

**Rapid Determination of High-resolution Protein Structures by  
Solution and Solid-state NMR Spectroscopy**

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

Jegannath Korukottu

aus Madurai, India

Göttingen 2008

D7

Referent: Prof. Dr. Annette Zippelius

Korreferent: Prof. Dr. Christian Griesinger

Tag der mündlichen Prüfung:

# Summary

NMR spectroscopy provides high-resolution structural information of biomolecules in near-physiological conditions. Structural studies of proteins and nucleic acids are critical for understanding biological processes at the molecular level. Although significant improvements were achieved in NMR spectroscopy in the last 20 years, the increase in genome sequencing data has created a need for rapid and efficient methods of NMR-based structure determination. NMR data acquisition can be accelerated significantly, when sensitive spectrometers are combined with new methods for sampling chemical shifts in multidimensional NMR experiments. Therefore, data analysis and in particular the requirement to assign side chain chemical shifts to specific atoms is the major bottleneck of rapid NMR-based structure determination. In chapter 2, a method termed FastNMR (*FAst STRucture determination by NMR*), is described in detail, which enables automatic, high-resolution NMR structure determination of domain-sized proteins starting from unassigned NMR data. Using FastNMR the *de novo* structure of the 65-residue cone snail neurotoxin conkunitzin-S2 was determined automatically.

Large classes of proteins, such as membrane proteins and insoluble aggregates of peptides and more complex systems, cannot be investigated with the above method, because the proteins cannot be made soluble for liquid-state NMR. Therefore, there is a considerable interest in the development of methods for protein structure determination that do not have these limitations. In chapter 3 and 4 of this thesis, it is demonstrated that, combining the knowledge obtained in solution-state NMR, a rapid determination of high-resolution protein structure of globular proteins, such as, potassium channel blocker, Kaliotoxin existing in free form and also in complex with KcsA-Kv1.3, from solid-state NMR data could be obtained. Also in chapter 4, an improved model of KTX-KcsA-Kv1.3 complex is proposed based on functional and solid-state NMR data.

Finally, chapter 5 sheds light on understanding the mechanism of alignment of proteins and efforts in improving the accuracy of prediction of charge-induced molecular alignment from the protein's known 3D structure, by employing more atomistically detailed electrostatic models. Preliminary results suggest that the accuracy in predicting RDCs and magnitude of alignment using detailed electrostatics might improve in comparison with the simplified model implemented in PALES.

# Zusammenfassung

NMR-Spektroskopie ermöglicht die Bestimmung hoch-aufgelöster Strukturen von Biomolekülen unter nahezu physiologischen Bedingungen. In den letzten 20 Jahren wurden erhebliche Fortschritte in der NMR-Spektroskopie erzielt, durch die Vielzahl sequenzierter Genome werden jedoch Hochdurchsatzverfahren zur Bestimmung der Tertiärstruktur von Proteinen immer wichtiger. Die Datenaufnahme kann erheblich beschleunigt werden, wenn moderne NMR-Spektrometer mit Methoden kombiniert werden, welche effizient chemische Verschiebungen in mehrdimensionalen NMR-Experimenten messen. Daher sind die Datenanalyse und insbesondere die Notwendigkeit, chemische Verschiebungen sequenzspezifisch den Atomen der Seitenketten zuzuordnen, die Haupthindernisse für eine schnelle NMR-basierte Proteinstrukturbestimmung. In Kapitel 2 wird die Methode FastNMR (*FAST* *Structure determination by NMR*) beschrieben, welche ausgehend von nicht zugeordneten NMR-Daten die automatische Bestimmung hoch-aufgelöster Strukturen von Proteinen - die aus einer Domäne bestehen - ermöglicht. Mittels FastNMR wurde automatisch die *de novo* Struktur des aus 65 Aminosäuren bestehenden, aus Kegelschnecken stammenden Neurotoxins Konkunitzin-S2 bestimmt.

Eine große Zahl von Proteinen, wie z.B. Membranproteine oder unlösliche Aggregate von Peptiden und komplexeren Systemen, läßt sich allerdings nicht mit den zuvor beschriebenen Methoden untersuchen, da diese Proteine nicht in Lösung gebracht werden können, um mit Hilfe von Lösungs-NMR untersucht zu werden. Daher besteht ein großes Interesse an der Entwicklung von Methoden zur Proteinstruktur-Aufklärung, die nicht auf Lösungs-NMR beschränkt sind. In Kapitel 3 und 4 dieser Arbeit wird gezeigt, dass mit Hilfe der in Lösung-NMR gewonnenen Erkenntnisse eine schnelle Aufklärung von globulären Proteinen, wie z.B. dem Kaliumkanal-Blocker Kalitoxin - in der freien Form als auch in der im Komplex mit KcsA-Kv1.3 gebunden Form - durch Festkörper-NMR möglich ist. Ebenfalls in Kapitel 4 wird ein verfeinertes Model des KTX-KcsA-Kv1.3 Komplexes vorgestellt auf der Grundlage von biochemischen Daten und Festkörper-NMR Ergebnissen.

Im fünften und letzten Kapitel wird ein besseres Verständnis des Orientierungsmechanismus von Proteinen erarbeitet und erste Ansätze für eine verbesserte Vorhersage der ladungsinduzierten Orientierung von Proteinen vorgestellt. Ausgehend von einer bekannten dreidimensionalen Struktur des Moleküls

wird dazu ein verbessertes elektrostatisches Modell angewendet. Erste Ergebnisse deuten an, daß die Vorhersagekraft durch ein detailliertes elektrostatisches Modell gegenüber der des einfachen und in PALES implementierten Modells sich leicht verbessern könnte.

# Acknowledgements

Foremost, I would like to thank my supervisors Prof. Christian Griesinger and Dr. Markus Zweckstetter for providing me with the opportunity to complete my PhD thesis at the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany. I am very grateful to Prof. Christian Griesinger for providing me with an excellent scientific environment, an outstanding infrastructure and also for his interest, encouragement and useful remarks. I especially want to thank my advisor on the spot, Dr. Markus Zweckstetter, whose support and guidance made my thesis work possible. He has been actively interested in my work and has always been available to advise me. I am very grateful for his patience, motivation, enthusiasm, and immense ideas and knowledge in NMR and structural biology that, taken together, make him a great mentor.

I would also like to thank Prof. Annette Zippelius for accepting me as her student in the Faculty of Physics, Georg-August Universität, Göttingen, Germany and Prof. Reiner Kree for their valuable advice and inputs. I would also like to thank other members of the thesis committee, Prof. Helmut Grubmüller, Prof. Tim Salditt and Dr. Marcus Müller for their support, advice and for evaluating my thesis.

I also take this opportunity to thank Nils-Alexander Lakomek for his generous help, Mrs. Sigrid Silberer and Dr. Jochen Junker for their help in various aspects through out my stay in the department. Much respect to my officemates and good friends, Hai-young Kim, Min-kyu Cho, Fernando Rodriguez-Castañeda for putting up with me for almost three years.

During this work I have interacted and benefited from many discussions with Dr. Young-Sang Jung, Dr. Monika Bayrhuber, Dr. Adam Lange, Dr. Vinesh Vijayan, Robert Schneider, Dr. Pierre Montaville, Dr. Karel Kubicek, Dr. Venkatesh Ramakrishnan, Dr. Dirk Bockelmann, Dr. Stefan Becker, Dr. Marc Baldus, Dr. Christophe Fares, Dr. Sigrun Rumpel, Dr. Peter Haberz, Marcel Reese, Marco Mukrasch, Dr. Edward d'Auvergne, for whom I have great regard, and I wish to extend my warmest thanks to all those who have helped me with my work in the Department of NMR based structural biology at Max Planck Institute for Biophysical Chemistry.

Additional crucial energy and vitality for this research was provided externally through my involvement in playing and promoting Cricket in Germany. Without which life wouldn't have been special and wouldn't have had opportunity to rub shoulders with great friends such as Saju Balakrishnan, Raghavendran Laxmi Narayanan, Rosan Ariyanayagam, Naeem Misdaq, Asad Ali Khan, Philip Brown, Alastor Premnath, Markus Nitsch and Dhushan Ekanayake.

Finally, I am forever indebted to my brother Devan and his family for their affection, love and constant support when it was most required.

## Publication list:

This thesis is based on the following papers:

### Chapter 2

[1] M. Bayrhuber, V. Vijayan, M. Ferber, R. Graf, J. Korukottu, J. Imperial, J. E. Garrett, B. M. Olivera, H. Terlau, M. Zweckstetter, S. Becker. Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. *Journal of Biological Chemistry*, **2005**, 25, 23766-70.

