).

Although a model of YadA-M and an X-ray structure of Hia are available, yet they fail to give conclusive information about autotransport mechanism. Point mutation of conserved G72 has shown decreased trimer stability in YadA-M (Grosskinsky, Schutz et al. 2007). However, the YadA-M model suggests that larger side-chain residues could be accommodated at the position of this glycine (Grosskinsky, Schutz et al. 2007). To understand why large side-chain residues in this position result in a reduced protein's stability and autotransport activity, demands a high-resolution X-ray crystallography or a nuclear magnetic resonance (NMR) molecular structure of YadA-M (Schutz, Weiss et al. 2010). Solid-state NMR is an excellent tool to get information about any dynamic or flexible domains which, in case of YadA-M, would most likely be involved in the autotransport.

2 Materials and Methods

2.1 YadA-M biosynthesis, purification and crystallization

YadA-M construct used in the NMR studies was prepared in Max Planck Institute for Developmental Biology. In this chapter, the general culture conditions and protein purification to crystallization steps are briefly described. The large-scale expression and purification were carried out as described by Wollman et al., however, some modifications were necessary. Labeling and crystallization were done by Dirk Linke (unpublished work). For details of the preparation, the reader is directed to the reference (Wollmann, Zeth et al. 2006).

Culture conditions and protein purification

YadA-M was overexpressed in the outer membranes of a special strain of *E. coli* ($CaCl_2$-competent *E. coli* BL21 DE3 Omp8). The expression strain did not readily grow in minimal medium, therefore fully enriched ^{13}C-^{15}N-labelled medium (BioExpress, Cambridge Isotope Laboratories) was used. The regular strain of *E. coli* could grow very well in the minimal medium; unfortunately, it could not express the protein in large amounts. Therefore the preparation of YadA-M was a trade-off between low expression with cheap labeling, and the large expression with expensive labeling strains. OmpA signal peptide for protein transport and an N-terminal strep-tag (15 residues) for protein purification were introduced using the pASK-IBA system. The outer membranes of *E. coli* were separated from inner membrane proteins and lipids by ultracentrifugation and differential solubilization with N-lauroylsarcosine. YadA-M was purified by cation exchange chromatography and phase separation as described (Wollmann, Zeth et al. 2006; Arnold and Linke 2007). 5 L of medium yielded approximately 200 mg wet weight of the uniformly labelled microrystalline material.

Crystallization

Microcrystals of YadA-M were obtained after dialysis of purified protein fractions from cation exchange chromatography (Wollmann, Zeth et al. 2006) against distilled water (ddH2O). Salts and the detergent C8POE (the non-ionic detergent is relatively easily removed by dialysis) were removed until the contents of the 25 kDA cutoff dialysis tubing

(Roth, Germany) became turbid. The microcrystals were harvested by centrifugation and kept wet in ddH2O.

Electron microscopy

Microcrystalline material was resuspended and 5 µl were spotted on freshly prepared EM grids. After leaving the drop for 1 min to adsorb the crystals, water was removed using a filter paper, and crystals were stained with 1% Uranylacetate solution for 10-20 sec. The staining solution was removed and the crystals were washed with water, and were air-dried after draining all residual water, again using filter paper.

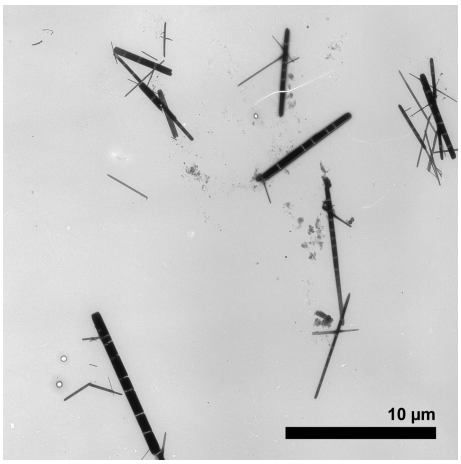


Figure 2.1 - Electron micrograph of microcrystals of YadA-M. The needle-shaped crystals have a length of 5-10 µm. The crystals are stable in ddH_2O for long time periods.

2.2 Solid-state MAS NMR

All solid-state MAS NMR experiments were performed at 275 K using samples from a single, ^{13}C, ^{15}N uniformly labelled YadA-M batch preparation. Data were obtained on 400, 600, 700, 850 and 900 MHz AVANCE spectrometers (Bruker, Karlsruhe, Germany). Spectrometers were equipped with double ($^{1}H/^{13}C$) and triple-resonance ($^{1}H/^{13}C/^{15}N$) CP/MAS probes (Bruker). The MAS frequency was adjusted to 10 or 13 kHz, for MAS rotors with a diameter of 4.0 and 3.2 mm, respectively. Except otherwise mentioned, magnetization transfer from ^{1}H to the ^{13}C or ^{15}N spins was achieved with ramped cross-polarization (CP) (Schaefer and Stejskal 1976; Metz, Wu et al. 1994); the ramp was kept relatively shallow (typically between 75 and 100% of the spin-lock field strength). Typical recycle delays between individual transients were kept between 2.7-3.0 s, to avoid sample heating. High-power proton decoupling with radio-frequency (rf) field strengths of 75-90 kHz using the TPPM (Bennett, Rienstra et al. 1995) or SPINAL-64 scheme (Sinha, Grant et al. 2005) was applied during evolution and detection periods. Typical ^{1}H and ^{13}C $\pi/2$ pulse lengths were 3.0 and 3-4 µs, respectively, typical ^{15}N $\pi/2$ pulses were longer, 6-7 µs.

2.2.1 Homonuclear ^{13}C-^{13}C correlation spectra

Homonuclear 2D ^{13}C-^{13}C correlation spectra with dipolar-assisted rotational resonance **(RAD/DARR)** (Takegoshi, Yano et al. 2001; Morcombe, Gaponenko et al. 2004) proton-driven spin diffusion (PDSD) (Szeverenyi, Sullivan et al. 1982; Suter and Ernst 1985) and DREAM (dipolar recoupling enhancement through amplitude modulation) mixing schemes (Verel, Baldus et al. 1998; Pauli, Baldus et al. 2001) were recorded at 900 MHz. For the RAD experiments, the following conditions were used: experiments were obtained at different MAS frequencies, i.e., 11, 12, and 13 kHz; initial ^{13}C magnetization was created using a 3.0 ms ramped CP (75 – 100%) with average rf-field strengths of 63 and 53 KHz, for ^{1}H and ^{13}C, respectively. Mixing times of 15, 25, 50, 100, 200, 300 and 500 ms were used, in order to probe residue specific, sequence specific, medium range and long range transfers. A 3.0 s recycle delay was used. Acquisition times were typically 18.0 and 9.6 ms in t_2 and t_1, respectively.

The ***J*-decoupled DARR** spectra were recorded at 900 MHz with a MAS frequency of 13 kHz. Mixing times of 15, 40 and 70 ms were used. *J*-decoupling was obtained in the indirect dimension using a combination of hard and band-selective (Gaussian) 180° pulses in the Cα region (Straus, Bremi et al. 1996). The PDSD experiments were recorded under similar conditions, but with a CP contact of 1.0 ms (75 – 100% ramp on carbons), and acquisition times in the direct and indirect dimensions of 15.0 and 4.0 ms, respectively. Two **PDSD** spectra with 15 and 100 ms mixing periods were recorded.

2D ^{13}C-^{13}C DREAM spectra were obtained at 13 kHz MAS; a CP contact of 1 ms (75 – 100% ramp on ^{13}C) was used. During the DREAM mixing period of 1.5 ms, the carrier was placed either between the Cα and Cβ regions (at ~50 ppm, Cα-Cβ spectrum) or between the Cβ and Cγ regions (at ~30 ppm, Cβ-Cγ spectrum). Acquisition times in the direct and indirect dimensions were 15.0 and 8.2 ms. The experiments were recorded with 1024 increments, 48 scans per increment and a recycle delay of 3.0 s, yielding a total experimental time of 41 hours per spectrum.

2D ^{13}C-^{13}C proton-assisted recoupling (PAR) (De Paepe, Lewandowski et al. 2008) correlation spectra were acquired at 850 MHz. PAR mixing times of 2.25, 6.0 and 15.0 ms were applied to achieve long range exchange. Initial ^{13}C magnetization was created with a 600 μs CP transfer from ^{1}H to ^{13}C (80–100% ramp on ^{1}H; 73 kHz average field strength). During evolution and acquisition, protons were decoupled using SPINAL-64 with 80 kHz field strength. Acquisition times of 20.0 ms and 13.2 ms were used in the direct and indirect

dimensions, respectively. The recycling delay was set to 2.8 s. The experiment time for one spectrum was approximately 18 hours.

2.2.2 Heteronuclear ^{15}N, ^{13}C correlation spectra

2D NCA and **NCO** experiments were recorded at 400 MHz, using a MAS frequency of 8 kHz. Initial magnetization was created using a 1 ms CP transfer from ^{1}H to ^{15}N (75 – 100% on ^{15}N). Selective transfer from ^{15}N to ^{13}C was achieved using an adiabatic CP transfer of 4 ms. The effective acquisition times were 20 ms and 12.8 ms for ^{13}C and ^{15}N, respectively. In total 128 increments were recorded in the ^{15}N dimension, with 64 scans per increment. Using a recycle delay of 3.0 s, the experimental time per spectrum was about 7 hours. The same experiments were extended to 2D NCACX and NCOCX with a PDSD mixing of 35 ms.

3D NCACX and 3D NCOCX (Castellani, van Rossum et al. 2003) experiments were recorded at 600 MHz, at a MAS frequency of 10 kHz. A CP contact of 2 ms was employed, with 60 and 43 kHz rf-field on ^{1}H and ^{15}N, respectively, with a 75 – 100% ramp on protons. Polarization was transferred from ^{15}N to $^{13}C\alpha$ or $^{13}C'$ with an adiabatic CP contact of 4 ms, with rf-fields of 35 and 25 kHz on ^{13}C and ^{15}N, respectively. For the NCACX spectra, PDSD mixing times of 35, 100, 200 and 500 ms were used; for the NCOCX spectra, PDSD mixing times of 35 and 200 ms were applied. The 3D spectra were obtained as data matrices of 1372 x 48 x 32, with effective acquisition times of 12 , 5.2 and 6.4 ms for ^{13}C, $^{13}C\alpha/^{13}C'$ and ^{15}N, respectively. For each increment, 104 scans were averaged with a 2.8 s recycle delay. Hence, the total experimental time for each 3D spectrum amounted 124 hours.

The **3D NCACB** experiment was recorded under similar conditions, however with a DREAM mixing scheme of 3 ms to selectively exchange between $C\alpha$ and $C\beta$ (Verel, Baldus et al. 1998; Pauli, Baldus et al. 2001). The NCACB experiment was recorded as data matrix of 2484 x 64 x 48, with effective acquisition times of 20.0, 5.1 and 5.4 ms for $^{13}C\beta$, $^{13}C\alpha$ and ^{15}N, respectively. A recycle delay of 3 s was used, and 56 scans were recorded per increment, leading to a total acquisition time of 144 hours.

The ^{15}N-^{13}C **TEDOR** (transferred echo double resonance) (Hing, Vega et al. 1992; Jaroniec, Filip et al. 2002) experiment was recorded at a field of 9.4 T (400 MHz), at 8 kHz MAS and at a temperature of 275 K. The two **REDOR** (rotational echo double resonance) (Gullion and Schaefer 1989) mixing periods were 1.0 ms each. Carbon and nitrogen 180° pulses were 6.2 μs and 13.4 μs, respectively. Initial ^{13}C magnetization was created with a 2.5 ms ramped ^{1}H-

^{13}C cross-polarization (75 – 100% ramp on the ^{13}C channel). During mixing and acquisition, ^{1}H-heteronuclar decoupling was applied with a TPPM scheme (Bennett, Rienstra et al. 1995), at a moderately high ^{1}H rf-field strength of 75 kHz. The spectrum was recorded with 224 scans, 128 increments and effective acquisition times of 22 ms and 16 ms, in the ^{13}C and ^{15}N dimensions, respectively. With a recycle delay of 2.8 s, the experimental time for the 2D experiment amounted 23 hours. For medium and long range restraints 2D ^{13}C-^{15}N TEDOR with REDOR mixing times of 2.24, 6.0 and 12.0 ms were recorded at 850 MHz. The ^{13}C and ^{15}N $\pi/2$ pulse lengths used in these spectra were 3.25 and 6 µs, respectively. The ^{15}N REDOR π-pulse length was 12 µs. SPINAL-64 ^{1}H decoupling with 90 kHz rf-field strength was applied during t_1 and t_2, and during the REDOR-mixing steps. They were recorded with 64 transients and the recycling delay was set to 4.0 s. The acquisition times in the ^{13}C and ^{15}N dimensions were 20 and 9.3 ms, respectively.

2.2.3 Methyl-filtered spectra

^{13}C-^{13}C Methyl-filtered Post-C7 experiments were obtained at 900 MHz, at a spinning frequency of 8 kHz. Magnetization is prepared by using a methyl filter (Jehle, Hiller et al. 2006); for this, a long ^{1}H-^{13}C cross-polarization (CP) contact of 2 ms is directly followed by a short polarization inversion (PI) of 70 µs, achieved with a 180° phase change of the ^{1}H spin-lock pulse (CPPI) (Wu, Burns et al. 1994). A polarization inversion period of 60 µs was found to be sufficient for 'depolarizing' non-methyl, protonated nuclei (Opella and Frey 1979; Opella, Frey et al. 1979). Following this preparation, a post-C7 (permutationally offset stabilized C7) scheme was used to obtain a 2D ^{13}C-^{13}C correlation spectrum (2Q/1Q) (Hohwy, Jakobsen et al. 1998). Acquisition time in the direct dimension was 12.0 ms, in the indirect (2Q) dimension 280 experiments with an increment of 32 µs were recorded.

^{13}C-^{13}C Methyl-filtered RAD/DARR experiments were obtained at 900 MHz by use of CP, directly followed by polarization inversion (PI), achieved with a 180° phase change of the ^{1}H spin-lock pulse (CPPI) (Wu, Burns et al. 1994) prior to t_1 evolution. A polarization inversion period of 60 µs was found to be sufficient for 'depolarizing' non-methyl, protonated nuclei.

A double methyl-filtered ^{13}C-^{13}C RAD/DARR experiment was obtained by addition of a 40 µs dipolar dephasing period (Opella and Frey 1979; Opella, Frey et al. 1979) immediately after DARR mixing, which was set to 300 ms. Single and double filtered spectra

were recorded with 96 transients and a recycle delay of 3 s. Acquisition times in the direct and indirect dimensions were 24.4 and 6.6 ms, respectively.

2.2.4 ChhC and NhhC

2D **ChhC** and 2D **NhhC** (Lange, Becker et al. 2005) correlations were recorded at 700 MHz. The ChhC experiments were acquired using a 1.5 ms CP contact from ^{1}H to ^{13}C, with average rf nutation frequencies of 50 and 60 kHz, respectively. For the ChhC experiment, the second and third rectangular CP contacts were kept short at 60 μs, to avoid mixing between non-covalently bonded carbons and protons. Acquisition times were 12 ms in t_1 and 7 ms in t_2. Various ^{1}H-^{1}H mixing times were used (35, 50, 80, 150, 200, 300 and 500 μs) to collect distance restraints in different ranges. During evolution and detection, SPINAL-64 decoupling with an rf-field strength of 85 kHz was applied. A 3.0 s recycle delay was used. The NhhC spectra were obtained under similar conditions, except for the ^{1}H-^{1}H mixing times, which were set to 35, 50, 100 and 200 μs. Initial ^{15}N magnetization was created with a 1.5 ms CP transfer (from ^{1}H to ^{15}N); the second and third rectangular CP were set to 310 μs (^{15}N to ^{1}H) and 60 μs (^{1}H to ^{13}C).

2.3 Data processing and analysis

All solid-state MAS NMR data were processed with TOPSPIN 2.1 (Bruker, Germany) and was analysed using SPARKY 3.113 (T. D. Goddard and D. G. Kneller, Sparky 3, University of California, San Francisco). Secondary structure from the primary sequence was determined by PSIPRED (McGuffin, Bryson et al. 2000). The backbone dihedral angles were obtained from the ^{15}N, Cα, Cβ and C` resonances using TALOS+ (Shen, Delaglio et al. 2009). The predicted dihedral angles were filtered to include only residues with more than 50% confidence of being in canonical secondary structure. This filtering resulted in 60 phi/psi restraints among which 56 were “Good" predictions and 4 had to be taken with caution (“Warn”category). The structure calculation was carried out using ISD (Rieping, Habeck et al. 2005) and ARIA (Rieping, Habeck et al. 2007).

3 Chemical Shift Assignment of YadA-M

3.1 Introduction

As introduced in Chapter 1, membrane proteins are involved in many important cellular processes, they play a key role in many diseases and are in prime focus of pharmaceutical industry, as they represent 90% of all drug targets. For a better understanding of the biological functions of membrane proteins, knowledge about their structure is essential. Despite their biological relevance, membrane proteins are heavily underrepresented in the Protein Data Bank (PDB), and the deposited high-resolution structures of membrane proteins amount less than 1% of all entries. As described in Chapter 1, this low representation in the PDB is largely brought about by difficulties to study them with established methods in structural biology, such as X-ray crystallography and solution NMR. Solid-state NMR, on the other hand, does not rely on availability of high-quality crystals, nor on rapid molecular tumbling, hence serves as a promising method for structural studies of disordered materials such as membrane proteins (McDermott 2009).

The chemical shift assignment forms an essential first step towards structure determination of proteins. In this chapter, is reported the ^{13}C and ^{15}N chemical shift assignment of the membrane anchor domain of YadA (YadA-M) by solid-state MAS NMR. It was possible to sequentially assign YadA-M almost completely, using a single, uniformly ^{13}C, ^{15}N labelled micro-crystallized sample as prepared by Wollmann et al. (Wollmann, Zeth et al. 2006). The YadA-M construct used in these studies consists of the trimeric membrane anchor domain and the first part of the stalk. Each protomer in the trimer has 105 residues, of which the first fifteen residues are part of a highly flexible strep-tag introduced to aid in protein purification. Attempts to prepare high-quality single crystals of YadA-M diffracting well enough for structure studies with X-ray crystallography were thus far unsuccessful (Wollmann, Zeth et al. 2006). However, during the efforts to optimize crystallization conditions, protocols were found to prepare microcrystalline material. Such 'messy' crystals are highly suitable for structure studies with solid-state magic-angle spinning (MAS) NMR, since this technique neither relies on availability of high-quality crystals, nor on rapid molecular tumbling. In fact, the high degree of structural order of the bulk material within

the microcrystals aids to reduce inhomogeneous broadening of the NMR lines (Martin and Zilm 2003).

3.2 YadA-M, an ideal membrane protein for solid-state MAS NMR

The microcrystalline preparation of YadA-M gives highly resolved lines which helped to accomplish its assignment. The spectral quality is illustrated with a one-dimensional (1D) ^{1}H-^{13}C cross-polarization (Metz, Wu et al. 1994) MAS NMR spectrum recorded from uniformly ^{13}C,^{15}N-labelled YadA-M (Figure 3.1A). The spectrum is well-dispersed as demonstrated by the many fine details in the generally poorly resolved carbonyl region around 172 ppm (shown as inset in Figure 3.1A). The overall high dispersion is partly the result from the secondary structure of TAA membrane anchors, which for YadA-M is about half α-helix / half β-sheet (Wollmann, Zeth et al. 2006).

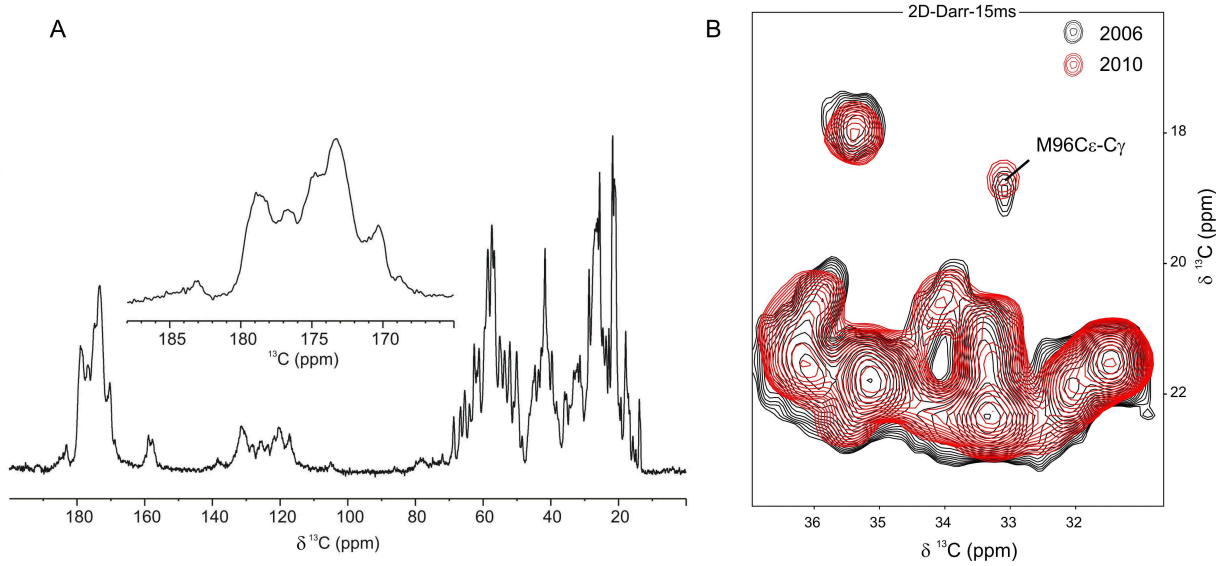


Figure 3.1 – One-dimensional (1D) ^{13}C and two-dimensional (2D) ^{13}C-^{13}C spectra recorded on YadA-M. The spectra are recorded at a field of 900 MHz, a spinning frequency of 13 kHz and at a temperature of 275 K, using a standard ^{1}H-^{13}C cross-polarization (75 – 100% ramp on ^{13}C). **(A)** 1D ^{1}H-^{13}C cross-polarization spectrum with well resolved lines. The dispersion in the carbonyl region is shown in the inset. **(B)** Overlay of two contour plots of 2D ^{13}C-^{13}C DARR spectra (mixing time 15 ms), recorded of the same uniformly ^{15}N,^{13}C-labelled YadA-M sample, but with a time interval of ~4 years. There is no sign of any significant change in the position of the methionine Cε-Cγ correlation. Since the methyl chemical shift of methionine is highly sensitive to any change in the sample due to oxidation of sulphur, this indicates that no oxidation of M96 had taken place, and that sample integrity was maintained during the course of solid-state MAS NMR experiments.

In addition, Fourier-Transform Infrared (FTIR) studies on YadA-M have shown that the main part of the protein is extremely rigid, leading to a very high spectral resolution also in FTIR (Wollmann, Zeth et al. 2006). The (micro-) crystalline environment and the overall rigidity of the protein will contribute to the structural homogeneity and hence, to a narrow line width. In contrast, the strep-tag that is part of the YadA-M preparation required for the protein purification is known to be highly mobile (colour-coded in grey, see Figure 3.4 below) (Wollmann, Zeth et al. 2006). As a result, it can be expected that the MAS NMR signals of residues in the strep-tag will be very weak or not observed at all. The presumption that all strong signals arise from residues in the rigid part of the protein was proven during the course of the assignment procedure by the self-consistency of the assignment. As a quality check and to monitor the integrity of the sample, 1D spectra were recorded directly before and after any multidimensional experiment. During the course of the work, no indication for change or degradation of the sample was observed. Sample integrity was also monitored by following the Cε-Cγ correlation of the (single) methionine in the sequence (M96), which did not show any sign of oxidation (Figure 3.1B). Hence, these findings imply that the YadA-M trimer is remarkably stable over a longer period of time and therefore is a very suitable system to be studied by solid-state MAS NMR.

3.3 Residue specific assignment of YadA-M

In the assignment procedure, a table is constructed in which each side-chain nuclear spin is uniquely linked to its backbone. This procedure begins with identifying individual spin systems in two-dimensional (2D) ^{13}C homonuclear correlation spectra. Every amino-acid residue gives a unique fingerprint pattern in these spectra, in which the backbone and side-chain carbon nuclei are correlated. This is demonstrated by a schematic representation in Figure 3.2A, which shows hypothetical cross peaks for a threonine and alanine spin system. In Figure 3.2B, a protein fragment is shown with arrows illustrating the through-space magnetization exchange among aliphatic carbon nuclei. DARR (Takegoshi, Yano et al. 2001) and PDSD (Suter and Ernst 1985) schemes are commonly used to record correlation spectra; here, the mixing time plays a crucial role to see different correlations of interest. For residue-specific assignment, short mixing times (typically less than 25 ms) are sufficient to observe most of the side-chain/backbone correlations. A contour plot from the aliphatic region of a 2D ^{13}C homonuclear correlation spectrum recorded on YadA-M with 25 ms DARR mixing is

shown in Figure 3.2C. To exemplify the assignment procedure, threonine and alanine are assigned and are highlighted with black and grey solid lines, respectively. Such 'fingerprint patterns' of the various types of amino acids in 2D carbon-carbon correlation experiments are extremely helpful for identification of different spin systems in the assignment procedure.

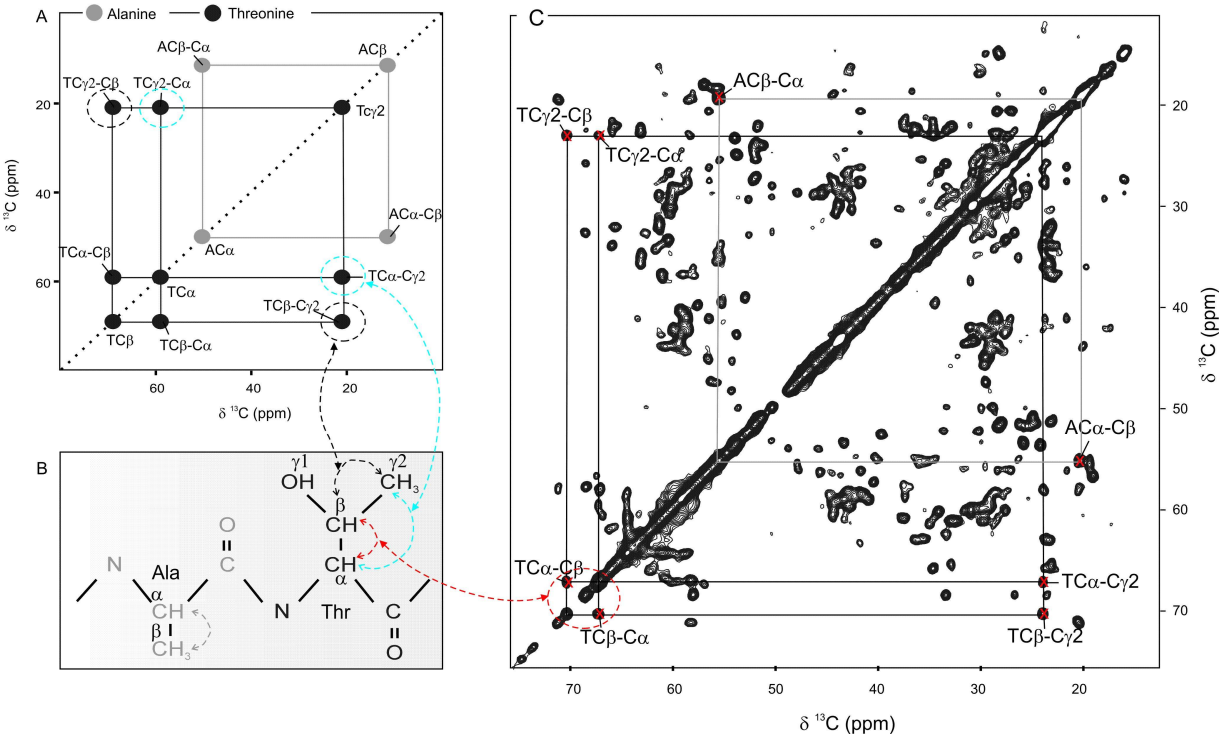


Figure 3.2 – Illustration of residue-specific assignment. (A) The characteristic fingerprint pattern of an alanine (grey) and threonine (black) in the aliphatic region is shown. Spins are interacting with each other through dipolar coupling which is utilized to produce correlation peaks in solid-state MAS NMR spectra. **(B)** The sequence of an alanine and threonine are drawn. Arrows illustrate the through-space magnetization transfer within the spin system. **(C)** Aliphatic part of 2D DARR spectrum. Black and grey solid lines show correlation peaks emerging from alanine and threonine spin systems, respectively.

Figure 3.3 shows the aliphatic and carbonyl regions of a 2D ^{13}C-^{13}C homonuclear correlation spectrum of YadA-M. The correlations were obtained by using a DARR/RAD scheme with 25 ms mixing (Takegoshi, Yano et al. 2001; Morcombe, Gaponenko et al. 2004). The spectrum is fairly well resolved and several amino-acid types can be directly identified upon their characteristic correlation pattern and unique chemical shifts, such as threonines, serines and isoleucines; other residues give fingerprint patterns in the more congested regions of the spectra and are not readily identified due to ambiguity; still, in many cases, side-chain correlation patterns can be traced putatively and their assignment is achieved by checking for self-consistency in the different spectra of the data set.

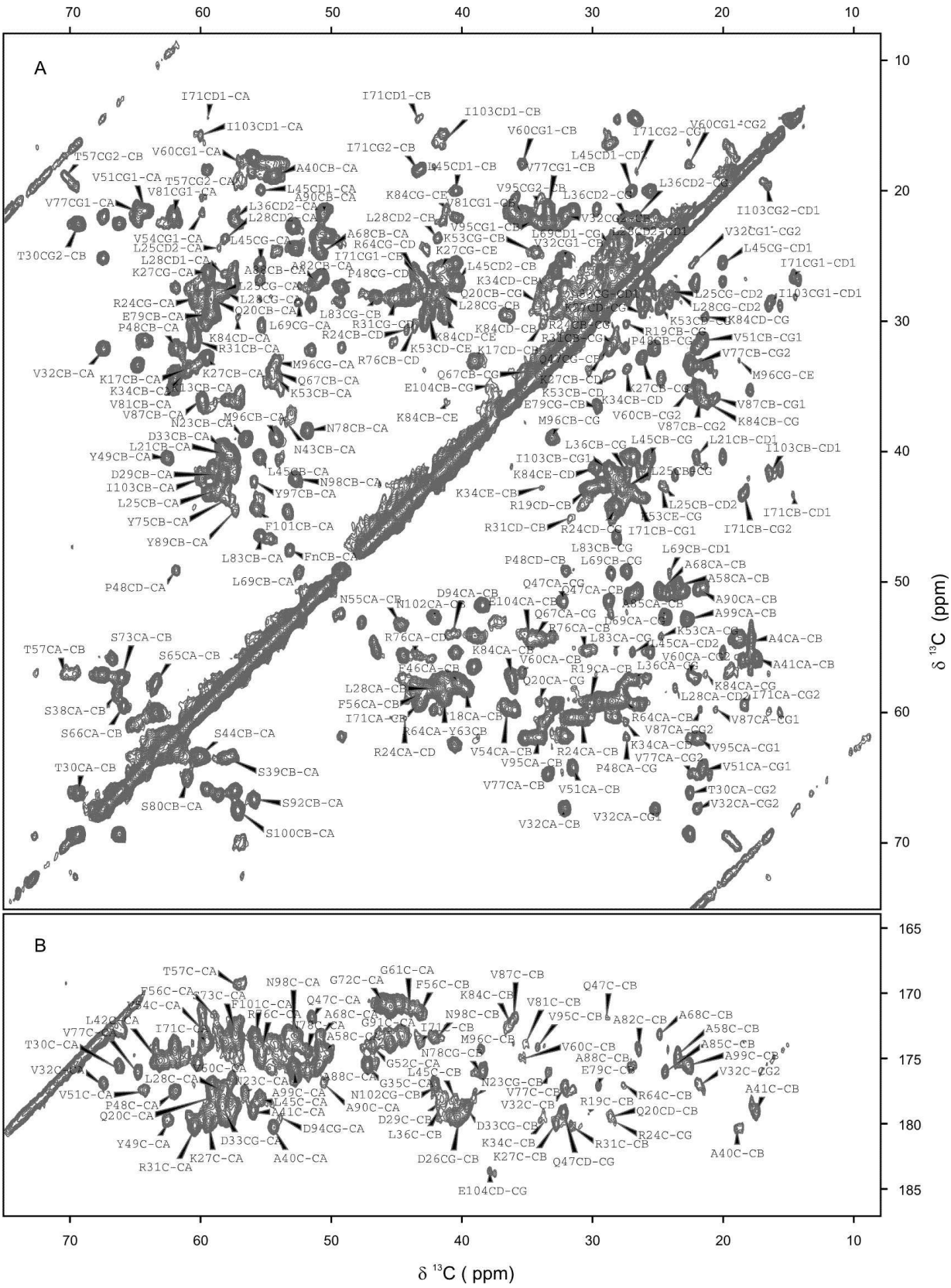


Figure 3.3 – 2D ^{13}C-^{13}C correlation spectrum of YadA-M. The spectrum was recorded on uniformly ^{13}C, ^{15}N labelled YadA-M on a 900 MHz Bruker spectrometer at 275 K. The spinning frequency was set to 12 kHz. Magnetization between ^{13}C spins was exchanged by use of a 25 ms DARR mixing period. The sequence-specific assignment of spin systems is included, based on analysis of a data set containing multidimensional heteronuclear NCACX and NCOCX spectra (*vide infra*).

Glycines do not have side chains, hence, they do not provide cross peaks in the aliphatic region. However, they are readily identified upon their unique upfield-shifted Cα-C` correlations (Figure 3.3B).

3.3.1 Starting points for sequence specific assignment

The identification of fingerprint patterns yields residue-specific, but mostly not sequence-specific assignments. However, when these patterns are distinctive either on the basis of well-dispersed chemical shifts or arise from residues that are unique or appear in a very little number, they provide good starting points for the sequential assignment. YadA-M has a highly repetitive primary sequence of mainly alanines (13), glycines (12), serines (12), leucines (8), valines (8), asparagines (6), aspartates (6), lysines (6), arginines (5), glutamines (5), tyrosines (5) and phenylalanines (5); together these residues constitute 91 out of 105 in total for YadA-M. In total eight unique spin systems were found as suitable starting points for the sequential assignment. This is illustrated in Figure 3.4, which shows the primary sequence of YadA-M; colour-coded in red are observable amino acids that appear either only once (H16, P48, M96) or twice in the sequence (T30 and T57; I71 and I103). From these starting points, the neighbouring residues can be assigned (colour-coded blue). There are three glutamic acids in YadA-M (E12, E79 and E104). E12 resides in the strep-tag and is not observed. Since E104 is identified by its neighbouring I103, the remaining glutamic acid (E79) is considered unique and colour-coded red as well. In addition, pair-wise residues that form unique doublets with at least one of the members easily identifiable are coloured green. 19 of such pairs were found which additionally provided unambiguous starting points for the sequential assignment.

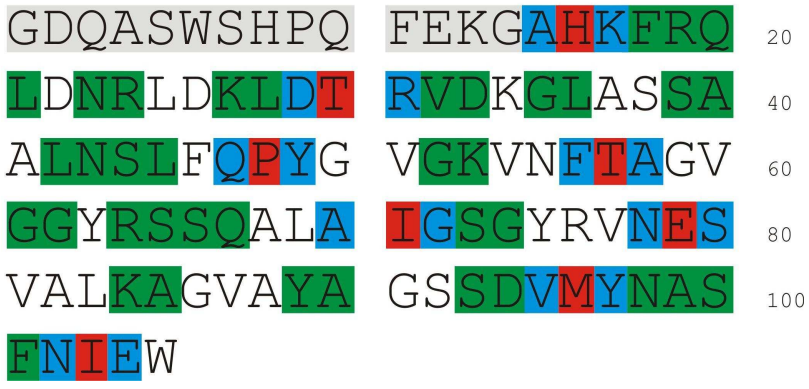


Figure 3.4 – Primary sequence of YadA-M, colour-coded with potential starting points for sequential assignment. Residues that appear only once (H16, P48, M96) or twice (T30 and T57; I71 and I103) are coloured red, their sequential neighbours blue, and unique sequential pairs green.

3.4 Sequence-specific assignment of YadA-M

Sequential relationships are established by linking ^{13}C resonances to the backbone-amide ^{15}N using multidimensional experiments. This was achieved using a data set consisting of several 2D and three-dimensional (3D) homo- and heteronuclear correlation experiments (*vide infra*). For the sequence-specific assignment of the residues, it is mandatory to know the backbone ^{15}N chemical shift for each residue. For this purpose, band-selective heteronuclear ^{13}C-^{15}N correlation spectra are recorded. Here, the essential building block is a selective polarization transfer step between the amide ^{15}N and the $^{13}C\alpha$ (which is called NCA transfer) or the C′ (NCO transfer) (Baldus, Petkova et al. 1998), achieved with adiabatic cross-polarization optimized for either the NCA condition or the NCO condition (Hediger, Meier et al. 1994). The NCA transfer is int**ra**residual, i.e., it connects between $^{15}N_i$ and $^{13}C\alpha_i$ of the same residue; the NCO transfer is int**er**residual and involves a crossing of the peptide bond to connect $^{15}N_i$ with $^{13}C'_{i-1}$. Figure 3.5 shows a 2D NCA (right) and NCO spectrum (left). In the NCA spectrum, certain residues give rise to fingerprint signals. For instance, glycines and alanines provide easily identifiable correlations due to their upfield shifted $^{13}C\alpha$ signals, and upfield (glycine) or downfield (alanine) shifted ^{15}N signal. Valines are characterized by a downfield shifted $^{13}C\alpha$ and ^{15}N resonance. Serines have an upfield ^{15}N signal and an intermediate $^{13}C\alpha$ signal (between 55 - 60 ppm). These identification ‘rules’ can only be applied in one direction: whereas it is true that in the typical glycine, alanine or valine NCA region no other residues are found, the other way round it is not true, as demonstrated by the shifts of, e.g., G52, A41, A99 and V87. In the NCO spectrum (Figure 3.5, left), carbonyl spins are correlated with backbone nitrogen spins. Note that in the NCO experiment no fingerprint patterns are found, since the correlations connect between spins of different amino acids. The two spectra are linked by the ^{15}N frequency of the backbone amide, providing sequential connectivities. Several examples to illustrate this pair-wise linking are highlighted in red in Figure 3.5. In the NCA spectrum for instance, the backbone $C\alpha$ of G59 is correlated with its amide nitrogen at ^{15}N=104.3 ppm. In the NCO spectrum, the carbonyl signal for A58 should appear at the same nitrogen shift. It is indeed observed that at ^{15}N=104.3 ppm in the NCO spectrum, the carbonyl signal of A58 is present. To determine the amide chemical shift for A58, the NCA spectrum is scanned and at ^{15}N=125.4 ppm, the A58

Cα is identified. In such a way, backbone ^{15}N chemical shifts of various amino acids are determined by establishing sequential links.

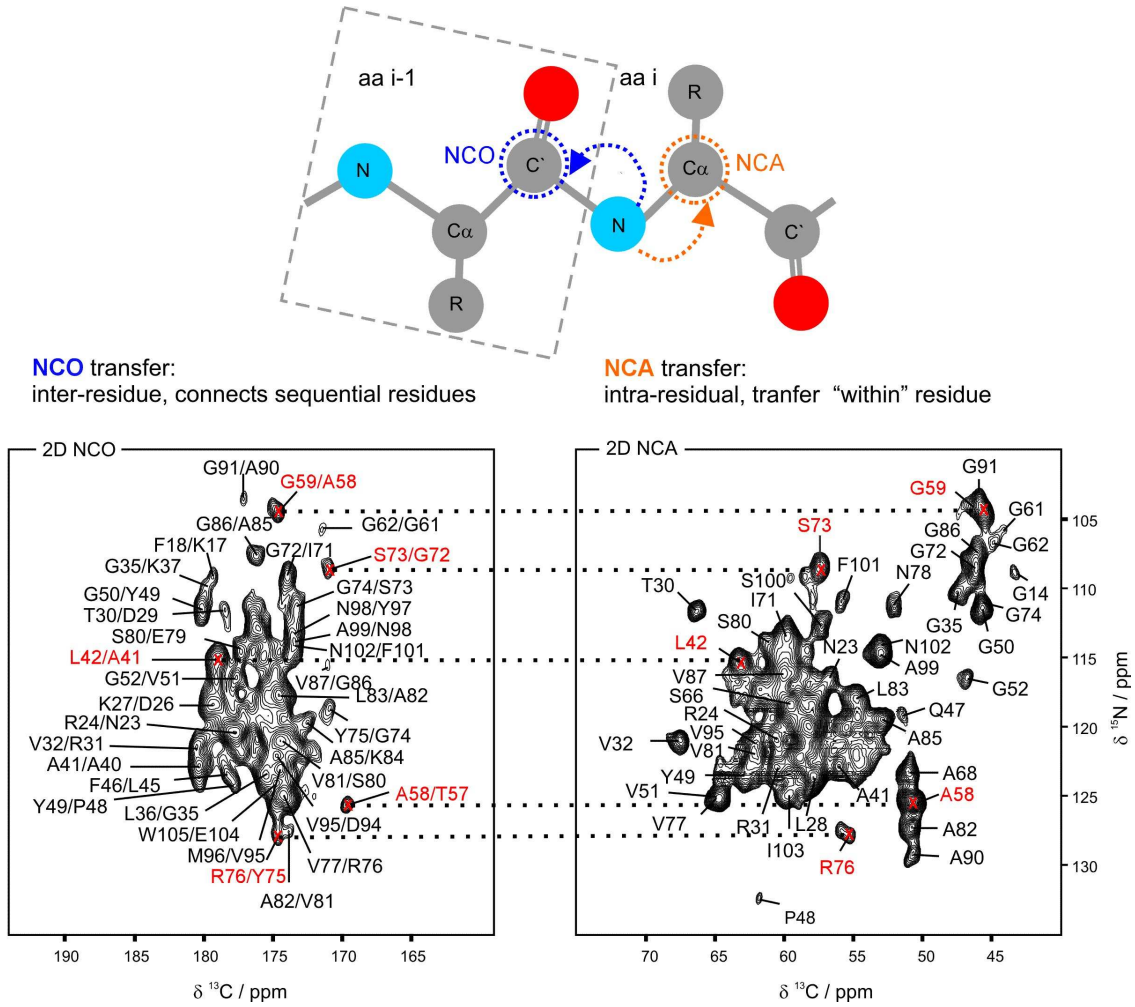


Figure 3.5 – Sequential assignments using two-dimensional ^{15}N, ^{13}C heteronuclear correlation spectra. Contour plots of 2D heteronuclear NCA and NCO (bottom right and left, respectively) showing correlation of ^{15}N with Cα of the same and C` of the preceding residue. Spectra were recorded at a field of 400 MHz, a spinning frequency of 8 kHz and at a temperature of 275 K. Polarization exchange between ^{1}H and ^{15}N was obtained by use of a 1 ms standard cross-polarization (CP) period; selective transfer from ^{15}N to ^{13}C' (NCO) or ^{13}Cα (NCA) was achieved by a 4 ms adiabatic CP. A schematic representation of intra-residual NCA and inter-residual NCO magnetization transfer is shown on top part of the diagram.

3.4.1 NCACX and NCOCX

In general, complete assignment walks can only be established if the NCA and NCO correlations are 'dressed' with the shifts of the side-chain spins, which is usually done by extending the NCA or NCO experiments with a non-specific ^{13}C-homonuclear transfer step (e.g., by employing a proton-driven spin diffusion (PDSD) transfer step (Szeverenyi, Sullivan

et al. 1982; Suter and Ernst 1985), or DARR/RAD mixing). The new experiments are referred to as either NCACX or NCOCX (where CX denotes 'any' carbon) (Pauli, Baldus et al. 2001). The assignment procedure is then to identify the same side-chain resonance pattern correlated to its own backbone ^{15}N (NCACX) and the backbone ^{15}N of the next residue in the sequence (NCOCX).

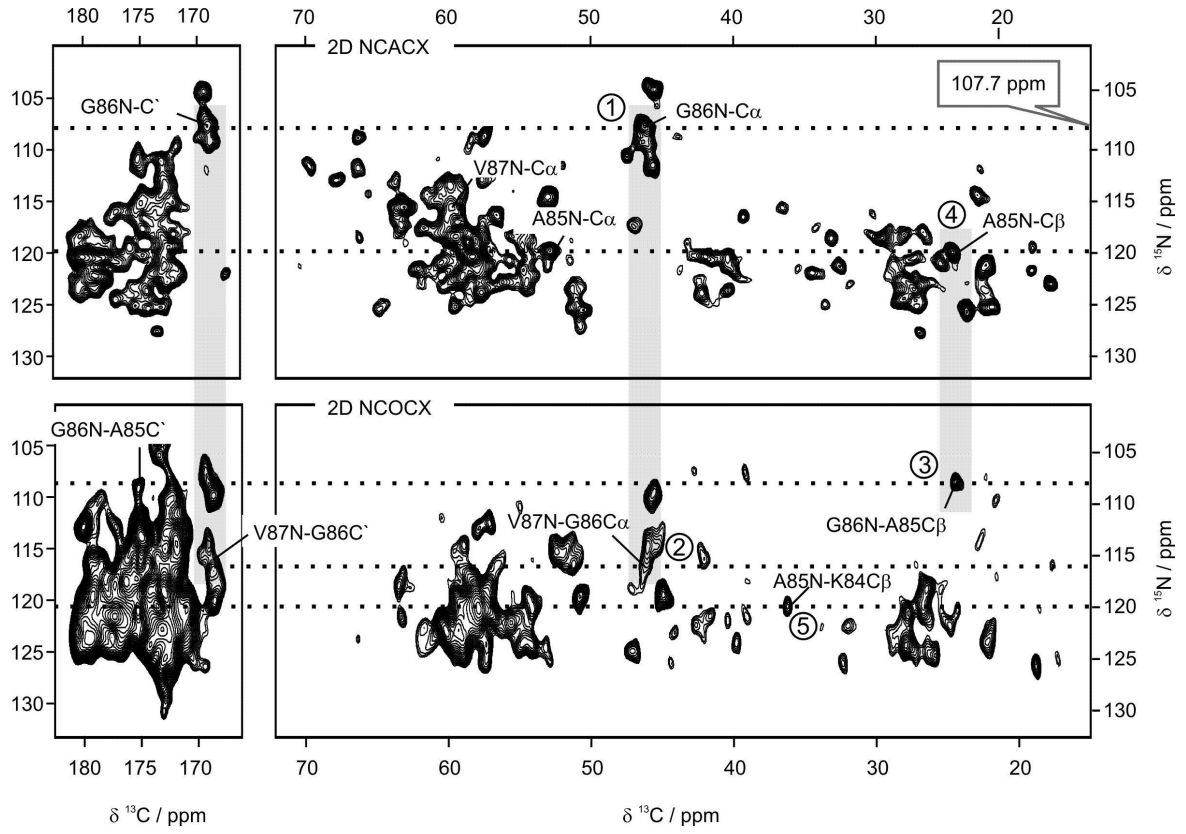


Figure 3.6 – Heteronuclear ^{15}N, ^{13}C 2D NCACX (top) and NCOCX (bottom). These experiments differ from NCA/NCO by addition of an extra ^{13}C-homonuclear mixing step. Spin systems are better identified on the basis of their backbone/side-chain correlations. Sequential links between amino-acid residues K84-V87 are shown.

Sequential correlation of a stretch of four residues (K84-V87) in YadA-M is illustrated in Figure 3.6. It starts with the identification of G86 spin system at ^{15}N=107.7 ppm, in the NCACX spectrum (labelled with 1). The backbone Cα and C` peaks for G86 are then searched in NCOCX spectrum and are observed at $^{15}$N=115.3 ppm (labelled with 2). The backbone Cα and C` cross peaks for G86, in both spectra, are highlighted with vertical grey bars. The ^{15}N chemical shift for residue following G86 in the primary sequence, i.e., V87, therefore, is ^{15}N=115.3 ppm and its cross peaks in NCACX should be observed at this nitrogen chemical shift. The Cα peak for V87 appears at ^{15}N=115.3 ppm whereas its side-

chain cross peaks are missing and show up only in extended mixing time spectra. Going backwards from G86, as its nitrogen chemical shift is at 107.7 ppm, the preceding residue (A85) is searched at 107.7 ppm in NCOCX spectrum. Indeed the Cβ and C` peaks for an alanine spin system are present at this chemical shift (Cα is in the noise region but was observed in other spectra), and therefore assigned to A85 (labelled with 3). To determine the ^{15}N chemical shift for A85, its cross peaks are identified in NCACX at ^{15}N=119.7 ppm (labelled with 4). For clarity, the Cβ peaks for A85 in both spectra are highlighted with grey bar. Using the same strategy, at ^{15}N=119.7 ppm, K84 is assigned (labelled with 5) and the procedure of recognizing and sequentially assigning spin systems could in principle be continued. However, as can be judged from Figure 3.6, overlap of peaks in 2D spectra allows only identification of a handful of spin systems. Especially in the alpha (50-63 ppm) and carbonyl (170-180 ppm) regions of both spectra, peaks are highly crowded and cannot be assigned without ambiguity. Therefore, it is obvious that these spectra should be recorded as 3D data to lift overlap between side-chain patterns by the additional ^{15}N frequency. In Figure 3.7, several 2D strips extracted from a 3D NCACX (red contours) and a 3D NCOCX (cyan contours) are shown. Both 3D spectra were recorded with 35 ms DARR mixing following the adiabatic ^{15}N-^{13}C transfer to exchange polarization between backbone and side chains. The NCOCX correlation is more dispersed than the NCACX correlation, since the amide ^{15}N and the ^{13}C' signals do not cluster in fingerprint regions as is the case for the NCA correlations. The strips illustrate the assignment procedure. In strip **A** from the NCACX experiment, a threonine spin system is identified based on its fingerprint pattern; the threonine can either be T30 or T57, as we have only two threonines in the sequence. The 3D NCOCX spectrum is scanned for the same pattern of side-chain signals and at ^{15}N=122.9 ppm, a similar cross-peak pattern is found (shown in strip **B**). In the next step, a strip is extracted at ^{15}N = 122.9 ppm from the 3D NCACX experiment (strip **C**). In this strip, a cross-peak pattern is found that is attributed to an arginine rather than to an alanine, from which the sequential connection can be establish between T30 and R31 rather than between T57 and A58. To confirm this assignment, the next residue should be a valine (V32), of which the nitrogen frequency can be found in the NCOCX experiment at 121.0 ppm (strip **D**). Indeed, the cross-peak pattern observed at a ^{15}N frequency of 121.0 ppm in the NCACX spectrum is highly characteristic for a valine (cf. strip **E**) and thus confirms the sequential assignments made. By searching for the V32 side-chain signals in the NCOCX, the ^{15}N shift of D33 is found at 121.6 ppm (cf. strip **F**).

This procedure is continued and the stretch of amino acids from T30 to K34 is assigned as shown in strips **A**-**J** in Figure 3.7.

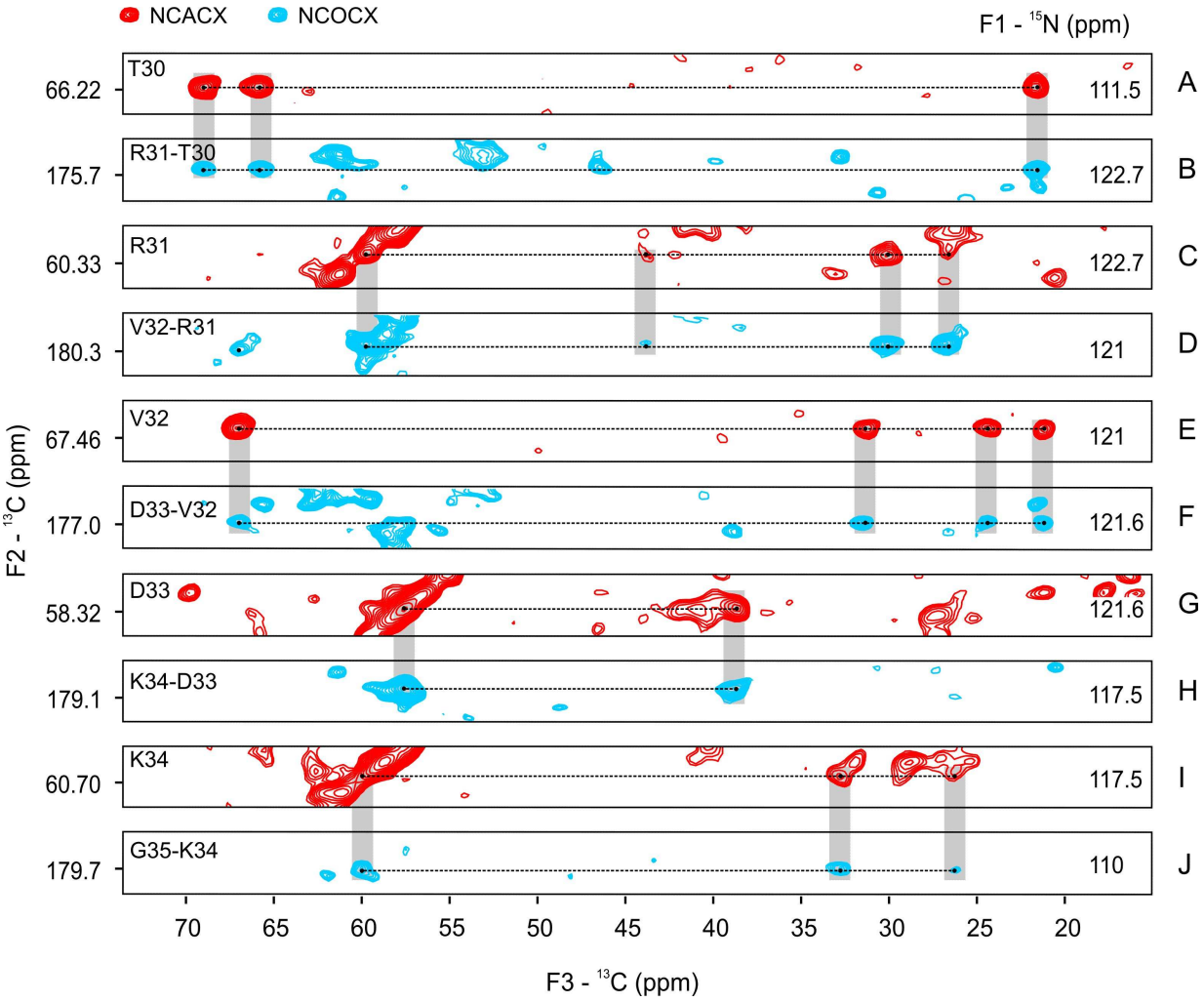


Figure 3.7 – Strips extracted from 3D NCACX and 3D NCOCX experiments, illustrating the sequential assignment strategy. The 3D experiments were recorded at a field of 600 MHz, a MAS frequency of 10 kHz and at a temperature of 275 K. Initial ^{15}N polarization was created with a standard ^{1}H-^{15}N CP (2 ms, 75 – 100% ramp on ^{1}H), followed by selective transfer from ^{15}N to ^{13}C′ or ^{13}Cα using an adiabatic CP. A PDSD (proton-driven spin diffusion) sequence of 35 ms was applied to exchange magnetization between the ^{13}C spins. Introduction of an additional ^{13}C dimension reduces the peak overlap, making identification and assignment of signals much easier. The strategy to sequentially link different residues is based upon identifying the same spin systems in both NCACX and NCOCX spectra, as highlighted by grey bars. 2D strips from NCACX and NCOCX are colour-coded as red and cyan, respectively. F1 is the ^{15}N dimension and values on right side of each strip are those at which each strip is extracted. F2 is the indirect ^{13}C dimension representing Cα and C′ frequencies for NCACX and NCOCX (shown on left side of each strip), respectively. F3 represents the direct ^{13}C dimension where backbone/side-chain signals of individual spin systems are observed.

3.4.2 Selective transfers using DREAM mixing

In cases of repetitive spin pairs or when cross peaks appear in the more congested regions of the spectra, multiple assignment possibilities may exist. In order to be able to assign residues unambiguously, additional spectra were recorded. For example, a helpful addition to the data set was a 3D NCACB spectrum (Figure 3.8, bottom), where selective exchange between ^{13}Cα and ^{13}Cβ was achieved with a DREAM transfer scheme (Verel, Baldus et al. 1998; Pauli, Baldus et al. 2001). In this spectrum, Cα-Cβ correlations of Ala, Val, Lys, Leu and Asp could be unambiguously resolved without overlap with correlations from other residues. This spectrum not only helped to assign new peaks, but also was used to cross-check already assigned spin systems. A comparison of two strips extracted from a 3D NCACX and 3D NCACB at ^{15}N= 121.3 ppm is shown in Figure 3.8.

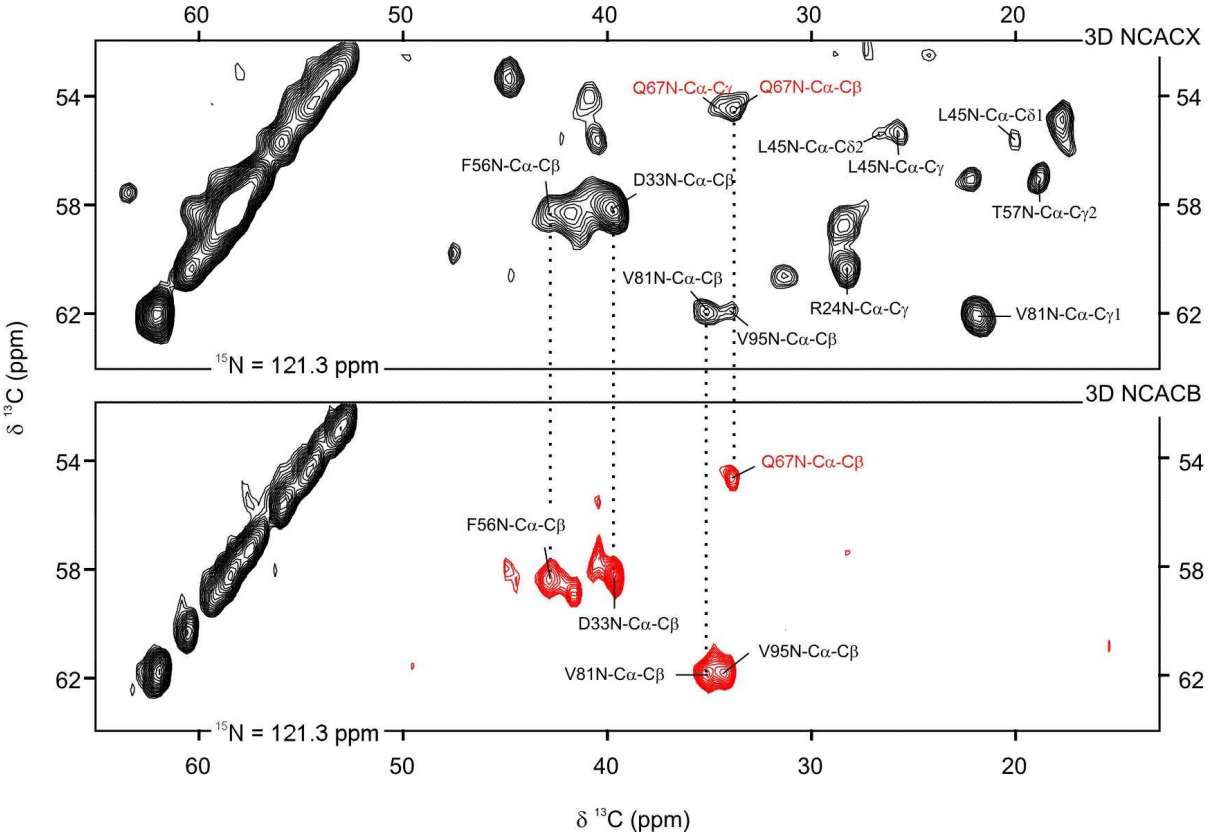


Figure 3.8 – Comparison of 3D NCACX and 3D NCACB. 2D ^{13}C-^{13}C strips extracted from 3D NCACX (600 MHz, ω / 2π=10 kHz, 35 ms PDSD mixing) and 3D NCACB (900 MHz, ω / 2π=12 kHz, 3 ms DREAM mixing) recorded on YadA-M. Single-bond double-quantum transfer in NCACB results in negative cross-peak intensities which are coloured red. Two-bond transfers are basically absent. Distinguishing between one-bond and two-bond transfers can lift specific ambiguity. In this case, it corroborates the assignment of Q67Cα-Cβ and Q67Cα-Cγ (assignments highlighted in red).

Along the same lines, two 2D ^{13}C correlation spectra were recorded, employing a DREAM transfer scheme optimized for the Cα-Cβ and Cβ-Cγ spectral regions (Figure 3.9). The advantage of the DREAM experiment is that it can be set up so that it predominantly provides transfers between spins that are directly connected. It was experienced that transfer events over two bonds occur less frequently and are mainly observed for spin pairs where at least one of the two carbon atoms is only weakly coupled to ^{1}H. Examples are two-bond couplings involving carbons in methyl groups (like leucine $C\delta_1$-$C\delta_2$, leucine Cβ-Cδ, valine $C\gamma_1$-$C\gamma_2$ and isoleucine Cβ-Cδ) or carbons in flexible side chains (like lysine Cε-Cγ and arginine Cδ-Cβ). Such occasional two-bond transfers can be easily distinguished from one-bond transfers since both have opposite sign in cross-peak intensity.

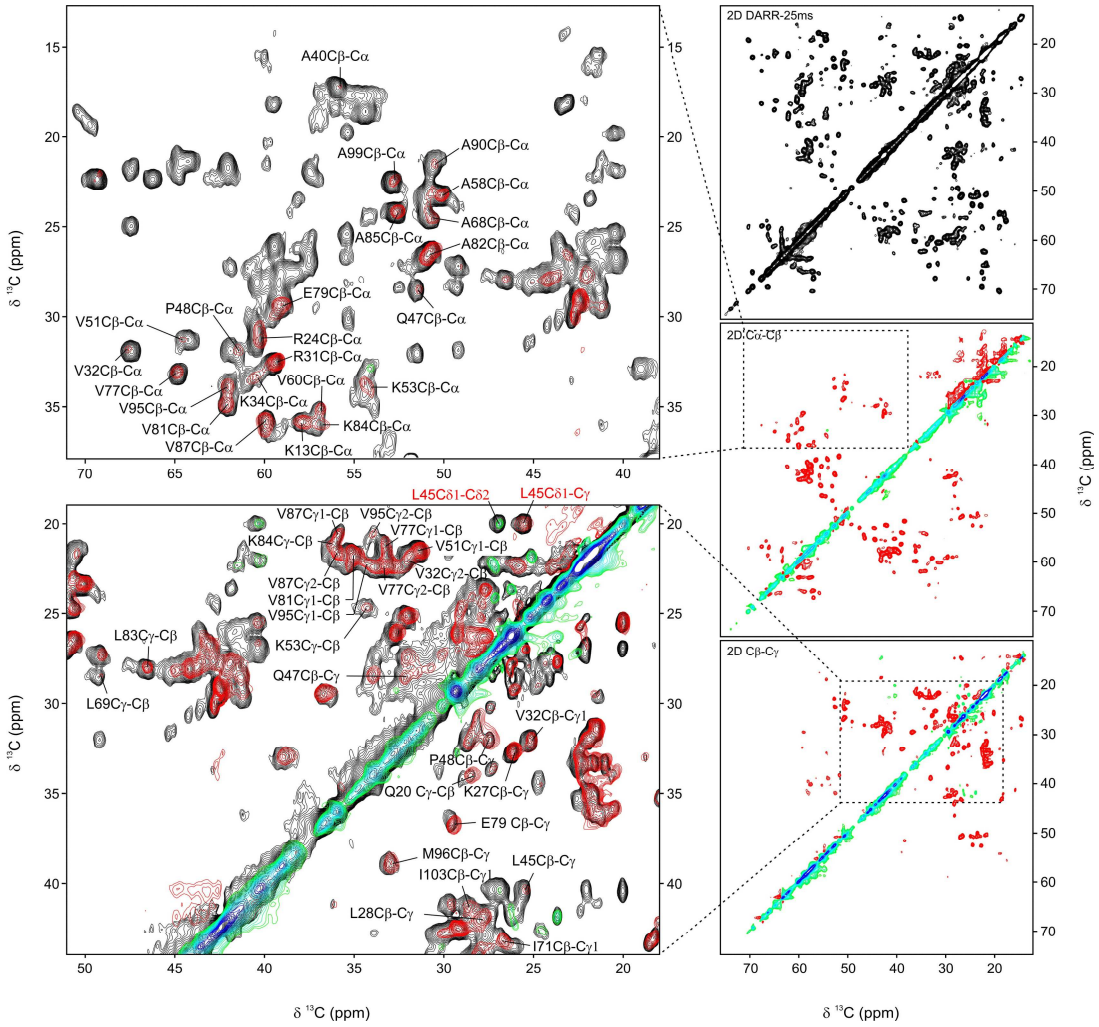


Figure 3.9 – Comparison of 2D DARR and DREAM spectra. The DARR spectrum was recorded with a mixing time of 25 ms (right, top panel). Two DREAM spectra were recorded, optimized for the Cα-Cβ region (right, middle panel) and the Cβ-Cγ region (right, bottom panel). On the left-hand side, the DREAM spectra are superimposed on the DARR spectrum (top: Cα-Cβ; bottom: Cβ-Cγ). In the DREAM spectra, one-bond transfers give negative cross peaks (red contour lines), zero-bond (diagonal) and two-bond transfers are positive (blue-green contour lines).

Note that in highly congested regions of the spectrum a partial cancellation of cross peaks may occur. In the 2D Cα-Cβ DREAM spectrum, correlations between spins that are more than one bond away are virtually suppressed (Figure 3.9, top left). This is instrumental to distinguish Cα-Cβ cross peaks of valines, lysines, arginines and glutamic acids from two-bond Cα-Cγ correlations from other residues that appear in the same spectral region (e.g., from leucines, lysines and arginines). By optimizing the DREAM transfer around a more upfield frequency, one-bond transfers, mainly between Cβ and Cγ carbons, can be distinguished from two-bond transfers involving methyl groups (Figure 3.9, bottom left). This is particularly useful for distinguishing Cγ and Cδ resonances of leucines. This is exemplified by cross peaks for L45 (assignments are highlighted in red) which have opposite signs in the spectrum (Figure 3.9, bottom). In leucines, the Cγ is directly bonded to Cδ1 and therefore has negative sign (red). On the other hand, Cδ1 and Cδ2 are separated by two bonds, thus show a positive peak (green). This differentiation was critically useful in making out several spin systems.

3.4.3 Consistency check of the sequence-specific assignment

The sequence-specific assignment relies on sequential contacts between the neighbouring residues. These contacts may either be of the type backbone/backbone, backbone/side-chain or side-chain/side-chain, depending upon the specific mixing time and mixing sequence used for magnetization transfer. In an NCACX transfer (35 ms PDSD), the most probable sequential contact to be observed is between carbonyl carbon atom of the "i-1" residue and the nearby Cα of the residue "*i*". At longer mixing times (100-200 ms), however, one can observe the complete "i-1" spin system at the ^{15}N frequency of residue "*i*". Similarly, in an NCOCX transfer (35 ms PDSD), the Cα of the residue "*i*" close to the carbonyl of the "i-1" residue will most likely provide a sequential contact (cf. Figure 3.5, top). At longer mixing times, observed backbone or side-chain correlation networks with the preceding or following residues will make the assignment unambiguous. Therefore, 3D NCACX and 3D NCOCX with longer mixing times (100 and 200 ms PDSD transfer) were recorded to achieve maximum number of sequential contacts. Controversially, it was found that such spectra helped to remove ambiguity, even though the total amount of cross peaks and hence spectral crowding is higher. The explanation for this is, whereas in NCACX with short CC mixing times magnetization is only exchanged intra-residually ('forward transfer'), with longer

mixing times the $^{13}C\alpha$ also exchanges with the backbone and side-chain spins of the previous residue in the sequence ('backward transfer'); this provides a great means for checking the self-consistency of the sequential assignment without need to revert to the intermediate $^{15}N_i$ and $^{13}C'_{i-1}$ shifts. In complete analogy, NCOCX spectra with longer CC mixing times do not only provide backward transfer between $^{13}C'$ and the side chain of the same residue, but also forward transfer, directly linking $^{13}C'_{i-1}$ with the side chain of residue i. An example of multiple transfer pathways mutually linking the pair S44-L45 is shown in Figure 3.10. In the top part of the Figure, magnetization transfer events are shown schematically. Red arrows represent the transfer of ^{13}C polarization within a spin system, blue arrows the exchange between residues, using the $^{13}C\alpha$ (left) or $^{13}C'$ (right) as starting point. The sequential assignment of the residue pair S44 and L45 is shown, aided by inter-residue correlations.

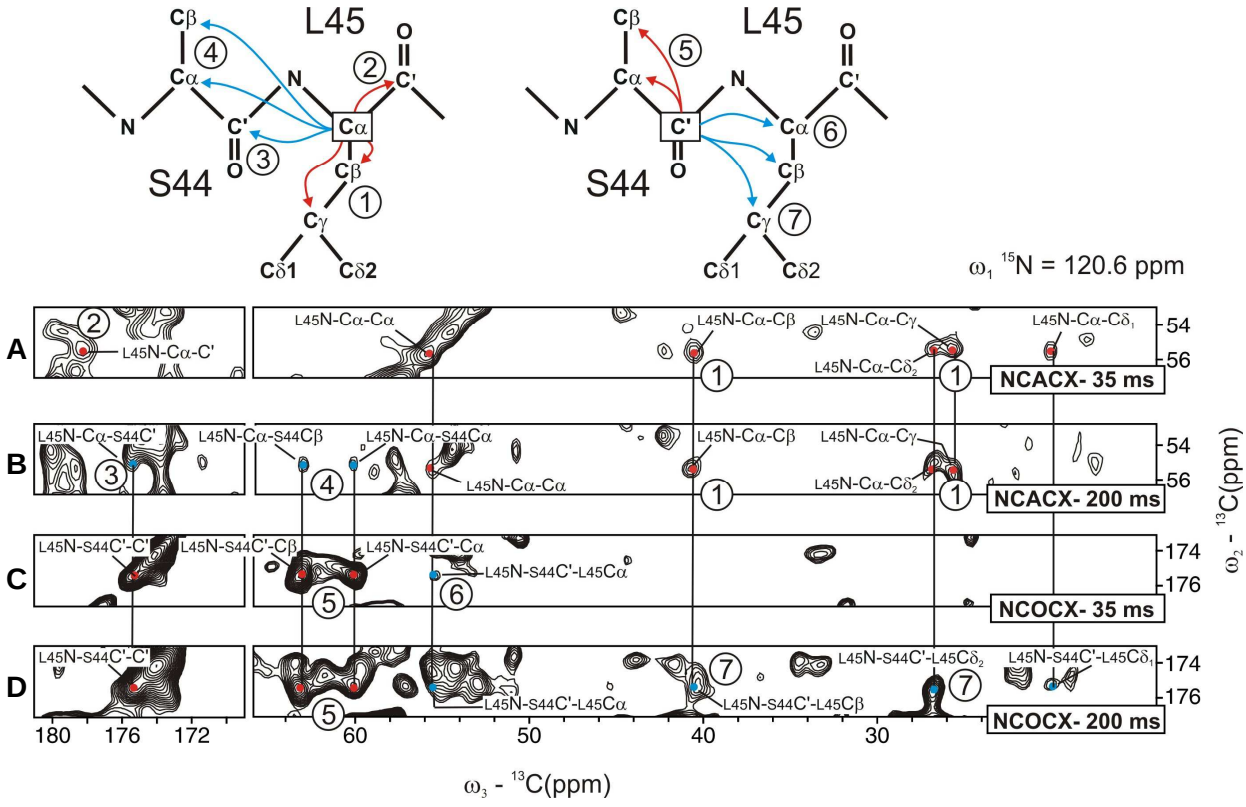


Figure 3.10 – Corroboration of sequential assignment. Schematic representation of magnetization transfer (top); red arrows represent intra-residue transfer of ^{13}C polarization, blue arrows inter-residue ^{13}C exchange, starting from $^{13}C\alpha$ (left) or $^{13}C'$ (right). 2D strips extracted from 3D NCACX (strips **A** and **B**) and 3D NCOCX spectra (strips **C** and **D**, bottom); the 3D spectra were recorded with proton-driven spin diffusion mixing of 35 ms (strips **A** and **C**) and 200 ms (strips **B** and **D**).

In the NCACX experiment recorded with long PDSD mixing time of 200 ms, polarization can be transferred from $^{13}C\alpha$ of residue i to residue i-1. This is shown schematically in the top-left

part of Figure 3.10 where blue arrows indicate the backward transfer from Cα of L45 to C′, Cα and Cβ of S44 (labelled with 3 and 4). The corresponding cross peaks are labelled accordingly in strip **B**. In contrast, in the NCACX experiment recorded with a short mixing time of 35 ms, only intra-residue cross peaks for L45 are observed (strip **A**). Likewise, in the NCOCX experiment with long PDSD mixing, polarization can be exchanged between $^{13}C'_{i-1}$ and residue i (forward transfer, Figure 3.10 top right, blue arrows). This is demonstrated in strip **D** which shows both intra-residue (labelled with 5) and inter-residue correlations (labelled 6 and 7). Interestingly, in the NCOCX obtained with a short mixing of 35 ms, apart from strong intra-residue cross peaks, weak transfer from S44 C′ to L45 Cα can be observed (cf. strip **C**).

Finally and along the same lines, a further powerful check for the consistency of the sequence-specific assignment is achieved by analysis of direct correlations between the ^{13}Cα spins. Also here, this transfer does not involve the intermediate backbone amide and carbonyl ^{13}C′ spins; hence, any potential mistake made during the alignment procedure of strips as described above would lead to inconsistent results.

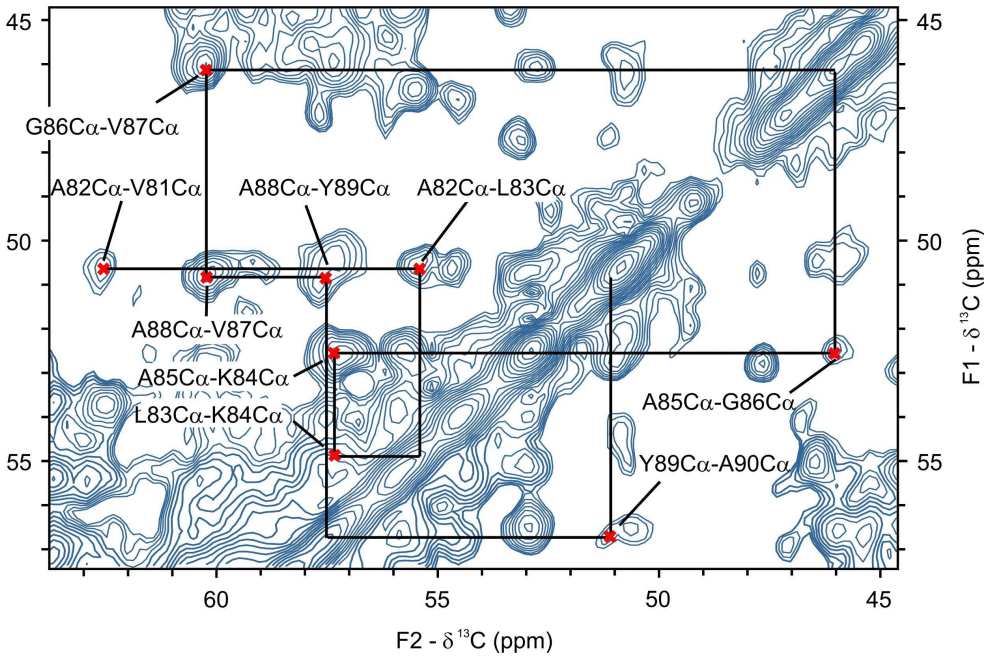


Figure 3.11 – Sequential walk of A82-L83-K84-A85-G86-V87-A88-Y89-A90. Contour plot taken from a 100 ms 2D PDSD, recorded on YadA-M. Presence of Cα-Cα correlation peaks corroborates the sequential assignment. The walk starts at V81 and ends at A90.

Sequential Cα-Cα correlations are readily observed in ^{13}C-^{13}C spin diffusion spectra with extended DARR or PDSD mixing times (typically 100 - 200 ms). As an example, the

sequential walk connecting the A82-A90 subsequence via the Cα's is shown in Figure 3.11. In such a 2D spectrum, Cα peaks of both "i+1" and "i-1" residues are most likely to appear at the Cα chemical shift of residue "i". For instance, at the Cα chemical shift of A82 (F1- 50.11 ppm), appear Cα cross peaks for both V81 (62.08 ppm) and L83 (55.18 ppm). A line can be drawn by connecting the residual and sequential peaks and is called sequential walk (Figure 3.11, sequential walk for residues V81-A90).

Since the relevant information is obtained from a relatively narrow spectral region, these experiments perfectly lend themselves for *J*-decoupling in the indirect dimension (Straus, Bremi et al. 1996). Plots of the Cα-Cα region in a standard and *J*-decoupled DARR spectrum are shown in Figure 3.12. Such *J*-decoupled spectrum is well resolved and is rich in Cα-Cα correlation peaks.

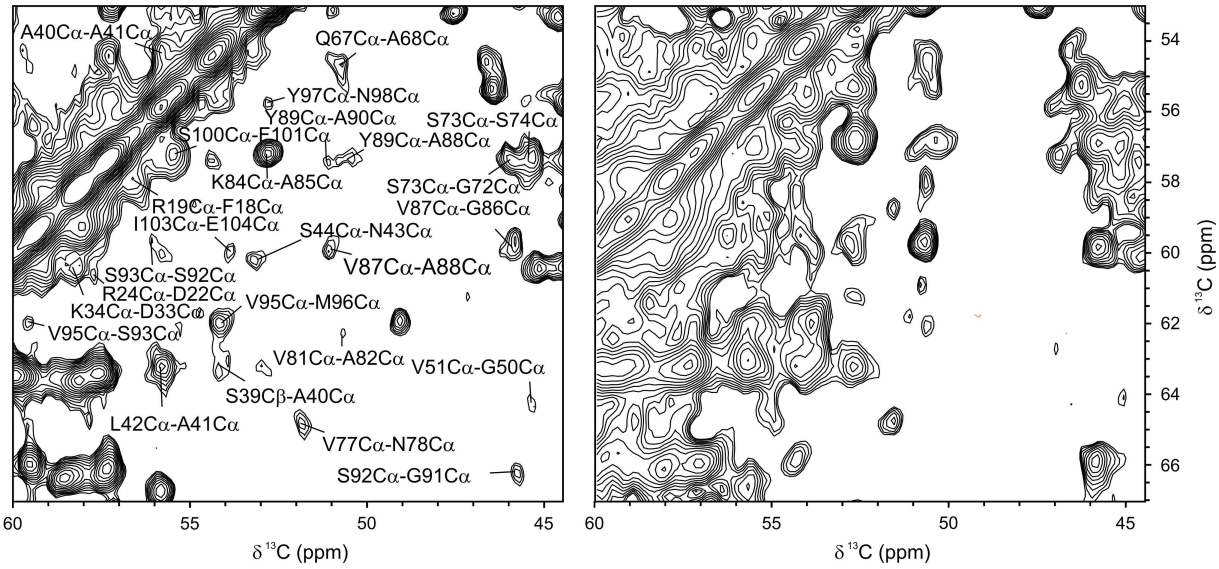


Figure 3.12 – The effect of *J*-decoupling on the resolution. The plot on the right shows a 2D proton-driven spin-diffusion experiment with 100 ms mixing time, the plot on the left shows the same spectral region taken from a *J*-decoupled DARR spectrum with a mixing time of 70 ms. *J*-decoupling was achieved with a combination of hard and band-selective (Gaussian) 180° pulses in the Cα region.