[2] J. Korukottu, M. Bayrhuber, P. Montaville, V. Vijayan, Y. S. Jung, S. Becker, M. Zweckstetter. Fast high-resolution protein structure determination by using unassigned NMR data. *Angewandte Chemie International Edition English*, **2007**, 46, 1176-9.

### Chapter 3

[3] J. Korukottu, A. Lange, V. Vijayan, R. Schneider, O. Pongs, S. Becker, M. Baldus, M. Zweckstetter. High-resolution 3D structure determination of Kaliotoxin by solid-state NMR spectroscopy. *Submitted*.

# Table of contents

<b>Abbreviations</b>	<b>12</b>
<b>Chapter 1 General Introduction</b>	<b>15</b>
1.1 Nuclear Magnetic Resonance (NMR)	15
1.1.1 The nuclear spin Hamiltonian	15
1.1.2 Magnetization and pulses: 1D and multidimensional NMR	20
1.2 Solution NMR	22
1.2.1 Protein structure determination <i>via</i> solution NMR	22
1.2.2 Protein NMR experiments	23
1.2.3 Assignment of resonances	27
1.2.4 Collection of structural restraints	28
1.2.4.1 Distance restraints	28
1.2.4.2 Dihedral angles restraints	30
1.2.4.3 Residual Dipolar couplings (RDCs)	32
1.2.5 Structure calculation <i>via</i> Simulated annealing	36
1.3 Solid-state NMR	42
1.3.1 Protein structure determination <i>via</i> solid-state NMR	45
1.3.2 Collection of structural restraints	45
1.3.3 Structure calculation <i>via</i> simulated annealing	47
<b>Chapter 2 Fast High-resolution Protein Structure Determination by using Unassigned NMR data</b>	<b>48</b>
2.1 Introduction	48
2.2 Material and Methods	53
2.2.1 Input of unassigned NMR data	53
2.2.2 Backbone fold and assignment	56
2.2.3 Side chain assignment and structure	56



2.2.4 Refinement of 3D structure to high-resolution	58
2.3 Results and Discussions	59
2.3.1 Backbone assignment and fold determination by FastNMR	63
2.3.2 Side chain assignment and assignment analysis	66
2.3.3 FastNMR produces correct high-resolution structure	67
2.3.4 FastNMR tolerates missing and wrong backbone assignments	67
2.3.5 Influence of intermediate backbone conformation on convergence to a high-resolution structure	69
2.3.6 FastNMR tolerates spurious peaks and multiple conformation	70
2.4 Conclusions	75
<b>Chapter 3 High-resolution 3D structure determination of kaliotoxin by solid-state NMR spectroscopy</b>	<b>77</b>
3.1 Introduction	77
3.2 Materials and methods	80
3.3 Results and Discussions	83
3.3.1 High-resolution solid-state structure of KTX	83
3.3.2 Comparison of high-resolution solution-state NMR structure of KTX	91
3.3.3 Reliability of high-resolution solid-state structure of KTX	93
3.4 Conclusions	98
<b>Chapter 4 High-resolution 3D structure of kaliotoxin bound to KcsA-Kv1.3 and improved complex (KTX-KcsA-Kv1.3) structural model by solid-state NMR spectroscopy</b>	<b>99</b>
4.1 Introduction	99
4.2 Materials and methods	101
4.2.1 Solid-state NMR data	101
4.2.2 Cross-peak assignment and structure calculation	101

4.2.3 Modeling of KcsA-Kv1.3	102
4.2.4 Docking of KTX with KcsA-Kv1.3	103
4.3 Results and Discussions	105
4.3.1 Solid-state NMR structure of KTX in complex with KcsA-Kv1.3	105
4.3.2 Comparison with high-resolution solution-state and solid-state NMR structures of KTX	110
4.3.3 Improved model of KTX-KcsA-Kv1.3	111
4.4 Conclusions	116
<b>Chapter 5 Understanding and Prediction of alignment and residual dipolar couplings of a protein from its known three-dimensional structure</b>	<b>117</b>
5.1 Introduction	117
5.1.1 Residual Dipolar Couplings (RDCs)	117
5.1.2 Electrostatics	123
5.2 Materials and methods	127
5.2.1 Experimental residual dipolar couplings	127
5.2.2 Coordinates and structural models	127
5.2.3 Model of protein/bilayer	128
5.2.4 Alignment tensor prediction	130
5.2.5 Electrostatic calculations	130
5.3 Results and Discussions	133
5.3.1 Calculation of electrostatic free energies	133
5.3.2 Prediction of molecular alignment and RDCs	136
5.4 Conclusions	146
5.5 Outlook	147
<b>Bibliography</b>	<b>148</b>

<b>Appendix</b>	<b>164</b>
<b>A Chemical shift assignment of KTX in solution</b>	<b>164</b>
<b>B Distance restraints for KTX in solid-state (free form)</b>	<b>167</b>
<b>C Chemical shift assignments of Conk-S2</b>	<b>170</b>
<b>D NMR restraints of Conk-S2</b>	<b>176</b>
D.1 Distance constraints	176
D.2 Dihedral angles	187
D.3 Residual dipolar couplings	190
<b>Curriculum Vitae</b>	<b>193</b>
<b>Lebenslauf</b>	<b>194</b>

# Abbreviations:

## A

<b>ADR</b>	ambiguous distance restraints
<b>ARIA</b>	ambiguous restraints for iterative assignment

## B

<b>BMRB</b>	biological magnetic resonance bank
-------------	------------------------------------

## C

<b>CANDID</b>	combined automated NOE assignment and structure determination module
<b>CBCA(CO)NH</b>	NMR experiment observing peptide $^{15}\text{N}$ , $^1\text{H}^N$ , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$
<b>Conk-S1</b>	Conkunitzin-S1
<b>Conk-S2</b>	Conkunitzin-S2
<b>COSY</b>	correlation spectroscopy
<b>CSA</b>	chemical shift anisotropy
<b>CTX</b>	charybdotoxin

## D

<b>DARR</b>	dipolar assisted rotational resonance
<b>DHPC</b>	dihexanoyl phosphatidylcholine
<b>DMPC</b>	dimyristoyl phosphatidylcholine
<b>DNA</b>	deoxyribonucleic acid
<b>DYANA</b>	dynamics algorithm for NMR applications

## F

<b>FastNMR</b>	fast structure determination by NMR
<b>FDPB</b>	finite difference Poisson-Boltzmann
<b>FID</b>	Free induction decay

## G

$\gamma$	nuclear gyromagnetic ratio
<b>H</b>	
<b>h</b>	Planck's constant
<b>HNCA</b>	NMR experiment observing peptide $^{15}\text{N}$ , $^1\text{H}^N$ , and $^{13}\text{C}_\alpha$
<b>HNCACB</b>	NMR experiment observing peptide $^{15}\text{N}$ , $^1\text{H}^N$ , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$
<b>HNCO</b>	NMR experiment observing peptide $^{15}\text{N}$ , $^1\text{H}^N$ , and previous $^{13}\text{CO}$
<b>HN(CO)CA</b>	NMR experiment observing peptide $^{15}\text{N}$ , $^1\text{H}^N$ , and previous $^{13}\text{C}_\alpha$
<b>HSQC</b>	heteronuclear single quantum correlation
<b>I</b>	
<b>IPAP</b>	inphase antiphase
<b>K</b>	
<b>K<sup>+</sup></b>	potassium
<b>KNOWNOE</b>	automated assignment of NOESY NMR spectra using a knowledge based method
<b>KTX</b>	Kaliotoxin
<b>L</b>	
<b>M</b>	
<b>MARS</b>	automatic backbone resonance assignment program
<b>MAS</b>	magic angle spinning
$\mu_0$	magnetic permeability of vacuum
<b>N</b>	
<b>NMR</b>	nuclear magnetic resonance
<b>NOAH</b>	automatic program to assign NOESY cross-peaks
<b>NOE</b>	nuclear Overhauser effect

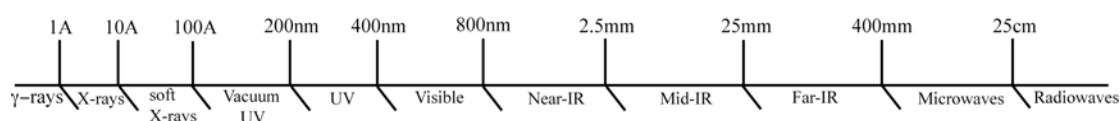
<b>NOESY</b>	nuclear Overhauser effect spectroscopy
<b>P</b>	
<b>PALES</b>	prediction of alignment from structure
<b>PASD</b>	probabilistic assignment algorithm for automated structure determination
<b>PC</b>	phosphatidylcholine
<b>PDB</b>	protein data bank
<b>PDSF</b>	proton driven spin diffusion
<b>Pf1</b>	filamentous bacteriophage
<b>PS</b>	phosphatidylserine
<b>R</b>	
<b>RDC</b>	residual dipolar coupling
<b>RF</b>	radio frequency
<b>RMSD</b>	root mean square deviation
<b>RNA</b>	ribonucleic acid
<b>S</b>	
<b>SA</b>	simulated annealing
<b>ssNMR</b>	solid-state NMR
<b>SVD</b>	singular value decomposition
<b>T</b>	
<b>T<sub>1</sub></b>	longitudinal or spin-lattice relaxation time
<b>T<sub>2</sub></b>	transversal or spin-spin relaxation time
<b>TALOS</b>	torsion angle likelihood obtained from shifts and sequence similarity
<b>TOCSY</b>	total correlation spectroscopy
<b>TROSY</b>	transverse relaxation optimized spectroscopy

# Chapter 1 General Introduction

## 1.1 Nuclear Magnetic Resonance (NMR)

### 1.1.1 The nuclear spin Hamiltonian

NMR is a technique for investigating the chemical and spatial structure of compounds, by exploiting a property, that some nuclei have, called *spin*. The nuclear spin is a quantum effect, and the associated quantum numbers  $I$  are multiple of  $1/2$ . The number of quantum states is  $2I+1$ . In a simplified model, we can imagine the nucleus with spin  $1/2$  spinning around an axis with only two possible orientations, *up* and *down*. Since it is a moving charge, the spinning nucleus generates a magnetic field. When there is no external magnetic field present, there is no energy difference between up and down spin states. If an external magnetic field  $B_0$  is applied, then the spin state that gives rise to a nuclear magnetic field aligned to the external magnetic field has a different, lower, energy from the other spin state. The energy difference  $\Delta E$  is small and falls in the range of radio frequencies (Figure 1.1).



**Figure 1.1:** Range of frequencies exploited by different spectroscopies to investigate the matter

The interaction between a nuclear spin  $I$  and a static magnetic field  $B_0$  is called *Zeeman interactions* and can be described by a second-rank tensor ( $3 \times 3$  matrix)  $\hat{Z}$ :

$$H_{\text{Zeeman}} = I\hat{Z}B_0 \quad (1.1)$$

The energy separation between two different Zeeman states (*Larmor frequency*) is given by:

$$\Delta E = \hbar\nu B_0 \quad (1.2)$$

and is directly proportional to the static magnetic field applied, where  $\hbar$  is the Planck constant and  $\nu$  is the frequency. The population differs for the two states and the ratio is given by Boltzmann distribution. The lowest-energy state, corresponding to a parallel orientation of nuclear spins along the magnetic field, is slightly more populated. As a result, the vector sum of all nuclear spins originates a macroscopic magnetic moment, named *magnetization*.

By irradiation with a suitable energy, it is possible to convert a nucleus from one spin state to the other. Such irradiations with radiofrequencies of short duration are usually named *pulses*.

The energy interaction between the spin  $I$  and the RF pulse is, analogously:

$$H_{\text{rf}} = \mathbf{I} \hat{Z} \mathbf{B}_1 \quad (1.3)$$

The power of NMR spectroscopy relies on the fact that nuclei in different parts of the molecule experience different local magnetic fields according to the molecule's structure, and consequently *resonate* (adsorb energy) at different frequencies. It means that similar atoms in different environments, such as carbon-bonded or oxygen-bonded hydrogen atoms, show different frequencies. This effect of the chemical environment on the absorbed frequency  $\nu$  is called *chemical shift*. By defining a reference frequency  $\nu_0$ , it is possible to express chemical shifts as dimensionless numbers, such that the value is not a function of the external magnetic field  $B_0$ , which allows direct comparison of spectra from NMR spectrometers with different magnetic field strengths and operating frequencies:  $\delta = \frac{\nu - \nu_0}{\nu_0}$ . Chemical

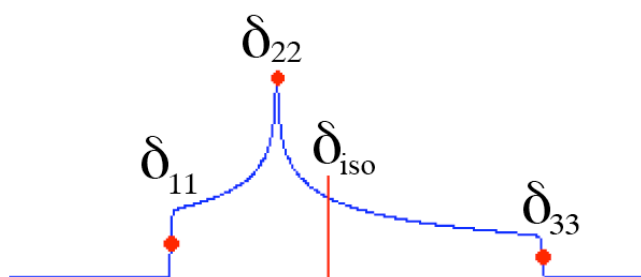


shifts are a potent means of identifying different nuclei within the protein and discriminating between them. These effects of the environment on the nuclear spin depend on the orientation of the sample with respect to the magnetic field and are therefore described by a second-rank *chemical shift* (or *shielding*) tensor  $\hat{\Sigma}$ :

$$H_{CS} = \gamma \mathbf{I}(\hat{\Sigma} \mathbf{B}_0) \quad (1.4)$$

where  $\gamma$  is the gyromagnetic ratio of the spin  $\mathbf{I}$  and  $(\hat{\Sigma} \mathbf{B}_0)$  represents the effective magnetic field experienced locally by the spin  $\mathbf{I}$ .

This dependence of the chemical shift on the orientation is referred to as *chemical shift anisotropy* (CSA). Generally, it is possible to express the chemical shift tensor in a coordinate system, so that all off-diagonal elements vanish. In this principal axes system, the chemical shift tensor is fully described by the three diagonal elements – the principal components ( $\delta_{11}$ ,  $\delta_{22}$  and  $\delta_{33}$ ) – and the three eigenvectors or Euler angles describing the orientation of the principal axes with respect to an arbitrary frame.



**Figure 1.2:** NMR Absorption line under anisotropic conditions for a powder.

Thus,  $\delta_{11}$  corresponds to the direction of least shielding, with the highest frequency, while  $\delta_{33}$  corresponds to the direction of highest shielding, with the lowest frequency:

$$\delta_{11} \geq \delta_{22} \geq \delta_{33} \quad (1.5)$$

The isotropic values,  $\delta_{iso}$ , are the average values of the principal components, and

correspond to the centre of gravity of the line shape:

$$\delta_{iso} = \frac{(\delta_{11} + \delta_{22} + \delta_{33})}{3} \quad (1.6)$$

Under isotropic conditions (in solution, or, in the solid state, upon rapid sample spinning (§ 1.3)), an absorption line in correspondence to  $\delta_{iso}$  will be observed and the Hamiltonian term simplifies into:

$$H_{CS} = \gamma \mathbf{I}(\sigma_{iso} \mathbf{B}_0) \quad (1.7)$$

where  $\sigma_{iso}$  is now a scalar.

Nuclear spins within a molecule may interact with each other. Several interactions of different nature are usually simultaneously active. The *scalar coupling* (or *J coupling*) represents the interaction between nuclei relayed through the electrons in the chemical bonds. The strength of the scalar interaction between two spins  $\mathbf{I}$  and  $\mathbf{S}$  is represented by a coupling constant  $J$ :

$$H_J = \mathbf{I} J \mathbf{S} \quad (1.8)$$

Homonuclear (proton) couplings between two atoms distant by two bonds are largest (~15 Hz), smaller for three-bond couplings (5-10 Hz), and smallest for long-range couplings (~1 Hz). Heteronuclear couplings are substantially larger. For example, proton-nitrogen couplings are around 90 Hz and proton-carbon couplings are on the order of 140 Hz.

A second interaction between nuclear spins that takes place is the *dipolar coupling*. Nuclear spins behave indeed as magnetic moments, and consequently interact with each other through space. The dipolar interaction is orientation-dependent and is therefore represented by a second-rank dipolar tensor  $\hat{D}$ :

$$H_D = \mathbf{I} \hat{D} \mathbf{S} \quad (1.9)$$

The dipolar tensor is a traceless tensor, that is, the sum of its diagonal elements is zero. Therefore, unlike the chemical shift, dipolar interactions are averaged to zero in

solution due to the isotropic tumbling of molecules (§ 1.2.4.3 *Residual Dipolar Coupling*) or in the solid-state under efficient sample spinning at the magic angle. However, the magnetic field that is generated from the dipolar coupling can have large effects on the spin-lattice relaxation rates of the spins (§ 1.2.4.1, *Nuclear Overhauser effect*).

In short, it is common to classify the NMR interactions as a total spin Hamiltonian,

$$\hat{H} = \hat{H}_{Zeeman} + \hat{H}_{rf} + \hat{H}_{CS} + \hat{H}_J + \hat{H}_D + \hat{H}_Q$$

which is simply a sum of all the terms described above. For spins  $I = \frac{1}{2}$  quadrupolar interaction  $\hat{H}_Q$  applies only for nuclei with spins larger than  $\frac{1}{2}$  and is therefore not considered further in this thesis.