3.4.4 Assignments from 15N – 13C TEDOR

Prolines are difficult to assign on the basis of 2D or 3D NCACX experiments because the lacking amide proton renders the first ^{1}H-^{15}N CP very inefficient. In the 2D NCA experiment shown above, only a relatively weak cross peak with $^{13}C\alpha$ is observable for P48 (cf. Figure 3.5, right); however, in the 2D NCO experiment (Figure 3.5, left), and in the 2D or 3D NCACX and NCOCX experiments, no correlations involving P48 were observed. Prolines

provide, however, a highly characteristic correlation pattern in 2D ^{15}N-^{13}C TEDOR experiments (Jaroniec, Tounge et al. 2001; Jaroniec, Filip et al. 2002), consisting of pair-wise cross peaks involving the Cα and Cδ in the α-region with the downfield-shifted backbone amide. As shown in Figure 3.13, a proline fingerprint pattern appears at a ^{15}N shift of 132.2 ppm, which is assigned to P48 (connected with a dashed line). Note that the other proline in the sequence (P9) resides in the mobile strep-tag and is indeed not observed in Figure 3.13. It is remarkable that out of 11 observable glycines, 9 can readily be identified in this spectrum and sequentially connected with the NCO region of the TEDOR. Alanines in the β-sheet region are well dispersed and their ^{15}N and Cα chemical shifts are easily assignable. Glycines, alanines and P48 which were assigned using this spectrum, are highlighted with grey boxes.

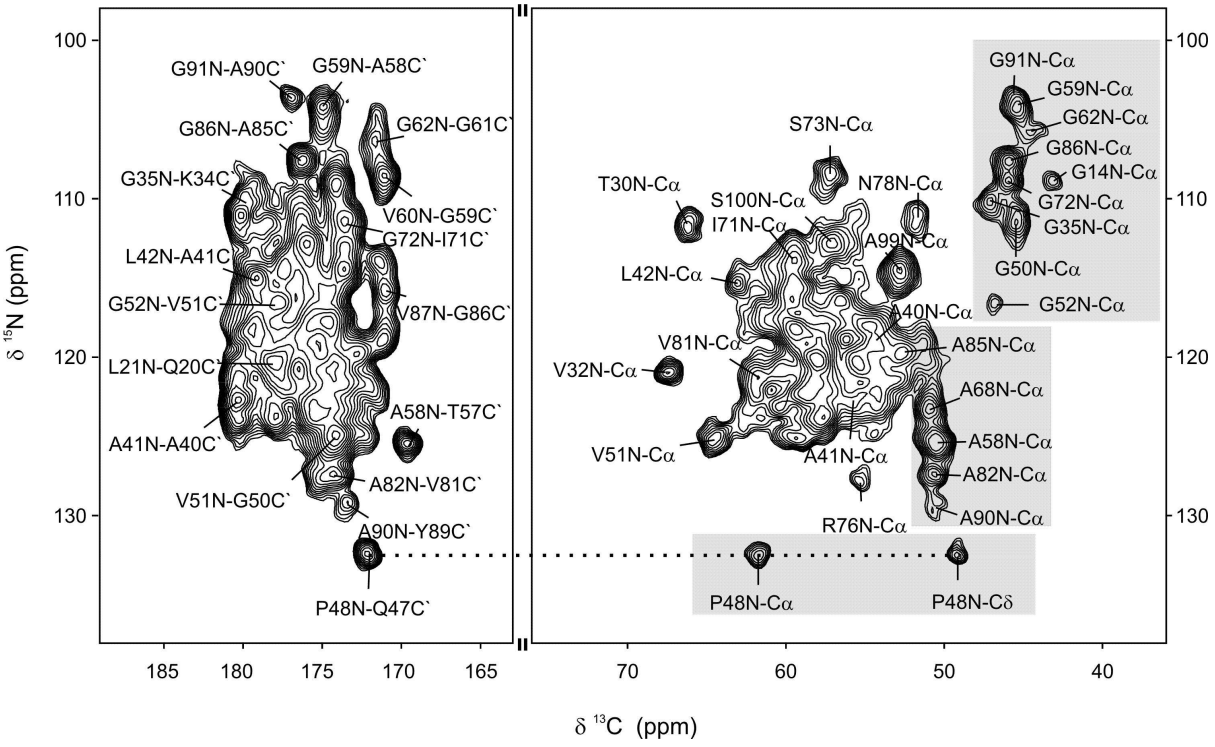


Figure 3.13 – 2D ^{15}N-^{13}C TEDOR. Contour plot of the NCO-region (left panel) and NCA region (right panel) of a TEDOR (transferred echo double resonance) (Jaroniec, Tounge et al. 2001; Jaroniec, Filip et al. 2002) experiment, recorded at a field of 9.4 T, 8 kHz MAS and at a temperature of 275 K. The experiment contains two REDOR (rotational echo double resonance) (Gullion and Schaefer 1989) mixing periods of 1 ms each, to transfer magnetization back-and-forth between carbon and nitrogen. The spectrum was recorded to assign prolines in YadA-M. Out of two prolines, only one is observable, while the other (P9) is in the N-terminal strep-tag and does not show up. The grey boxes highlight the fingerprint regions for ^{15}N-^{13}C correlations involving glycines, alanines and proline.

3.4.5 Methyl-filtered post-C7

The last experiment in the data set was a methyl-filtered post-C7 experiment. Methyl filtering can be achieved spectroscopically by a straightforward combination of cross-polarization with phase inversion (CPPI) and a dipolar dephasing step (Jehle, Hiller et al. 2006).

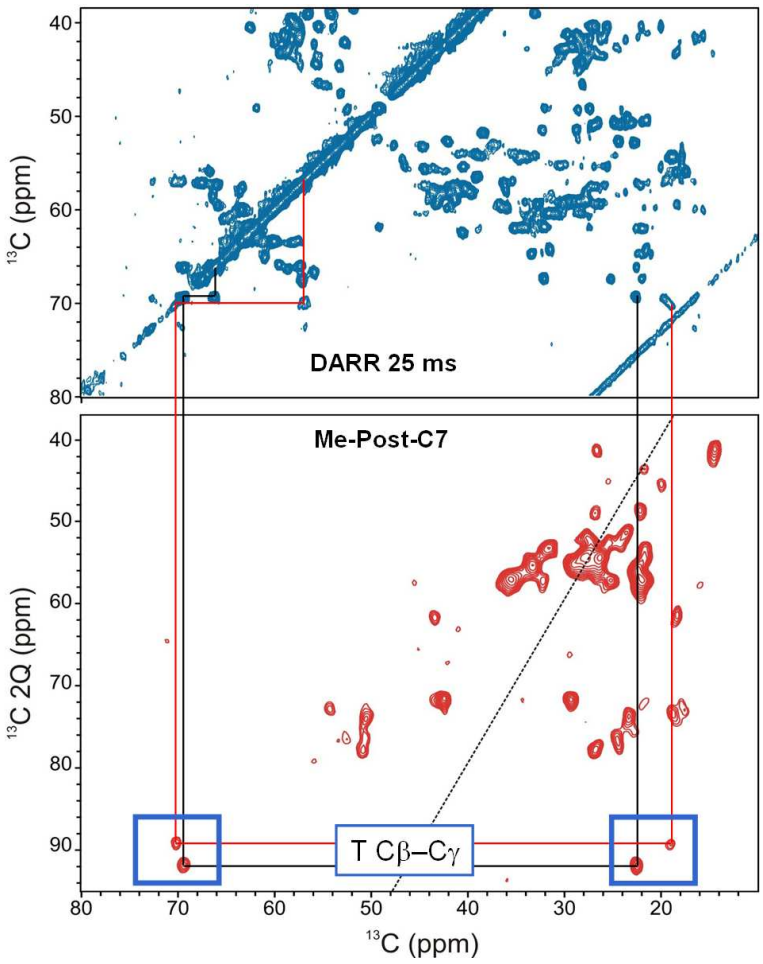


Figure 3.14 – Methyl-filtered Post-C7. Assignment of threonines was aided by a combined evaluation of a standard 2D ^{13}C-^{13}C DARR experiment (top) and a methyl-filtered Post-C7 experiment (bottom).

The filter selects carbon atoms that are relatively weakly coupled to ^{1}H, primarily targeting at the methyl-carbons, but also non-protonated carbonyls and carbons in flexible side chains. In combination with the Post-C7 (Hohwy, Jakobsen et al. 1998), the spectrum is highly instrumental for the assignment and/or corroboration of the assignment of methyl-containing amino acids (Ala, Ile, Leu, Val, and Thr). More specifically, it helped to unambiguously identify T57, which has an unusual upfield-shifted ^{13}Cα; as a result, the Cα-

Cβ cross peak of T57 is found in the serine Cα-Cβ fingerprint region and as such was not identified straightforwardly (cf. Figures 3.14). Additionally, in the carbonyl part of Post-C7, Cβ-Cγ correlations for Asn, Asp and Cγ-Cδ correlations for Gln, Glu can be readily identified.

3.5 Summary

In summary, from evaluation of several homonuclear and heteronuclear solid-state MAS NMR spectra (cf. Table 3.1), all residues in the subsequence 14-104 of YadA-M could be unambiguously assigned (BMRB entry 18108) (cf. Appendix II). Despite the highly repetitive sequence of YdaA-M, several unique spin systems and a fair number of pair-wise unique spin systems were found that served as starting points of the assignment (cf. Figure 3.4). Glycines and alanines, even though relatively abundant in YadA-M, proved very useful for sequential assignment. Their backbone $^{13}C\alpha$ and ^{15}N are readily identifiable from 2D NCA type spectra (cf. Figure 3.5) and can be used as sequential linkers in a parallel evaluation of 3D NCACX and NCOCX spectra (cf. Figure 3.6, 3.7). Serine residues, on the other hand, turned out to be less helpful; although serines are generally easily identifiable by virtue of their downfield Cβ chemical shift, six out of the 12 serines in YadA-M form SS pairs, which made it difficult to link them sequentially. Selective transfers using DREAM mixing in 3D NCACB, 2D CACB and 2D CBCG helped identifying several spin systems by reducing the information content of spectra (cf. Figure 3.8, 3.9). It has been shown by subtilisin treatment that the strep-tag is extremely mobile and renders fair degree of flexibility to the N-terminal residues (Wollmann, Zeth et al. 2006). As a consequence, most of the residues in the strep-tag were not observed in any of the spectra of the data set, while some others gave very weak and incomplete cross-peak patterns. For instance, as noted above, only one strong proline signal set was detected, whereas the proline that resides in the strep-tag was not observed (cf. Figure 3.13). The assigned residues in the subsequence 14-104 can completely account for the cross peaks detected in the various spectra of the data set; the sequential assignment is highly self-consistent (e.g., cf. Figure 3.10) and no strong cross peaks were 'left over' as unassigned, which strongly supports the initial presumption that the observable signals predominantly arise from the rigid part of the protein. Also for the rigid part, for several regions the cross peaks appeared systematically weaker, suggesting a higher degree of (local) mobility. Among these are the subsequence K53-V54-N55, the region around A37 and the C-

terminus (W105). Although a tentative assignment for W105 was possible, the sequential link was poor and W105 was not considered as part of the assigned residues. A double-quantum methyl filtered Post-C7 was used for assignment and/or corroboration of Ala, Ile, Leu, Val, and Thr (cf. Figure 3.14). In the carbonyl part of the Post-C7, presence of Cβ-Cγ correlations for Asn, Asp and Cγ-Cδ correlations for Gln, Glu make this spectrum extremely useful. Aromatic amino-acid residues show relatively weak sequential correlations with neighbouring spin systems; their aromatic ring can act as a magnetization sink, resulting in weaker Cα-Cβ intra-residue correlations. Phenylalanines were identified on the basis of their aromatic Cγ cross peaks (140-142 ppm). The Cζ of both tyrosines and arginines appear at around 159 ppm and helped confirming assignment of these residues. Tyrosines were distinguished from arginines by virtue of their aromatic Cε1, Cε2 peaks (116-117 ppm). In a nutshell, the virtually complete NMR assignment of a membrane protein by solid-state MAS NMR emphasizes the strength of this technique and offers new perspectives in structural biology for research of this important class of proteins.

Table 3.1 – List of solid-state MAS NMR spectra used to achieve residue-specific and sequence-specific assignment of YadA-M

Experiment type	Mixing time/ms	Nuclei involved	Proton frequency/MHz	MAS/kHz	No. of assignments	Information obtained
2D DARR	15	^{13}C-^{13}C	900	12	545	Identification & assignment of spin systems
2D DARR	15	^{13}C-^{13}C	900	12		Quality control
2D DARR	25	^{13}C-^{13}C	900	13		Identification & assignment of spin systems
2D DARR	50	^{13}C-^{13}C	900	11	76	Sequential correlations
2D DARR	100	^{13}C-^{13}C	900	11	120	Sequential correlations
J-decoupled DARR	15	^{13}C-^{13}C	900	13		Quality Control
J-decoupled DARR	40	^{13}C-^{13}C	900	13	39	Cα-Cα sequential correlations
J-decoupled DARR	70	^{13}C-^{13}C	900	13	273	Cα-Cα sequential correlations
2D PDSD	15	^{13}C-^{13}C	900	13	265	Sequential assignment
2D PDSD	100	^{13}C-^{13}C	900	13	536	Sequential assignment
2D DREAM	1.5	^{13}C-^{13}C	900	13	142	Residue-specific assignment
2D DREAM-AB	1.5	^{13}C-^{13}C	900	13	74	Gives predominantly Cα-Cβ peaks
2D DREAM-BG	1.5	^{13}C-^{13}C	900	13	96	Gives predominantly Cβ-Cγ peaks
2D Me-only	70	^{13}C-^{13}C	900	8	55	Identification of A, V, L, I, M, T, K
Post-C7 Me	70	^{13}C-^{13}C	900	8	44	Identification of A, V, L, I, T
2D TEDOR	1	^{15}N-^{13}C	400	8	49	Identification of P, G, A
2D NCA		^{15}N-^{13}C	400	8	49	Correlates backbone ^{15}N with Cα
2D NCO		^{15}N-^{13}C	400	8	35	Correlates backbone ^{15}N with C`
2D NCACX	25	^{15}N-^{13}C	400	8	44	Correlates backbone ^{15}N with carbon nuclei
2D NCOCX	25	^{15}N-^{13}C	400	8	25	Sequential assignment
3D NCOCX	35	^{15}N-^{13}C-^{13}C	600	10	275	Sequential assignment
3D NCOCX	200	^{15}N-^{13}C-^{13}C	600	10	410	Sequential assignment
3D NCACB	3	^{15}N-^{13}C-^{13}C	900	13	102	Identification of ^{15}N and Cβ
3D NCACX	35	^{15}N-^{13}C-^{13}C	600	10	281	Residue-specific assignment
3D NCACX	100	^{15}N-^{13}C-^{13}C	600	10	66	Sequential Assignment
3D NCACX	200	^{15}N-^{13}C-^{13}C	600	10	254	Corroborates sequential assignment
3D NCACX	500	^{15}N-^{13}C-^{13}C	600	10	44	Sequential and long-range assignment

4 The Structure of YadA-M

4.1 Introduction

In chapter 3 the strategies to achieve unambiguous chemical shift assignment of YadA-M by solid-state MAS NMR are described. This chapter presents in detail the acquisition of distance restraints and determination of secondary and tertiary structure of YadA-M by solid-state MAS NMR.

The assigned ^{13}C and ^{15}N chemical shifts contain dihedral angle information (phi and psi torsion angles), which is routinely used to predict secondary structure elements using programmes like TALOS+ (Shen, Delaglio et al. 2009). Prior knowledge about the secondary structure of a protein is helpful in guiding the evaluation of medium and long-range distance restraints. For this purpose, multidimensional experiments with different pulse sequences have been designed and successfully employed in solid-state MAS NMR. The quality of the calculated structure depends upon the accuracy and number of the distance restraints. Assignment of distance restraints needs special care, as false assignments may pose complications in the structure calculation. However, structure calculation programmes are able to handle limited number of consistently violating restraints that can be reassigned or used as ambiguous restraints in the subsequent iterations. A schematic representation of different steps involved in a protein's structure calculation by solid-state MAS NMR is shown in Figure 4.1.

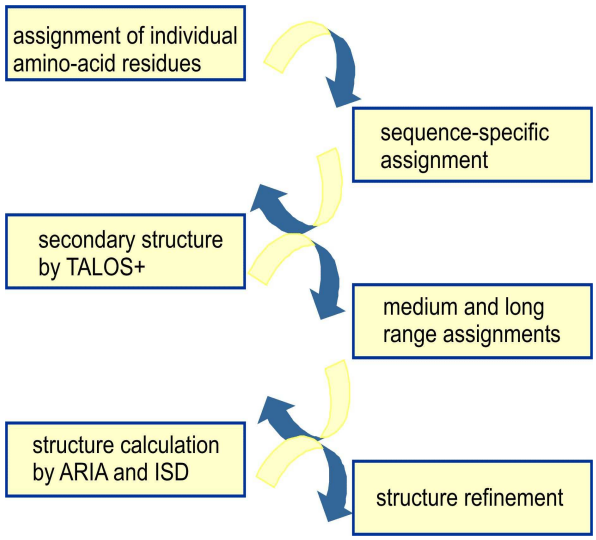


Figure 4.1 – Schematic representation of various steps involved in determining a protein structure by solid-state MAS NMR. Chemical shifts for backbone ^{13}C and ^{15}N are obtained by unambiguous sequential assignment. TALOS+ uses these chemical shifts to predict the dihedral angles. Information about the secondary structure assists evaluation of the data and helps generating structural restraints from extended mixing time spectra. ISD and ARIA are able to use unambiguous and ambiguous restraints along with predicted dihedral angles to calculate the tertiary structure. Structure calculation proceeds through various steps whereby every step refines the structure.

4.2 Secondary Structure of YadA-M

Several secondary structure prediction programmes are online available which, on basis of the primary sequence of a protein, predict its secondary structure. PSIPRED, a highly accurate method for secondary structure prediction (McGuffin, Bryson et al. 2000) was used to predict the secondary structure of YadA-M. The prediction results are shown in Figure 4.1A, in which secondary structure elements are highlighted in different colours. According to the PSIPRED prediction, YadA-M comprises of a single α-helix and four β-strands; these are connected with each other by short loops. An experimentally based secondary structure prediction can be done using the program TALOS+. In addition to the sequence of a protein, TALOS+ uses the backbone ^{13}C, ^{15}N chemical shifts (Shen, Delaglio et al. 2009). It compares the secondary chemical shifts and sequence of the query protein with those of proteins with known structure in the Protein Data Bank (PDB) database. Chemical shifts are influenced by the electronic distribution around nuclei and are directly correlated with the dihedral angles (Spera and Bax 1991). Therefore, ^{13}C (particularly Cα and Cβ) and ^{15}N chemical shifts are good monitors of secondary structure of a protein. The torsion angles and secondary structure of YadA-M as determined by TALOS+ are shown in Figure 4.2B. Based on the chemical shifts, an N-terminal α-helix (16-35, 40-42) and four β-strands, spanning from residues 54-63 (strand 1), 67-76 (strand 2), 81-90 (strand 3), and 95-104 (strand 4) are predicted. The primary sequence of YadA-M is shown in Figure 4.3C and is colour-coded according to the predicted secondary structure elements. The helix (red) is connected to the β-strands (grey) by a loop (blue). The four β-strands are linked with each other by relatively short loops (blue). First 15 N-terminal residues belong to the strep-tag (marine blue). TALOS+ could not predict the region "^{37}ASSA^{40}" (colour-coded green) with high confidence. The side-chain peaks for A37 remain unassigned; this residue was recognized by weak sequential contacts in spectra recorded with extended mixing times. At this stage, the position of strands relative to each other and to the helix is not known. Experimental evidence in the form of distance restraints is required to determine the exact position of the helix relative to the β-sheet. The knowledge about register and directionality (whether parallel or antiparallel) of β-strands requires careful assignment of the solid-state MAS NMR spectral restraints.

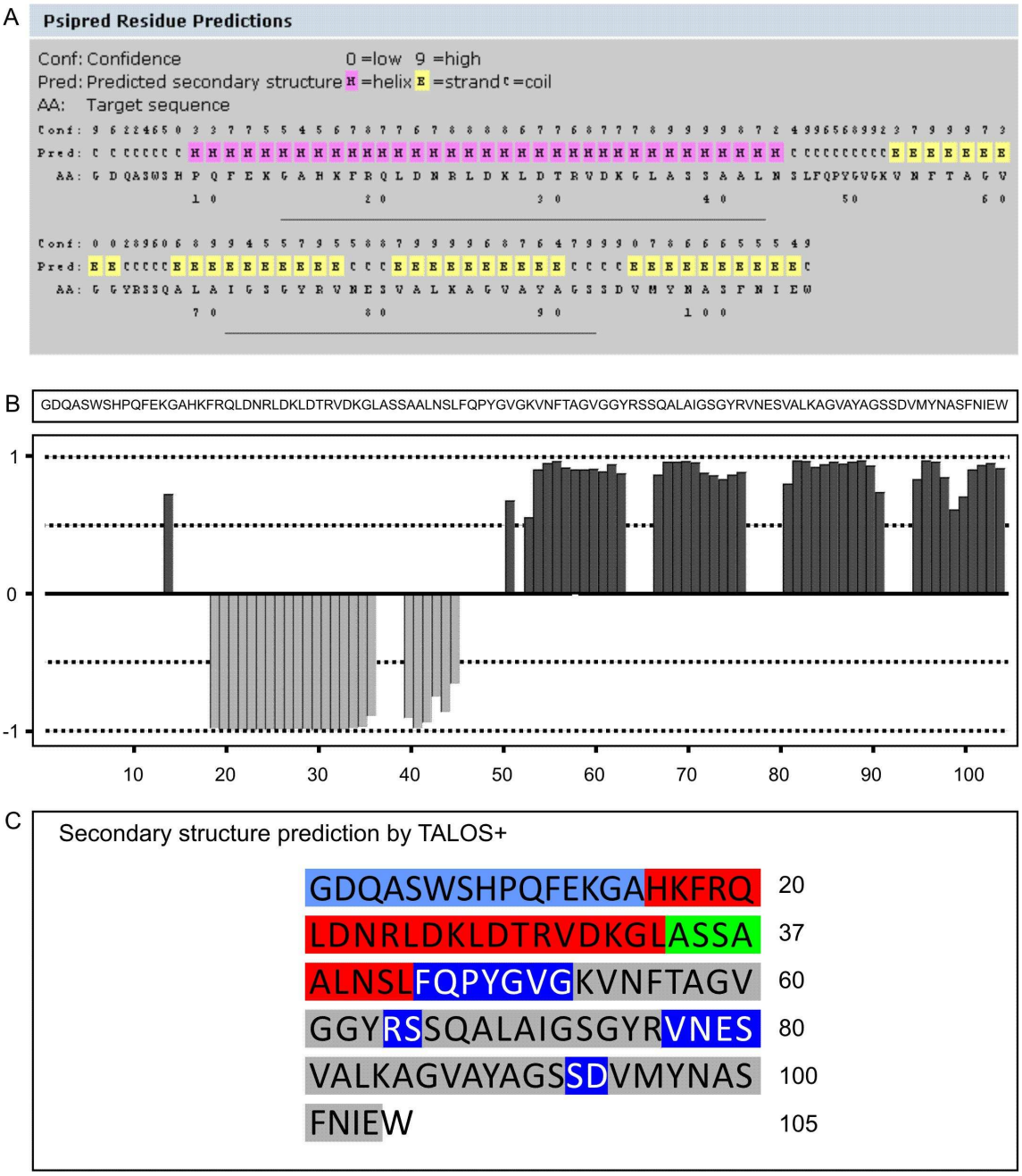


Figure 4.2 – (A) Secondary structure prediction of the YadA-M by PSIPRED. The results from PSIPRED protein structure prediction server for YadA-M are shown. `AA´ represents the amino-acid sequence of the YadA-M. Different secondary structures are labelled with C (coil), H (helix) and E (strand) and are coloured differently. The confidence of prediction is also shown for each individual residue with 0 for lowest and 9 for highest degree of confidence. **(B) Secondary structure prediction by TALOS+**. The bars with positive values indicate the β-strands while those with negative values show the α-helix. It is notable that instead of one single helix, as predicted by the PSIPRED, TALOS+ indicates a broken helix in YadA-M. The sequence and numbering of YadA-M are shown on top and bottom of the panel B, respectively. Values on Y-axis show confidence of the prediction. **(C) Sequence-based representation of the predicted secondary structure of YadA-M.** The primary sequence is shown, colour-coded according to secondary structure elements. First 15 residues are part of the strep-tag (marine blue) followed by an α-helix (red). Four β-strands (grey) are connected with each other by loops (blue). The residues A37-A40 were predicted as random coil by TALOS+ and are colour-coded as green.

4.3 Tertiary Structure

Side-chain/side-chain and side-chain/backbone medium and long-range contacts are required to determine the tertiary structure of YadA-M. For this purpose, a full suite of homo- and heteronuclear extended mixing time spectra (cf. Table 4.1) was recorded. Two-dimensional (2D) ^{13}C DARR spectra at extended mixing times provided a wealth of structural information. The number of spectral cross peaks and hence the information content increases with increasing mixing time (cf. Figure 4.3). In a 25 ms spectrum (coloured green in Figure 4.3), only intra-residue cross peaks were observed. Sequential backbone/backbone (mainly Cα-Cα) contacts were observed in a 100 ms DARR spectrum (blue). These sequential contacts are mandatory to confirm sequential assignment. At still higher mixing times, i.e., 200 ms (not shown), 300 ms (black) and 500 ms (red) DARR spectra, plentiful of distance restraints were observed. Several residue-specific peaks in the 500 ms DARR spectrum disappeared due to relaxation processes.

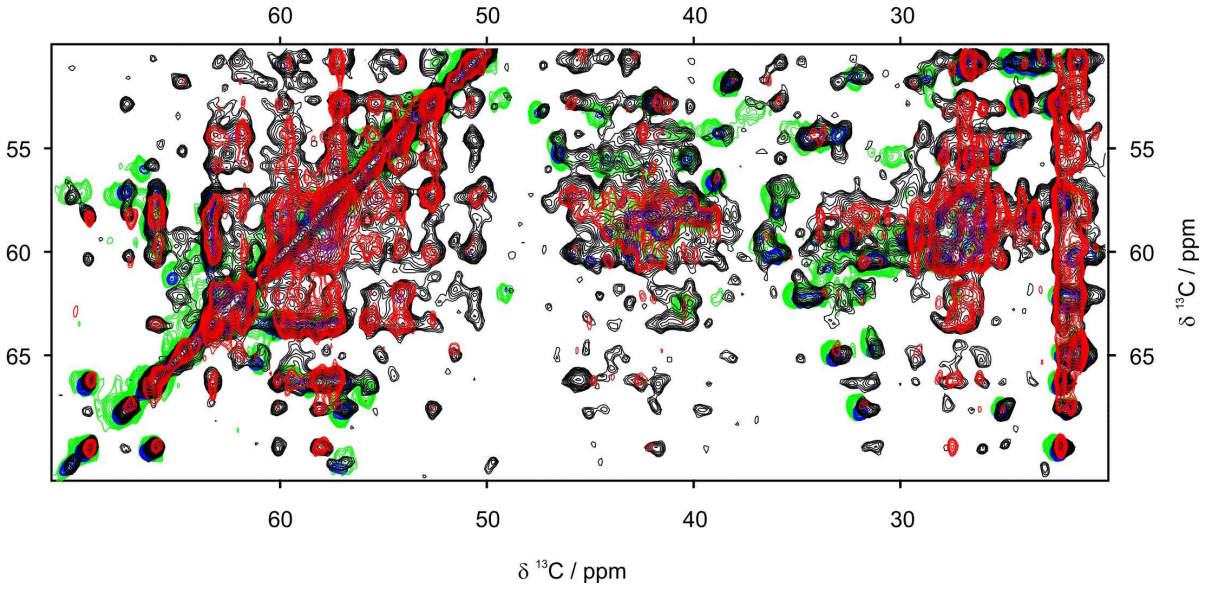


Figure 4.3 – The amount and type of spectral information depends upon the mixing time. A superposition of 2D ^{13}C DARR spectra with 25 ms (green), 100 ms (blue), 300 ms (black) and 500 ms (red) is shown.

4.3.1 Structure-defining NMR restraints

A large amount of data are achievable from solid-state NMR spectra of a uniformly ^{13}C, ^{15}N labelled sample. However, assignment of distance restraints in such spectra is difficult due to severe overlap of spectral peaks. To minimize the potential risk of assignment of short range

transfers to long range distance restraints, residue-specific and sequential assignment options – if present – were given preference over medium and long-range assignments; the order of priority was: residue-specific > sequential > short range > medium range > long range. Despite several potential long-range assignment possibilities, only those peaks were taken into account as long-range restraints which were unambiguous in both dimensions. The peaks which were unambiguous only in one dimension were left unassigned in the second dimension (partially assigned peaks). The peaks which had more than one possibilities of assignment in both dimensions were picked as ambiguous restraints (unassigned peaks). The ^{13}C-^{13}C distances help to define the secondary and tertiary structure of proteins. For defining the secondary structure of an α-helix, short-range restraints are needed. For example, in a solid-state MAS NMR spectrum, contacts of the type "i to i ± 3" for a stretch of residues are signatures of an α-helix. These contacts may emerge from the backbone/backbone and backbone/side-chain carbons of neighbouring residues. On the other hand, determination of tertiary structure of a protein requires backbone/side-chain or side-chain/side-chain contacts of the long-range type. As an illustration of this, typical α-helical and β-sheet distances between backbone C` and Cα in i to i+n (n=1 - 5) residues are tabulated in Figure 4.4.

secondary structure	backbone carbon	distance b/w residue i and i+1 / Å	distance b/w residue i and i+2 / Å	distance b/w residue i and i+3 / Å	distance b/w residue i and i+4 / Å	distance b/w residue i and i+5 / Å
α-helix	CO to Cα	2.4	4.5	4.5	5.1	7.4
β-strand	CO to Cα	2.4	5.4	8.8	12	15.3

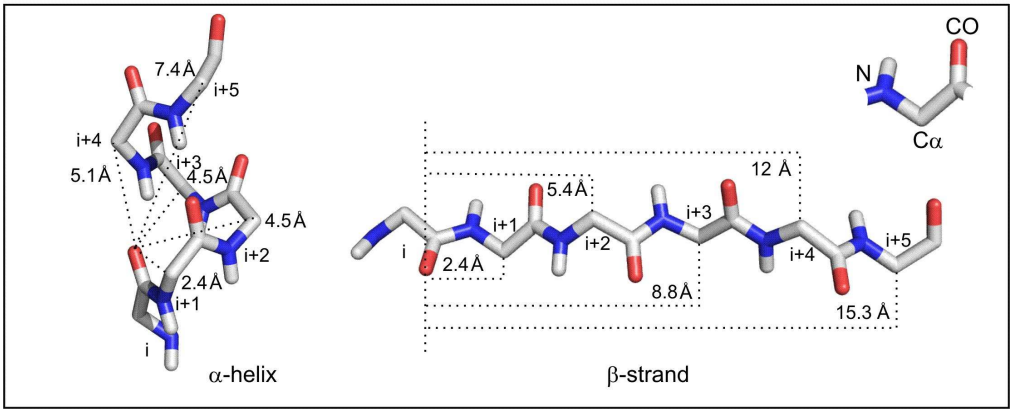


Figure 4.4 – Distance restraints between the backbone Cα and C`. Typical values of the backbone carbon distances in the β-strand and the α-helix are tabulated. A schematic representation of α-helix (left side) and a β-strand (right side) is shown in the bottom part of the diagram. For sake of clarity, only the backbone nuclei are shown and are highlighted with different colour-coding, i.e., blue for amide nitrogen, white for backbone Cα, and red for carbonyl oxygen.

Using solid-state MAS NMR spectra with different mixing times, the α-helix and β-sheet of YadA-M were distinguished from each other on the basis of their structural distances. Several short and medium-range distance restraints were observed which uniquely defined the N-terminal α-helix. The contour plot taken from a 2D PDSD spectrum recorded with 100 ms mixing (Figure 4.5) shows several carbon-carbon correlations involving residues L28 to G35 that help to define the α-helix. Typically, the majority of α-helical contacts were observed between "i" to "i ± 3" residues. As shown in the spectrum, T30 and V32 have well dispersed chemical shifts and show many sequential, short and medium range contacts. On the other hand, it is unlikely to observe similar contacts for residues in a β-strand (cf. Figure 4.4). However, long-range restraints between different β-strands in a parallel or anti-parallel β-sheet are very likely to be detected for this mixing time. For example, the cross peak between residue F56 and S73 (cf. Figure 4.5) is favourably a contact between two different β-strands, as these residues are lying faraway in the primary sequence. The register of four β-strands in YadA-M was determined on the basis of distance restraints from several spectra. Sequential and long-range peaks for β-sheet are highlighted in boxes in Figure 4.5.

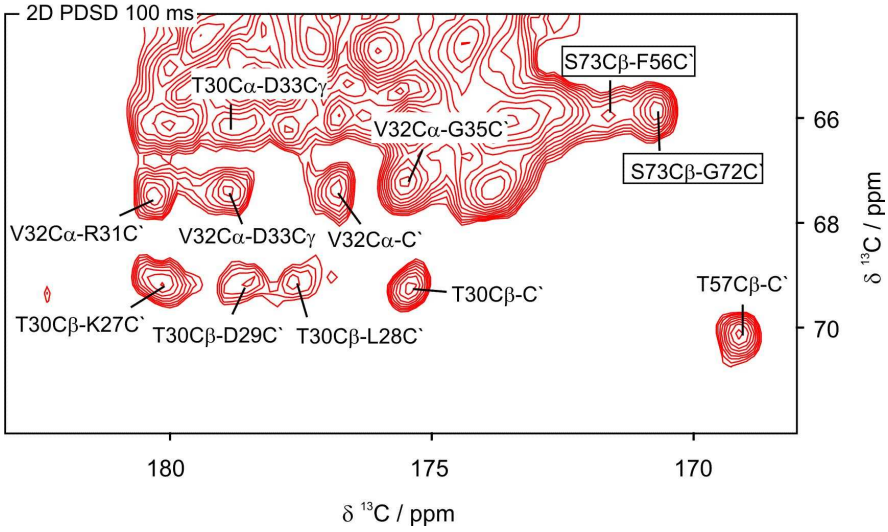


Figure 4.5 – Contour plot of a 2D PDSD spectrum recorded with 100 ms mixing time at a 900 MHz spectrometer. Several sequential, short and medium range assignments are shown for a stretch of residues in the N-terminal α-helix.

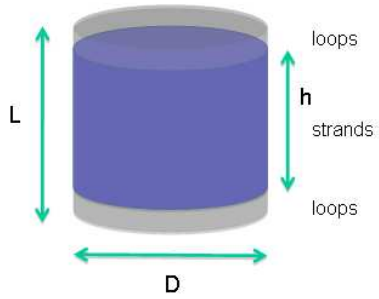
The amino-acid residues in the unstructured regions (i.e., random coil, β-turn) usually show weaker cross-peak intensities owing to increased flexibility. YadA-M possesses a single long loop that connects the N-terminal helix with the first β-strand (N43-G52); this region shows a fairly lower number of structural restraints. The unstructured behaviour of the loop region, the reduced number of distance restraints and the fact that it can adopt multiple conformations, will add up to make this region more difficult to be calculated; moreover,

any change in the loop orientation will affect the positioning of the helix relative to the sheet in the structure of YadA-M.

4.3.2 The shear number (S) for β-barrel of YadA-M

Several NMR contacts between strand β1 and β4 were observed which were used to determine the inter-monomer register of β-strands. The likeliness that these contacts have an intra-monomer origin was ruled out, because this would result into a highly improbable, narrow barrel comprised of only four strands. Therefore, these contacts were treated as inter-monomer restraints. In the symmetric YadA-M trimer, each of the three monomers contributes four β-strands to form a twelve-stranded β-barrel. The fact that loops between β-strands are short (~ 3 residues on average) strongly suggests that contacts between strands β1-β2, β2-β3, and β3-β4 are intra-monomer in nature. The shear number (S) plays a vital role in determining the structure of β-barrel proteins where β-strands are tilted relative to the barrel axis in such a way that first and last strands are hydrogen bonded to form a closed barrel. The protomers in YadA-M can slide in steps of two-residue with respect to each other in order to satisfy amide-carbonyl hydrogen bonds between the antiparallel β-strands (i.e., between strand 1 of protomer A and strand 4 of protomer B). Shear number for one protomer would either be zero, two, four or six if neighbouring protomers are aligned exactly parallel, two residues, four residues or six residues apart from each other. Because of its trimeric nature, the shear number for YadA-M should be multiple of three, i.e., would either be 0, 6, 12 or 18. Increasing the shear number increases the width and decreases the height of the barrel. From the NMR restraints between strand 1 and 4 of neighbouring protomers, the determined shear number for YadA-M was 12. The dimensions of a barrel with this shear number are sufficient to accommodate three helices in the barrel pore, and for the barrel to pass through the outer membrane. This is illustrated in Figure 4.6 where four hypothetical β-barrels with shear number 0, 6, 12 and 18 are shown. On top part of the Figure is shown a cylinder where "h" represents the height of the β-barrel without loops, "L" the total height, and "D" the width of the barrel. In the Table, values of these dimensions for barrels with different shear numbers are shown which were calculated according to Murzin et al. (Murzin, Lesk et al. 1994). According to the calculations, the barrels with shear number 0 and 6 have a diameter of 17 and 18.2 Å, respectively. Both of these barrels do not have sufficient room to accommodate a triple-helix inside. On the contrary, the diameter for the barrel with

shear number 18 is sufficiently large (i.e., 25.6 Å), however, its height (i.e., 22.6 Å) is not enough to pass through the outer membrane. Therefore, logic also supports the NMR-determined shear number 12 for YadA-M which results into a reasonable diameter and height of the barrel.