Compared to other spectroscopy, a major advantage with NMR is the possibilities to modify at will the Hamiltonian, with few restrictions, adapting it to the special requirements of the problem to be solved. The ease with which the nuclear spin Hamiltonian can be modified depends on the fact that the nuclear interactions are very weak compared to the interactions exploited in other spectroscopic techniques, such as infrared spectroscopy or VIS-UV spectroscopy. In order to override an interaction, an alternative, competitive perturbation of the system has to be applied. To be effective, the energy of the perturbation has to be significantly larger than the interaction to manipulate. Examples of this are *spin decoupling* and *sample spinning*, vital tools in solution and solid-state NMR.

In general, by playing with the energy terms of the Hamiltonian, it is possible to design a large number of NMR experiments, which provide different information: e.g., it is possible to obtain spectra which establish through-bond correlations between nuclei showing scalar coupling, or rather through-space correlations between nuclei

showing dipolar coupling.

### **1.1.2 Magnetization and pulses: 1D and multidimensional NMR**

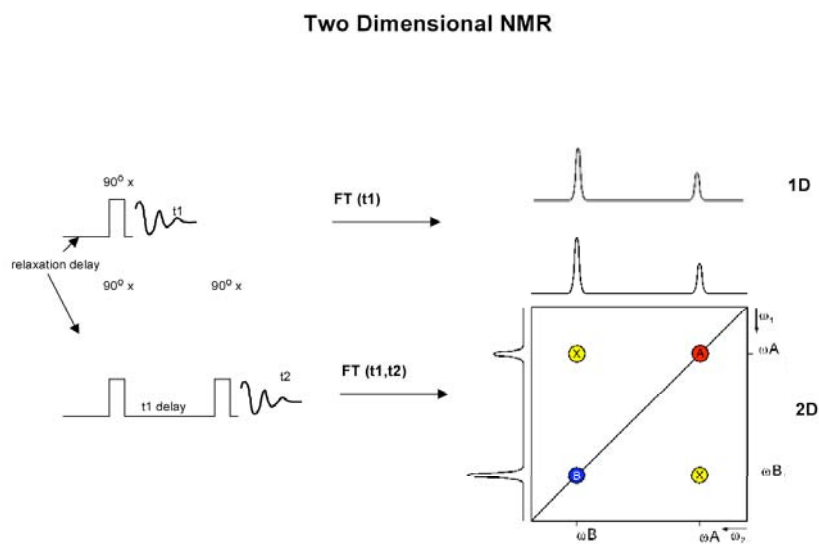
Upon application of appropriate pulses, rotations of the magnetization can be induced. A major property of the magnetization is that its component in the plane perpendicular to the axis of the external magnetic field rotates around this axis with an angular velocity proportional to the Larmor frequency, that is the frequency  $\nu$  of electromagnetic radiation required to excite transitions between Zeeman levels (Equation 1.2). If we now introduce a frame that rotates with this frequency, the magnetization appears static. However, spins are also evolving under their chemical shift (Equation 1.7), which depends on the gyromagnetic ratio ( $\gamma$ ) and the shielding by the chemical environment. The precession of the several components of the magnetization along the static magnetic field constitutes a time-varying magnetic field, which in turn has the property to induce an electromotive force (according to Maxwell it is an electrical field) in a coil appropriately located close to the sample.

The measured intensity of the generated alternating current as a function of time is called FID (Free Induction Decay). In the case of a protein, the bulk magnetic moment is originated by nuclear spins that have in general different chemical shifts. Hence, the FID contains all different resonance frequencies of the nuclei, corresponding to their chemical shifts. By Fourier transformation<sup>a</sup>, the FID can be converted into a frequency function, that is, the NMR spectrum. To study biopolymers like proteins, DNA and RNA, 1D NMR spectroscopy is unable to resolve the frequency of the individual nuclear spins. Hence, additional spectral dimensions were introduced, like, 2D, 3D and even 4D spectroscopic techniques, to

---

<sup>a</sup> Fourier transformation is an important mathematical operation, which allows transforming time domain data into the frequency domain and vice versa.

increase resolution and to extract additional information. The invention of multidimensional spectra was a major leap in NMR spectroscopy. Every 2D experiment, for example, can be described with a simple basic scheme, consisting of a *preparation* period, an *evolution* period  $t_1$  (during which the spins are labeled according to their chemical shift), a *mixing* period where spins are correlated to each other, and finally a *detection* period. For measuring a 2D spectrum, many FIDs are acquired for incremented values of the  $t_1$  delay (*evolution period*) to generate a second frequency dimension. The recorded FIDs are then Fourier transformed with respect to both  $t_1$  and  $t_2$  (as illustrated in Figure 1.3). Signal with two different frequency coordinates (*cross-peaks*) indicate a correlation between two nuclei. The length of the mixing period is commonly referred to as *mixing time*. The intensity of each cross-peak as a function of the length of the mixing time represents the *build-up curve* for the cross-peak.



**Figure 1.3:** Principles of 2D NMR spectroscopy.

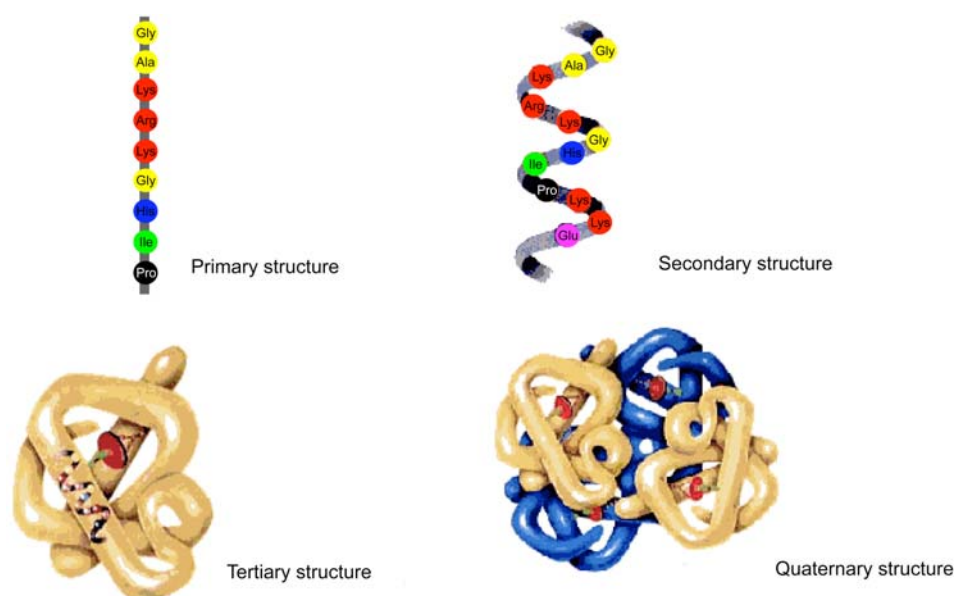
In 3D or 4D spectra, the 2D peaks are dispersed along one or two more orthogonal axes using the chemical shifts of one or two bonded heteronuclei.

## 1.2 Solution NMR

Over the last two decades, solution NMR has become a major technique in structural biology. It has proven to be a powerful technique to investigate protein structures in solution at atomic definition. Compared to X-ray crystallography, solution NMR allows not only to investigate the structures of biopolymers in a nearly physiological environment, but also to determine their dynamic properties. Hence, solution NMR is a vital tool also in protein biophysics, allowing for the study of protein-ligand and protein-protein interactions, protein folding, kinetics and catalysis.

### 1.2.1 Protein structure determination *via* solution NMR

Proteins are an important class of biological macromolecules present in all biological organisms, made up of elements such as carbon, hydrogen, nitrogen, oxygen and sulfur. Proteins are biopolymers of combination of 20 amino acids, characterized by four structural levels as shown in Figure 1.4: (1) primary structure given by amino acid sequence, (2) secondary structure defined by the local conformation of the



**Figure 1.4:** Structural levels of proteins: primary, secondary, tertiary and quaternary

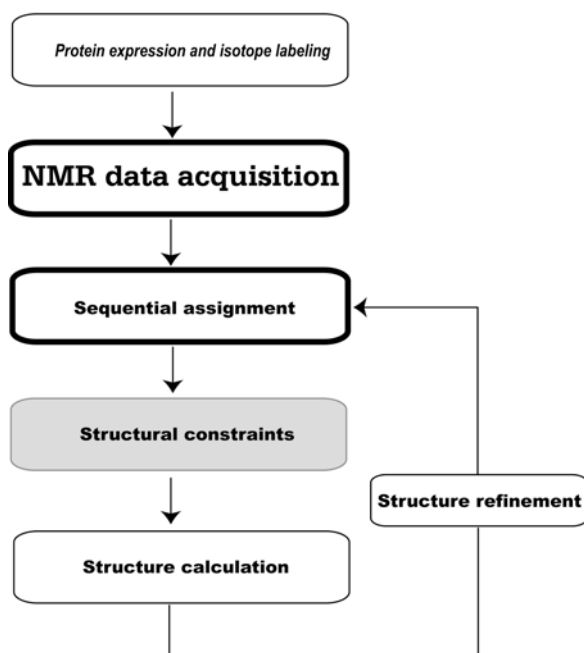
backbone, (3) tertiary structure represented by the spatial proximity of the secondary structural elements and (4) quaternary structure that specifies the packing of several polypeptide chains. To be able to perform their biological function, protein folds into one, or more, specific spatial conformations, driven by number of noncovalent interactions such as hydrogen bonding, ionic interactions, Van der Waals' forces and hydrophobic packing.