Shear (s)	0	6	12	18
<D> (A)	17	18.2	21.3	25.6
h (A)	34	31.8	27.2	22.6
L (A)	~ 41	~ 38.8	~ 34.2	~ 29.6

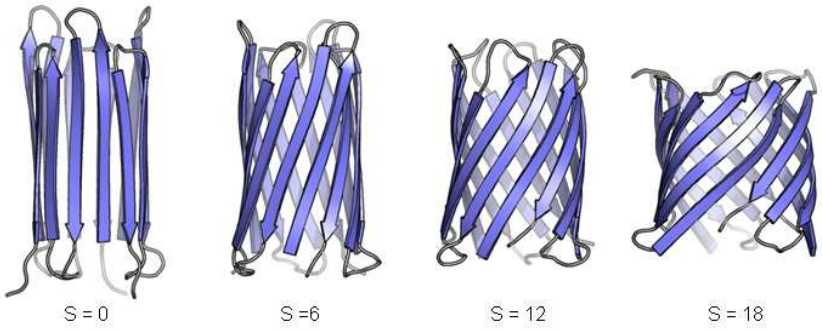


Figure 4.6 – Schematic representation of β-barrel dimensions with different shear numbers. YadA-M has a shear number 12 which was measured from the NMR data.

In spectra of a uniformly labelled sample, a distance restraint can be of various types (sequential, short range, medium range or long range). Distinguishing inter-strand peaks from others, therefore, is not an easy task. 2D NhhC and ChhC (Lange, Becker et al. 2005) spectra were recorded, tuned to achieve distance restraints between β-strands. In both of these experiments, proton-proton contacts between the antiparallel strands are indirectly determined. In NhhC/ChhC experiments, magnetization is first transferred from nitrogen/carbon to directly-bonded protons via a short CP time (cf. Section 2.2.4) In the next step, protons are allowed to exchange magnetization using a variable mixing time. Finally the proton magnetization is transferred and acquired on the carbons where the detected signals give indirect information about the protons involved in the exchange. The inter-strand peaks can be observed by selecting a short mixing time which should be just enough to ensure a transfer between alpha protons on two adjacent β-strands. In Figure 4.7 a contour plot taken from a 2D NhhC is shown recorded with a mixing of 200 µs. The assignments of cross peaks from inter-strand origin are highlighted in red colour. In addition to deriving

inter-strand information from this spectrum, the many sequential cross peaks were utilized to corroborate sequential assignment. As shown on the left side of Figure 4.7, The distance between amide proton of the residue "i" and alpha proton of the residue "i-1" is ~2.2 Å (highlighted in blue) and is smaller than the distance between inter-strand protons (amide proton of residue "i" and alpha proton of residue "j+1", typically ~2.8 Å) Therefore, sequential peaks (assignments in black) are dominant over inter-strand peaks in such spectra (cf. Figure 4.7).

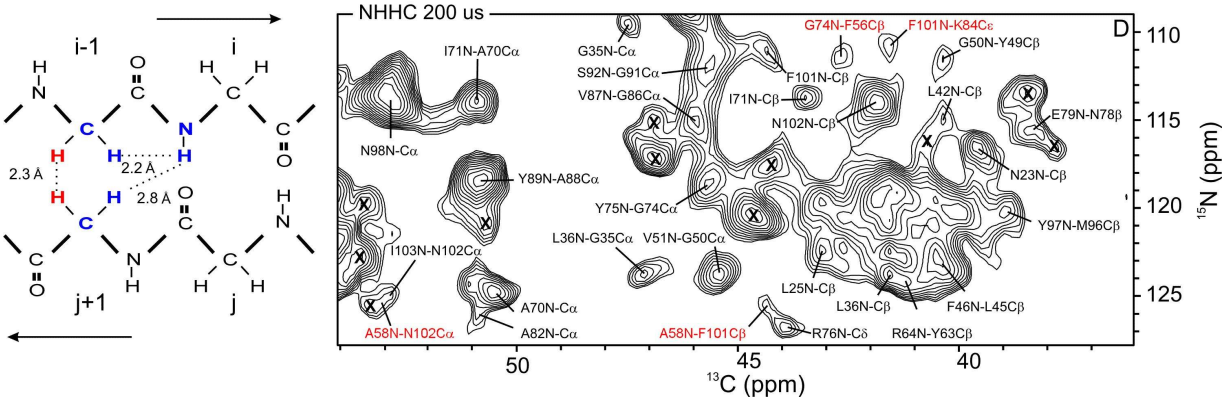


Figure 4.7 – Contour plot taken from 2D NhhC 200 µs showing indirect sequential and inter-strand H-H transfer. Two antiparallel β-strands are drawn (with only backbone atoms) on left side of the Figure with typical inter-proton distances in case of NhhC (blue) and ChhC transfer (red). Both of these experiments were extremely helpful to determine the topology of β-sheet in YadA-M.

2D ChhC spectra with various mixing times (35, 50, 80, 150, 200, 300, 500 µs) were recorded on YadA-M to achieve additional inter-strand contacts. A typical through-space distance between alpha protons on adjacent anti-parallel β-strands is ca. 2-3 Å (highlighted in red in Figure 4.7, left side). This is smaller than the distance between protons on sequential residues (~ 4.5 Å). Therefore, cross peaks in a short mixing time ChhC spectrum (~ 35-100 µs) should preferably be assigned to inter-strand restraints. Thus, it was possible to attain several long-range inter-strand restraints for β-sheet of YadA-M on the basis of these spectra. An overlay of a 2D ^{13}C-^{13}C DARR and ChhC spectra is shown in Figure 4.8. Contacts between the helix and hydrophilic residues of the β-sheet were observed that proved to be crucial in defining the position of α-helix relative to the β-sheet. For example, a contact between S44 and S100 is shown in the spectrum (Figure 4.8, right). The information that this interaction is between two different protomers comes from the structure calculation, i.e., such assignments were used as ambiguous restraints that could be allocated both to intra- or inter-protomer contacts during structure calculation. In general, due to its extreme stability, engineering of mixed,

complementary labelling schemes is not feasible for YadA-M; taking apart the trimer is only possible after complete denaturation of the proteins, which afterwards cannot be reassembled into trimers. This fact makes it also almost impossible to retrieve inter-helical contacts from the coiled-coil region, for which pairs of residues provide highly similar intra- and inter-helical distances. Therefore, all α-helical contacts were treated as intra-helix restraints.

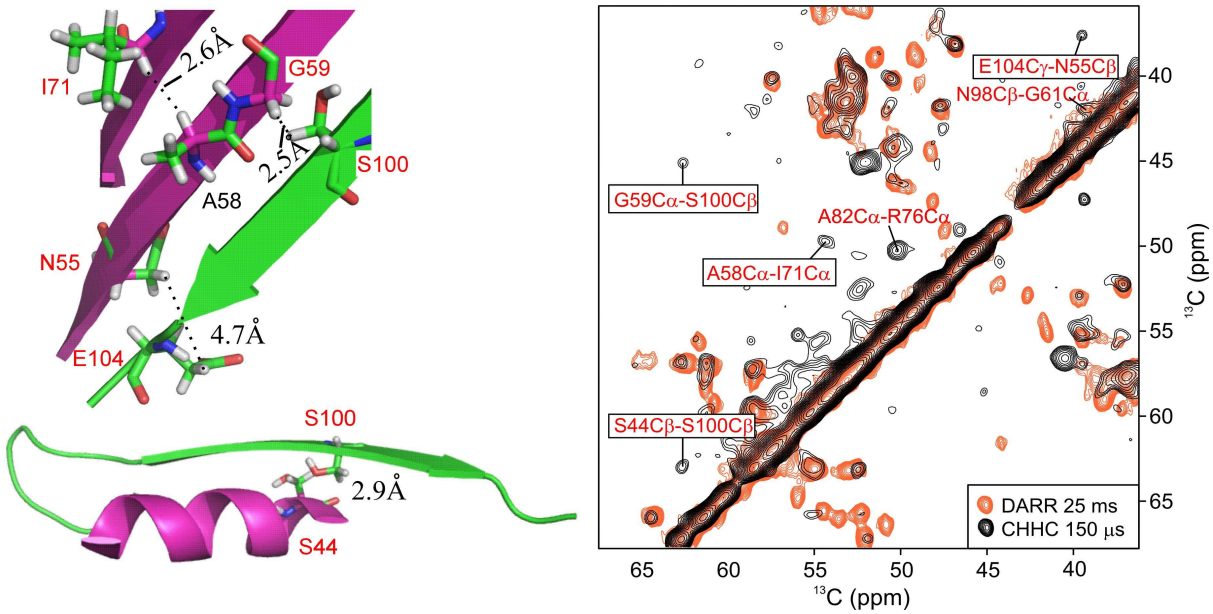


Figure 4.8 – (Right) Overlay of a 2D ^{13}C-^{13}C ChhC spectrum (150 µs mixing, black contours) and a 2D ^{13}C-^{13}C DARR spectrum (25 ms mixing, orange contours). Inter-strand peaks (both intra- and inter-protomer) are highlighted in boxes. (Left) Inter-strand restraints with average inter-proton distances are shown by cartoon representation of β-strands. Colour of the β-strands is different for different protomers. Also the relative position of α-helix of 'protomer A' with respect to a β-strand of 'protomer B' is shown.

2D ^{13}C DARR experiments with 200, 300 and 500 ms mixing times were recorded to get distance restraints over 7 Å. These spectra were rich in structural information and hundreds of diverse distance restraints were obtained from them. Although it is not a strict rule, the intensity of peaks can be used as indicator to differentiate between distance ranges. For intermediate mixing times, the peaks emerging from long-range contacts are generally less intense than those from medium and sequential contacts. A contour plot taken from 2D DARR 300 ms is shown in Figure 4.9, strip B. Long-range inter-strand peaks are highlighted in red colour; their intensity is lower than those of residue-specific and sequential peaks (assignments with black colour). At still higher mixing times, the situation becomes different. The peaks from short distances become weaker as a result of relaxation, and sometimes

disappear. For example, a 2D DARR experiment with 500 ms mixing shows far less number of peaks than 2D DARR with 200 and 300 ms mixing (cf. Figure 4.3, 500 ms DARR).

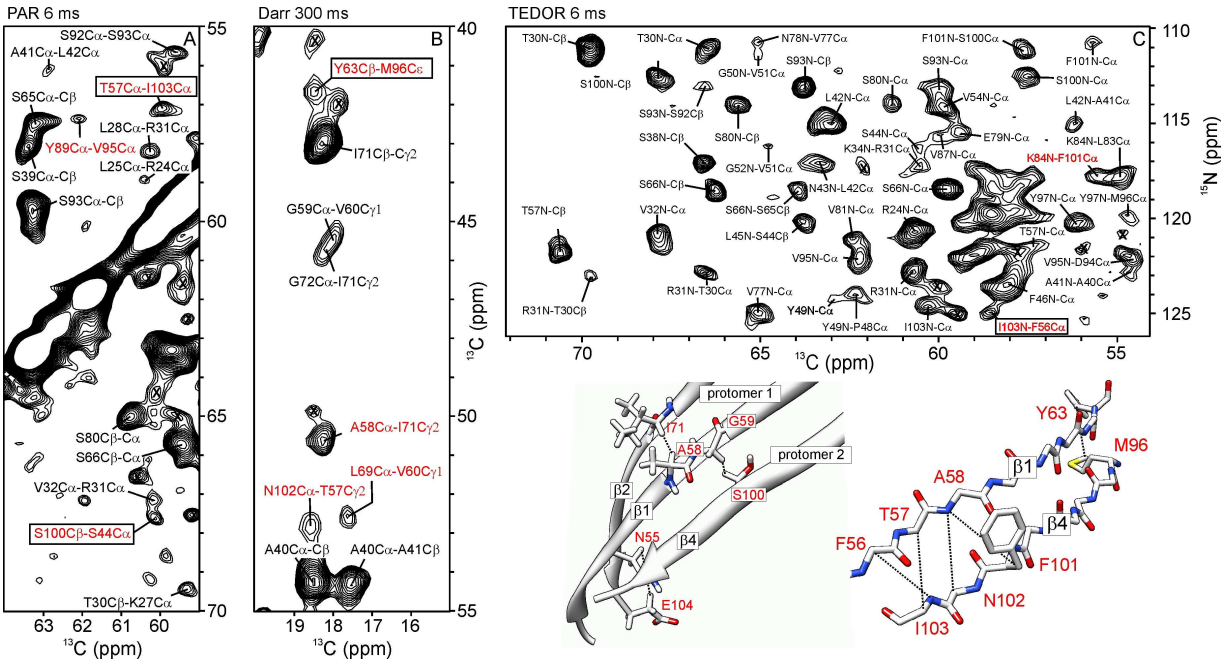


Figure 4.9 – Long-range interactions observed in solid-state NMR spectra of YadA-M (A) Contour plots showing sections of a 2D ^{13}C-^{13}C PAR spectrum recorded using a mixing time of 6 ms; **(B)** 2D ^{13}C-^{13}C DARR spectrum recorded with a mixing time of 300 ms; **(C)** 2D ^{15}N-^{13}C TEDOR recorded with mixing time of 6 ms. For all spectra, correlations are labelled as follows: long-range, structure-defining interactions are labelled in red; interactions that are depicted as restraints on the structural models (lower right) are highlighted with an additional box. Black crosses label NMR correlations that were used as 'ambiguous restraints' in the structure calculations.

Along the same lines, 2D homonuclear ^{13}C PAR (proton assisted recoupling) with different mixing times were recorded. The new spectra were found very efficient in ^{13}C-^{13}C long-range transfers and were rich in medium- and long-range distance restraints. In contrast to DARR spectra, in which long mixing times (i.e., on the order of 200-300 ms) are needed to see long-range distance restraints, similar information is present in a 6 ms PAR spectrum. Similarly, 2D heteronuclear ^{15}N-^{13}C TEDOR spectra were recorded to obtain long-range ^{15}N and ^{13}C restraints. These spectra are not only information-rich but also much better resolved than 2D NhhC spectra (cf. 4.7, right). Figure 4.9 shows a composition of excerpts from 2D PAR, DARR and TEDOR spectra, demonstrating the quality of the NMR data and some of the important restraints that helped to define the tertiary structure. The determined topology of intra- and inter-monomer β-strands on the basis of these restraints, is shown on the bottom right of the Figure 4.9.

4.3.3 Filtered restraints

Previously, 2D ^{13}C-^{13}C methyl-filtered spectra have been effectively used to distinguish side-chain methyl shifts from other nuclei (Jehle, Hiller et al. 2006). By combining extended ^{13}C-^{13}C mixing times (300 ms) with similar spectral editing schemes (Wu, Burns et al. 1994), it was possible to identify many long-range restraints, originating predominantly from methyl nuclei. The cross-polarization with a polarization inversion (CPPI) step, introduced after the first CP (cf. Section 2.2.3), quickly nulls or 'de-cross-polarizes' carbon signals which are strongly coupled to protons (i.e., –CH and $–CH_2$ groups), whereas signals from methyl groups and highly mobile side-chain nuclei (mostly Lysine Cε signals) escape this filter. Before the DARR mixing step, therefore, the main reservoir of magnetization comprises of methyl and flexible side-chain nuclei. By allowing this filtered magnetization to mix with surrounding nuclei non-specifically for 300 ms, a number of important medium- and long-range correlations between methyl groups and other carbons were observed. This is illustrated in the bottom-middle of Figure 4.10. For comparison, a 2D DARR with 300 ms is also shown (bottom-left). Introduction of the methyl-filter reduces the amount of information and helped in the assignment of restraints. Two zooms of this spectrum (i.e., A and B) are shown on top part of the Figure with long-range assignments; note that one of the interacting nuclei is a methyl group. In zoom A, both side-chain methyl groups of I71 (I71Cδ1 and I71Cγ2) show a long-range inter-strand contacts with A58. Moreover, a contact between methyl of T57 and L45 was also identified. Lysines have relatively long and mobile side chains. Indeed, in zoom B, two peaks originating from side-chain Cε of K34 and K53 are observed. These long-range peaks helped to determine the alignment of helix with respect to the β-sheet, as shown in the top left part of the Figure 4.10. In a second experiment, an additional dipolar-dephasing step (Opella and Frey 1979; Opella, Frey et al. 1979) was introduced immediately before acquisition. This "double" methyl-filter (one filter employed before and the other after DARR mixing) results in a spectrum with long-range cross peaks mostly from methyl and mobile carbon nuclei (Figure 4.10, bottom right). The spectrum significantly alleviates signal overlap and helped identifying useful long-range restraints. The carbonyl region of both the methyl-filtered spectra contained several backbone/side-chain long-range contacts.

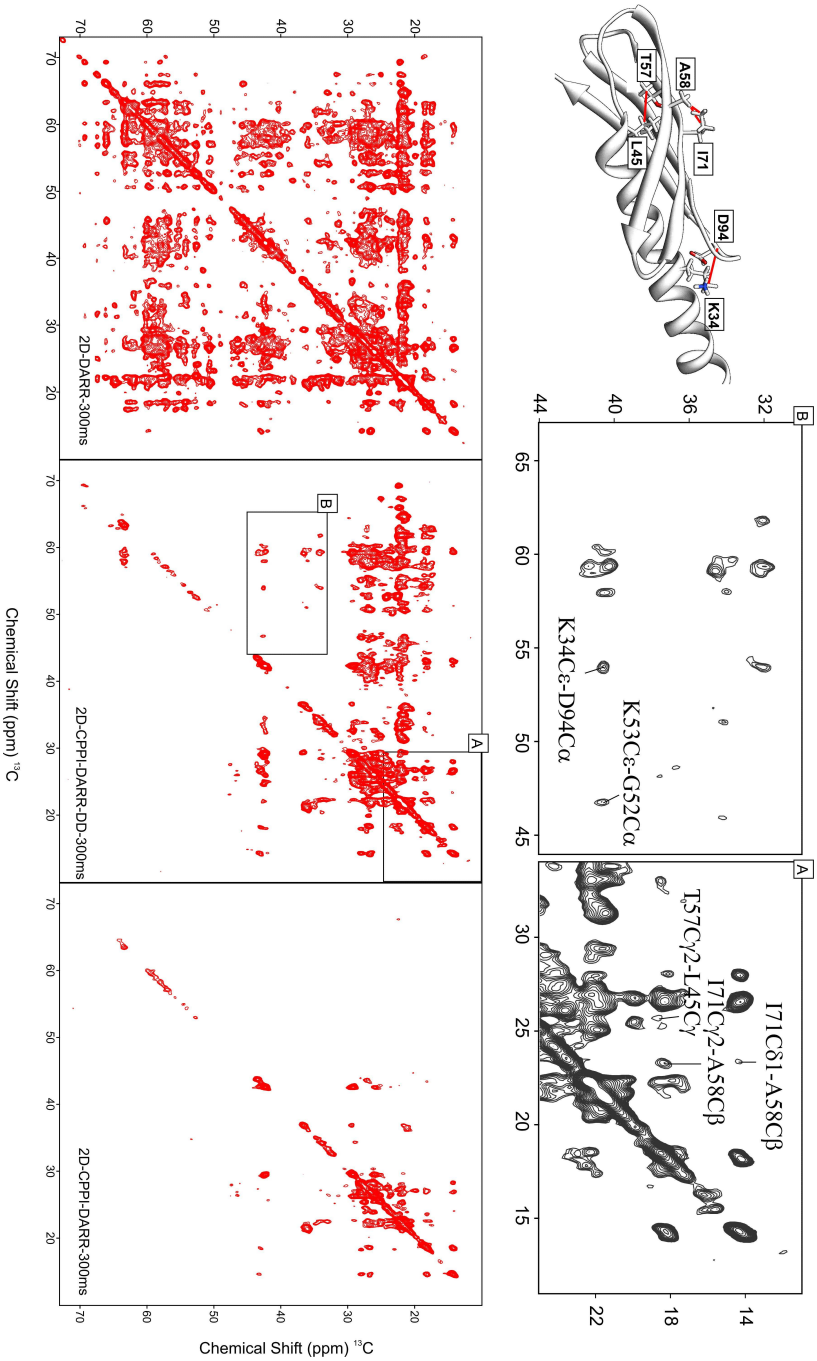


Figure 4.10 – 2D methyl-filtered spectra to achieve filtered restraints. A comparison of the information content in three different spectra, i.e., 2D DARR (left), 2D CPPI-DARR (with one methyl-filter; middle) and 2D CPPI-DARR-DD (with two methyl filters; right) is shown in the bottom part of the Figure. The regions labelled 'A' and 'B' in the (middle) CPPI-DARR spectrum are shown in more detail at the top. Structure defining inter-strand and strand-helix assignments are shown with graphical representation on top left part of the Figure.

The structure defining restraints discussed above are just a few examples to illustrate the information content in the full data set of homo- and heteronuclear MAS NMR correlation experiments. A detailed overview of the data set and summary of the structural information derived from it that served as input for the structure calculations is given in Table 4.1.

Table 4.1 – List of solid-state MAS NMR spectra, together with number of different distance restraints which were used in the YadA-M structure calculation

Experiment type	Mixing time	Nuclei involved	Proton frequency/ MHz	MAS/ kHz	Total picked peaks	Total manual assignments	Sequential assign-ments	short range assign-ments	medium range assign-ments	long range assign-ments	Inter-molecular assignments	Upper bounds /Å
DARR	200 ms	^{13}C-^{13}C	900	13	887	669	230	32	6	58	24	7.50
DARR	300 ms	^{13}C-^{13}C	900	13	979	736	242	78	6	60	24	7.80
DARR	500 ms	^{13}C-^{13}C	900	12	624	403	116	37	2	20	7	7.80
CPPI-DARR-DD	300 ms	^{13}C-^{13}C	900	12	129	75	9	1	0	6	0	7.80
CPPI-DARR	300 ms	^{13}C-^{13}C	900	12	453	319	79	23	2	13	11	8.0
PDSD	15 ms	^{13}C-^{13}C	900	13	302	266	43	1	1	2	1	4.50
PDSD	100 ms	^{13}C-^{13}C	900	13	653	536	165	24	0	3	4	6.25
PAR	2.25 ms	^{13}C-^{13}C	850	13.33	278	246	19	0	0	6	1	5.0
PAR	6 ms	^{13}C-^{13}C	850	13.33	477	362	68	15	4	23	4	6.50
PAR	15 ms	^{13}C-^{13}C	850	13.33	290	225	47	5	0	13	2	8.0
TEDOR	2.2 ms	^{15}N-^{13}C	850	13.33	220	190	62	2	0	1	2	5.50
TEDOR	6 ms	^{15}N-^{13}C	850	13.33	236	212	81	11	0	2	1	6.50
TEDOR	12 ms	^{15}N-^{13}C	850	13.33	48	43	10	0	0	1	0	7.50
ChhC	35 µs	^{13}C-^{13}C	700	10	28	20	0	0	0	1	1	3.75
ChhC	50 µs	^{13}C-^{13}C	700	10	37	33	2	0	0	1	1	4.0
ChhC	80 µs	^{13}C-^{13}C	700	10	66	56	6	0	0	4	1	4.50
ChhC	150 µs	^{13}C-^{13}C	700	10	116	84	10	2	0	4	4	5.75
ChhC	200 µs	^{13}C-^{13}C	900	13	69	61	2	1	1	3	1	6.0
ChhC	300 µs	^{13}C-^{13}C	900	13	97	68	10	3	0	4	1	6.50
ChhC	500 µs	^{13}C-^{13}C	900	13	177	103	19	11	2	7	4	7.0
NhhC	35 µs	^{15}N-^{13}C	700	10	68	44	22	0	0	0	0	3.50
NhhC	50 µs	^{15}N-^{13}C	700	10	59	45	18	0	0	0	0	4.0
NhhC	100 µs	^{15}N-^{13}C	700	10	124	70	25	0	0	0	0	5.0
NhhC	200 µs	^{15}N-^{13}C	700	10	124	84	33	0	0	5	4	6.0

Table 4.1 – An overview of the 24 spectra of the data set, including the number of picked peaks and number of assignments that provided structure-defining restraints. A brief explanation of the different columns is as follows:

Experiment type: Type of pulse sequence used to exchange magnetization between different nuclei.
Mixing time: Time allowed for magnetization exchange; as a rule of thumb, short mixing times mainly provide intra-residue or sequential exchange, while extended mixing times are required for long-range transfers.
Proton frequency/MHz: Proton Larmor frequency of the experiment.
MAS/kHz: Magic-angle spinning frequency of the experiment.
Total picked peaks: Total number of both manually assigned and ambiguous peaks for each experiment; ambiguity can be both in one or two dimensions.
Sequential assignments: Backbone/backbone, backbone/side-chain and side-chain/side-chain correlations between residue (i) and (i ± 1).
Short range assignments: Transfer between nuclei of residues that are separated by more than two - but less than three – amino-acid residues in the primary sequence. These types of restraints are particularly helpful for defining the secondary structure of α-helices.
Medium range assignments: Transfer between nuclei of residues that are separated by more than three - but less than five – amino-acid residues in the primary sequence; these restraints are particularly helpful in defining β-turns connecting the β-strands.
Long range assignments: Transfer between nuclei of residues that are separated by more than five amino-acid residues in the primary sequence. These restraints are very important for defining the exact geometry of the molecule. Long-range restraints were used to define the register of β-sheet and helped to position the N-terminal α-helix relative to the β-sheet in YadA-M.
Inter molecular assignments: Distance restraints between protomers of the symmetrical trimer. Restraints between strand β-1 of protomer A and β-4 of protomer B are required to define the correct shear number of the β-barrel. The β-strands are tilted with respect to the N-terminal α-helices; hence the position of the N-terminal α-helix relative to the β-sheet is defined by several intermolecular restraints between the α-helix of protomer A and the β-sheet of protomer A and/or B.
Upper bounds in Å: Maximum distance between interacting nuclei used by ARIA and ISD for structure calculation.

4.4 *De novo* structure determination of YadA-M

As already described in the introduction chapter ("aims of the project"), structure determination of YadA-M was a joint project in which sample preparation and structure calculation were carried out in close collaboration with Max Planck Institute for Developmental Biology, Tübingen. In collaboration with Michael Habeck (MPI Tübingen), solid-state MAS NMR data were used for calculating the structure of YadA-M using ISD. In this chapter, different steps involved in structure calculation of YadA-M by the ISD are described. In addition, the structure of YadA-M was determined using ARIA, in collaboration with Benjamin Bardiaux (FMP Berlin), and a brief summary of the protocol is included at the end of this chapter.

ISD (inferential structure determination) (Rieping, Habeck et al. 2005) was used to calculate the structure of YadA-M from available solid-state NMR data. ISD is able to facilitate structure calculations from sparse and hybrid structural data (Bayrhuber, Meins et al. 2008). This method has successfully been applied in solution-state NMR structure determinations and the work described here is the first attempt to determine protein structure from solid-state NMR data. The structure determination of YadA-M was a challenging task for two reasons: Firstly, the solid-state MAS NMR distance restraints in general, are poorly defined as compared to those from typical solution NMR data. Secondly, the very fact that YadA-M is a symmetrical trimer makes it difficult to distinguish between inter- and intra-monomer contacts, since complexes with complementary isotope labelling couldn't be engineered. ISD tackled these problems by using an iterative strategy where ambiguity of restraints was progressively removed in multiple iterations. In the first step, an approximate structure of the monomer was determined to infer intra-monomer hydrogen bonds. The inferred hydrogen bonds were used in the next iterations, resulting in a better defined monomer structure. The final structure of monomer was then assembled into a trimer using strict C3 rotational symmetry. That way, it was possible to determine the structure of YadA-M entirely on the basis of solid-state NMR data.

4.4.1 Distance restraints used by ISD

Distance restraints were derived from 24 different two-dimensional solid-state MAS NMR spectra comprising a total of 2033 manually assigned cross peaks. These include sequential, short-range, medium-range and long-range peaks (Table 4.2). A list of 1192 non-redundant

restraints was obtained by combining all these spectra. The duplicated restraints appearing in more than one spectrum were merged into one with the smallest upper bound. The maximum distance bounds used for each spectrum are listed in Table 4.1. An overview of all 24 spectra with the total number of restraints used by ISD in the final structure calculation is shown in Table 4.3. In addition to the 1192 non-redundant restraints, the table includes an additional of 146 peaks obtained from automatic assignment, yielding a total of 1338 restraints.

Restraint type	**Full set**	**Non-redundant set**
long-range (**$\|i–j\| > 5$**)	340	219
medium-range (**$4 \leq \|i–j\| \leq 5$**)	25	19
short-range (**$2 \leq \|i–j\| \leq 3$**)	290	219
sequential	1378	735
intra-residual	0	0
Total	**2033**	**1192**

Table 4.2 – Number of distance restraints derived from manually assigned cross peaks from 24 solid-state MAS NMR spectra. A total of 2033 peaks in various distance regimes were manually assigned. Peaks which appear in multiple spectra were merged to a single distance restraint with smallest distance bound, thus yielding a set of 1192 non-redundant restraints.

Experiment	**Mixing times**	**Number of restraints**	**Total**
ChhC	35/50/80/150/200/300/500 µs	6/4/14/16/2/15/43	100
NhhC	35/50/100/200 µs	32/12/28/39	111
DARR	200/300/500 ms	266/163/87	516
PDSD	15/100 ms	55/182	237
TEDOR	2.24/6/12 ms	48/60/2	110
PAR	2.25/6/15 ms	35/106/31	172
CPPI-DARR	300 ms	71	71
CPPI-DARR-DD	300 ms	21	21
Total			**1338**

Table 4.3 – Number of restraints used by ISD in the final structure calculation. The restraints include 1192 non-redundant distance bounds that were extracted from manually assigned cross peaks and an additional set of 146 restraints from automated assignment.

YadA-M is a homo trimer which means that each monomer should contribute 4 strands to form a 12-stranded β-barrel. Based on the fact that the connecting loops are relatively short, contacts between strands 1-2, 2-3 and 3-4 are assumed to be intra-monomer restraints. As described earlier in this chapter, the restraints between strand 1 and 4 were considered as inter-monomer restraints (a highly improbable, narrow β-barrel would result from a four-

stranded beta sheet). Using these assumptions, all non-redundant distance restraints were classified into 525 intra-monomer, 48 inter-monomer and 619 ambiguous restraints. A contact map representation of these restraints shows an anti-parallel arrangement of the four β-strands (Figure 4.11A). In the contact map, the inter-monomer restraints between β1- β4 are coloured green, while the intra-monomer restraints between β1- β2, β2- β3 and β3- β4 are coloured black. All ambiguous restraints are coloured blue in the contact map. A comparison of upper bound statistics between a solution data set of ubiquitin (1D3Z) and the solid-state restraints from YAD-M (Figure 4.11B) shows how loose are restraints in solid-state NMR data set.

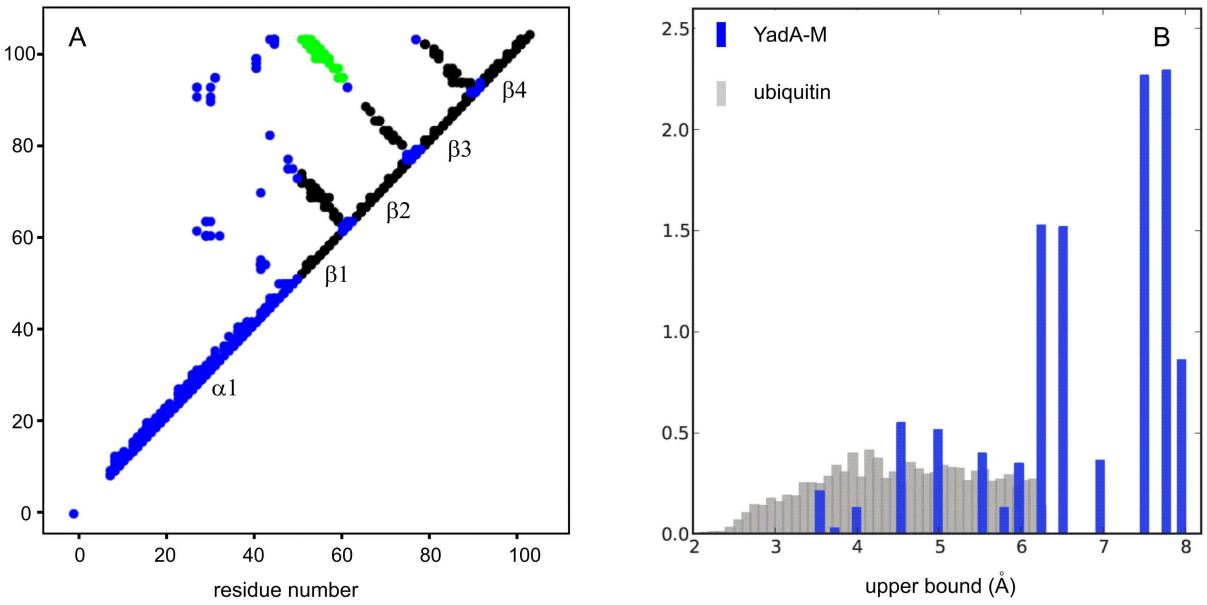


Figure 4.11 – Contact map showing the distance restraints obtained from 24 solid-state MAS NMR spectra. Colours indicate if a restraint is treated as intra-monomer (black), inter-monomer (green), or ambiguous (blue) restraint; (b) Comparison of upper bound statistics between a solution data set of ubiquitin (1D3Z) and the solid-state MAS NMR restraints of YadA-M.

4.4.2 Structure of the monomer (simulation 1-3)

Structure of the monomer was calculated using 619 ambiguous and 525 intra-monomer solid-state MAS NMR restraints. Although the 48 inter-monomer restraints were removed in the monomeric structure calculation, the presence of possibly unassigned inter-monomer restraints in the data set needed a soft error-tolerant restraint potential. Unlike solution NMR spectral peaks, in ssNMR spectra, intensity-distance relationship is not well defined; therefore, a coarse grain classification into different distance classes was used for different mixing time spectra. The ISD model for distance data evaluates the probability of finding an

upper and a lower bound distance for each unknown experimental distance. Keeping in mind that the correct distances are unknown, distance restraints were treated as nuisance parameters and were integrated using Monte Carlo sampling. Unknown experimental distances were related to the calculated distances in the structure by log-laplace distribution. The structure of YadA-M trimer was achieved after six simulations. The effective restraint potential and a tabulated summary of all six simulations are shown in Figure 4.12. Positive and negative signs indicate presence and absence of restraints used in each individual simulation.

No.	Structure	Distances	phi/ psi	H-bonds	B-weighted positions
1	Monomer	+	+	-	-
2	Monomer	+	+	+	-
3	Monomer	-	-	-	+
4	Trimer	+	-	-	+
5	Trimer	+	+	+	-
6	Trimer	+	+	+	-

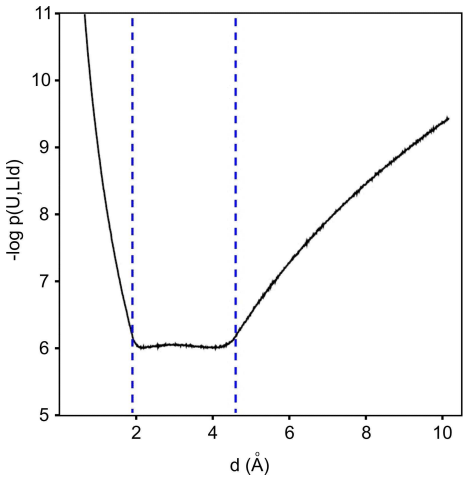


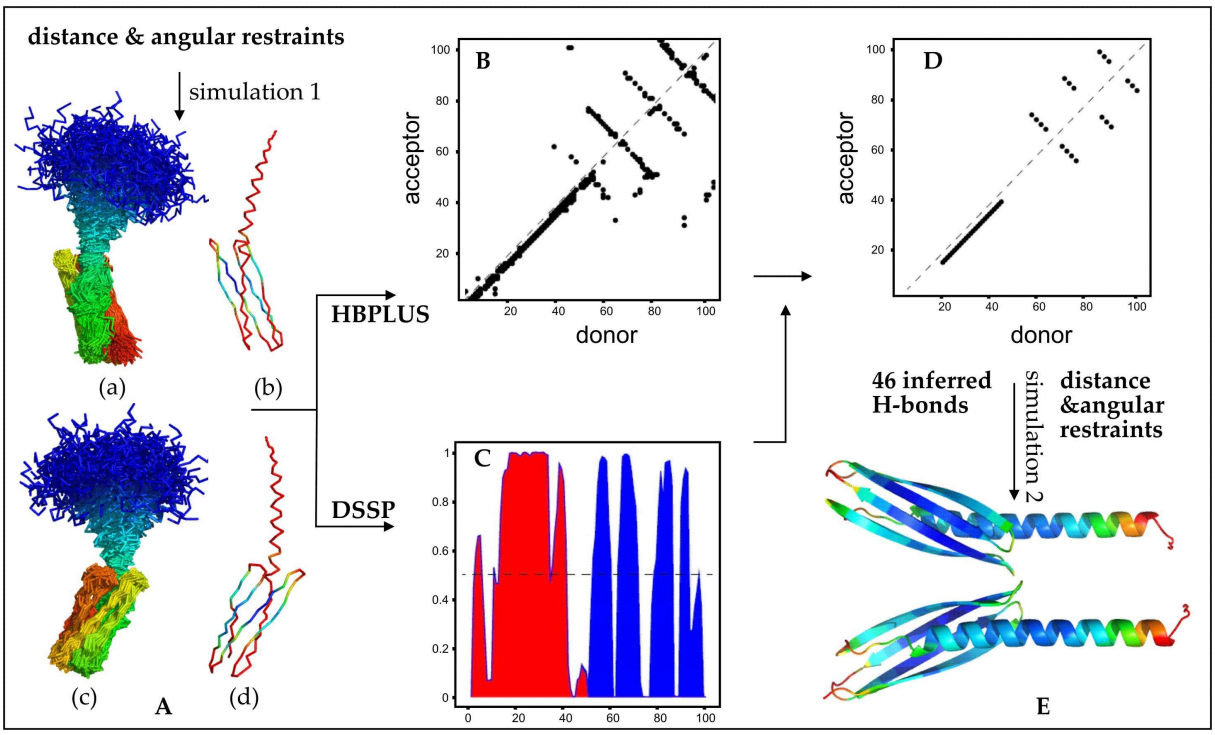
Figure 4.12 – Overview of the ISD structure calculation. "±" indicates inclusion/omission of restraints. On right side is shown the effective restraint potential resulting from an upper bound U = 4.5 Å and a lower bound L = 1.8 Å (indicated as blue dashed lines) at a weight k = 4.0.