In order to understand the functions of proteins at a molecular level, it is often necessary to determine the three dimensional structure of proteins. This is the topic of the scientific field of structural biology, that employs techniques such as X-ray crystallography and NMR spectroscopy, to determine the structure of proteins. There are also many *ab initio* methods which predict the protein structure from the primary sequence based on similarity and comparison. The commonly employed procedure for structure determination of proteins based on solution NMR consists of several consecutive steps, covering sample preparation, NMR data acquisition, peak-picking, resonance assignment, collection of distance- and additional structural restraints (such as Residual dipolar couplings, hydrogen bonds, disulphide bridges) and, finally, structure calculation and structure refinement (shown in Figure 1.5) [Wuthrich 1986]. Till date, this is the most widely followed strategy for determination of protein structures in solution, despite several alternative promising approaches that mostly aim at omission of the cumbersome and biologically irrelevant resonance-assignment [Grishaev and Llinas 2002a].

### **1.2.2 Protein NMR experiments**

There are three 2D spectra that are widely used for the structure determination of proteins with a mass of up to 10kD: 2D COSY, 2D TOCSY and 2D NOESY

[Wuthrich, Wider 1982] are the most commonly used homonuclear (signals of the same isotope (usually  $^1\text{H}$ ) are detected during the evolution periods) correlation experiments exploiting J (scalar) couplings. Apart from protons a protein contains other magnetic active nuclei. For NMR of proteins,  $^{15}\text{N}$  and  $^{13}\text{C}$  are of special importance. The use of these hetero nuclei allows some new features in NMR, which facilitate the structure determination especially of larger proteins (> 100 amino acids).



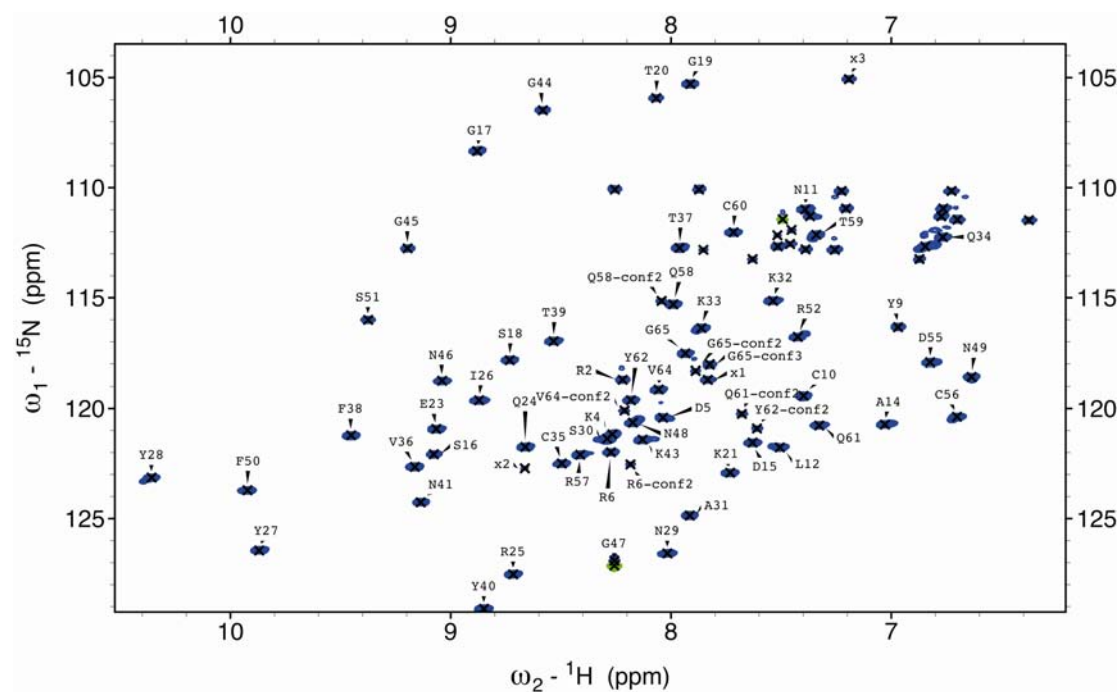
**Figure 1.5:** Strategy of NMR protein structure determination.

The most important heteronuclear experiment is HSQC (Heteronuclear Single Quantum Coherence) (Figure 1.6). Each signal in a HSQC spectrum represents a proton that is bound to a nitrogen atom. The spectrum contains the signals of the  $\text{H}^{\text{N}}$  protons in the protein backbone. Since there is only one backbone  $\text{H}^{\text{N}}$  per amino acid, each HSQC signal represents on single amino acid.

2D spectra of proteins are often crowded with signals. Therefore, these spectra are spread out in a third dimension, so that signals are distributed in a cube instead of a plane. This spread out is achieved by combining HSQC and NOESY in a single 3D



experiment: The NOESY is extended by an HSQC step. The resulting experiment is called 3D NOESY-HSQC. In a similar way, a TOCSY-HSQC can be constructed by combining the TOCSY and the HSQC experiment.

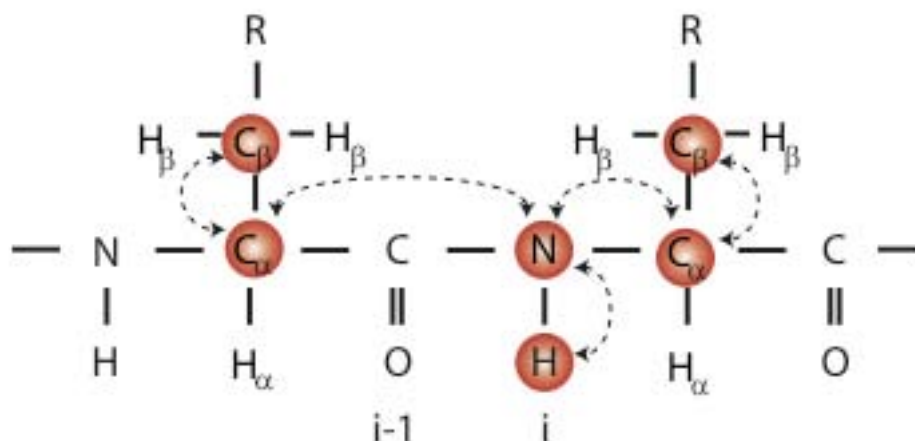


**Figure 1.6:** 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC of Conkunitzin-S2 with some assignments marked

Triple resonance experiments are the method of choice for the sequential assignment of proteins. These experiments are called ‘triple resonance’ because three different nuclei ( $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$ ) are correlated. The experiments are performed on doubly labeled ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) proteins. The names of the triple resonance experiments are very descriptive. The names of all nuclei, which are used for magnetization transfer during the experiment, are listed in the order of their use, bracketing the names of nuclei which are used only for transfer and whose frequencies are not detected. HNCACB, CBCA(CO)NH are the two key experiments to establish correlations along the protein backbone *via* heteronuclear correlations.