In the first simulation **(simulation 1)**, intra-monomer solid-state MAS NMR distance restraints along with 120 angular restraints obtained from TALOS+ were used. The simulation was carried out using a replica-exchange Monte Carlo algorithm, (Habeck, Nilges et al. 2005). The conformational prior (i.e., the already established information about the system; cf. Section 1.6.3 for details) included the standard soft-repulsive non-bonded force field (PROLSQ potential form with parallhdg5.3 parameters (Linge and Nilges 1999; Linge, Habeck et al. 2003) and interactions involving hydrogen atoms switched off) and additional phi/psi and hydrogen bonding potential for backbone hydrogen bonds (Scheme 4.2A). A total of 11980 replica transitions are simulated which after convergence yield 4980 structures in the final ensemble. The weight of the distance restraints is estimated and shared between intra and ambiguous restraints using Gibbs sampling (Habeck, Rieping et al. 2006). The estimated weight is quite low (3.76 ± 0.14) due to the inter-monomer restraints which cannot

be accommodated in a single monomer and result in an elevated "noise" level. On average, 79.7 ± 6.4 distance restraint violations (> 0.5 Å) were observed within the monomer ensemble (intra-monomer: 21.0 ± 3.9, ambiguous: 58.0 ± 5.3). The force constants of the phi and psi restraints were estimated and clipped to values of 20 (corresponding to a circular variance of 13°). The ensemble despite being heterogeneous (see Scheme 4.1, Figure A) clearly shows a four-stranded beta sheet and an N-terminal alpha helix. The monomer ensemble from simulation 1 was used to infer the secondary structure and register of beta strands using DSSP (Kabsch and Sander 1983). The secondary structure was assigned according to the 'maximum a posteriori estimate' derived from the posterior probability (cf. Figure 4.1C) with additional smoothing in the helix region, resulting in the following secondary structure: 13-42 helix, 54-62 strand 1, 65-74 strand 2, 81-89 strand 3, and 93-100 strand 4. The hydrogen bonds were also calculated for all 4980 structures by HBPLUS (Mcdonald and Thornton 1994). Posterior probability of secondary structure and contact map of main-chain hydrogen bonds are shown in Figure C and B, respectively, in Scheme 4.1. The hydrogen bonds which are consistent with the inferred secondary structure are filtered for the next simulation (Figure D). For the hydrogen bonds in the alpha helix, only those were kept which follow the canonical (i + 4, i) pattern. The register of neighbouring strands was inferred by selecting the subset of restraints that are consistent with an anti-parallel pattern (i,j), (i+2, j-2), etc. In this way, a total of 46 hydrogen bonds were obtained (converted to 92 hydrogen bonding restraints) among which 26 are helical bonds and 20 are bonds between the anti-parallel strands 1-2, 2-3 and 3-4. Additional support for the boundaries and relative register of hairpins comes from hairpins detected with SimShift (Ginzinger and Fischer 2006).

In the second simulation **(simulation 2),** the 46 hydrogen bonds were implemented as distance restraints between donor and acceptor (2.9 Å) and amide hydrogen and acceptor (1.9 Å) using a fixed 100 kcal/mol force constant. They were used along with angular and distance restraints and a forcefield with a more realistic non-bonded energy function adapted from the Rosetta software (Kuhlman, Dantas et al. 2003; Habeck 2011) (cf. Scheme 4.2B). The simulation started from an extended structure which after 3000 replica transitions, started to converge, yielding an ensemble of 1500 structures (total replica transitions were 4500). The inferred hydrogen bonds used in the second simulation resulted in a slightly higher (3.85 ± 0.14) estimated weight of the distance restraints. The number of restraint violations was also reduced to 64.0 ± 6.3 (intra-monomer: 16.0 ± 3.6, ambiguous: 47.0 ± 4.6).

An average structure as well as its local B factor was calculated from the ensemble. This is shown in Figure E of Scheme 4.1, where B-factor colouring indicates the local variability in the structure ensemble ranging from rigid (blue) to highly flexible (red). Additional intra-monomer hydrogen bonds were inferred from the average monomer structure. An average monomer structure was determined in the third simulation **(simulation 3)**. The simulation was relatively short (620 replica-exchange transitions) in which an average structure as close as possible to the ensemble from simulation 2 was obtained using B-factor weighted positional restraints (Scheme 4.2C). The resulting structure obeyed the covalent restraints and showed good non-bonded interactions.



Scheme 4.1 – Summary of monomeric structure calculation by ISD (simulation 1 & 2). In the first simulation, a structure ensemble is obtained using 525 intra-monomer and 619 ambiguous restraints**.** The ensemble is shown in Figure A where (a), (c) represent front and back views of the super-imposed structure ensemble (colour changes from blue to red for N- and C-terminus, respectively), and (b), (d) correspond to front and back views of the average structure obtained with robust superposition (colour indicates local variance, i.e., blue for rigid and red for highly flexible) (Mechelke and Habeck 2010). This ensemble is used to predict secondary structure information and main-chain hydrogen bonds using HBPLUS and DSSP, respectively. Inferred secondary structure of YadA-M is shown in Figure C where red and blue colours represent helix and β-strands, respectively. Observed main-chain hydrogen bonds are indicated by black dots in Figure B. Those filtered hydrogen bonds which are consistent with the inferred secondary structure are converted into distance restraints (Figure D) and are used in simulation 2 along with distance and angular restraints to achieve a refined monomer structure. In Figure E two views of the mean structure obtained from simulation 2 are shown. The local B-factor was also calculated and variability within the structure ensemble is indicated by different colouring in Figure E.

4.4.3 Assembly of the trimer (simulation 4-5)

The mean monomer structure obtained from simulation 3 was used to assemble the trimer in simulation 4. The simulation used distance restraints, B-factor weighted positional restraints, and strict C3 symmetry. The Lennard-Jones potential for non-bonded interactions between non-hydrogen atoms was used as force field. Exact rotational symmetry was implemented in ISD using a single monomer structure parameterized in dihedral angles that was subjected to forces and interactions with its "symmetry mates" generated by the symmetry operator.

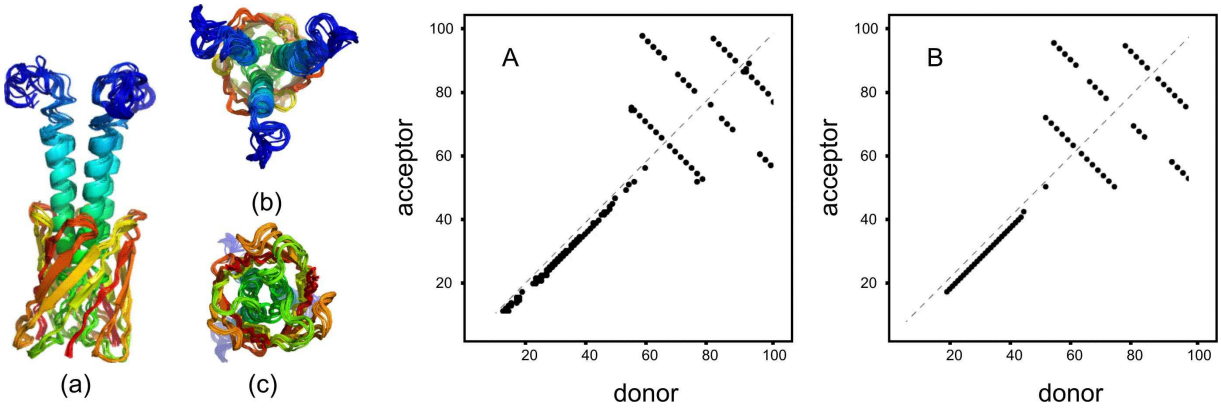
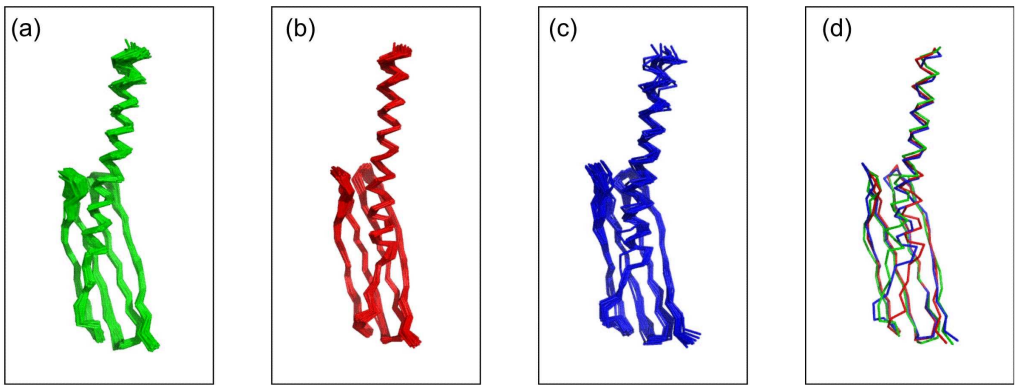


Figure 4.13 – (Left) The structure ensemble obtained with simulation 4 showing side (a), top (b), and bottom (c) views. **(Right)** Hydrogen bonds inferred from the trimer are shown in A. Hydrogen bonds consistent with the inferred secondary structure were converted into distance restraints and are shown in B.

Ambiguous distance restraints were calculated as r^{-6} averages over all possible intra- and inter-monomer distances. The simulation started with 3520 replica transitions which started to converge after 2500 transitions (Scheme 4.2D). Ambiguous restraints that were violated in the previous simulation were satisfied in the trimer, which, in combination with a high force constant, resulted in a significantly reduced number of violations (> 0.5 Å) 19.0 ± 2.3 (intra-monomer: 7.0 ± 1.1, inter-monomer: 1.0 ± 1.8). The assembled trimer structure shows a β-barrel with the trimeric helix passing through the pore (Figure 4.13, left). Interestingly, the ensemble converged to a unique register between strand 1 and 4. From this ensemble, 61 intra-monomer and 10 inter-monomer (between strand 1 and 4) hydrogen bonds were derived which, along with distance and angular restraints, were used in the next replica simulation **(simulation 5).** In addition, a purely repulsive force field as the prior distribution was used (Scheme 4.2E). 40 distance restraints between atoms separated by one or two covalent bonds only, were removed (intra-monomer: 24, ambiguous: 16 restraints). The

simulation started with 2400 transitions which converged after 1000 transitions giving an ensemble of 1400 structures. Because of the lower force constant (10.98 ± 1.13), an increase in distance restraint violation (39.0 ± 6.1) was observed (intra-monomer: 15.0 ± 2.7, inter-monomer: 1.0 ± 0.8, ambiguous: 22.0 ± 5.0). Cluster analysis (Hirsch and Habeck 2008) of the 1400 monomer structures resulted in three main conformers in the ensemble with population of 54, 32 and 14%, and ensemble RMSDs of 0.74, 0.60, and 1.0 Å, respectively (Figure 4.14). The region spanning residues G35 to Y49 showed the largest backbone structural differences (local RMSD > 2 Å). Conformer 1 has a straight helix, whereas conformers 2 and 3 have a kink in the helix. A detailed statistics for three conformers from simulation 5 is shown in the table of Figure 4.14. The statistics table shows that cluster 2 violated the distance restraints more than cluster 1 and 3. This was reflected in a slightly lower distance weight for cluster 2 (10.2 kcal/mol) than that for cluster 1 and 3 (~11.4 kcal/mol). Cluster 1 is not only the most populated conformer but also showed the best average dihedral angle energy and lowest average non-bonded energy among the three clusters.



Statistics	Conformer 1	Conformer 2	Conformer 3
Population size	54%	32%	14%
Heterogeneity (13-105, monomer) Å	0.74	0.60	1.0
Total violations (< 0.5 Å)	36.0 ± 4.5	45.0 ± 4.8	36.0 ± 4.3
Intra-monomer violations (< 0.5 Å)	15.0 ± 2.8	15.0 ± 2.1	12.0 ± 2.0
Inter-monomer violations (< 0.5 Å)	1.0 ± 0.9	1.0 ± 0.8	1.0 ± 0.6
Ambiguous violations (< 0.5 Å)	19.0 ± 2.7	28.0 ± 3.2	23.0 ± 4.0
Circular variance phi [10^{-3}]	46.2 ± 7.4	51.9 ± 9.9	57.8 ± 12.9
Circular variance psi [10^{-3}]	46.8 ± 11.1	46.5 ± 6.8	57.1 ± 11.3
1st generation packing quality	-3.51 ± 0.24	-2.77 ± 0.24	-3.35 ± 0.22
2nd generation packing quality	-3.37 ± 0.23	-3.14 ± 0.23	-3.61 ± 0.23
Ramachandran plot appearance	-3.51 ± 0.60	-3.02 ± 0.37	-4.14 ± 0.31
Backbone conformation	0.33 ± 0.29	0.24 ± 0.39	-0.09 ± 0.38
chi-1/chi-2 rotamer normality	-8.13 ± 0.24	-8.03 ± 0.24	-8.27 ± 0.34

Figure 4.14 – Structure ensembles from simulation 5 and the statistics. Three main conformers are shown in green (54%), red (32%), and blue (14%); (d) shows the mean structure with same colour-coding. The 12 first N-terminal residues which are disordered are not shown. Conformer statistics for simulation 5 is summarized in the table.

4.4.4 Final Refinement (simulation 6)

Consistently violating restraints in the ensemble of the first conformer were identified and were removed to improve the structure which was selected for **simulation 6**. The peak lists were searched for those new restraints which were consistent with the structure. For each distance restraint all possible realizations (*rAA*, *rAB*, *rBA)* were evaluated and counted the number of times each realization was smaller than the upper bound within a tolerance of 1 Å. In this way, frequencies (*pAA, pAB, pBA*) consistent with the respective distance were obtained. Following probabilities were constructed to classify the distance restraints into intra-monomer, inter-monomer and ambiguous restraints.

• Probabilities for unique assignments:

$$\Pr(AA) = pAA\ (1 - pAB)(1 - pBA) \quad (1)$$

$$\Pr(AB) = pAA\ (1 - pAA)\ pAB\ (1 - pBA) \quad (2)$$

$$\Pr(BA) = (1 - pAA)(1 - pAB)\ pBA \quad (3)$$

• Probabilities for ambiguous assignments:

$$\Pr(AA \text{ or } AB) = pAA\ pAB\ (1 - pBA) \quad (4)$$

$$\Pr(AA \text{ or } BA) = pAA\ (1 - pAB)\ pBA \quad (5)$$

$$\Pr(AB \text{ or } BA) = (1 - pAA)\ pAB\ pBA \quad (6)$$

$$\Pr(AA \text{ or } AB \text{ or } BA) = pAA\ pAB\ pBA \quad (7)$$

• Probabilities for wrong assignments:

$$\Pr(\text{wrong}) = (1 - pAA)(1 - pAB)(1 - pBA) \quad (8)$$

Depending upon the probabilities obtained from above equations, a restraint was classified either as intra-monomer [Pr(*AA)* > 0.9], inter-monomer [Pr(*AB)* > 0.9], or was completely removed [Pr(wrong) > 0.9]. In all other cases, the restraint was used as completely

ambiguous. A final list of restraints obtained using this scheme is shown in Table 4.4 and Figure 4.15.

Number of restraints (per monomer)		**Structural quality**	
Distance restraints			
Intra monomer **=**	***1064***	***Ramachandran Statistics***	
• Intra-residual (**$\|i-j\| = 0$**)	16	• Most favoured regions	77.58 ± 1.38
• Sequential (**$\|i-j\| = 1$**)	631	• Allowed regions	16.38 ± 2.09
• Medium-range (**$2 \geq \|i-j\| \geq 5$**)	222	• Generously allowed regions	3.31 ± 1.40
• Long-range (**$\|i-j\| > 5$**)	195	• Disallowed regions	2.70 ± 0.91
Inter monomer **=**	***81***	***WHATIF Z-scores***	
Ambiguous **=**	***193***		
		• Backbone conformation	0.47 ± 0.44
Total distance restraints **=**	***1338***	• 2nd generation packing-quality	-2.88 ± 0.18
Dihedral angle restraints (ΦΨ) **=**	***120 (60/60)***	• Ramachandran plot-appearance	-3.24 ± 0.55
Hydrogen bonds restraints-(intra/inter) **=**	***71 (61/10)***	• χ1/ χ2 rotamer normality	-3.80 ± 0.61
Restraints statistics		**Structural precision**	
• Number of distance-violations > 0.5 Å	9 ± 2	• Backbone atoms-(monomer)	0.74 ± 0.26 Å
• Number of distance-violations > 0.3 Å	32 ± 2	• Heavy atoms-(monomer)	1.38 ± 0.58 Å
• RMS of distance-violations	0.118 ± 0.004 Å	• Backbone atoms (trimer)	0.84 ± 0.32 Å
		• Heavy atoms (trimer)	1.45 ± 0.61 Å

Table 4.4 – Restraints and structure statistics of the final YadA-M ensemble. Ramachandran statistics were calculated with PROCHECK.

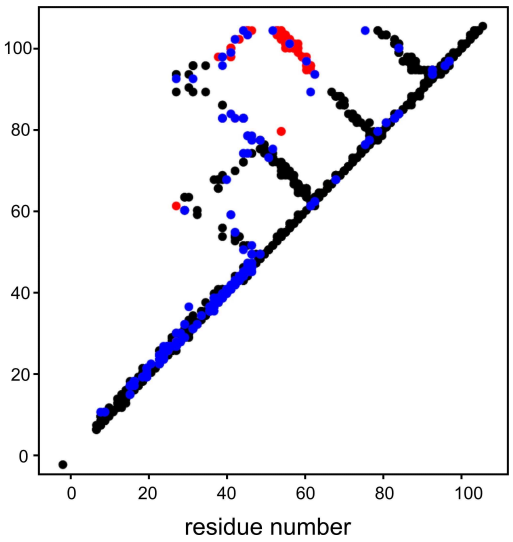


Figure 4.15 – Contact map showing distance restraints used in the final refinement. Different colours indicate whether a restraint was treated as intra-monomer (black), inter-monomer (red) or ambiguous (blue).

This list of restraints (Table 4.4) was used in the final structure calculation **(simulation 6)** of YadA-M. The first replica exchange was run with 4970 transitions based on Lennard-Jones force field without hydrogen interactions. In a subsequent simulation, hydrogen interactions were switched on and an additional potential for side-chain dihedral angles taken from the CHARM force field was used (Scheme 4.2F). The number of transitions was 1330 which were converged to the final ensemble of 250 conformational samples. The final YadA-M structure was selected from this ensemble. The estimated weight of the structure was 11.1 ± 0.3. The final structure of YadA-M (monomer and trimer) with different orientations is shown in Figure 4.16.

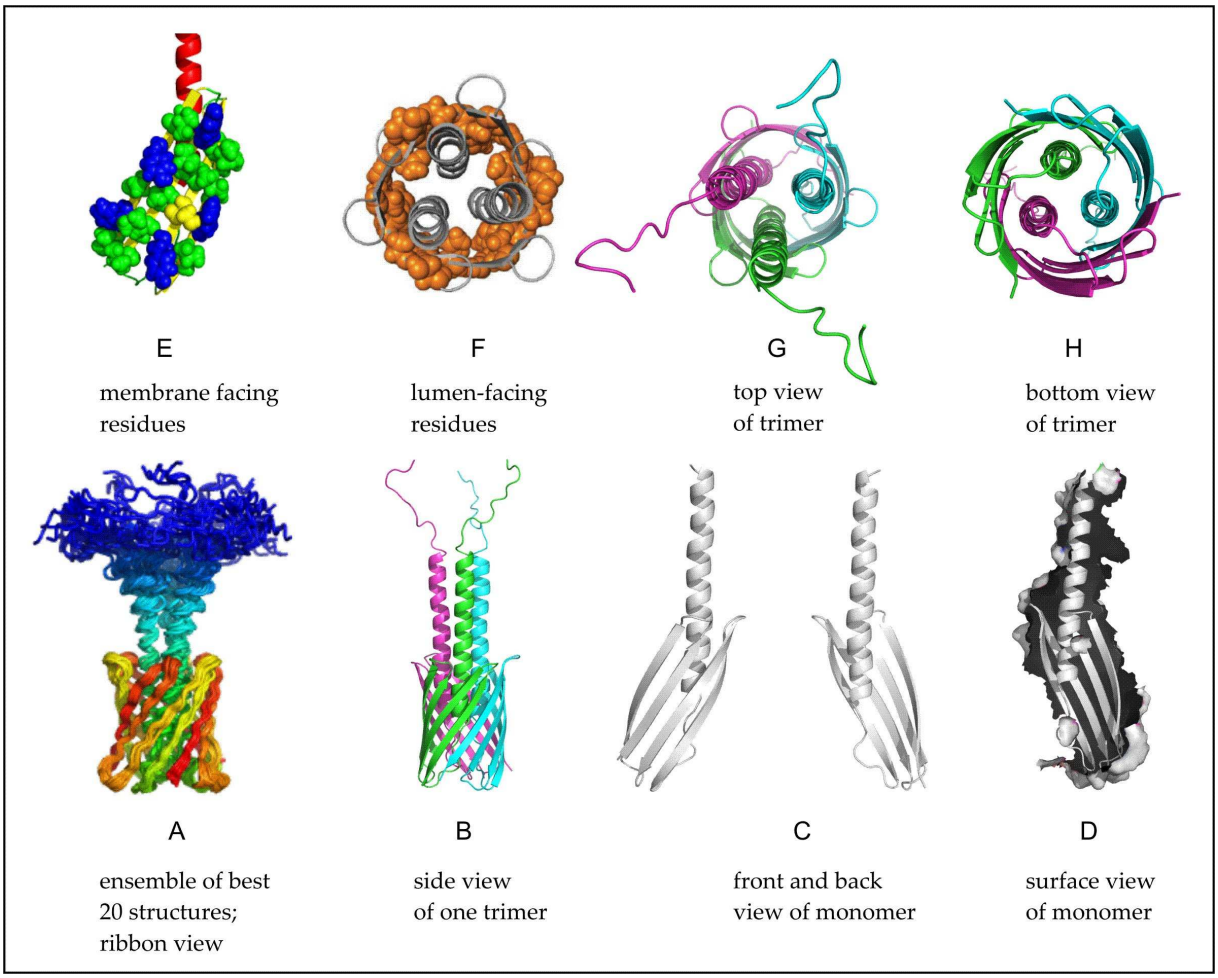
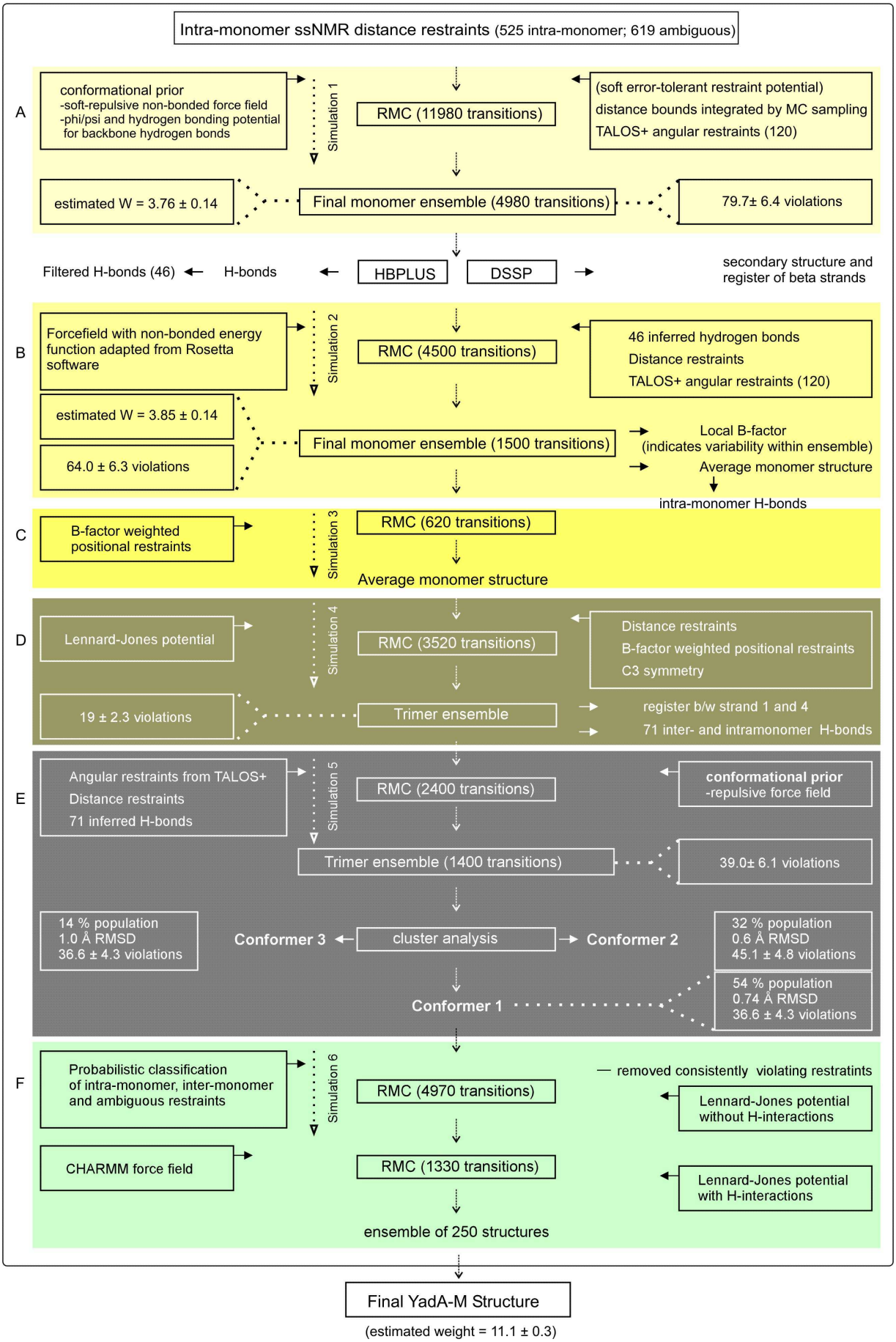


Figure 4.16 – Solid-state MAS NMR structure of YadA-M determined *de-novo* by ISD. (A) side view of the final structure ensemble comprising 20 representative conformers; **(B, G, H)** side, top and bottom views of one trimer of YadA-M; **(C)** side views of the protomer from front and back; **(D)** surface representation of one protomer; **(E)** side view of protomer with membrane-facing hydrophobic (green) and aromatic (blue) residues; **(F)** ring of small residues facing the barrel interior highlighted as orange spheres (β1 : N55, T57, G59, G61; β2 : A66, A68, A70, G72; β3 : G86, A88; β4 : N98, S100) seen from top along the barrel axis.

Scheme 4.2 - Summary of ISD calculation



4.5 Structure of YadA-M by ARIA

Using the same NMR data, structure of YadA-M was determined by Benjamin Bardiaux using conventional iterative ARIA protocol (Rieping, Habeck et al. 2007). The general scheme of structure calculation carried out by ARIA and structure obtained are shown in Figure 4.17.

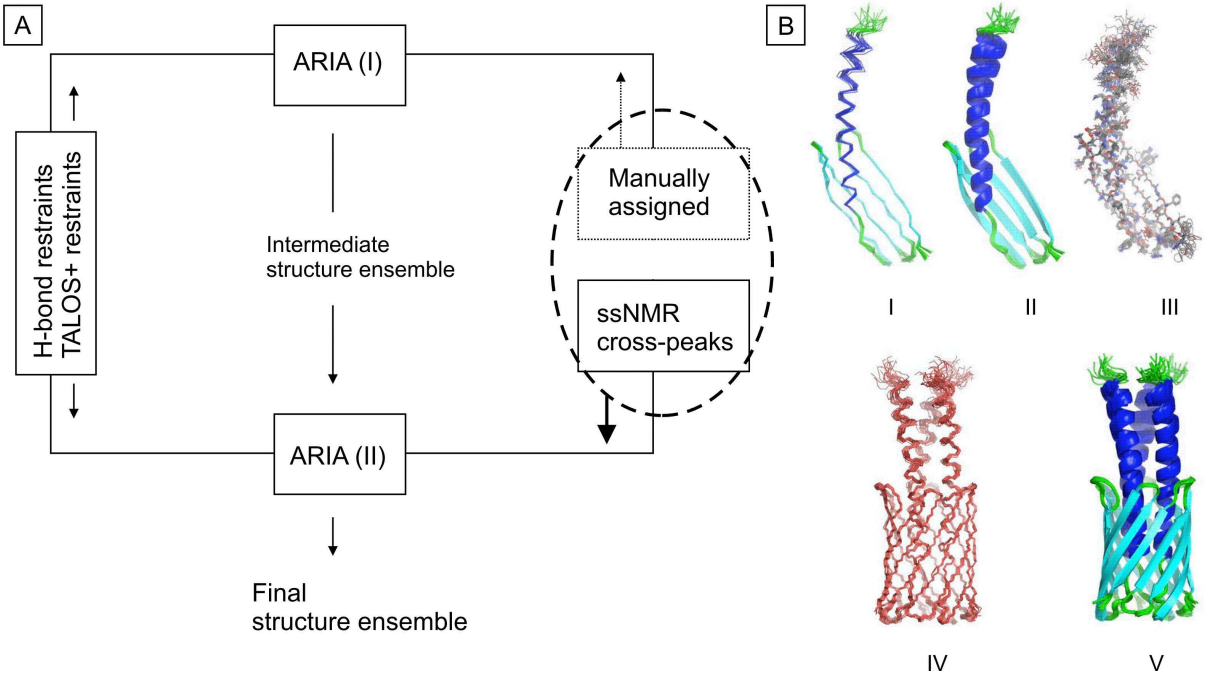


Figure 4.17 – (A) ARIA scheme used to calculate the structure of YadA-M. **(B)** Solid-state MAS NMR structure of YadA-M determined by ARIA; (I-III) final structure ensemble of the monomer in ribbon, cartoon and heavy atom views; (IV-V) side views of the final structure ensemble of the trimer with backbone and cartoon views. Ribbon and cartoon views in I, II and V are coloured by secondary structure (i.e., helix-blue, sheet-cyan, loop-green).

Structure calculation by ARIA was achieved in two steps. In the first step (ARIA I), all manually assigned cross peaks, hydrogen bonds and dihedral angle restraints were used to get a first ensemble of trimeric YadA-M. All peaks except those between β-strand 1 and β-strand 4 were considered as ambiguous due to symmetric nature of the molecule. The structure ensemble is well defined with an RMSD of 1.1 Å. This intermediate ensemble served as a template for calculating a high resolution structure in the second step (ARIA II), where all available NMR solid-state MAS NMR cross peaks were used. The final structure shows a high precision with an RMSD of 0.7 Å and 1.2 Å for backbone and heavy-atoms, respectively.

4.6 Description of the structure

The solid-state MAS NMR structure of YadA-M (PDB entry 2LME), determined by ISD, gives the following structural information: YadA-M is a trimeric β-barrel (Figure 4.16A, B). N-terminal helical domain, which forms a trimeric coiled-coil, passes through the barrel as shown in top and bottom views of YadA-M (Figure 4.16G, H). There are two major structural elements, an α-helix (G14-S44) and a β-sheet (β1: K53-Y63; β2: S66-V76; β3: V81-A90; β4: D94-E104) in YadA-M. The β-strands are connected to each other with relatively shorter β-turns. The helix is connected to the first β-strand with a larger flexible loop spanning from residue L45 to G52 (Figure 4.16C). β-strand 1 and 4 of neighbouring monomers are hydrogen bonded with each other to form a closed barrel with the β-strands tilted with respect to the pore axis. Amino-acid residues which are facing the membrane are largely hydrophobic (I, L, V) or aromatic (Y, F) with bulky side chains pointing into the membrane. Like most β-barrels, YadA-M has aromatic girdles at the top and bottom rim of the barrel. This is illustrated in Figure 4.16E where only membrane-facing residues are shown as green (hydrophobic) and blue (aromatic) spheres. The interior of the barrel comprises small side-chain residues (A, N, G, S, T) that leave enough space to accommodate the trimeric helical bundle (Figure 4.16F). The ARIA structure consists of a similar β-barrel through which N-terminal trimeric helix is passing. However, very minute differences were observed in both structures. In the ARIA structure, coiled-coil is more "coiled" whereas in ISD structure this is more straight. The major differences in both structures lie in a region from residues H16-L25, L42-G52, and A85-S93.

4.7 Conclusion

The structure of the membrane-anchor domain of YadA was determined *de novo* from solid-state MAS NMR. Unambiguous chemical shift assignment plays a crucial role in determining the secondary, and later on the tertiary structure of a protein. Structural restraints of different distances were acquired from 2D homo- and heteronuclear spectra with variable mixing times. The structural information obtained from these spectra provided input for the automated structural calculation, using the programmes ISD and ARIA. The structure of YadA-M was also calculated by ARIA, which gave a highly similar structure.

5 Discussion

5.1 Different interactions that stabilize YadA-M

YadA-M forms an extremely stable trimeric β-barrel and resists denaturation even at extreme buffer and temperature conditions (Wollmann, Zeth et al. 2006). Several interactions (based on the occurrence of solid-state NMR spectral correlations) are tabulated in Appendix I which, in addition to other factors, may contribute to the stability of YadA-M. The minimum and maximum distances for each interaction as well as its type (i.e., whether intra-monomer or inter-monomer) are based upon the final ensemble of the YadA-M structure. These interactions are classified into four categories, which are: aromatic, charged, polar and hydrophobic. The contribution of these interactions to the stability of proteins in general, is backed up by previous studies.

5.1.1 Interactions involving aromatic residues

In aromatic residues, electrons are localized on both sides of the aromatic ring, giving the face of the ring a partial negative charge, while the hydrogen atoms confer a partial positive charge to the edges of the ring. Ring currents in the aromatic side chains of Phe, Trp and Tyr contribute to protein stability by forming weak interactions with residues close in space. It has been reported that polar interactions between side-chain aromatic rings and backbone-amide protons as well as with aliphatic side-chain nuclei, are involved in forming local structures and stabilizing secondary structure in peptides and proteins (Lovas, Toth et al. 2001). Positive or partially positive charged amino groups of Lys, Arg, Asn, Gln and His that are located close to the ring face of Phe, Tyr and Trp (within 6 Å), form van der Waals contacts with their π-electron cloud (Burley and Petsko 1986). Amino-acid residues with a negative charge can interact with the protons on edges of aromatic rings. Several of such structure-stabilizing interactions in which aromatic side chains are involved with charged (e.g., E104-F56, D33-Y63, K34-Y63 and K53-Y75) as well as polar residues (e.g., F56-N102, F56-S73, T57-F101) were observed in YadA-M and are listed in Table 7.1.1 and 7.1.2 of Appendix I.

Glycines and prolines are residues which can destabilize the β-sheet structure. However, when glycines present in parallel and/or anti-parallel β-sheets make cross-strand

pairs with aromatic residues; this interaction can compensate for the destabilization effect (Regan and Merkel 1998). YadA-M has six glycines in the β-sheet region, one of them (G62) is involved in an amide-carbonyl hydrogen bond with Y97. The configuration of the other five glycines is indeed such that an aromatic residue at positions "j+1/j-1" is blanketing them (cf. Figure 4.7 for positions i and j), i.e., G59/F101, G61/Y97, G72/F56, G74/F56 and G86/Y97. This is illustrated in Figure 5.1, which shows the position of aromatic ring of F101 with respect to the G59. The distance between the face of the aromatic ring and alpha protons of G59 is approximately 3 Å. Glycines in the β-sheet of YadA-M thus, are a good choice to allow convenient passage of N-terminal domain at the one hand and nullify the destabilization effect by cross-strand interactions on the other hand (Appendix I, Table 7.1.3).

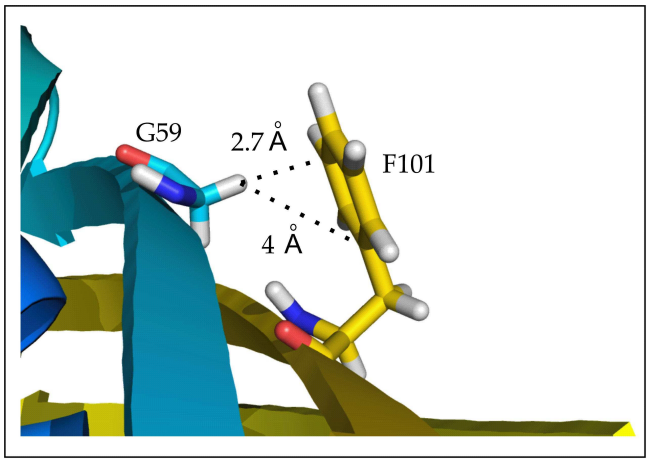


Figure 5.1 - Cross-strand interactions in YadA-M.The aromatic ring of F101 is covering G59.

5.1.2 Interactions involving charged and polar residues

At any pH there may exist positively and negatively charged groups in proteins which are involved in strong ionic interactions, called salt bridges. By definition, a salt bridge comprises of a hydrogen bond and electronic interactions, and is mostly present between carboxylate oxygen atoms of Asp or Glu and the side-chain amino groups of Lys or Arg. The distance between the hydrogen acceptor (carboxylate oxygen) and donor (amide nitrogen) should be no more than 3.9 Å, whereas the distance between acceptor and hydrogen atom should be 2.5 Å in order for a salt bridge to be established (Xu, Tsai et al. 1997). A few salt bridge interactions have been observed in YadA-M. On the periplasmic side of the β-barrel, a salt bridge was observed between the first β-strand of monomer 1 and the fourth β-strand of monomer 3 (cf. Figure 5.2). The distance between the Hζ of positively charged K53 and the Oδ of negatively charged E104 is 2 Å, thus a hydrogen bond as well as electronic interactions between them can be built. Other possible salt bridges are between residues D94-R64 (2.8 Å), K34-D94 (2.7 Å), and R31-D94 (2.9 Å) (Appendix I, Table 7.2.1). Similarly, polar and charged

residues if close in space can form hydrogen bonds, e.g., the carboxyamide side chain of N55 can make a hydrogen bond with carboxylate side-chain oxygens of E104 in YadA-M (Appendix I, Table 7.2.2).

It is known that proline residues play a vital role in stabilizing protein structure, because of their rigidity and nitrogen heteroatom. A dipole is created by the nitrogen heteroatom leading to non-covalent electrostatic interactions with other residues (Vondrasek, Biedermannova et al. 2008). In YadA-M, there is only one proline (P48), and potential interactions with other residues have been listed in Appenidx 1, Table 7.2.3.

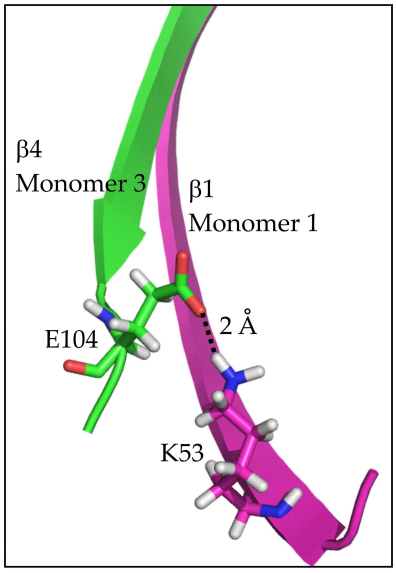


Figure 5.2 - Inter-monomeric salt bridge in YadA-M.

Amino-acid residues which are either charged, or participate in hydrogen bonding at physiological pH, are considered polar. Several interactions in YadA-M where either one or both of the residues were polar, are listed in Appendix I (Table 7.3.1 and 7.3.2).

5.1.3 Interactions involving hydrophobic residues

Hydrophobic interactions form the most abundant and important non-covalent forces involved in stabilizing protein structure. Indeed, the hydrophobic interactions induce protein folding to form a 3D structure. Val, Leu and Iso are the most hydrophobic residues but other residues may also be involved in hydrophobic interactions depending upon their total hydrophobicity (Kyte and Doolittle 1982). Among aromatic residues, Tyr and Trp have intermediate hydrophobicity. Interactions between hydrophobic residues, between hydrophobic and aromatic residues, and between hydrophobic and neutral residues are listed in Appendix I (Table 7.4.1, 7.4.2 and 7.4.3, respectively).

5.2 Crystal contacts observed in solid-state MAS NMR spectra

The purified YadA-M sample (uniformly ^{13}C, ^{15}N labelled) formed 5-10 µm long needle-shaped 3D microcrystals after dialysis against doubly-distilled water (ddH_2O) (cf. Chapter 2.2) (Wollmann, Zeth et al. 2006). Microcrystals are formed by a regular arrangement of myriad number of crystals; this arrangement is uniform throughout except at the boundaries

of the microcrystals. This implies that in a microcrystalline protein sample, the bulk arrangement of amino-acid residues in every microcrystal should be identical. Therefore, crystal contacts (contacts between different trimers) can be expected to appear in the solid-state NMR spectra of YadA-M. Indeed, some cross peaks could not be assigned and may be the result of such crystal contacts. The structure of YadA-M reveals that membrane-facing residues are largely hydrophobic with long side chains (i.e., Ile, Leu, Val, Phe; cf. Figure 4.16) therefore, these are potential candidates to show crystal contacts. The Cδ1 chemical shifts of I71 and I103 are 14.2 and 16.1 ppm, respectively and are the most upfield shifted resonances in the YadA-M spectra (Figure 5.3A). Any observed contact with the Cδ1 of both I71 and I103 should therefore be unambiguous with respect to isoleucine as one of the interaction partners. In a 100 ms 2D PDSD spectrum, cross peaks with aromatic side chains of F56 and F101 were observed at the Cδ shifts of I71 and I103 (Figure 5.3D).

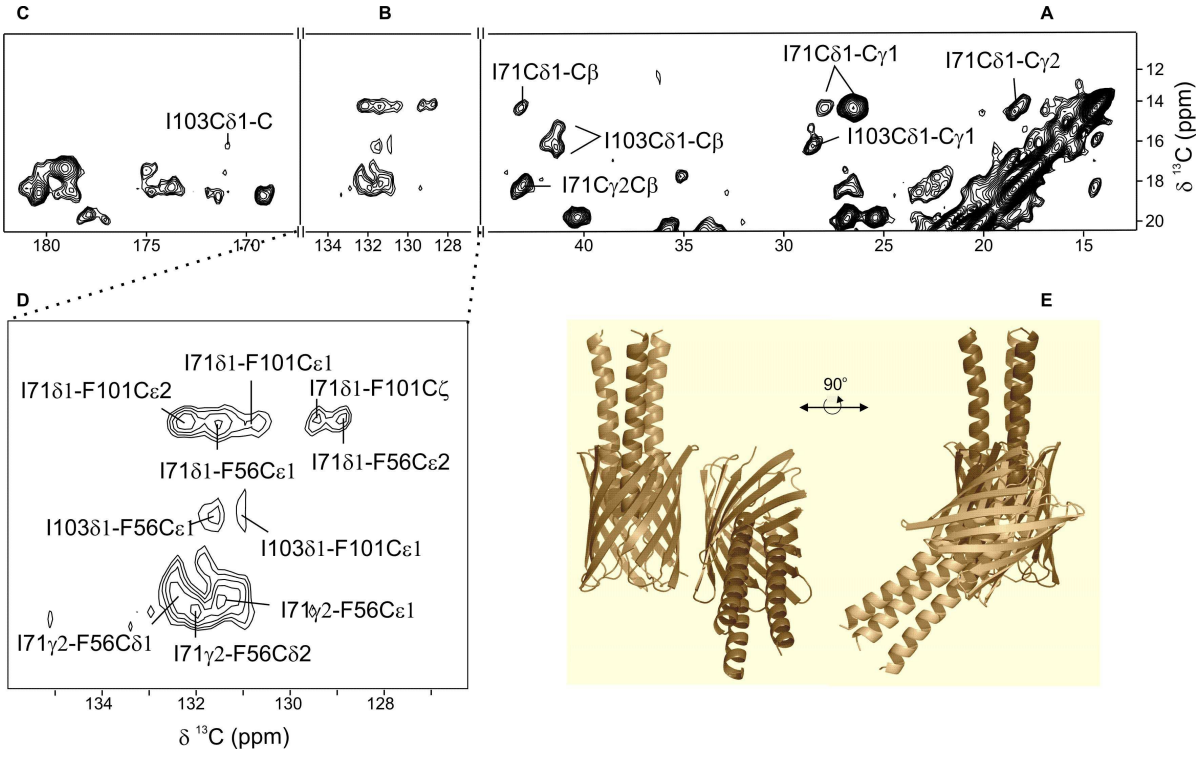


Figure 5.3 – Crystal contacts in microcrystalline YadA-M. (**A, B, C**) Contour plots showing the aliphatic, aromatic and carbonyl regions of a 2D PDSD 100 ms spectrum, respectively. The Cδ shifts of both isoleucines are the most upfield shifted signals and show peak multiplicity (cf. strip A). Their correlation with the carbonyl signal is either missing (I71Cδ1-C`) or appears as a weak signal (I103Cδ1-C`) (strip C). However, strong correlations with the aromatic side chains of F56 and F101 are present at Cδ shifts of I71 and I103 (strip B). (**D**) A close-up of the aromatic part of the spectrum, showing the most probable crystal contacts. (**E**) Organizational model of two YadA-M trimers in the crystal docking obtained by HADDOCK.

As shown in the spectrum, these peaks are relatively strong and imply a close interaction (~4 Å). However, close inspection of the YadA-M structure discloses that there is no aromatic side chain within an immediate vicinity of ~8 Å of both isoleucines. In addition, for the applied mixing time, the magnetization exchange is relatively short range, as demonstrated by either the absence (for I71) or the only weak signal intensity (for I103) of the intra-residual Cδ1-C` cross peak (cf. strip C). This strongly suggests that the strong, unassigned correlations are a result of intermolecular (inter-trimer) contacts. Using these peaks, the contact interface between two trimers was modelled with HADDOCK (Dominguez, Boelens et al. 2003; De Vries, van Dijk et al. 2007). The most reasonable arrangement where the intermolecular restraints could be satisfied is illustrated in Figure 5.3E. The main axes of two neighbouring molecules make a relative angle of ~130° and are stacked in an up-and-down organization. The interface is made up mostly of exposed hydrophobic residues on the membrane-facing side of the barrel. The internal symmetry of the trimer requires each YadA molecule in the crystal to be surrounded by three other trimers. No clashes were observed in the reconstruction of ten adjacent YadA-M molecules in the putative lattice.

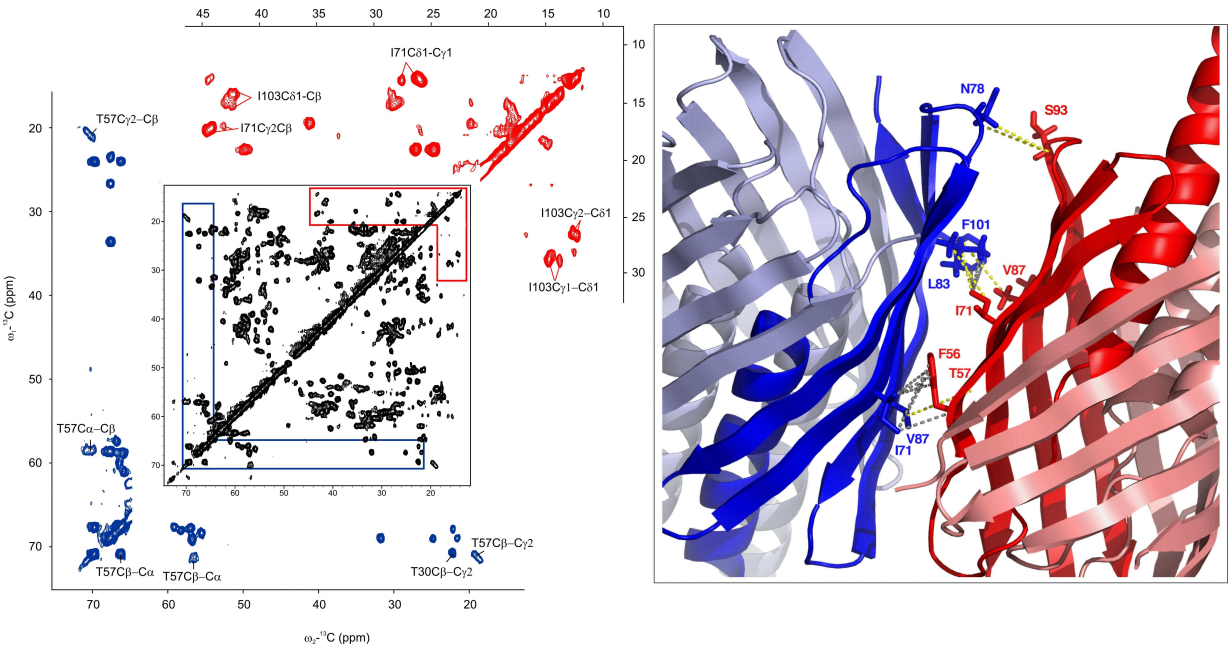


Figure 5.4 – Crystal contacts in microcrystalline YadA-M, and peak multiplicity of I71, I103 and T57. On left side a contour plot taken from a 2D ^{13}C DARR experiment with 25 ms mixing is shown. Red and marine-blue boxes are enlarged to show heterogeneity of peaks for isoleucines and T57, respectively. On the right side of the diagram observed intermolecular interactions are depicted. Also HADDOCK assigned crystal contacts are shown which were rejected during the structure calculation (e.g., from residues N78, S93, T57, V87 and L83).

There were additional unassigned peaks in the long-range mixing time spectra, which too could be attributed to the inter-trimer contacts, however, they were in a more congested region and were left unassigned. Peak multiplicity was observed for I71, I103 and T57 (Figure 5.4, left). Crystal contacts may trap isoleucines into defined but multiple conformations which can explain the peak multiplicity. For T57, it was initially thought that its peak multiplicity is because of the surrounding phenylalanine rings (i.e., F46, F56 and F101). The ring flipping may cause a heterogenous electronic environment for T57. This cage of aromatic rings contributes to the unusual upfield shift for T57. However, some of the cross peaks rejected during the calculation of the trimer could be later assigned as crystal contacts; surprisingly, peaks originating from crystal contacts, involving T57 were among them (Figure 5.4, right), supporting the assumed role of crystal-packing effects in the observed peak multiplicity. Moreover, very weak signals observed for V87 (cf. Section 3.4.1, Figure 3.6) can also be due to the heterogeneity of peaks. The assumption that peak multiplicity is the consequence of crystalline contacts can explain the results but also needs further experimental support. Nevertheless, the work reported here provides preliminary evidence that solid-state NMR is capable of reflecting the crystal-packing information and that the residues involved in these interactions may show heterogeneous spectral peaks.

5.3 Role of glycines in YadA-M

Glycines, because of their small size are frequently found in flexible loops and in β-turns. In YadA-M there are 12 glycines out of which six are present in the β-sheet, five in loop regions (two in the strep-tag), and one in the α-helix.

YadA-M	other TAAs
G59	A, N
G61	A, S
G72	A, S, T, N
G74	A, S, I, Q
G86	S, N, Q
G62 (membrane)	T, N, L, I, R, F, Y

Table 5.1 – The glycines in the β-sheet of YadA-M showing the degree of conservation for small side-chain residues. The table illustrates that while lumen-facing glycines are replaced with small side-chain residues in other TAA members, the membrane-facing G62 shows no preference for such conservation.

All except one of the β-sheet glycines are facing the pore lumen and thus offer minimum hindrance to the passenger domain. The lumen-facing glycine residues offer grooves for hydrophobic side chains to fit into, and stabilize the tertiary structure (see Appendix I for all interactions of glycines with other residues). These five glycines are conserved and replaced only with relatively small side-chain residues (e.g., Ala, Ser, Asn) in other TAAs. On the other hand, the single membrane-facing glycine (G62) shows no preference for such conservation and is replaced with much bulkier residues (R, F, and Y) in other TAAs. This supports the hypothesis that glycines in YadA-M are preferred in the beta sheet to offer least interference to the passenger domain. This is illustrated in Table 5.1 which was based on the sequence alignment (cf. Figure 1.3, Chapter 1) in reference (Grosskinsky, Schutz et al. 2007).

5.3.1 Significance of highly conserved G72

G72 is the most highly conserved amino acid among all glycines in YadA-M. Series of mutations of G72 with residues with longer side chain and polar amino-acid residues have shown that G72 enjoys extraordinary significance in the autotransport and stability of TAAs (Grosskinsky, Schutz et al. 2007). From the solid-state MAS NMR structure of YadA-M, it appears that methyl groups of L45, another highly conserved residue, are pointing towards G72 (cf. Figure 5.5).

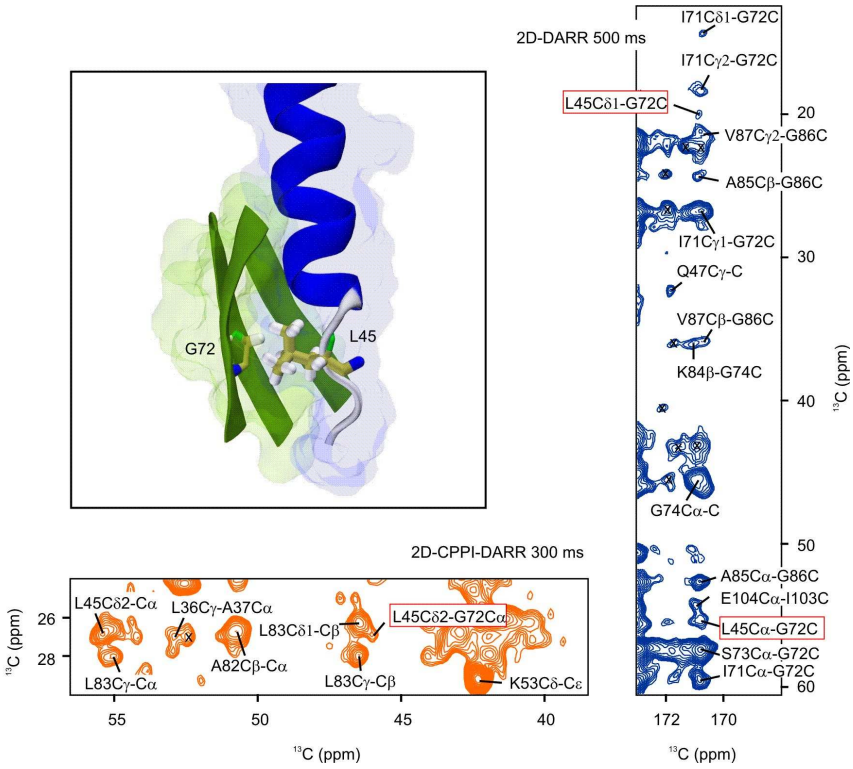


Figure 5.5 – Interactions between L45 and G72. 2D Contour plots from a 2D DARR 500 ms (marine blue) and 2D-CPPI-DARR 300 ms (orange) spectrum, showing long-range hydrophobic contacts between the highly conserved residues L45 and G72. L45 lies in the loop region and needs to interact with barrel wall to form a stable conformation (see section 5.4).

In the crystal structure of Hia, a similar leucine is facing the conserved glycine residue. From the fact that L45 shows strong preference for conserved hydrophobicity (cf. Appendix III), it is tempting to assume that G72 is involved in interactions with a hydrophobic residue (L45 in case of YadA-M) from the N-terminal helix, thus conferring extra stability to the sheet-helix alignment. Point mutations of G72 with polar residues break the hydrophobic interactions, obstruct the formation of a stable trimer and hinder the autotransport in YadA. The long-range spectral correlations between G72 and L45 in YadA-M are shown in Figure 5.5.

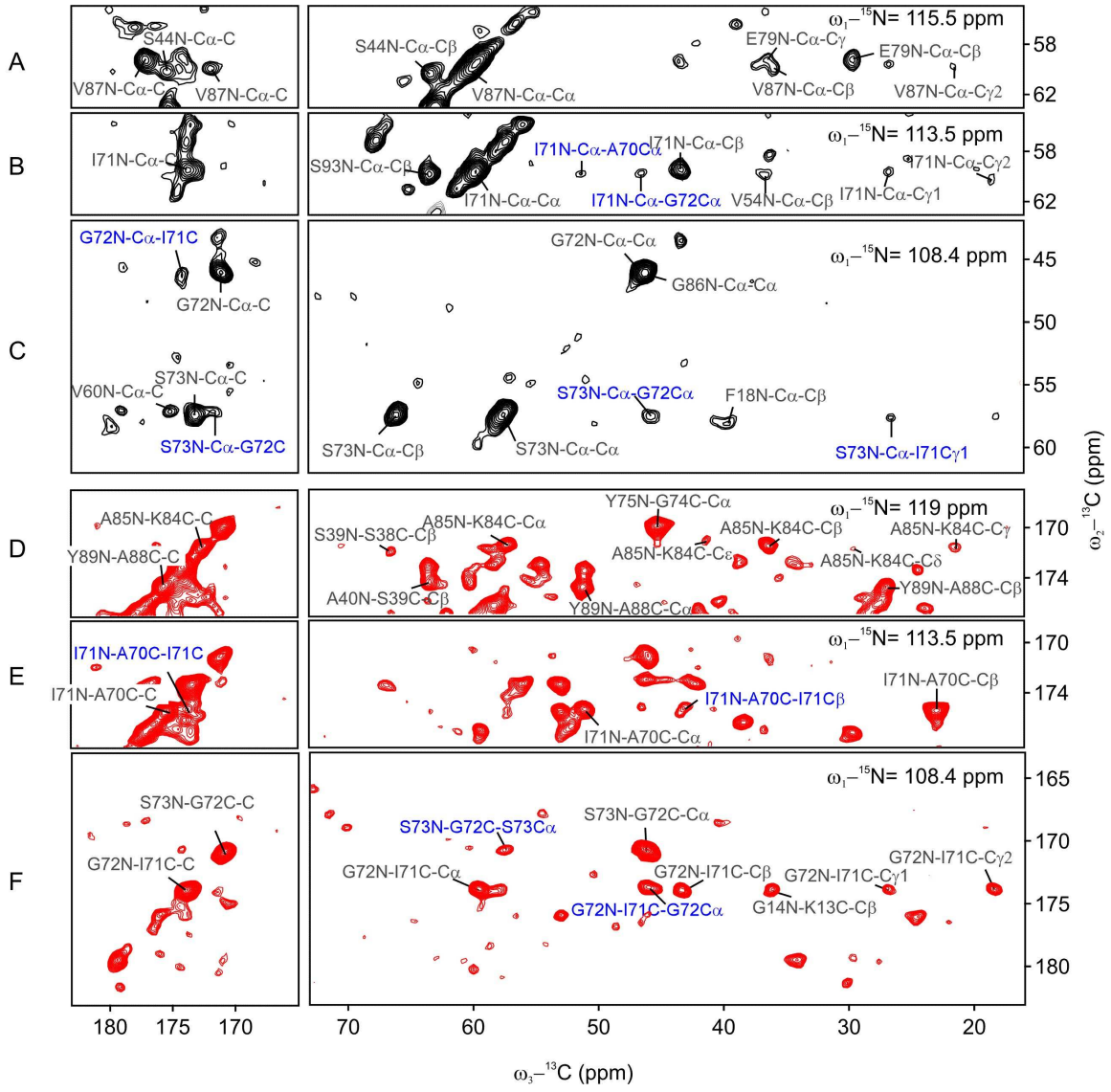


Figure 5.6 – 2D strips of 3D NCACX and 3D NCOCX recorded with a 35 ms PDSD mixing scheme, showing the sequential correlation of I71-G72-S73. Panels A, B, C are from a 3D NCACX spectrum; panels D, E, F from a 3D NCOCX spectrum. Peaks with blue labels represent sequential cross peaks. The amide nitrogen shifts at which the 2D strips are extracted are shown at the top right of each panel. For the used mixing time, strong sequential correlations were not observed for other residues.

From the NMR data, another significant feature for G72 was observed. Contrary to other residues, G72 and its sequential neighbours (i.e., I71 and S73) display much stronger residue-specific and sequential cross peaks, reflecting that this triplet (^{71}IGS^{73}) possesses a comparative low degree of flexibility. Solid-state NMR support for this rigidity is demonstrated in Figure 5.6, which shows different strips from 3D NCACX (black contours) and NCOCX (red contours) spectra recorded with 35 ms PDSD mixing; intra-residual and sequential assignments are labelled in black and blue, respectively, for I71, G72 and S73. In panels B, C, E and F several sequential contacts within the triplet are shown. From the structure of YadA-M, it is evident that some of these transfers are over relatively large distances (e.g., the distance between S73NCα-I71Cγ1 (panel C) is ~6.7 Å). Such strong residue-specific and sequential peaks were not observed for other residues in the same data. For comparison, panel A and D are shown from the same NCACX and NCOCX spectra, respectively. In panel A, cross peaks for V87 are shown, a residue which is located exactly above I71 in the β-sheet and it is reasonable to expect it having a similar degree of rigidity. However, there are no sequential correlations for this residue, even the intra-residue cross peaks are weak in intensity (Cγ2 cannot be seen in the spectrum). This analysis supports the assumption that the IGS triplet forms a relatively rigid part of the β-sheet in YadA-M. The rigidity of the conserved region may contribute to stronger G72-L45 interaction.

5.4 ASSA region

As already mentioned in chapter 3, cross peaks for A37 did not show up in 2D and 3D solid-state MAS NMR spectra recorded with short mixing times. Moreover, the observed chemical shifts for the alpha-helical residues S38 and S39 suggest a random-coil conformation. Another interesting feature is the high degree of conservation for A37 and A41. Henceforth, this stretch of the α-helix (i.e., ^{37}ASSA^{40}) is referred to as the "ASSA" region. The order parameters (S^2) and secondary structure of YadA-M were predicted from the backbone chemical shifts using TALOS+ (Shen, Delaglio et al. 2009). Low order parameters were observed for the loop region (i.e., the region between N-terminal helix and β1-strand), β-turns (short loops connecting the β-strands) and the ASSA region of YadA-M (Figure 5.7A). In addition, TALOS+ was not able to predict the secondary structure for the first three residues of the ASSA region with great confidence (green vertical bars in Figure 5.7B).

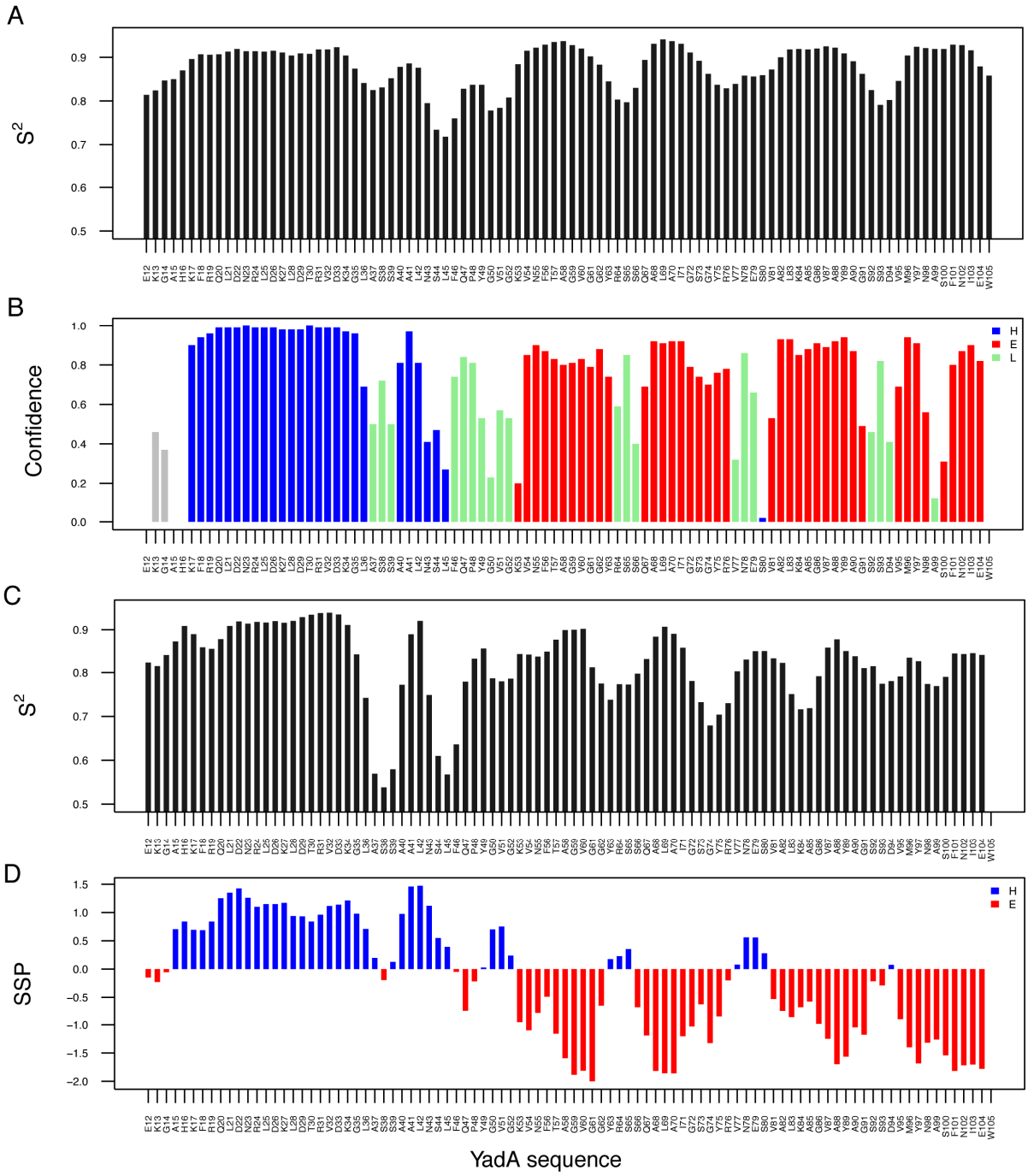


Figure 5.7 – Secondary structure propensities and order parameters for YadA-M as predicted from the backbone ^{13}C and ^{15}N chemical shifts. (A) Order parameter and **(B)** secondary structure prediction confidence reported by TALOS+. The order parameter for loops is generally lower than those from structured regions. The results are in good agreement except for the helical ASSA region for which a lower order parameter (panel A), and low confidence for secondary structure prediction (panel B) is observed. **(C)** The results from RCI which show a pronounced decrease in the order parameter for ASSA region. **(D)** The results for secondary structure propensity calculated by SSP, where red bars represent the sheet and blue bars, the helix. Again, a markedly low helix propensity for the ASSA region is predicted. Indeed, it shows ASSA to be a helix-breaker.

The Random Coil Index (RCI) analysis (Berjanskii and Wishart 2005) provides an empirical correlation between the secondary chemical shifts and flexibility of a protein. Using the backbone chemical shifts, RCI calculates the random coil index from which model-free order parameter and RMSF of residues can be estimated. RCI predicted high random-coil values for the ASSA region, whilst the predicted order parameter was very low; in fact, the ASSA region witnessed the lowest order parameter in YadA-M (Figure 5.7C). Furthermore, a low helical propensity for this region was detected by the secondary structure propensity (SSP) analysis (Marsh, Singh et al. 2006) (Figure 5.7D). All of these observations are pointing into the same direction: the ASSA region is either a helix-breaker or a part of helix with high flexibility, or both. The latter was further backed up by markedly reduced Cα-Cβ cross-peak intensities in 2D ^{13}C-^{13}C DARR spectra of YadA-M (Figure 5.8B).

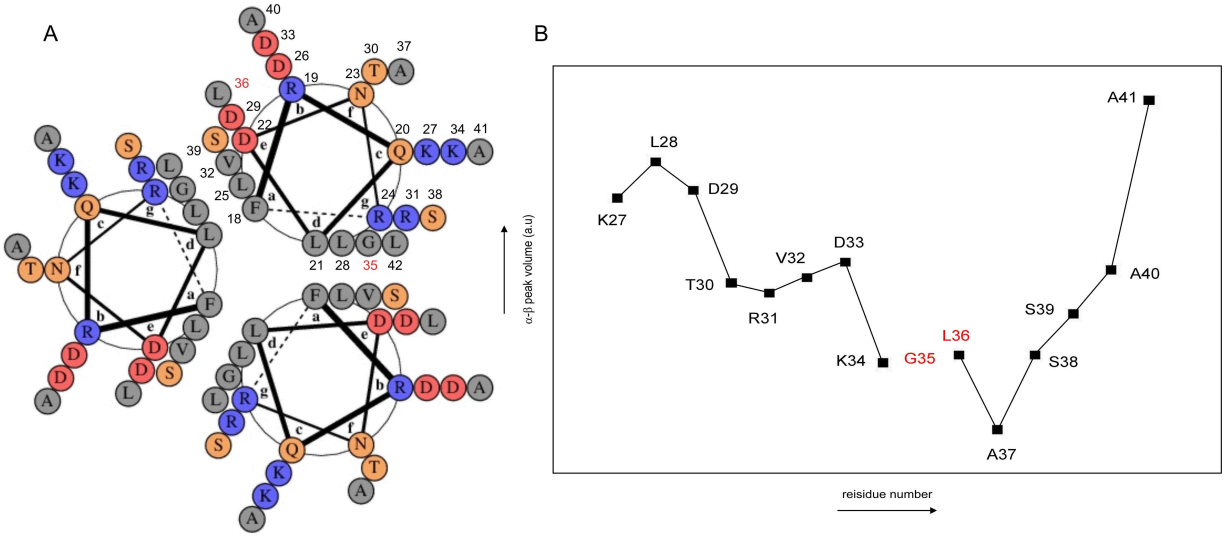


Figure 5.8 – Coiled-coil pattern for trimeric YadA-M and drop in peak volume for the ASSA region. **(A)** Coiled-coil prediction for N-terminal α-helix of YadA-M. The helical wheel diagram was drawn by DrawCoil 1.0. The residues at *a* and *d* positions are generally hydrophobic, which is also here the case, except for G35 (highlighted with red number). Different types of residues are coloured differently, i.e., grey (hydrophobic except G35), pink (negatively charged), blue (positively charged) and orange (polar). **(B)** Integrated intensities of alpha-beta cross peaks for residues K27-A41 in the α-helix, plotted against residue number. A gradual intensity drop is observed in the region around A37, which corroborates an increased flexibility for this stretch of the α-helix. Note that G35 does not have a beta carbon and has been left out of the plot. The Cα-Cβ correlation for A37 was missing in 2D DARR 25 ms spectrum, therefore a zero value was used for this peak.

In cross-polarization based experiments, the peak volume for mobile residues is generally lower; indeed, weaker cross-peak intensities were observed for residues in the loop regions of YadA-M. A graph for peak volume plotted against residue number is shown for alpha-helical residues K27 to A41 in Figure 5.8B. This gradual decrease of peak volume in the ASSA region is another indicator of increased flexibility in this region.

YadA has a coiled-coil stalk where the helices are inter-wound with each other in a heptad repeat fashion. A prediction for the coiled-coil for the YadA-M helix (from residues F18–L36) is shown in Figure 5.8A. The residues in positions *a* and *d* are hydrophobic (coloured grey) and make the core of the coiled-coil. It is important to note that A37 and A41 are the only 'hydrophobic residues' in the α-helix whose side chains are pointing towards the β-sheet; all other hydrophobic side chains are facing the core of the coiled-coil (cf. Figure 5.8A). According to the prediction, amino-acid residues from F18-K34 show a classical coiled-coil pattern which presumably is breaking just before ASSA region as there should be a hydrophobic residue in place of G35. Interestingly, G35 is the *only* glycine residue present in the N-terminal α-helix (out of a total of 12 glycines in YadA-M sequence) and is located in an unexpected position in the coiled-coil (the *d* position), which also points to a special function of the ASSA region.

5.4.1 Evolutionary analysis of YadA-M

Coiled-coil and evolutionary studies for YadA-M were carried out in close collaboration with the Max Planck Institute for Developmental Biology, Tübingen. An alignment for several members of TAA family with 70% sequence identity to YadA-M was produced. According to the alignment results, the most conserved residues were G35, A37, L42 and G72 (cf. Appendix III). Analysis of hydrophobicity of the TAA members was also part of the alignment studies. Socket server (Walshaw and Woolfson 2001) was used to analyze the 3D structures of YadA-M (PDB ID code 2LME) and Hia (PDB ID code 2GR7). The results are shown in Appendix VA; according to these studies, the part of the helix following residue G35 stretches and adopts an almost extended conformation. Instead of G35 which should be part of the next *d* layer, L36 (highlighted with red number in Figure 5.8A) shows knobs-into-holes packing (cf. Appendix VB). The results suggest that in both YadA-M and Hia, there is a switch from seven-residue (heptad) to an eleven-residue (hendecad) repeat (residues L25, L28, V32 and L36 make the core layers of hendecad repeat in YadA-M). Moreover, L42,

despite being hydrophobic and located in the *d* position for canonical heptad repeat, is no longer part of the heptad repeat and forms contacts to the barrel wall. In addition, L42 showed somewhat unusual cross-peak intensities noticed in the NMR spectra; it showed strong backbone Cα and C` peaks, whereas the side-chain peaks were very weak, hidden in the noise region. Strong conservation of hydrophobic residues at position *a* and *d* (i.e., F18, L21, L25, L28 and V32) was observed throughout the TAAs family. This indicates that all members of the family show canonical coiled-coil structure in this region. Surprisingly, the residue at the putative *d* position of G35 shows no conservation for hydrophobicity, rather and in accord with YadA, the next residue (i.e., L36) shows absolute conservation for hydrophobicity. The results for hydrophobicity studies are shown in Appendix IV and further corroborate the assumption that heptad repeat is followed by a hendecad repeat. The fact that residues G35 and S39 (which would occupy a *d* and *a* position in an otherwise heptad repeat) show no preference for hydrophobicity conservation, strongly suggests that a switch from 7-repeat to an 11-repeat, which ends at L36 right before the ASSA region, is conserved across all TAAs. Evolutionary analysis also revealed that the residues facing the barrel wall tend to be small in all TAAs. Small side-chain and a low helix propensity was also a characteristic feature of ASSA region (cf. Appendix IV). Finally, a comparison was made between the solid-state MAS NMR structure of the YadA-M and the recently solved crystal structures of the stalk domain of YadA (Alvarez, Gruber et al. 2010). Hernandez Alvarez et al., (2010) studied several constructs of the YadA stalk fragments that included the ASSA region. The authors could not resolve the C-terminal part of the stalk (where the ASSA region is located) and introduced mutations in this region to overcome the low helix propensity and to force this region into a heptad packing. In appendix VC, the superimposed structures of the coiled-coil of YadA-M and crystal structure of the stalk domain of YadA are shown. Even in the mutant constructs, the structure could not be resolved for all residues; the polypeptide chain did not build beyond the ASSA region. The coiled-coil structure of YadA-M superimposes well on to the stalk constructs except in the ASSA region. The heterogeneity in the structure of the YadA stalk substantiates the finding that the heptad repeat is lost after G35, and that the ASSA and residues beyond (i.e., L42, L45) need to interact with the barrel wall in order to adopt a stable conformation. Indeed, several solid-state MAS NMR interactions of these residues with the barrel inner wall were found (cf. sections 5.31, 5.42, and 5.43).

5.4.2 Interactions of ASSA residues with barrel interior

Contacts between the stretch of small helical residues in ASSA region with the ring of small barrel residues are corroborated by NMR correlations between A37-A68 and A41-G61. This is illustrated in Figure 5.9. The panels (A-D) show 2D contour plots of aliphatic and carbonyl regions extracted from a 2D ^{13}C-^{13}C DARR experiment recorded with 300 ms mixing on YadA-M. Transfers involving conserved small residues in the α-helix and β-sheet are highlighted with red boxes. Mutual transfers between residues A68 and G61 in adjacent β-strands are also highlighted with red boxes. Strips A and C show the Cα and C′ regions with correlations for alanines in the β-sheets. Strip B shows sequential and long-range correlations involving the Cβ of alanines, strip D shows a correlation between the Cα of G61 and the C′ of A41.

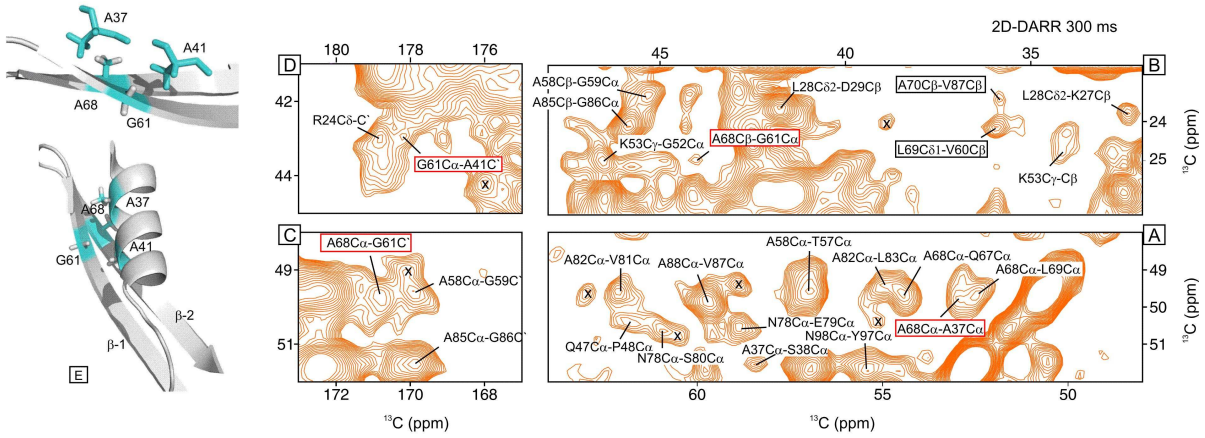


Figure 5.9 – Interactions between conserved residues. Regions extracted from a 2D DARR experiment recorded with 300 ms mixing on YadA-M (A-D, right). Long-range transfers involving conserved small residues in the α-helix and β-sheet, and between G61 (β1) and A68 (β2) are highlighted with red boxes. The conserved residues A37, A41, A68 and G61 are coloured in cyan in the ribbon model of YadA-M (E, left).

The conserved residues A37, A41, A68 and G61 are colour-coded in cyan in the ribbon model of YadA-M, which is shown in the left part of the Figure 5.9. For clarity, the α-helix has been left out in the Figure on the top, to obtain a better view of the position of A37 and A41 relative to G61 and A68 in β-strands 1 and 2, respectively. These and the L45-G72 interactions (cf. section 5.3.1) support the finding that a break in heptad repeat leads to interaction of α-helical residues with the interior of the barrel.

5.4.3 Non-covalent S···O interactions

From the solid-state MAS NMR structure of YadA-M, it was found that the sulphur atom of M96 in strand β4 is close to the oxygen atoms of the helical ASSA serines, i.e, S38-S39 (S···O distances ~3.5 Å, see Figure 5.10). The divalent electrophilic sulphur of M96 could in principle form a non-covalent interaction with the nucleophilic side-chain hydroxyl oxygens from S38 and S39 (see below). Such an interaction would explain why no degradation of the YadA-M sample was observed by monitoring the (single) methionine, since the "pseudo-oxidation" of the sulphur by the serine oxygens protects it from oxidation (cf. Chapter 3, Figure 3.1B).

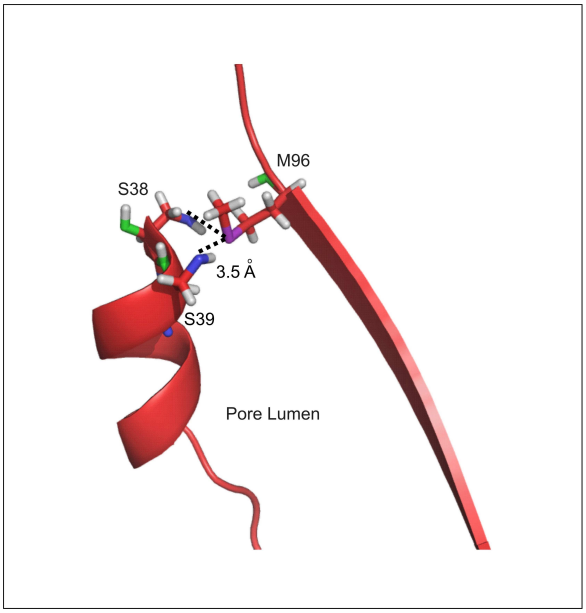


Figure 5.10

Intramolecular S···O interaction between the sulphur of M96 in the β-sheet and side-chain oxygens of S38 and S39 in the α-helix. Distances between O and S are about 3.5 Å. The long side chain of M96 in the pore-lumen can bridge the gap between the sheet and the helix, and can form a non-covalent S···O interaction. The amino-acid residues involved in this interaction lie on the extracellular side, at the edge of the barrel.