## HNCACB

The HNCACB experiment correlates  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  resonances of an amino acid with the  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  resonances of the same residue and  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  resonances of the neighboring residue (Figure 1.7). In this experiment N(i), HN(i),  $\text{C}_\alpha$ (i) and  $\text{C}_\beta$ (i)



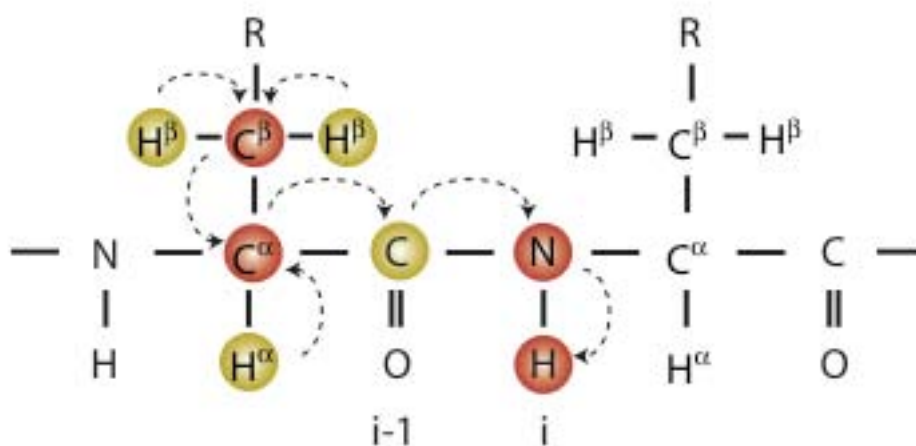
**Figure 1.7:** HNCACB experiment: The magnetization is transferred from the  $^1\text{H}^{\text{N}}(i) \rightarrow ^{15}\text{N}(i) \rightarrow ^{13}\text{C}_\alpha(i)/^{13}\text{C}_\alpha(i-1)$  and then from there to  $^{13}\text{C}_\beta(i)/^{13}\text{C}_\beta(i-1)$ . After which it comes back to  $^1\text{H}^{\text{N}}(i)$  along the same path. The frequencies of  $^1\text{H}^{\text{N}}(i)$ ,  $^{15}\text{N}(i)$ ,  $^{13}\text{C}_\alpha(i)$ ,  $^{13}\text{C}_\beta(i)$ ,  $^{13}\text{C}_\alpha(i-1)$  and  $^{13}\text{C}_\beta(i-1)$  (red) are observed.

resonances of the  $i^{\text{th}}$  amino acid and  $^{13}\text{C}_\alpha(i-1)$  and  $^{13}\text{C}_\beta(i-1)$  resonances of the preceding amino acid are observed. For a medium-sized protein ( $\sim 15\text{KDa}$ ), this experiment alone can provide virtually complete sequential assignment of the  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  resonances, because in addition to the sequential connectivities, the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts provide information on the amino acid type.

## CBCA(CO)NH

The CBCA(CO)NH experiment correlates both the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  resonances of an amino acid residue with the  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  resonances of the preceding residue (Figure 1.8). In this experiment,  $^{13}\text{C}_\beta(i-1)$ ,  $^{13}\text{C}_\alpha(i-1)$ ,  $^{15}\text{N}(i)$  and  $^1\text{H}_\text{N}(i)$  resonances are

observed. This experiment is useful to circumvent the degeneracy between the intra-residue ( $^{13}\text{C}_\beta(i)$  and  $^{13}\text{C}_\alpha(i)$ ) and inter-residue ( $^{13}\text{C}_\beta(i-1)$  and  $^{13}\text{C}_\alpha(i-1)$ ) chemical shifts.



**Figure 1.8:** CBCA(CO)NH experiment: The magnetization is transferred from the  $^1\text{H}_\alpha(i-1)/^1\text{H}_\beta(i-1) \rightarrow ^{13}\text{C}_\alpha(i-1)/^{13}\text{C}_\beta(i-1) \rightarrow ^{13}\text{C}'(i-1) \rightarrow ^{15}\text{N}(i) \rightarrow ^1\text{H}^{\text{N}}(i)$ . The  $^1\text{H}_\alpha$ ,  $^1\text{H}_\beta$  and  $^{13}\text{C}'$  (yellow) act only as relay nucleus, their frequency are not detected. The frequencies of  $^1\text{H}^{\text{N}}(i)$ ,  $^{15}\text{N}(i)$ ,  $^{13}\text{C}_\alpha(i-1)$  and  $^{13}\text{C}_\beta(i-1)$  (red) are observed.

### 1.2.3 Assignment of resonances

The aim of the analysis of NMR spectra is to extract all available structural information of the proteins, such as, inter-atomic distances, torsion angles etc. A plethora of 2D and 3D experiments is present in the literature designed for assigning each observed chemical shift to a nuclear spin, for instance,  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  etc., within the protein. This procedure is referred to as *resonance assignment*. The assignment of the resonances is usually done in two different separate steps. First, resonances within each single amino acid are assigned. As a result, chemical shifts are grouped in different spin systems, one for each amino acid. Second, the different spin systems are connected with each other by exploiting scalar couplings between heteronuclei in the backbone, or, alternatively, through-space correlations between protons [Wuthrich 1986].

From the combination of CBCA(CO)(NH and HNCACB experiments backbone

resonance and the sequential connectivities are obtained. These experiments provide the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts to establish the sequential link between neighboring residues. Furthermore, when both  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts are provided at the same time, it gives important information about the amino acid type and secondary structure (e.g.  $\alpha$ -Helix and  $\beta$ -strand). However, the quality of the spectra frequently makes the assignment process difficult, because of the ambiguity of sequential connections, missing chemical shifts, additional artifact peaks and isolated connectivities due to either missing chemical shifts or the occurrence of prolines in the amino acid sequences, which are not observable due to lack of the  $^1\text{H}^{\text{N}}$  atom. Assigning the resonances is the critical step in the strategy of structure determination, due to limited resolution and spectral overlap. In the end, the quality of the determined structures depends on the number of correct assignments.

#### **1.2.4 Collection of structural restraints**

In protein NMR spectroscopy, structure calculations are usually carried out using, (1) distance between two specific atoms (NOE) obtained from multidimensional NOESY spectra, (2) dihedral angle constraints derived from chemical shifts and J coupling constants. (3) residual dipolar couplings (RDCs). In some cases, disulphide and/or hydrogen bond distance constraints from other experimental data are also included.

##### **1.2.4.1 Distance restraints**

In solution NMR, molecular tumbling produces an isotropic Hamiltonian, where only isotropic chemical shifts and J couplings are left. The dipolar coupling vanishes in solution; nevertheless it affects nuclear spin relaxation. In particular, the relaxation of a spin is influenced *via* dipolar coupling by the presence of another close spin, whose

spin population has been perturbed. These effects on spin relaxation are called Nuclear Overhauser effect (NOE) [Jeener, Meier 1979;Macura and Ernst 1980;Solomon 1955], and have the important property of depending on internuclear distances, hence they provide a way to measure them. NOESY is the 2D/3D solution NMR experiment, which directly exploits this effect to correlate nuclei that are close in space (distance smaller than 5Å). NOE is by far the most important NMR-observable used in determining protein structures. While chemical shifts and J-coupling constants give local structural information, NOE data can relate atoms that are far apart in the series of chemical bonds connecting the biomolecule, but are close in space.

Using a first-order approximation, the NOE cross-peak intensities or volume  $V$  is proportional to the relaxation rates and may be expressed as [Jeener, Meier 1979]:

$$V = \langle r_{ij}^{-6} \rangle f(\tau_c) \quad (1.10)$$

It depends on the distance between the two nuclei  $i$  and  $j$ ,  $r_{ij}$ , as well as on the rotational correlation time  $\tau_c$  (describing the Brownian tumbling motion of the protein in solution). By measuring cross-peak intensities  $V$ , *distance restraints* can be derived. It can be estimated in the 2D NOESY, 3D  $^{15}\text{N}$ -NOESY HSQC and 3D  $^{13}\text{C}$ -NOESY-HSQC spectra.

Distances are derived from the spectra after calibration against NOE signals for known distances (such as distances in elements of secondary structure elements) and grouped into few classes. An upper and lower bound of distance is assigned to each class. The lower bound is often set to the sum of the van der Waals radii of the two protons. In this procedure, all non-sequential signals that are visible in the NOESY spectra have to be assigned, the number of which easily exceeds 1000 in a medium sized protein (ca. 120 amino acids). It is distinguished between cross peaks of protons

NOE class	Distance [Å]	Upper bound [Å]
very strong	2.3	2.5
strong	2.8	3.1
medium	3.1	3.4
weak	3.5	3.9
very weak	4.2	5.0