There are only six lumen-facing hydrophobic residues in the β-sheet of YadA-M. Except methionine all others are alanines that partake in hydrophobic interactions with the helix whilst giving minimum hinderance to the passenger domain. On the other hand, methionine has a relatively bulky side chain, and its presence seems a non-obvious choice of nature as it would occupy more space and impede the translocation of passenger domain, *unless* it has an important function inside the pore-lumen. It is suggested that S···O interaction stabilizes the β-sheet and α-helix tertiary structure of YadA-M.

X-ray crystallographic analyses have shown that in their crystallographic structures, sulfoxide complexes show intramolecular non-covalent S···O close interactions. Moreover, intermolecular S···O interactions were reported between sulfoxides and amides in solutions

(Nagao, Miyamoto et al. 2006). Similar intra- and intermolecular interactions involving the divalent sulphur of methionine (Met) residues were characterized in biomolecules (Pal and Chakrabarti 2001). The sulphur atom of methionine is reported to interact with the backbone carbonyl group, the carboxylate side chain or the ring of an aromatic residue. In these interactions, the divalent sulphur acts as an electrophile while the other electron-rich atom behaves as a nucleophile. The distance for such close contacts should be no more than 4 Å and both sulphide and carbonyl planes should have fixed orientations with respect to each other. This study also reported that 3% of Met residues interact with more than one oxygen atoms and that the sulphur atom shows a tendency to simultaneously engage both with oxygen and aromatic atoms. According to another study, there is a small fraction of oxygen or nitrogen atoms in proteins which, despite being buried in the protein core, do not form hydrogen bonds with either the solvent or the protein molecules (McDonald and Thornton 1994). The S···O interactions supply an elegant mechanism that reduces the number of unsatisfied hydrogen bonds.

5.4.4 Three serine pairs in YadA-M

YadA-M has three serine pairs in its primary sequence, i.e., ^{38}SS^{39}, ^{65}SS^{66} and ^{92}SS^{93}. The second and third pairs are part of short turns that connect beta strand 1-2 and 3-4, respectively. The first serine pair, despite being a part of the ^{37}ASSA^{40} helical region, shows random-coil chemical shifts (cf. Figure 4.2C). It is interesting to note that the majority of serines (12 in total) is in the loop regions. The first two (S5, S8) are part of the flexible, unstructured strep-tag (cf. Figure 4.16). The position of S44 is at the N-terminus of alpha helix and functions as an adapter between the relatively rigid helix and flexible loop. S73 and S100 are part of beta strand 2 and 4, respectively. Leaving S38, S39 aside, only 2 out of 10 serines are in the secondary structure. The random-coil chemical shifts for ^{38}SS^{39} and the preference of SS pairs to be part of the loop regions (^{65}SS^{66} and ^{92}SS^{93}) suggest that the serine pair in the ^{37}ASSA^{40} region has two functions, i.e., it can form a sharp turn or hairpin, and it can act as a helix-breaker. The assigned chemical shifts for 10 observable serines are shown in the contour plot taken from a 2D ^{13}C-^{13}C DARR experiment with 25 ms mixing (Figure 5.11). The assignments for the three serine pairs are highlighted in green colour. The position of these pairs is shown in green at the left of the diagram, with top and side views of the YadA-M structure.

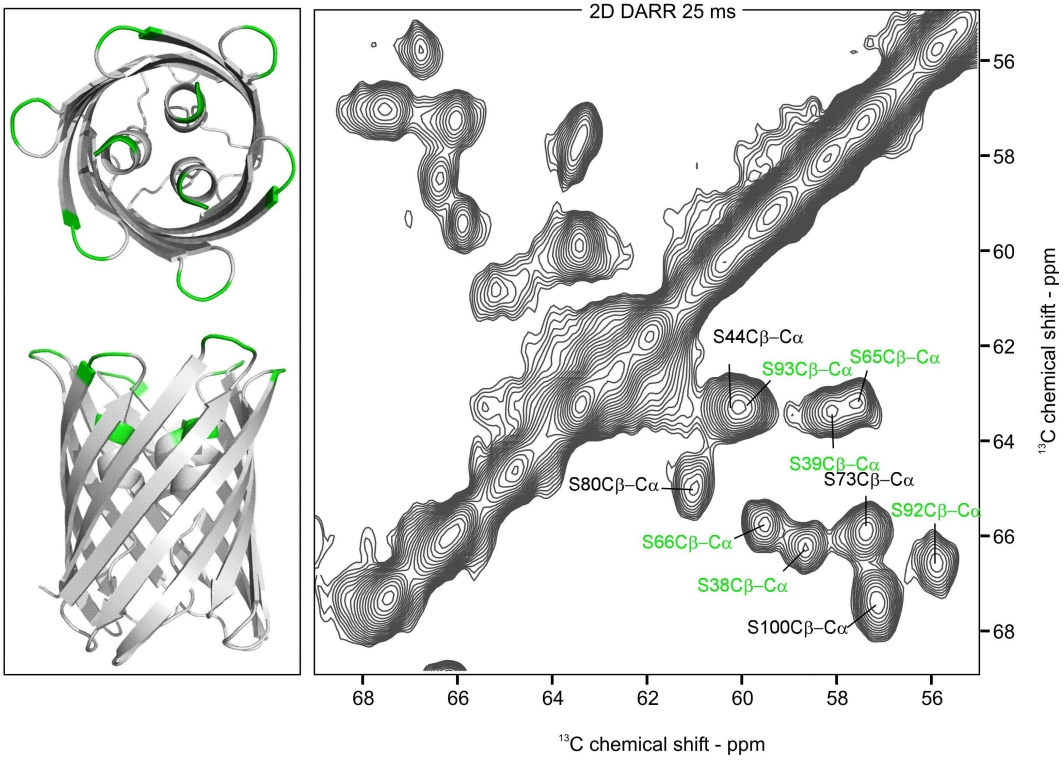


Figure 5.11 – Contour plot of a 2D ^{13}C-^{13}C DARR spectra showing the serine region of the spectrum. All ten observable serines are shown (two serines, i.e., S5 and S8 are in mobile strep-tag). Assignments for serine pairs are highlighted in green. Approximately 80% of serines in YadA-M are in loop regions. In the inset are shown two orientations of YadA-M (coiled-coil region is cut to S38-S39 for clarity). Part of the structure with serine pairs is coloured green.

5.5 Proposed mechanism of autotransport in YadA and TAAs

The hairpin model is one of the several anticipated hypotheses about the autotransport mechanism in TAAs (cf. Section 1.3). The findings from solid-state MAS NMR structural data and evolutionary studies strongly support the hairpin model of autotransport in YadA and other members of the TAA family. The transport of the passenger domain proceeds through a stepwise but concerted manner in which the ASSA region is identified as the core of the hairpin. A pictorial illustration for the suggested mechanism of autotransport in YadA is shown in Figure 5.12. In the first step, the C-terminal domain of YadA gets inserted into the OM and starts folding with the help of the Bam machinery and probably unknown periplasmic chaperones. In monomeric autotransporters, the transport of the passenger domain is thought to proceed via a similar hairpin, which may already be formed during the membrane insertion of the barrel (Ieva, Skillman et al. 2008). Therefore, it is quite possible that the C-terminal domain of YadA may start folding prior to the membrane insertion in order to bring the membrane-facing hydrophobic residues in the required alignment. In the

intermediary stages, the last approximately 68 C-terminal residues from each monomer (i.e., from A37-W105) are in a partially folded state, and the ASSA region makes the kink (highlighted in red colour) allowing the unfolded polypeptide to loop back into the periplasm (step B, Figure 5.12).

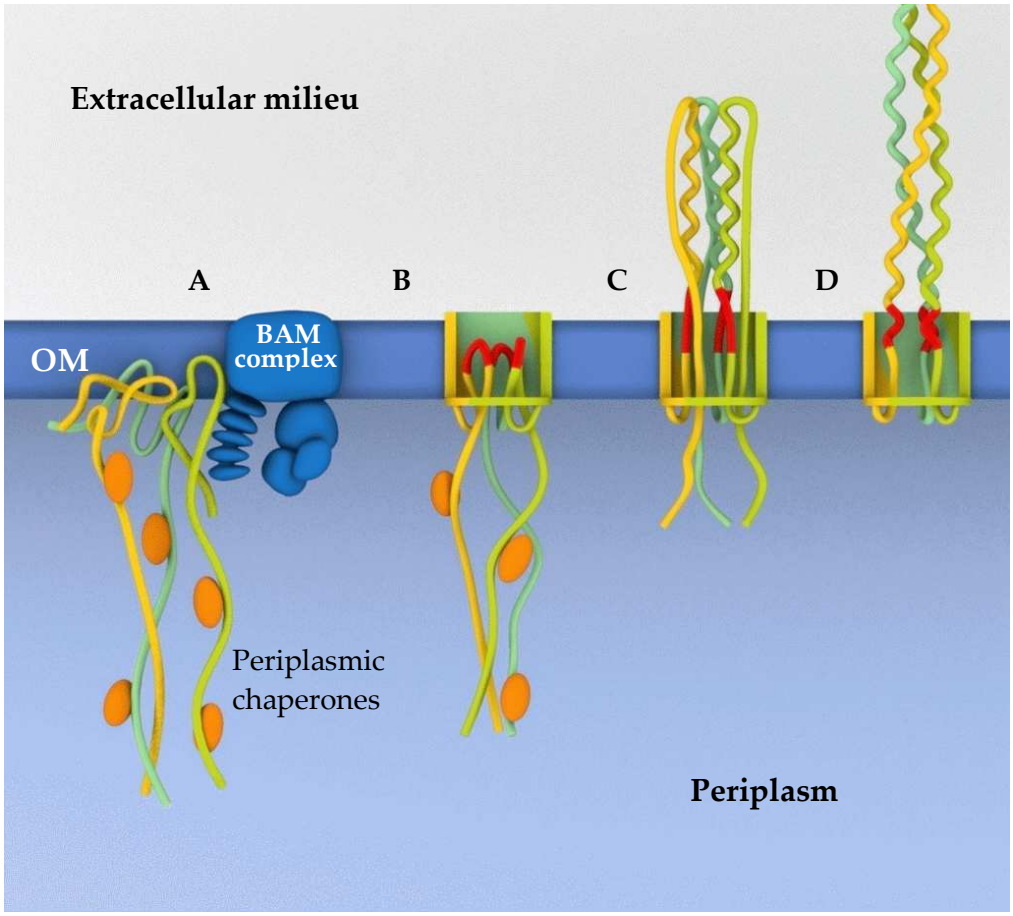


Figure 5.12 – Schematic representation of the proposed mechanism of autotransport in TAAs. (A) The C-terminal domain inserts into the outer membrane which is assisted by Bam complex and periplasmic chaperones. **(B)** An interim, unstable phase showing the barrel and the hairpin (ASSA) region. The hairpins are painted red, while the unfolded chains are hanging in the periplasm. **(C)** Export of the passenger domain with C- to N-terminal directionality; sequential folding of the coiled-coil supplies the required energy. The hairpin segment adopts an extended conformation to allow more room for the moving polypeptide. **(D)** The pore is occluded after the transport is accomplished. The ASSA region forms interaction with the barrel interior residues to settle down. The barrel depicted in steps B and C could be more loose and flexible and may also be unfolded at some places before the transport is complete.

In the following step, the unfolded chains start folding in a C- to N-terminus direction and moving upside into the extracellular environment (Step C). It has been proposed that in classical autotransporters, this sequential folding provides the energy required to drive the export process (Junker, Besingi et al. 2009; Peterson, Tian et al. 2010). Therefore, it is highly

plausible that the huge enthalpy gain during the coiled-coil formation provides the transport energy in TAAs. Additionally, the coiled-coil packing advocates a zipper-like, highly synchronized autotransport mechanism of three protein chains. During the transitional stages, the hairpin region remains in the extended conformation due to its low folding propensity (cf. Step C) and folds only after the transport is complete. This provides more space inside the barrel for the simultaneous export of three unfolded polypeptides. The structure of YadA-M shows that there are small side-chain residues in the interior of the barrel (A, N, G, S, T); this conservation for the lumen-facing residues should be for the very same reason. However, it is assumed that this conservation to match space requirements is more for the intermediate structure, formed during autotransport rather than that of the final structure. The most essential conservation in this respect seems to be for G72, as it is highly conserved across TAAs and mutation studies have shown that even an alanine in this position considerably impairs the autotransport (Grosskinsky, Schutz et al. 2007). Additionally, it is expected that the intermediate barrel is undergoing breathing motions without breaking the hydrogen bonds resulting in a wider pore, as, e.g., in usher pores (Phan, Remaut et al. 2011).

On account of the helical twist of the coiled coil, which typically is in the range of 150-200 Å (Lupas and Gruber 2005), it is thinkable that the unfolded N-terminal chains may entangle with the folding structure during the C-terminal coiled-coil formation. A recent study has revealed a remarkable strategy used by nature to prevent such an entanglement; a transition from right- to left-handed coiled coil has been observed in the region close to the YadA membrane anchor which intelligently resolves the problem of entanglement (Alvarez, Gruber et al. 2010). Similar structural elements that switch from left- to right-handed orientation are witnessed in other TAAs, e.g., the neck and saddle domain of EibD (Leo, Lyskowski et al. 2011).

5.5.1 Evidence of the suggested autotransport model

The proposed model in which the ASSA region forms the hairpin is supported concurrently from solid-state MAS NMR data, the determined YadA-M structure, the evolutionary analysis, the conserved interactions, and the previous studies.

A) Solid-state MAS NMR data

Following observations in solid-state MAS NMR spectra of YadA-M suggest a direct involvement of ASSA region in the autotransport:

i) Increased flexibility of ASSA region as observed by a drop in peak intensity and missing NMR signal for A37.
ii) Low order parameter for ASSA as predicted from the chemical shifts by TALOS+ and RCI.
iii) Non-helical chemical shifts for S38 and S39.

B) The structure of YadA-M

i) The determined structure displays small side-chain residues in the interior of the barrel (i.e., A, N, G, S, T) and the hairpin (ASSA) region.
ii) The preference of serines to be in flexible regions; and of serine pairs to be in loop regions.

C) The evolutionary analysis

i) High degree of conservation for small side-chain residues in ASSA region.
ii) A conserved low helix propensity for ASSA in YadA as determined by SSP.
iii) Loss of coiled-coil packing at G35 which, despite being into a *d* position shows no preference for conserved hydrophobicity. This results into a switch from a classical 7-residue repeat to a straighter 11-residue repeat which ends at ASSA region, a feature conserved for TAAs.

D) The interactions of ASSA and the residues beyond with the interior of the barrel are required in order to overcome the loss of coiled-coil packing. The following interactions of the ASSA region are certainly corroborated from the NMR data:

i) Contacts between A37-A68 and A41-G61. These residues (both in the ASSA and in the barrel) are conserved, suggesting that these interactions are conserved.
ii) Non-covalent S···O interaction involving S38, S39 and M96, which stabilizes the hairpin after the autotransport.
iii) Interaction between highly conserved L45 and G72. The rigid IGS region provides a better L-G interaction.

E) Previous studies have shown that at least 76 C-terminal residues in Hia (Surana, Cutter et al. 2004) and 70 C-terminal residues in YadA (Roggenkamp, Ackermann et al. 2003) are required in order for barrel formation and autotransport to occur. The failure to show trimerization and autotransport in constructs that lack these C-terminal residues (i.e., without ASSA region), confirms the involvement of hairpin region in the autotransport.

5.5.2 Possible explanation of the autotransport mechanism

Most probably, when the last 68 C-terminal residues enter the lipid bilayer, four β-strands and part of the helix (W105-A37, where ^{37}ASSA^{41} forms the kink) are folded simultaneously, leaving the rest part of the passenger domain (i.e., ~L36-G1) still unfolded into the periplasm. The β-barrel formed by the inter-strand hydrogen bonding between three monomers would be in a transition state with non-native long range strand-strand and strand-helix hydrophobic interactions. An unstable, flexible and wider pore is formed due to the presence of folded and unfolded chains. This transition state with high free energy has to be surpassed by every protein during folding progression. The free energy of unfolded and folded states are far less than the intermediate, partially folded states (Sosnick and Barrick 2011). Hence, to attain stability, the passenger domain has to get outside of the cell using energy of folding; or otherwise the whole structure collapses, leading to a failed autotransport. In the latter case, the protein is degraded by DegP (cf. Figure 1.4) (Grosskinsky, Schutz et al. 2007). The bottleneck for a successful autotransport should therefore be the formation of this transition state in which a preliminary barrel and part of the helix, up to the hairpin region are 'structured', whereas the N-terminal polypeptide is still in the unfolded state. The rest of the job is done by ultra-fast readjustment of repetitive hydrophobic and hydrophilic residues in the passenger domain. A careful analysis of the sequence alignment of YadA-M with its other family members indicates that most of the amino-acid residues which are fully or partially conserved, are hydrophobic or neutral in nature. Therefore, the β-sheet should align its lumen-facing hydrophobic residues (e.g., G61, A68, G72) with hydrophobic residues of the helix (e.g., A37, A41, L45). The assumption that conserved residues are involved in autotransport is further backed up by the non-conserved stretch of amino acids from L46-V51 which connects the first β-strand with the helix. It is a flexible loop that functions as non-residue specific place holder, and presumably does not make interactions necessary for the autotransport mechanism. The results reported in this thesis strongly support the hairpin model for the autotransport. Moreover, they provide, on the basis of above mentioned findings a plausible explanation for the fact that in order trimerization and autotransport in YadA to be successful, the last 68 C-terminal residues (i.e., A37-W105) are required and that the process will fail for any shorter construct.

6 Bibliography

Alvarez, B. H., M. Gruber, et al. (2010). "A transition from strong right-handed to canonical left-handed supercoiling in a conserved coiled-coil segment of trimeric autotransporter adhesins." Journal of Structural Biology **170**(2): 236-245.

Arnold, T. and D. Linke (2007). "Phase separation in the isolation and purification of membrane proteins." Biotechniques **43**(4): 427-+.

Baldus, M., A. T. Petkova, et al. (1998). "Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems." Molecular Physics **95**(6): 1197-1207.

Bardiaux, B., A. Bernard, et al. (2009). "Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA." Proteins-Structure Function and Bioinformatics **75**(3): 569-585.

Barnhart, M. M. and M. R. Chapman (2006). "Curli biogenesis and function." Annual Review of Microbiology **60**: 131-147.

Bayrhuber, M., T. Meins, et al. (2008). "Structure of the human voltage-dependent anion channel." Proceedings of the National Academy of Sciences of the United States of America **105**(40): 15370-15375.

Bennett, A. E., C. M. Rienstra, et al. (1995). "Heteronuclear Decoupling in Rotating Solids." Journal of Chemical Physics **103**(16): 6951-6958.

Berjanskii, M. V. and D. S. Wishart (2005). "A simple method to predict protein flexibility using secondary chemical shifts." Journal of the American Chemical Society **127**(43): 14970-14971.

Berjanskii, M. V. and D. S. Wishart (2005). "A simple method to predict protein flexibility using secondary chemical shifts." Journal of the American Chemical Society **127**(43): 14970-14971.

Bernstein, H. D. (2007). "Are bacterial 'autotransporters' really transporters?" Trends in Microbiology **15**(10): 441-447.

Bingle, L. E. H., C. M. Bailey, et al. (2008). "Type VI secretion: a beginner's guide." Current Opinion in Microbiology **11**(1): 3-8.

Bolin, I., L. Norlander, et al. (1982). "Temperature-Inducible Outer-Membrane Protein of Yersinia-Pseudotuberculosis and Yersinia-Enterocolitica Is Associated with the Virulence Plasmid." Infection and Immunity **37**(2): 506-512.

Brunger, A. T. and M. Nilges (1993). "Computational Challenges for Macromolecular Structure Determination by X-Ray Crystallography and Solution Nmr-Spectroscopy." Quarterly Reviews of Biophysics **26**(1): 49-125.

Burley, S. K. and G. A. Petsko (1986). "Amino-Aromatic Interactions in Proteins." Febs Letters **203**(2): 139-143.

Cascales, E. and P. J. Christie (2003). "The versatile bacterial type IV secretion systems." Nature Reviews Microbiology **1**(2): 137-149.

Castellani, F., B. J. van Rossum, et al. (2003). "Determination of solid-state NMR structures of proteins by means of three-dimensional 15N-13C-13C dipolar correlation spectroscopy and chemical shift analysis." Biochemistry **42**(39): 11476-11483.

China, B., M. P. Sory, et al. (1993). "Role of the YadA protein in prevention of opsonization of Yersinia enterocolitica by C3b molecules." Infect Immun **61**(8): 3129-3136.

Chou, K. C., L. Carlacci, et al. (1990). "Conformational and Geometrical Properties of Idealized Beta-Barrels in Proteins." Journal of Molecular Biology **213**(2): 315-326.

Cornelis, G. R. and F. Van Gijsegem (2000). "Assembly and function of type III secretory systems." Annual Review of Microbiology **54**: 735-774.

Cover, T. L. and R. C. Aber (1989). "Yersinia-Enterocolitica." New England Journal of Medicine **321**(1): 16-24.

De Paepe, G., J. R. Lewandowski, et al. (2008). "Proton assisted recoupling and protein structure determination." J Chem Phys **129**(24): 245101.

De Vries, S. J., A. D. J. van Dijk, et al. (2007). "HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets." Proteins-Structure Function and Bioinformatics **69**(4): 726-733.

Delepelaire, P. (2004). "Type I secretion in gram-negative bacteria." Biochimica Et Biophysica Acta-Molecular Cell Research **1694**(1-3): 149-161.

Dominguez, C., R. Boelens, et al. (2003). "HADDOCK: A protein-protein docking approach based on biochemical or biophysical information." Journal of the American Chemical Society **125**(7): 1731-1737.

El Tahir, Y. and M. Skurnik (2001). "YadA, the multifaceted Yersinia adhesin." International Journal of Medical Microbiology **291**(3): 209-218.

Foster, T. J. and M. Hook (1998). "Surface protein adhesins of Staphylococcus aureus." Trends in Microbiology **6**(12): 484-488.

Ginzinger, S. W. and J. Fischer (2006). "SimShift: Identifying structural similarities from NMR chemical shifts." Bioinformatics **22**(4): 460-465.

Goodsell, D. S. (1996). "Our Molecular Nature." 3-7.

Grosskinsky, U., M. Schutz, et al. (2007). "A conserved glycine residue of trimeric autotransporter domains plays a key role in Yersinia adhesin A autotransport." J Bacteriol **189**(24): 9011-9019.

Gullion, T. and J. Schaefer (1989). "Rotational-Echo Double-Resonance Nmr." Journal of Magnetic Resonance **81**(1): 196-200.

Guntert, P., C. Mumenthaler, et al. (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." Journal of Molecular Biology **273**(1): 283-298.

Habeck, M. (2011). "Statistical mechanics analysis of sparse data." Journal of Structural Biology **173**(3): 541-548.

Habeck, M. (2012). "Inferential Structure Determination from NMR data (a book chapter from) Bayesian Methods in Structural Bioinformatics.": 287-309.

Habeck, M., M. Nilges, et al. (2005). "Bayesian inference applied to macromolecular structure determination." Physical Review E **72**(3).

Habeck, M., M. Nilges, et al. (2005). "Replica-exchange Monte Carlo scheme for Bayesian data analysis." Physical Review Letters **94**(1).

Habeck, M., W. Rieping, et al. (2006). "Weighting of experimental evidence in macromolecular structure determination." Proceedings of the National Academy of Sciences of the United States of America **103**(6): 1756-1761.

Hediger, S., B. H. Meier, et al. (1994). "Nmr Cross-Polarization by Adiabatic Passage through the Hartmann-Hahn Condition (Aphh)." Chemical Physics Letters **223**(4): 283-288.

Henderson, I. R., F. Navarro-Garcia, et al. (2004). "Type V protein secretion pathway: the autotransporter story." Microbiol Mol Biol Rev **68**(4): 692-744.

Hing, A. W., S. Vega, et al. (1992). "Transferred-Echo Double-Resonance Nmr." Journal of Magnetic Resonance **96**(1): 205-209.

Hirsch, M. and M. Habeck (2008). "Mixture models for protein structure ensembles." Bioinformatics **24**(19): 2184-2192.

Hohwy, M., H. J. Jakobsen, et al. (1998). "Broadband dipolar recoupling in the nuclear magnetic resonance of rotating solids: A compensated C7 pulse sequence." Journal of Chemical Physics **108**(7): 2686-2694.

Hoiczyk, E., A. Roggenkamp, et al. (2000). "Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins." Embo Journal **19**(22): 5989-5999.

Hubner, C. A. and T. J. Jentsch (2002). "Ion channel diseases." Human Molecular Genetics **11**(20): 2435-2445.

Jaroniec, C. P., C. Filip, et al. (2002). "3D TEDOR NMR experiments for the simultaneous measurement of multiple carbon-nitrogen distances in uniformly (13)C,(15)N-labeled solids." Journal of the American Chemical Society **124**(36): 10728-10742.

Jaroniec, C. P., B. A. Tounge, et al. (2001). "Frequency selective heteronuclear dipolar recoupling in rotating solids: accurate (13)C-(15)N distance measurements in uniformly (13)C,(15)N-labeled peptides." Journal of the American Chemical Society **123**(15): 3507-3519.

Jehle, S., M. Hiller, et al. (2006). "Spectral editing: selection of methyl groups in multidimensional solid-state magic-angle spinning NMR." Journal of Biomolecular Nmr **36**(3): 169-177.

Judge, P. J. and A. Watts (2011). "Recent contributions from solid-state NMR to the understanding of membrane protein structure and function." Curr Opin Chem Biol **15**(5): 690-695.

Junker, M., R. N. Besingi, et al. (2009). "Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion." Molecular Microbiology **71**(5): 1323-1332.

Kabsch, W. and C. Sander (1983). "Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features." Biopolymers **22**(12): 2577-2637.

Kapperud, G. and J. Lassen (1983). "Relationship of virulence-associated autoagglutination to hemagglutinin production in Yersinia enterocolitica and Yersinia enterocolitica-like bacteria." Infect Immun **42**(1): 163-169.

Koretke, K. K., P. Szczesny, et al. (2006). "Model structure of the prototypical non-fimbrial adhesin YadA of Yersinia enterocolitica." J Struct Biol **155**(2): 154-161.

Kuhlman, B., G. Dantas, et al. (2003). "Design of a novel globular protein fold with atomic-level accuracy." Science **302**(5649): 1364-1368.

Kyte, J. and R. F. Doolittle (1982). "A Simple Method for Displaying the Hydropathic Character of a Protein." Journal of Molecular Biology **157**(1): 105-132.

Laird, W. J. and D. C. Cavanaugh (1980). "Correlation of autoagglutination and virulence of yersiniae." J Clin Microbiol **11**(4): 430-432.

Lange, A., S. Becker, et al. (2005). "A concept for rapid protein-structure determination by solid-state NMR spectroscopy." Angew Chem Int Ed Engl **44**(14): 2089-2092.

Laskowski, R. A., M. W. Macarthur, et al. (1993). "Procheck - a Program to Check the Stereochemical Quality of Protein Structures." Journal of Applied Crystallography **26**: 283-291.

Laskowski, R. A., J. A. C. Rullmann, et al. (1996). "AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR." Journal of Biomolecular Nmr **8**(4): 477-486.

Lehr, U., M. Schutz, et al. (2010). "C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA." Molecular Microbiology **78**(4): 932-946.

Lehr, U., M. Schutz, et al. (2010). "C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA." Molecular Microbiology **78**(4): 932-946.

Leo, J. C., I. Grin, et al. (2011). Philos Trans R Soc Lond B Biol Sci (**in press)**

Leo, J. C., A. Lyskowski, et al. (2011). "The Structure of E. coil IgG-Binding Protein D Suggests a General Model for Bending and Binding in Trimeric Autotransporter Adhesins." Structure **19**(7): 1021-1030.

Leo, J. C. and M. Skurnik (2011). "Adhesins of human pathogens from the genus Yersinia." Biological Reactive Intermediates Vi **715**: 1-15.

Linge, J. P., M. Habeck, et al. (2003). "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics **19**(2): 315-316.

Linge, J. P. and M. Nilges (1999). "Influence of non-bonded parameters on the quality of NMR structures: A new force field for NMR structure calculation." Journal of Biomolecular Nmr **13**(1): 51-59.

Linke, D., T. Riess, et al. (2006). "Trimeric autotransporter adhesins: variable structure, common function." Trends in Microbiology **14**(6): 264-270.

Lovas, S., G. Toth, et al. (2001). "Stabilization of local structures by pi-CH and aromatic-backbone amide interactions involving prolyl and aromatic residues." Protein Engineering **14**(8): 543-547.

Lyskowski, A., J. C. Leo, et al. (2011). "Structure and Biology of Trimeric Autotransporter Adhesins." Bacterial Adhesion: Chemistry, Biology and Physics **715**: 143-158.

Marsh, J. A., V. K. Singh, et al. (2006). "Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation." Protein Science **15**(12): 2795-2804.

Martin, R. W. and K. W. Zilm (2003). "Preparation of protein nanocrystals and their characterization by solid state NMR." J Magn Reson **165**(1): 162-174.

McDermott, A. (2009). "Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR." Annual Review of Biophysics **38**: 385-403.

Mcdonald, I. K. and J. M. Thornton (1994). "Satisfying Hydrogen-Bonding Potential in Proteins." Journal of Molecular Biology **238**(5): 777-793.

McDonald, I. K. and J. M. Thornton (1994). "Satisfying hydrogen bonding potential in proteins." Journal of Molecular Biology **238**(5): 777-793.

McGuffin, L. J., K. Bryson, et al. (2000). "The PSIPRED protein structure prediction server." Bioinformatics **16**(4): 404-405.

McLachlan, A. D. (1979). "Gene duplications in the structural evolution of chymotrypsin." Journal of Molecular Biology **128**(1): 49-79.

Mechelke, M. and M. Habeck (2010). "Robust probabilistic superposition and comparison of protein structures." Bmc Bioinformatics **11**.

Meng, G., N. K. Surana, et al. (2006). "Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter." EMBO J **25**(11): 2297-2304.

Metz, G., X. L. Wu, et al. (1994). "Ramped-Amplitude Cross-Polarization in Magic-Angle-Spinning Nmr." Journal of Magnetic Resonance Series A **110**(2): 219-227.

Morcombe, C. R., V. Gaponenko, et al. (2004). "Diluting abundant spins by isotope edited radio frequency field assisted diffusion." Journal of the American Chemical Society **126**(23): 7196-7197.

Murzin, A. G., A. M. Lesk, et al. (1994). "Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis." Journal of Molecular Biology **236**(5): 1369-1381.

Nabuurs, S. B., C. A. E. M. Spronk, et al. (2003). "Quantitative evaluation of experimental NMR restraints." Journal of the American Chemical Society **125**(39): 12026-12034.

Nabuurs, S. B., C. A. E. M. Spronk, et al. (2006). "Traditional biomolecular structure determination by NMR spectroscopy allows for major errors." Plos Computational Biology **2**(2): 71-79.

Nagao, Y., S. Miyamoto, et al. (2006). "Highly stereoselective asymmetric Pummerer reactions that incorporate intermolecular and intramolecular nonbonded S...O interactions." Journal of the American Chemical Society **128**(30): 9722-9729.

Nilges, M. (1995). "Calculation of Protein Structures with Ambiguous Distance Restraints - Automated Assignment of Ambiguous Noe Crosspeaks and Disulfide Connectivities." Journal of Molecular Biology **245**(5): 645-660.

Nilges, M., G. M. Clore, et al. (1988). "Determination of 3-Dimensional Structures of Proteins from Interproton Distance Data by Dynamical Simulated Annealing from a Random Array of Atoms - Circumventing Problems Associated with Folding." Febs Letters **239**(1): 129-136.

Nilges, M., M. Habeck, et al. (2008). "Probabilistic structure calculation." Comptes Rendus Chimie **11**(4-5): 356-369.

Nummelin, H., M. C. Merckel, et al. (2004). "The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll." Embo Journal **23**(4): 701-711.

Opella, S. J. and M. H. Frey (1979). "Selection of Non-Protonated Carbon Resonances in Solid-State Nuclear Magnetic-Resonance." Journal of the American Chemical Society **101**(19): 5854-5856.

Opella, S. J., M. H. Frey, et al. (1979). "Detection of Individual Carbon Resonances in Solid Proteins." Journal of the American Chemical Society **101**(19): 5856-5857.

Pal, D. and P. Chakrabarti (2001). "Non-hydrogen bond interactions involving the methionine sulfur atom." J Biomol Struct Dyn **19**(1): 115-128.

Pauli, J., M. Baldus, et al. (2001). "Backbone and side-chain 13C and 15N signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 Tesla." Chembiochem **2**(4): 272-281.

Pauli, J., M. Baldus, et al. (2001). "Backbone and side-chain C-13 and N-15 signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla." Chembiochem **2**(4): 272-281.

Peterson, J. H., P. Tian, et al. (2010). "Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment." Proceedings of the National Academy of Sciences of the United States of America **107**(41): 17739-17744.

Phan, G., H. Remaut, et al. (2011). "Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate." Nature **474**(7349): 49-U71.

Regan, L. and J. S. Merkel (1998). "Aromatic rescue of glycine in beta sheets." Folding & Design **3**(6): 449-455.

Rieping, W., M. Habeck, et al. (2007). "ARIA2: Automated NOE assignment and data integration in NMR structure calculation." Bioinformatics **23**(3): 381-382.

Rieping, W., M. Habeck, et al. (2005). "Inferential structure determination." Science **309**(5732): 303-306.

Rieping, W., M. Nilges, et al. (2008). "ISD: a software package for Bayesian NMR structure calculation." Bioinformatics **24**(8): 1104-1105.

Roggenkamp, A., N. Ackermann, et al. (2003). "Molecular analysis of transport and oligomerization of the Yersinia enterocolitica adhesin YadA." J Bacteriol **185**(13): 3735-3744.

Roggenkamp, A., K. Ruckdeschel, et al. (1996). "Deletion of amino acids 29 to 81 in adhesion protein YadA of Yersinia enterocolitica serotype O:8 results in selective abrogation of adherence to neutrophils." Infection and Immunity **64**(7): 2506-2514.

Roggenkamp, A., K. Ruckdeschel, et al. (1996). "Deletion of amino acids 29 to 81 in adhesion protein YadA of Yersinia enterocolitica serotype O:8 results in selective abrogation of adherence to neutrophils." Infect Immun **64**(7): 2506-2514.

Sandkvist, M. (2001). "Type II secretion and pathogenesis." Infection and Immunity **69**(6): 3523-3535.

Schaefer, J. and E. O. Stejskal (1976). "C-13 Nuclear Magnetic-Resonance of Polymers Spinning at Magic Angle." Journal of the American Chemical Society **98**(4): 1031-1032.

Schoneberg, T., A. Schulz, et al. (2004). "Mutant G-protein-coupled receptors as a cause of human diseases." Pharmacology & Therapeutics **104**(3): 173-206.

Schutz, M., E. M. Weiss, et al. (2010). "Trimer stability of YadA is critical for virulence of Yersinia enterocolitica." Infect Immun **78**(6): 2677-2690.

Shen, Y., F. Delaglio, et al. (2009). "TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." Journal of Biomolecular Nmr **44**(4): 213-223.

Shen, Y., F. Delaglio, et al. (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." Journal of Biomolecular Nmr **44**(4): 213-223.

Simeone, R., D. Bottai, et al. (2009). "ESX/type VII secretion systems and their role in host-pathogen interaction." Current Opinion in Microbiology **12**(1): 4-10.

Sinha, N., C. V. Grant, et al. (2005). "SPINAL modulated decoupling in high field double- and triple-resonance solid-state NMR experiments on stationary samples." Journal of Magnetic Resonance **177**(2): 197-202.

Skurnik, M., I. Bolin, et al. (1984). "Virulence plasmid-associated autoagglutination in Yersinia spp." J Bacteriol **158**(3): 1033-1036.

Sosnick, T. R. and D. Barrick (2011). "The folding of single domain proteins--have we reached a consensus?" Curr Opin Struct Biol **21**(1): 12-24.

Spera, S. and A. Bax (1991). "Empirical Correlation between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts." Journal of the American Chemical Society **113**(14): 5490-5492.

Spronk, C. A. E. M., S. B. Nabuurs, et al. (2004). "Validation of protein structures derived by NMR spectroscopy." Progress in Nuclear Magnetic Resonance Spectroscopy **45**(3-4): 315-337.

Straus, S. K., T. Bremi, et al. (1996). "Resolution enhancement by homonuclear J decoupling in solid-state MAS NMR." Chemical Physics Letters **262**(6): 709-715.

Surana, N. K., D. Cutter, et al. (2004). "The Haemophilus influenzae Hia autotransporter contains an unusually short trimeric translocator domain." Journal of Biological Chemistry **279**(15): 14679-14685.

Suter, D. and R. R. Ernst (1985). "Spin diffusion in resolved solid-state NMR spectra." Phys Rev B Condens Matter **32**(9): 5608-5627.

Swendsen, R. H. and J. S. Wang (1986). "Replica Monte-Carlo Simulation of Spin-Glasses." Physical Review Letters **57**(21): 2607-2609.

Szeverenyi, N. M., M. J. Sullivan, et al. (1982). "Observation of Spin Exchange by Two-Dimensional Fourier-Transform C-13 Cross Polarization-Magic-Angle Spinning." Journal of Magnetic Resonance **47**(3): 462-475.

Takegoshi, K., T. Yano, et al. (2001). "Indirect high-resolution observation of 14N NMR in rotating solids." Journal of the American Chemical Society **123**(43): 10786-10787.

Veiga, E., E. Sugawara, et al. (2002). "Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains." EMBO J **21**(9): 2122-2131.

Veiga, E., E. Sugawara, et al. (2002). "Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains." Embo Journal **21**(9): 2122-2131.

Verel, R., M. Baldus, et al. (1998). "A homonuclear spin-pair filter for solid-state NMR based on adiabatic-passage techniques." Chemical Physics Letters **287**(3-4): 421-428.

Vondrasek, J., L. Biedermannova, et al. (2008). "Another role of proline: stabilization interactions in proteins and protein complexes concerning proline and tryptophane." Physical Chemistry Chemical Physics **10**(42): 6350-6359.

Vriend, G. (1990). "What If - a Molecular Modeling and Drug Design Program." Journal of Molecular Graphics **8**(1): 52-&.

Walshaw, J. and D. N. Woolfson (2001). "SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures." Journal of Molecular Biology **307**(5): 1427-1450.

Wang, J., S. Kim, et al. (2001). "Structure of the transmembrane region of the M2 protein H(+) channel." Protein Sci **10**(11): 2241-2250.

Wollmann, P., K. Zeth, et al. (2006). "Purification of the YadA membrane anchor for secondary structure analysis and crystallization." Int J Biol Macromol **39**(1-3): 3-9.

Wu, X. L., S. T. Burns, et al. (1994). "Spectral Editing in Cpmas Nmr - Generating Subspectra Based on Proton Multiplicities." Journal of Magnetic Resonance Series A **111**(1): 29-36.

Xu, D., C. J. Tsai, et al. (1997). "Hydrogen bonds and salt bridges across protein-protein interfaces." Protein Eng **10**(9): 999-1012.