**Table 1.1:** NOE distance bounds used in protein structure determination.

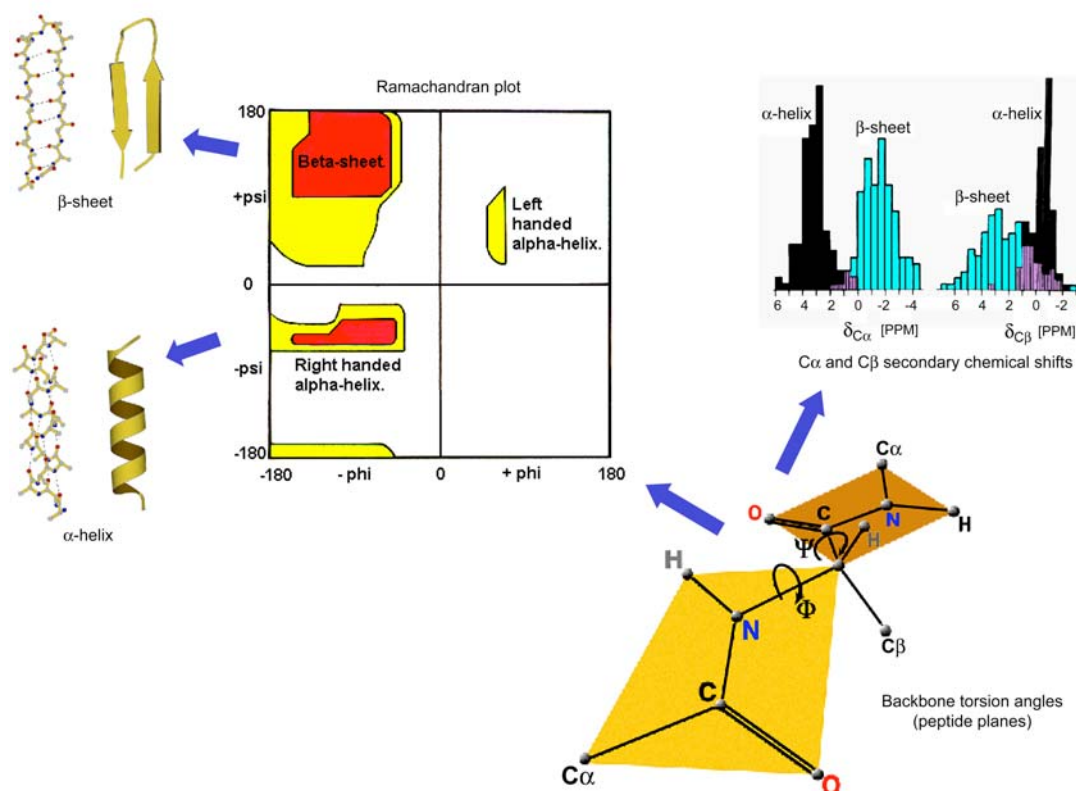
no more than five amino acids apart in the protein sequence (medium range NOE's) and those, which are more than five amino acids apart (long range NOE's). The former are mainly indicative of the protein backbone conformation and are used for secondary structure determination, whereas the latter are an expression of the global structure of the protein and therefore contain the main information used for tertiary structure calculation. NOE assignment is one of the most time consuming and difficult parts for the structure determination due to the ambiguity of the peaks and overlaps in the NOESY spectra.

#### 1.2.4.2 Dihedral angle restraints

*Dihedral angle restraints* represent a second important group of restraints that is possible to derive from NMR spectra. Dihedral angles in the protein backbone influence directly the three-bond scalar couplings  $^3J$  constants *via* Karplus equation [Karplus 1963]:

$$^3J(\tau) = A\cos^2(\tau) + B\cos(\tau) + C \quad (1.11)$$

The constants A, B and C depends on the particular nuclei involved in the covalent



**Figure 1.9:** Elements of protein secondary structure ( $\alpha$ -helix and  $\beta$ -sheet) defined by the backbone dihedral angles ( $\phi$  and  $\psi$ ) and empirical correlations with  $^{13}C_{\alpha}$  and  $^{13}C_{\beta}$  secondary chemical shifts [Cornilescu, Delaglio 1999; Wishart 1994].

bonds and change depending on which dihedral angle ( $\phi$ ,  $\psi$  or  $\chi_1$ ) the letter  $\tau$  stands for. Traditionally,  $\phi$  angles are obtained by measurement of  $^3J_{H^{\alpha}H^N}$  constants, whereby  $\chi_1$  angles from  $^3J_{H^{\alpha}H^{\beta}}$  constants. Hence, after measuring J couplings, it is possible to restrain the backbone torsion angles. Alternatively as shown in Fig. 1.9, dihedral angle restraints can be obtained exploiting the information contained in the secondary chemical shifts of the heteronuclei in the backbone (N,  $C_{\alpha}$  and CO), which in turn depend on  $\phi$  and  $\psi$ . The *secondary chemical shift* is defined as the difference between the measured chemical shift and the chemical shifts in a random-coil (e.g., unstructured) protein and represents that component of the chemical shift that is induced by the three-dimensional structure. Dihedral angle restraints can be derived automatically with the TALOS [Cornilescu, Delaglio 1999] program, which relates

the backbone secondary shifts with an internal database of high-resolution protein structures. TALOS strategy was used in this work to derive dihedral angle restraints.

### **1.2.4.3 Residual Dipolar Couplings**

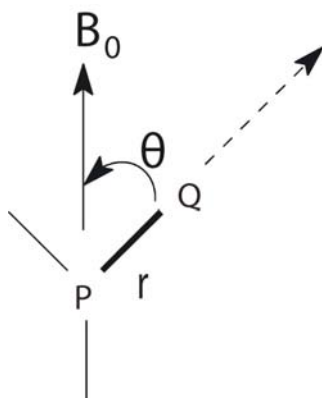
Recently, a different source of structural information has been used for structure determination: the magnetic dipole-dipole coupling between spin  $\frac{1}{2}$  nuclei (e.g.  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ). These dipolar couplings contain information on the orientation of inter-nuclear vector relative to the magnetic field and distance between the involved nuclei. They have proven invaluable for improving the accuracy of macromolecular NMR structure determination, for independently validating their accuracy, for refining crude homology of proteins and for defining intermolecular interactions. In addition, dipolar couplings can be used to search for homologous structures or substructures in a structure database and potentially could replace the time consuming regular NOE data collection and analysis process [Permi and Annila 2001; Meiler, Peti and Griesinger 200; Gaemers and Bax 2001].

The use of residual dipolar couplings (RDCs) to enhance the information available from high-resolution NMR spectra has a long history. Its roots can be traced to the substantial amount of NMR done in magnetic-field-aligned liquid crystals in the early 1960's by Saupe and Englert [Saupe and Englert 1963]. They demonstrated that an organic molecule dissolved in a nematic liquid crystalline phase exhibits quite strong alignment when placed in an NMR magnet. Thoughts about application to biomolecules in solution arose more than 15 years ago with the observation that isolated molecules with sufficient anisotropic susceptibilities would adopt slightly non-isotropic orientational distributions when placed in a high magnetic field. The major breakthrough with respect to any potentially routine use of dipolar couplings



for biomolecular structure determination was the demonstration that adjustable degrees of alignment could be achieved by placing the molecule under investigation into a dilute, aqueous liquid crystalline phase of dihexanoyl phosphatidylcholine (DHPC) and dimyristoyl phosphatidylcholine (DMPC) [Sanders and Prestegard 1993; Tjandra and Bax 1997].

The application described above relies on the fact that anisotropic contributions to nuclear interactions, e.g. dipolar interaction between pairs of magnetically active spin  $\frac{1}{2}$  nuclei, do not average to zero when the molecules of have a preferred orientation. The net alignment of the molecules of interest, which can be introduced by a liquid crystalline medium, is on the order of  $10^{-3}$  and is fundamental to the success of RDC based studies.



**Figure 1.10:** Dipolar coupled spin pair. The bond length  $r$  is assumed to be fixed and the primary variable is the angle  $\theta$  between magnetic field  $B_0$  and the internuclear vector.

Dipolar couplings are potentially quite large interactions, caused by the magnetic field produced by one nucleus (e.g. nucleus Q) affecting the energy of another nucleus (e.g. nucleus P) (Figure 1.10). The components orthogonal to the magnetic field  $B_0$  have a negligible effect on the total magnitude of the vector sum of the external and the dipolar field. Thus only the z component of the dipolar field of nucleus Q will change the resonance frequency of nucleus P by an amount that depends on the internuclear distance and on the orientation of the internuclear vector relative to  $B_0$ . For a fixed

orientation of the vector, nuclear spin P can decrease or increase the total magnetic field at nucleus Q, depending on whether P is parallel or antiparallel to  $B_0$ . In an ensemble of molecules half of the P nuclei will be parallel to  $B_0$  and the other half antiparallel, and Q will show two resonances separated in frequency by,

$$D^{PQ} = D_{\max}^{PQ} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \quad (1.12)$$

where  $\theta$  is the angle between the internuclear vector and  $B_0$ , the brackets  $\langle \rangle$  denote time or ensemble averaging, and

$$D_{\max}^{PQ} = -\frac{\mu_0}{4\pi} \gamma_P \gamma_Q \frac{\hbar}{2\pi} \left\langle \frac{1}{r_{PQ}^3} \right\rangle \quad (1.13)$$

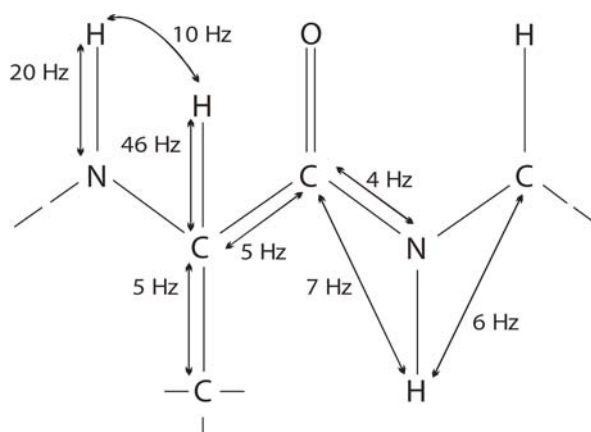
is the doublet splitting that applies for the case where  $\theta = 0$  and where  $\mu_0$  is the magnetic permeability in vacuum;  $\hbar = h/(2\pi)$ , in which  $h$  = Planck's constant;  $\gamma_P$ , gyromagnetic ratio of nucleus P;  $\gamma_Q$ , gyromagnetic ratio of nucleus Q;  $r_{PQ}$ , the internuclear distance between nucleus P and Q. Equation (1.12) shows the dipolar splitting,  $D^{PQ}$ , provides direct information on the angle  $\theta$ .