Yeo, H. J., S. E. Cotter, et al. (2004). "Structural basis for host recognition by the Haemophilus influenzae Hia autotransporter." Embo Journal **23**(6): 1245-1256.

7 Appendices

Appendix I – Interactions in YadA-M

7.1 Interactions involving aromatic residues

7.1.1 Interactions between aromatic and charged residues

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers (M) Involved
E104Cα-F56C`	2D DARR 200 ms, 2D DARR 300 ms	6.4	7	M1-M3
D33Cβ-Y63C`	2D DARR 300 ms	4.9	5.9	Intra-M
Y63C`-D33Cγ	2D DARR 200 ms	4.1	5.2	Intra-M
D33C`-Y63Cε1	2D DARR 300 ms	6.7	7.5	Intra-M
D33Cγ-Y63Cε2	2D DARR 200 ms	3.7	4.9	Intra-M
K34C`-Y63C`	2D CPPI-DARR 300 ms	7.5	8.4	Intra-M
K53Cβ-Y75Cβ	2D DARR 200 ms	7.6	8.3	Intra-M

7.1.2 Interactions between aromatic and polar residues

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
F56C`-N102Cα	2D PDSD 100 ms	4.7	5.2	M1-M2
N102Cα-F56Cβ	2D DARR 200 ms	6.1	7	M1-M2
F56C`-N102Cβ	2D DARR 300 ms	5.1	5.8	M1-M2
F56Cδ2-S73Cα	2D DARR 300 ms	4	7	Intra-M
F56Cε2-S73Cα	2D DARR 300 ms	4.1	7.4	Intra-M
S73Cα-F56C`	2D PAR 15 ms	4.7	5.2	Intra-M
S73Cα-F56Cβ	2D PAR 6 ms	4.9	5.7	Intra-M
S73Cα-F56Cε2	2D DARR 200 ms	4.1	7.4	Intra-M
S73Cβ-F56C`	2D PDSD 100 ms	5.3	6.2	Intra-M
S73Cβ-F56Cβ	2D DARR 200 ms; 300 ms	4.9	5.7	Intra-M
T57C`-F101C`	2D CPPI-DARR 300 ms	5.3	5.6	M1-M3

7.1.3 Interactions between aromatic and neutral residues

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
G86Cα-Y97Cα	2D DARR 200 ms	5.8	6.3	Intra-M
Y97Cα-G86C`	2D DARR 200 ms; 300 ms	4.5	5	Intra-M

7.2 Interactions involving charged and polar residues

7.2.1 Interactions where both residue are charged

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
R31Cβ-D94Cβ	2D DARR 200 ms, 2D PAR 6 ms	6	7.1	Intra-M
D94Cγ-R31Cδ	2D DARR 200 ms; 300 ms; 500, 2D CPPI-DARR 300 ms	3.3	5.1	Intra-M
K34Cε-D94Cα	2D DARR 500 ms, 2D DARR 500 ms, 2D PDSD 100 ms	5.3	6.8	Intra-M
R64Cα-D94Cα	2D DARR 300 ms	5.5	6.5	M1-M3
R64C`-D94Cβ	2D DARR 300 ms, 2D CPPI-DARR 300 ms	7.6	9	M1-M3

7.2.2 Interactions between charged and polar residues

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
N55Cα-E104C`	2D PAR 15 ms	4.9	5.7	M1-M3
N55Cα-E104Cβ	2D CPPI-DARR-DD 300 ms	4.9	5.7	M1-M3
E104Cβ-N55C`	2D CPPI-Darr 300 ms	6	6.7	M1-M3
N55C`-E104Cδ	2D DARR 200 ms	7.1	8.1	M1-M3
E104Cδ-N55Cα	2D DARR 200 ms, 2D DARR 300 ms	5.5	7.2	M1-M3
E104Cγ-N55Cβ	2D ChhC 150 µs	4.7	6.2	M1-M3
E104Cδ-Q47C`	2D DARR 200 ms	5.6	6.5	M1-M3
K84Cβ-S100C`	2D DARR 500 ms	4.6	5.4	Intra-M
K84Cβ-S100Cα	2D PAR 2.25 ms ; 6 ms	5	6	Intra-M
S100N-K84Cδ	2D NhhC 200 µs	6.1	7.5	Intra-M
K84Cβ-Q47C`	2D DARR 500 ms	6.9	7.7	Intra-M
K84Cδ-Q47C`	2D DARR 500 ms	5.8	7.2	Intra-M
S73Cβ -K84C`	2D DARR 500 ms	7.2	7.8	Intra-M
S66Cβ-K34Cδ	2D DARR 500 ms	5.1	6	Intra-M

7.2.3 Interactions between charged residues and P48

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
P48Cα-E104Cα	2D DARR 500 ms	7.7	8.9	M1-M3
P48Cβ-E104Cβ	2D DARR 200 ms, 2D DARR 300 ms	6.7	8.2	M1-M3
E104Cδ-P48Cβ	2D DARR 200 ms	4.3	5.9	M1-M3
P48Cβ-E104Cγ	2D DARR 200 ms; 300 ms	5.5	6.9	M1-M3
E104Cγ-P48Cδ	2D ChhC 500 µs	4.2	5.8	M1-M3

7.3 Interactions involving polar residues

7.3.1 Interactions where both residues are polar

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
S44Cα-N98Cα	2D DARR 300 ms; 500 ms	6.6	7.4	M1-M3
S44Cβ-N98Cα	2D DARR 500 ms	5.1	6	M1-M3
T57Cβ-N102Cα	2D CPPI-DARR 300 ms	5.4	5.9	M1-M3
T57Cγ2-N102Cα	2D DARR 200 ms, 2D DARR 300 ms	4.8	6.1	M1-M3
N102Cα-T57C`	2D PDSD 100 ms	5.5	5.8	M1-M3

N102Cα-T57Cα	2D ChhC 500 μs	4.7	5.1	M1-M3

7.3.2 Interactions between polar and neutral residues

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
G59Cα-S100Cα	2D PAR 2.25 ms, 3D NCACX 200 ms	3.2	3.6	M1-M3
S100Cβ-G59Cα	2D ChhC 200; 300 μs, 2D DARR 300 ms	3	3.6	M1-M3
G61Cα-N98Cα	2D ChhC 150; 200 μs, 2D DARR 300 ms	4.1	4.7	M1-M3
N98Cβ-G61C`	2D DARR 200 ms	5.4	6.8	M1-M3
N98Cβ-G61Cα	2D ChhC 150 μs	4.5	5.8	M1-M3
N98N-G61C`	2D TEDOR 2.24 ms	5.5	6	M1-M3
G72C`-T57C`	2D DARR 200 ms	4	4.5	Intra-M
T57C`-G72Cα	2D PDSD 15 ms, 2D DARR 300 ms	4.3	4.9	Intra-M
Q67Cγ-G62C`	2D DARR 200 ms	4.9	6.3	Intra-M
G62Cα-Q67Cα	2D ChhC 200ms; 500 μs	3.7	4.3	Intra-M
Q67Cγ-G61Cα	2D DARR 300 ms	4.4	5.9	Intra-M
S66Cβ-G62Cα	2D DARR 300 ms	6.1	6.4	Intra-M

7.4 Interactions involving hydrophobic residues

7.4.1 Interactions where both residues are hydrophobic

Interactions	Solid-state MAS NMR Spectra	d min. Å	d min. Å	Monomers involved
I103Cγ1-V81Cα	2D DARR 200 ms	5.4	6.4	Intra-M
I103Cδ1-V81Cγ1	2D DARR 300 ms	5.3	7.6	Intra-M
V81Cβ-I103Cα	2D DARR 200 ms	5	6.1	Intra-M
I103Cα-V81Cα	2D ChhC 50; 80; 500 μs, 2D PAR 2.25 ms, 2D DARR 200; 300 ; 500 ms	4.4	5.2	Intra-M
V60Cγ1-A99Cα	2D DARR 200 ms, 2D CPPI-DARR 300 ms, 2D PAR 15 ms	7.1	7.7	M1-M3
V87Cβ-M96C	2D DARR 200 ms	4.5	5	Intra-M
A88Cα-V95Cγ1	2D DARR 500 ms	4.9	5.9	Intra-M
V87Cβ-V95Cβ	2D DARR 200 ms; 300 ms	6	7	Intra-M
A88C-V95Cβ	2D DARR 200 ms, J-decoupled 2D DARR 70 ms	4.4	5.4	Intra-M
V95Cβ-A88Cα	2D DARR 200 ms	5.5	6.6	Intra-M
I71Cα-V87Cγ2	2D DARR 500 ms	7.2	8	Intra-M
I71Cβ-V87Cγ1	2D DARR 200 ms; 300 ms	5.1	6.4	Intra-M
I71Cγ1-V87Cγ1	2D CPPI-DARR 300 ms	6.2	7.5	Intra-M
A70Cβ-V87Cβ	2D DARR 300 ms	6.9	7.8	Intra-M
I71Cβ-V87Cβ	2D DARR 300 ms	6.5	7.3	Intra-M
I71Cγ1-V87Cβ	2D DARR 200 ms; 300 ms	7.4	8.2	Intra-M
I71Cγ2-V87Cβ	2D DARR 200 ms	5	6.4	Intra-M
V87Cβ-I71Cα	2D PAR 15 ms	6.1	7	Intra-M
V87Cβ-I71Cβ	2D DARR 200 ms	6.5	7.3	Intra-M
II03Cδ1-L83Cα	2D DARR 300 ms	6.8	8.7	Intra-M
V81Cβ-I103Cα	2D DARR 200 ms	5	6.1	Intra-M
V81Cα-I103C	2D DARR 200 ms; 300 ms; 500 ms, 2D PDSD 100 ms	5.1	5.7	Intra-M
V81Cα-I103Cα	2D PDSD 15 ms, 2D ChhC; 50; 80; 200; 300 μs, 2D PAR 2.25; 15 ms, 2D DARR 300; 500 ms	4.4	5.2	Intra-M
I103Cγ1-V81Cα	2D DARR 200 ms	5.4	6.5	Intra-M

V51Cα-V77Cβ	2D DARR 500 ms	5.1	6.4	Intra-M
G52Cα-V77Cβ	2D DARR 200 ms	3.9	5.6	Intra-M
V51Cα-V77C	2D DARR 500 ms	3.9	4.8	Intra-M
I71Cγ2-A58Cα	2D DARR 200 ms; 300 ms	6.1	6.6	Intra-M
A58Cα-I71Cγ1	2D PAR 2.25 ms; 6 ms	4.1	4.9	Intra-M
I71Cγ1-A58Cβ	2D PAR 2.25 ms	3.4	4.4	Intra-M
I71Cγ1-V87Cγ1	2D-CPPI-DARR 300 ms	6.2	7.5	Intra-M
A58C-I71Cδ1	2D DARR 200 ms	5.3	6.4	Intra-M
A58Cα-I71Cδ1	2D DARR 300 ms	4.2	5.4	Intra-M
I71Cδ1-A58Cβ	2D DARR 200; 300; 500 ms, 2D PAR 6; 15 ms, 2D-CPPI-DARR 300 ms	3.2	4.3	Intra-M
V87Cα-I71Cα	2D ChhC 80; 150; 200 μs, 2D PAR 6 ms	5.4	6	Intra-M
I71Cα-A58Cα	2D ChhC 500 μs	4.1	4.6	Intra-M
V87Cβ-I71Cα	2D PAR 15 ms	6.1	6.7	Intra-M
A70Cβ-V87Cβ	2D DARR 300 ms	6.9	7.8	Intra-M
A70Cα-V60Cγ2	2D DARR 500 ms	7.4	8	Intra-M
V60Cγ1-L69Cγ	2D DARR 300 ms	2.9	4.8	Intra-M
L69Cγ-A68Cα	2D PDSD 100 ms	5.3	6.3	Intra-M
V60Cα-L69Cδ1	2D DARR 200; 300 ms	5.1	6.4	Intra-M
L69Cδ1-V60Cβ	2D PAR 6 ms; 15 ms	4.5	6.1	Intra-M
L69Cδ1-V60Cγ2	2D PAR 6 ms; 15 ms	5.5	7.2	Intra-M
V60Cβ-L69Cβ	2D DARR 300 ms	4.3	5.4	Intra-M
L69Cβ-V60Cγ1	2D PAR 6 ms	3.1	4.5	Intra-M
V60N-G59C-L69Cα	2D NCOCX 200 ms	5	5.6	Intra-M
L69Cα-V60Cα	2D DARR 200; 300 ms, 2D PDSD 100 ms	4.1	4.7	Intra-M
L69Cα-V60Cγ1	2D DARR 300 ms	3.3	4.4	Intra-M

7.4.2 Interactions between aromatic and hydrophobic residues

Interactions	**Solid-state MAS NMR Spectra**	**d min. Å**	**d min. Å**	**Monomers involved**
V87Cβ-Y97C`	2D DARR 200; 300 ms	5.7	6.4	Intra-M
V87Cγ1-Y97C`	J-decoupled 2D DARR 70 ms	7.1	7.8	Intra-M
V87Cα-Y97Cα	2D DARR 300; 500 ms, 2D ChhC 300 μs; 500 μs	4.3	4.8	Intra-M
V87Cβ-Y97Cα	2D PAR 6 ms	4.5	5.3	Intra-M
Y97Cα-V87C`	2D DARR 500 ms	5.2	5.7	Intra-M
Y97Cα-V87Cγ2	2D DARR 500 ms	3.9	5.3	Intra-M
V87Cβ-Y97Cβ	2D DARR 200; 300 ms, J-decoupled 2D DARR 70 ms	4.8	5.7	Intra-M
M96C`-Y63Cζ	2D DARR 200 ms; 300 ms	7.7	8.9	M1-M3
Y63Cζ-M96Cβ	2D CPPI-DARR 300 ms	5.5	6.5	M1-M3
M96Cβ-Y63C`	2D DARR 500 ms	4.3	5.2	M1-M3
M96Cβ-Y63Cβ	2D DARR 200 ms; 300 ms; 500 ms, 2D PDSD 100 ms	3.5	4.7	M1-M3
Y63Cβ-M96Cε	2D PAR 6 ms, 2D DARR 200 ; 300 ms	5.1	6.3	M1-M3
M96Cε-Y63Cα	2D PDSD 100 ms, 2D DARR 200 ms; 300 ms, 2D CPPI-DARR 300 ms, 2D PAR 6 ms	5.3	6.5	M1-M3
M96Cε-Y63Cδ2	2D DARR 300 ms	5.6	6.5	M1-M3
M96Cε-Y63Cγ	2D PDSD 100 ms	4.6	5.5	M1-M3
Y63Cε1-M96Cγ	2D CPPI-DARR 300 ms	4.6	5.9	M1-M3
Y89Cα-V95Cα	2D ChhC 500 μs, 2D DARR 500 ms	4.3	5	Intra-M
V95Cα-Y89Cβ	2D DARR 300 ms	5.2	5.9	Intra-M
Y89Cε1-V95Cβ	2D CPPI-DARR 300 ms	4.3	6.3	Intra-M
Y89Cε2-V95Cγ1	2D DARR 200; 300 ms, 2D PDSD 100 ms	3.7	7.2	Intra-M
V95Cγ1-Y89Cα	2D DARR 500 ms	4	5.4	Intra-M

V95Cγ1-Y89Cδ1	2D DARR 200 ms ; 300 ms	3.3	5.3	Intra-M
V95Cγ1-Y89Cδ2	2D DARR 200 ms ; 300 ms	3.5	6.7	Intra-M
V95Cγ1-Y89Cγ	2D DARR 300 ms	3.6	5.7	Intra-M
L69Cβ-Y89Cα	2D DARR 200 ms	5.7	6.7	Intra-M
L69Cδ1-Y89Cα	2D PAR 15 ms	6.8	8.4	Intra-M

7.4.3 Interactions between hydrophobic and neutral residues

Interactions	**Solid-state MAS NMR Spectra**	**d min. Å**	**d min. Å**	**Monomers involved**
A99NCα-G86C	3D NCACX 200 ms	6.4	7.1	Intra-M
M96Cγ-G35Cα	2D DARR 200 ms	5.9	6.6	Intra-M
M96Cγ-G62C	2D DARR 200 ms	4.5	4.9	M1-M3
G35C-M96Cε	2D DARR 500 ms	4	5.2	Intra-M
G35Cα-M96Cε	2D DARR 500 ms	3.6	4.8	Intra-M
G62Cα-M96Cα	2D DARR 200 ms	5.5	6	M1-M3
L42Cδ1-G86C	2D DARR 200 ms	3.9	5.2	Intra-M
A85C-G72Cα	2D DARR 200 ms	4.1	4.4	Intra-M
G74Cα-L83Cα	2D PAR 6 ms	5.7	6.3	Intra-M
G74NCα-A85C	2D NCACX 200 ms	5.5	6.1	Intra-M
V54Cγ1-G74C	2D DARR 200 ms; 300 ms	5.8	6.8	Intra-M
G72N-A58Cα	2D NhhC 200 μs	4.7	5.1	Intra-M
A85C-G72Cα	2D DARR 200 ms	4.1	4.4	Intra-M
L45Cγ-G72C	2D DARR 200 ms; 300 ms	6.4	7.3	Intra-M
I71Cδ1-G59C	2D DARR 200 ms	7.2	8.3	Intra-M
G59C-A70Cα	2D DARR 200 ms	4.5	5.1	Intra-M
L69Cδ1-G59C	2D-CPPI-DARR 300 ms	4.8	7.4	Intra-M
A68Cβ-G61Cα	2D-CPPI-DARR 300 ms	5.8	6.5	Intra-M
A68Cα-G61C	2D DARR 300 ms	4.8	5.3	Intra-M

Appendix II – ^{13}C, ^{15}N chemical shifts of YadA-M as determined by solid-state MAS NMR

Residue	N	C`	Cα	Cβ	Cγ	Cγ1	Cγ2	Cδ	Cδ1	Cδ2	Cε	Cε1	Cε2	Cζ
E12	122.9	179.8	57.82	31.68	-	-	-	-	-	-	-	-	-	-
K13	113.6	174.1	58.03	35.96	21.33	-	-	-	-	-	-	-	-	-
G14	108.9	171.1	43.34	-	-	-	-	-	-	-	-	-	-	-
A15	-	-	53.98	-	-	-	-	-	-	-	-	-	-	-
H16	117.1	175.6	60.82	30.29	135.8	-	-	-	-	-	-	-	-	-
K17	-	179.6	60.3	33.5	-	-	-	29.99	-	-	42.95	-	-	-
F18	109	179.2	58.11	39.44	-	-	-	-	129.2	-	-	-	-	-
R19	117.8	179.5	59.33	30.11	27.45	-	-	42.86	-	-	-	-	-	-
Q20	119	178	59.09	28.44	33.96	-	-	179.7	-	-	-	-	-	-
L21	120.2	179.9	58.02	40.34	27.62	-	-	-	22.05	-	-	-	-	-
D22	123.1	178.8	57.76	40.04	-	-	-	-	-	-	-	-	-	-
N23	116.4	177.6	56.43	38.99	177.9	-	-	-	-	-	-	-	-	-
R24	120.7	180	60.39	30.85	28.05	-	-	44.12	-	-	-	-	-	160
L25	122.4	178.9	58.64	42.73	27.59	-	-	-	23.91	24.37	-	-	-	-
D26	118.9	179.4	57.93	40.11	179.6	-	-	-	-	-	-	-	-	-
K27	118.2	180.1	59.34	32.59	26.01	-	-	29.27	-	-	41.73	-	-	-
L28	124.1	177.7	58.13	41.93	27.76	-	-	-	26.28	23.66	-	-	-	-
D29	120	178.3	58.11	41.77	179.9	-	-	-	-	-	-	-	-	-
T30	111.6	175.7	66.2	69.33	-	-	22.3	-	-	-	-	-	-	-
R31	122.8	180.3	60.29	31.51	28.27	-	-	45.04	-	-	-	-	-	160.7

Residue	N	C`	Cα	Cβ	Cγ	Cγ1	Cγ2	Cδ	Cδ1	Cδ2	Cε	Cε1	Cε2	Cζ
V32	121	177	67.44	31.93	-	25.08	21.92	-	-	-	-	-	-	-
D33	121.6	179.1	58.28	39.6	178.8	-	-	-	-	-	-	-	-	-
K34	117.5	179.7	60.63	33.62	22.14	-	-	27.36	-	-	42.63	-	-	-
G35	110	175.5	47.19	-	-	-	-	-	-	-	-	-	-	-
L36	123.5	178.7	57.47	41.64	27.05	-	-	-	-	22.15	-	-	-	-
A37	115	174.8	52.89	-	-	-	-	-	-	-	-	-	-	-
S38	117.2	172.8	58.6	66.26	-	-	-	-	-	-	-	-	-	-
S39	119.1	178.7	57.94	63.45	-	-	-	-	-	-	-	-	-	-
A40	119	180.3	54.38	18.61	-	-	-	-	-	-	-	-	-	-
A41	122.7	179.1	55.84	17.45	-	-	-	-	-	-	-	-	-	-
L42	115.1	175.2	62.93	40.23	26.89	-	-	-	22.16	21.22	-	-	-	-
N43	117.1	175.9	53.55	38.01	178.8	-	-	-	-	-	-	-	-	-
S44	116.1	175.3	60.08	63.3	-	-	-	-	-	-	-	-	-	-
L45	120.5	178.1	55.42	40.37	25.48	-	-	-	19.9	26.82	-	-	-	-
F46	123.5	177.8	57.65	40.18	142.5	-	-	-	133.8	132.9	-	-	-	-
Q47	119.1	171.9	51.42	28.76	32.2	-	-	179.3	-	-	-	-	-	-
P48	132.3	177.7	61.8	31.9	27.25	-	-	49.03	-	-	-	-	-	-
Y49	124	180	62.49	40.4	132.4	-	-	-	135	133.3	-	117.2	-	158.6
G50	111.4	173.9	45.48	-	-	-	-	-	-	-	-	-	-	-
V51	124.9	177.5	64.44	31.34	-	21.47	-	-	-	-	-	-	-	-
G52	116.4	174.1	46.63	-	-	-	-	-	-	-	-	-	-	-
K53	117.6	175.6	54.3	34.36	24.64	-	-	29.06	-	-	42.48	-	-	-
V54	114.1	172.8	59.8	36.45	-	21.43	-	-	-	-	-	-	-	-
N55	121.6	173.4	53.2	44.46	-	-	-	-	-	-	-	-	-	-
F56	121.4	171.7	58.55	42.92	139.9	-	-	-	132.2	131.9	-	131.3	128.8	-

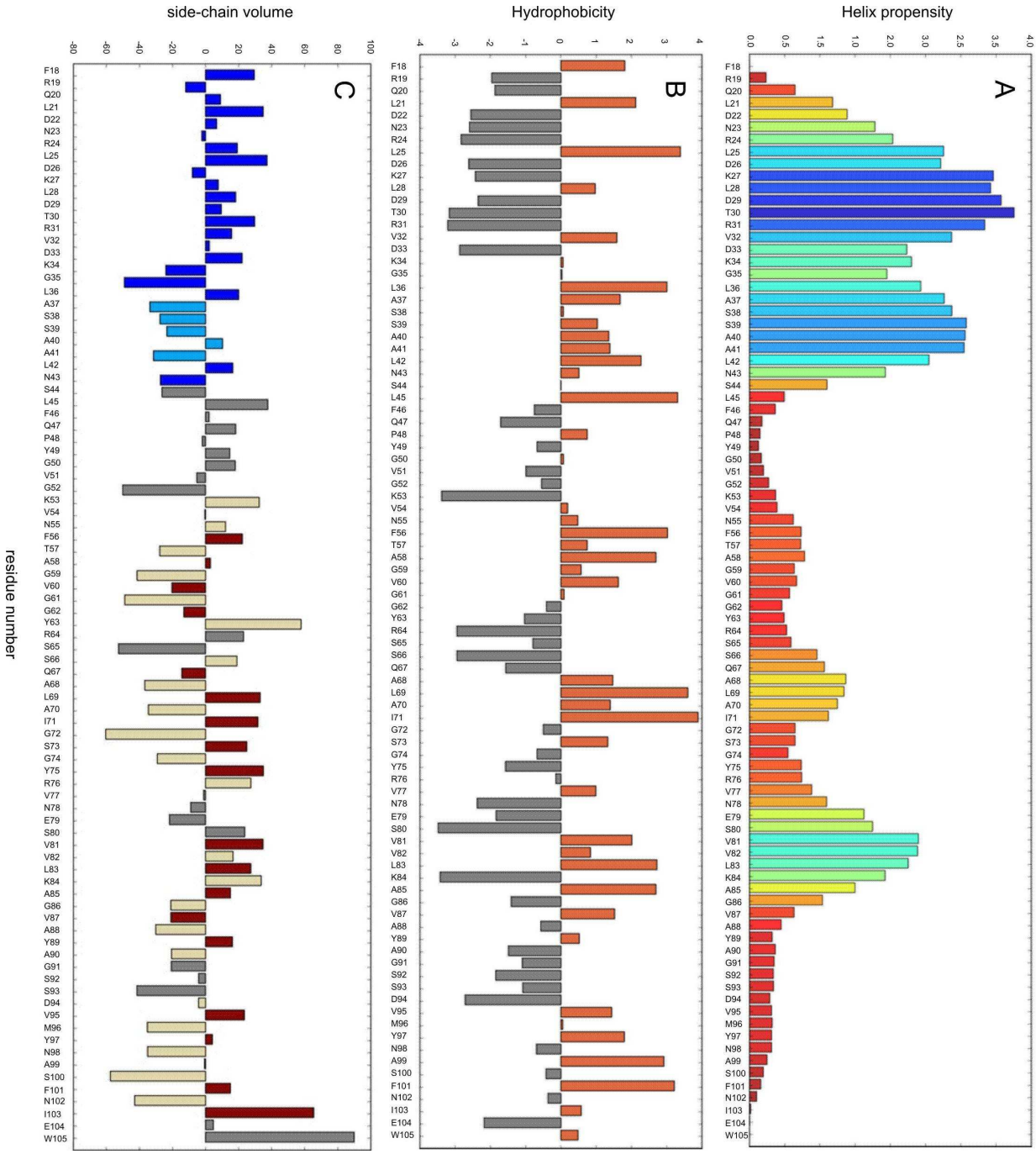
Residue	N	C`	Cα	Cβ	Cγ	Cγ1	Cγ2	Cδ	Cδ1	Cδ2	Cε	Cε1	Cε2	Cζ
T57	121.6	169.3	57.15	70.24	-	-	18.67	-	-	-	-	-	-	-
A58	125.4	174.7	50.3	23.33	-	-	-	-	-	-	-	-	-	-
G59	104.3	171	45.39	-	-	-	-	-	-	-	-	-	-	-
V60	108.3	174.9	56.93	35.44	-	17.91	22.5	-	-	-	-	-	-	-
G61	105.2	171.4	44.01	-	-	-	-	-	-	-	-	-	-	-
G62	106.2	170.7	44.64	-	-	-	-	-	-	-	-	-	-	-
Y63	118.7	173.6	58.35	41.46	129.4	-	-	-	132.5	132.8	-	117.9	117.5	158.3
R64	124.6	177.1	58.95	27.81	26.11	-	-	44.03	-	-	-	-	-	159.7
S65	119.6	174.1	57.63	63.38	-	-	-	-	-	-	-	-	-	-
S66	118.4	173.3	59.52	65.9	-	-	-	-	-	-	-	-	-	-
Q67	121.6	174.9	54.42	33.8	34.7	-	-	180.1	-	-	-	-	-	-
A68	123.2	173.2	50.76	24.72	-	-	-	-	-	-	-	-	-	-
L69	120.8	174.2	52.55	49.16	28.46	-	-	-	24.07	-	-	-	-	-
A70	125.5	175.4	50.83	23.23	-	-	-	-	-	-	-	-	-	-
I71	113.9	173.9	59.5	43.09	-	26.67	18.34	-	14.24	-	-	-	-	-
G72	108.8	170.8	45.93	-	-	-	-	-	-	-	-	-	-	-
S73	108.6	173	57.34	66.08	-	-	-	-	-	-	-	-	-	-
G74	111.2	170.9	45.51	-	-	-	-	-	-	-	-	-	-	-
Y75	118.6	175	57.72	43.48	-	-	-	-	133.4	132.4	-	117.1	118.4	158.3
R76	127.6	174.4	55.32	30.35	26.57	-	-	43.65	-	-	-	-	-	159.4
V77	125.1	176.1	64.76	33.18	-	21.53	22.27	-	-	-	-	-	-	-
N78	111.2	174.5	51.77	38.24	176.2	-	-	-	-	-	-	-	-	-
E79	115.8	177.2	59.17	29.39	36.65	-	-	183.5	-	-	-	-	-	-
S80	114	174.6	60.92	65.14	-	-	-	-	-	-	-	-	-	-
V81	121.2	174	62.08	34.95	-	21.76	-	-	-	-	-	-	-	-

Residue	N	C`	Cα	Cβ	Cγ	Cγ1	Cγ2	Cδ	Cδ1	Cδ2	Cε	Cε1	Cε2	Cζ
A82	127.4	174.4	50.67	26.48	-	-	-	-	-	-	-	-	-	-
L83	117.8	175.4	55.17	46.5	28.05	-	-	-	26.27	-	-	-	-	-
K84	117.7	172.5	57.17	36.13	21.48	-	-	29.53	-	-	41.18	-	-	-
A85	119.7	176	52.71	24.33	-	-	-	-	-	-	-	-	-	-
G86	107.5	170.8	45.98	-	-	-	-	-	-	-	-	-	-	-
V87	115.3	172	59.87	35.93	-	20.81	21.74	-	-	-	-	-	-	-
A88	124.8	175.7	50.87	26.89	-	-	-	-	-	-	-	-	-	-
Y89	118.9	173.3	57.4	44.54	132.5	-	-	-	132.1	133.1	-	117.9	118.1	159.7
A90	129.4	176.8	50.6	21.63	-	-	-	-	-	-	-	-	-	-
G91	103.6	173.3	45.72	-	-	-	-	-	-	-	-	-	-	-
S92	112.8	173.7	55.82	66.55	-	-	-	-	-	-	-	-	-	-
S93	112.9	174	59.66	63.33	-	-	-	-	-	-	-	-	-	-
D94	121.1	174.9	54.07	40.95	179.8	-	-	-	-	-	-	-	-	-
V95	122.1	174.6	61.94	34.11	-	22.17	20.6	-	-	-	-	-	-	-
M96	122.8	174.1	54.16	38.86	33	-	-	-	-	-	18.52	-	-	-
Y97	119.9	173	55.57	42.37	-	-	-	-	132.5	-	-	117.4	-	158.7
N98	114.1	173.4	52.71	42.01	177	-	-	-	-	-	-	-	-	-
A99	114.3	175.6	52.88	22.6	-	-	-	-	-	-	-	-	-	-
S100	112.6	173.8	57.12	67.49	-	-	-	-	-	-	-	-	-	-
F101	111	173.4	55.5	44.29	139.1	-	-	-	132.4	133.7	-	130.9	132.1	129.1
N102	114.4	173.9	52.75	41.89	177.2	-	-	-	-	-	-	-	-	-
I103	124.8	171.1	60.01	41.48	-	28.65	19.6	-	16.16	-	-	-	-	-
E104	122.5	175.1	54.25	34.66	37.72	-	-	183.6	-	-	-	-	-	-
W105	124	-	-	-	-	-	-	-	-	-	-	-	-	-

Appendix III – Sequence alignment of different members of TAAs

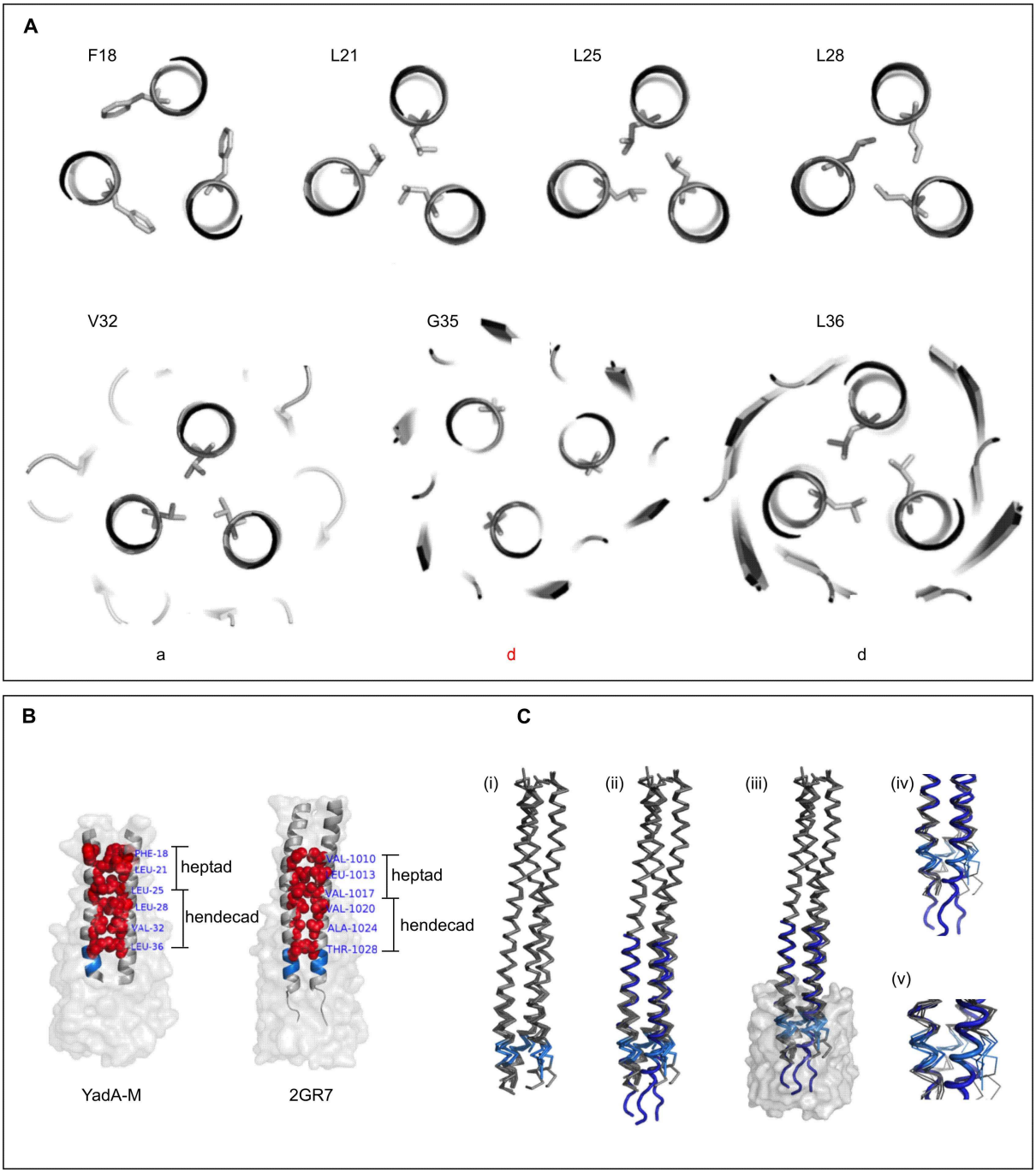
Sequence alignment of the TAA membrane anchor domain was produced in collaboration with Dirk Linke and Michael Habeck using the MPI toolkit (http://toolkit.tuebingen.mpg.de). Hydrophobic residues in the transmembrane β-strands are marked with asterisk; *a* and *d* represent the core positions of the coiled coil. The ASSA region is highlighted with red colour and comprises residues G35-L42. Highly conserved G72 and L45 are coloured green (see section 5.4.1 for details). Hydrophobicity and coiled-coil analysis was also part of the alignment. The Figures in Appendices III-V are adapted from reference Shahid et al., (submitted).

Appendix IV – Evolutionary analysis



(A) Helix propensity was calculated by Agadir (agadir.crg.es); the height and colour of bars illustrate the extent of helix propensity, where blue colour is for maximum and red for minimum helix propensity. **(B)** Conservation of hydrophobicity (Kyte-Doolittle scale); orange and grey bars indicate hydrophobic and hydrophilic residues (on an average scale), respectively. The pattern supports a hendecad repeat for L25, L28, V32 and L36 (orange bars). **(C)** Conservation of small side-chain residues in the ASSA region (highlighted in marine blue) and of small barrel residues facing the pore lumen (light yellow). Dark- and marine blue bars represent the N-terminal α-helix, grey bars, the loop region, and maroon and light yellow represent the β-sheet residues.

Appendix V – Coiled-coil analysis, switch from heptad to hendecad repeat, and comparison of YadA-M with the crystal structures of YadA-stalk



(A) Coiled-coil layers in YadA-M structure. G35 should be a *d* layer (highlighted in red), instead, its sequential neighbour, L36, shows the core packing. **(B)** Structures of YadA-M (2LME) and Hia (2GR7), illustrating the switch from 7 to 11 repeat right before ASSA region (highlighted in marine blue). **(C)** Comparison of the solid-state MAS NMR structure of YadA-M with the crystal structures of engineered version of YadA stalk. ASSA region is highlighted in marine blue; (i) shows an overlay of different engineered versions of the stalk; (ii) is the same as (i) but with the coiled-coil of YadA-M shown as blue ribbon; (iii) is the superimposed structure of YadA-M on stalk domain structures; (iv, v) close-ups of the hairpin region illustrating the heterogeneity of the crystal structures in ASSA region. None of the structures superimpose perfectly with the NMR structure. See details in Section 5.4.1.

Curriculum Vitae

Aufgrund von Datenschutz wird der Lebenslauf nicht veröffentlicht.

List of Publications

This thesis is partially based upon the following publications:

1. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals

Shakeel Ahmad Shahid, Benjamin Bardiaux, W Trent Franks, Ludwig Krabben, Michael Habeck, Barth-Jan van Rossum & Dirk Linke

Nature Methods (2012) doi:10.1038/nmeth.2248

2. Assignment and secondary structure of the YadA membrane protein by solid-state MAS NMR

Shakeel A. Shahid, Stefan Markovic, Dirk Linke & Barth-Jan van Rossum

Assignment of the YadA membrane protein by solid-state MAS NMR

Scientific Reports 2, (2012) doi:10.1038/srep00803

PDB entry

Solid-state NMR structure of the membrane anchor domain of the trimeric autotransporter YadA

PDB code = 2LME, DOI:10.2210/pdb2lme/pdb

Talks at Conferences

De novo structure of YadA membrane anchor domain by solid-state MAS NMR

3rd EU/CCPN NMR conference – ITQB Oeiras, Portugal; 28-30 September 2011

Posters at Conferences

Shakeel A. Shahid, Benjamin Bardiaux, Trent Franks, Stefan Markovic, Michael Habeck, Dirk Linke, Barth-Jan van Rossum.

De novo structure of YadA membrane anchor domain by solid-state MAS NMR

Presented at 52nd ENC conference, April 10-15, 2011, Pacific Grove, California.

Press Release

How bacteria attack their host cells with sticky lollipops

(Max-Planck-Campus Tübingen)

http://tuebingen.mpg.de/en/homepage/detail/how-bacteria-attack-their-host-cells-with-sticky-lollipops.html

Wie Bakterien ihre Wirtszellen mit klebrigen Lollis angreifen

(Leibniz-Institut für Molekulare Pharmakologie (FMP)

http://www.fmp-berlin.info/news/pressroom-test/pressroom-detail/article/wie-bakterien-ihre-wirtszellen-mit-klebrigen-lollis-angreifen.html