Knowing  $\theta$  for a bonded pair of nuclei, e.g.  $^1\text{H}$ - $^{15}\text{N}$ , can be very useful in defining a molecular structure. The brackets around the  $\theta$  dependent term, however, denote a time average. Normally in solution we assume isotropic sampling (that the time average results a molecular tumbling that uniformly samples directions in space). Consequently, for any pair of nuclei, the  $3\cos^2\theta - 1$  term averages to zero and no residual dipolar coupling can be measured [Prestegard 1998]. For this reason NMR spectroscopists were relegated to measure dipole-dipole interaction indirectly through spin relaxation based phenomena such as NOE until 1997, when Tjandra and Bax demonstrated the use of a liquid crystalline medium to introduce a tunable degree of alignment.

One important property of residual dipolar couplings between two nuclei P and Q that have scalar coupling is that dipolar coupling ( $D^{PQ}$ ) is added to the isotropic coupling constant ( $J^{PQ}$ ). Therefore, when measuring the effective dipolar coupling under anisotropic conditions ( $E^{PQ}$ ), the spin-spin couplings need to be subtracted from the measured coupling to obtain the desired dipolar coupling.

$$E^{PQ} = J^{PQ} + D^{PQ} \quad (1.14)$$

Often the one-bond dipolar couplings (e.g. NH,  $C_\alpha H_\alpha$ ,  $C_\alpha C'$ , C'N) are measured, but also two-bond and three-bond dipolar couplings can be measured. One-bond dipolar couplings are easier to interpret because the inter-atomic distance is known and the magnitude of the dipolar interaction is relatively large [De Alba and Tjandra 2002].



**Figure 1.11:** The predicted maximum absolute residual dipolar contributions for different nuclei in the protein main-chain at 0.1% levels of alignment.

The NMR methods used for measuring residual dipolar couplings can be divided into two general categories: frequency resolved methods (J-resolved) and intensity-based experiments (J-modulated). In frequency resolved methods the separation of the peaks is measured in a frequency domain. In intensity based experiments the coupling is extracted from the resonance intensity rather than from the experimental splitting. The Principle underlying the J-modulated experiments is to pass the observed signal

through a period in which the intensity is modulated by a known function of the spin-spin coupling [Brunner 2001; Prestegard, Al-Hashimi 2000]. An intensity-based experiment is particularly useful in case of overlap problems in the described coupled spectra because of the increased number of signals due to the doublet splitting. In addition, J-modulated spectroscopy is a good alternative when the coupling of interest is small compared with the line width, which is often the case with  $^3J$  couplings.  $^{15}\text{N}$ - $^1\text{H}$  splitting was measured under isotropic and anisotropic conditions using 2D IPAP- $^{15}\text{N}$ - $^1\text{H}$  HSQC experiments [Ottiger, Delaglio 1998].  $^1\text{H}$ - $^{15}\text{N}$  RDCs were extracted by subtraction of the  $^1J_{\text{NH}}$  scalar coupling measured for the isotropic sample [Tjandra and Bax 1997]. Alternatively,  $^1\text{H}$ - $^{15}\text{N}$  RDCs were obtained from the  $^{15}\text{N}$ - $^1\text{H}$  splittings of a modified 3D TROSY-HNCO experiment [Chou, Delaglio 2000]. The quantitative J correlation method was used to determine one-bond dipolar  $^{13}\text{C}'$ - $^{15}\text{N}$  couplings. Usually, these couplings are measured together with the  $^1\text{H}$ - $^{15}\text{N}$  RDCs [Vijayan and Zweckstetter 2005].  $^{13}\text{C}'$ - $^{13}\text{C}_\alpha$ -RDCs are measured from the splitting in the  $^{13}\text{C}'$  dimension of a  $^{13}\text{C}_\alpha$  coupled 3D HNCO experiment.

### **1.2.5 Structure calculation *via* Simulated Annealing**

Following the sequential assignment and assignment of NOESY spectra, NOE-derived distance restraints, dihedral angle restraints, residual dipolar couplings (additional restraints like hydrogen bonds, disulphide bridges, if present) are used to calculate an ensemble of structures. However, the experimentally determined distances and torsion angles by themselves are not sufficient to fully characterize a protein structure, as they are based on a limited number of proton-proton distances. Only the knowledge of empirical input data, such as bond lengths of all covalently attached atoms and bond angles, enables a reasonably exact structure determination.

The ultimate goal of structure calculation algorithms is to operate an optimization of the simultaneous agreement of an atomic model with the observed data and with *a priori* chemical information. Most algorithms used for structure calculation aim to find the global minimum of a hybrid energy function  $E$  (*target function*), which includes *a priori* chemical knowledge of the system (the force field, defining bond lengths, bond angles, improper angles and non-bonded interactions) and experimental data (the structural restraints):

$$\begin{aligned}
 E &= E_{chem} + E_{exp} = \sum_i w_i E_i = \\
 &w_{cov\,alent} E_{cov\,alent} + w_{angle} E_{angle} + w_{float} E_{float} + w_{vdW} E_{vdW} + \quad (1.15) \\
 &w_{NOE} E_{NOE} + w_{dihed} E_{dihed} + w_{RDC} E_{RDC} + \dots
 \end{aligned}$$

$E_{chem}$  in Equation 1.15 contains energy terms for covalent bonds, bond angles, chirality, planarity and nonbonded repulsion are all approximated by a harmonic function. In contrast, non-covalent van-der-Waals attraction forces and electrostatic interactions are simulated by an inharmonic Lennard-Jones potential or Coulomb potential, respectively. Nonbonded repulsions are described by a quartic potential [Stein, Rice 1997].  $E_{exp}$  contains energy terms describing the experimental structural restraints.

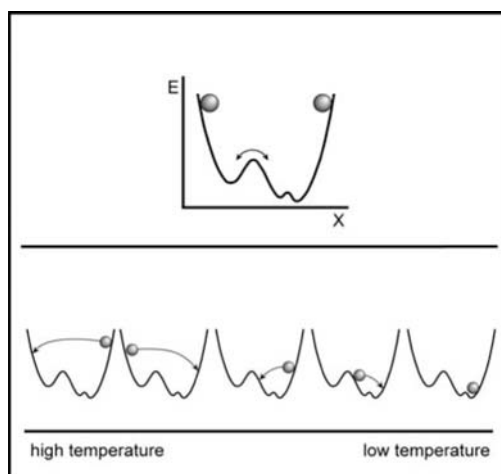
The target function  $E$  is a function of many molecular parameters, most importantly of atomic coordinates. The large number of variables makes this function very complex and originates the so-called ‘multiple minima problem’: the target function contains many local minima in addition to the global minimum. The standard minimization methods [Brunger and Nilges 1993] tend to steer the system into local minima and frequently fail to reach the global minimum if the starting model is far away from the correct one. The sampling of a larger conformational space can be achieved by using Simulated Annealing (SA) optimization techniques [Brunger, Adams 1997; Brunger and Nilges 1993; Kirkpatrick, Gelatt 1983], which has the

important property of enabling the system to overcome local minima (Figure 1.12). Monte Carlo-based and molecular dynamics-based SA have been used in structure prediction, molecular modeling, X-ray refinement and NMR structure determination for many years and have had a large impact in structural biology.

Structure calculation based on SA molecular dynamics consists in the integration of Newton's equation of motion:

$$\frac{d^2\vec{r}_i}{dt^2} = -\frac{c}{m_i} \frac{\nabla E}{\vec{r}_i}, \quad (1.16)$$

where  $\vec{r}_i$  is the vector of Cartesian coordinates,  $m_i$  the mass of the atom  $i$ ,  $c$  is a constant and  $\nabla E$  is the gradient of the target function. The global minimum of the target function  $E$  is searched by reducing the temperature of the system after a high-temperature phase during a molecular dynamics simulation. In this context, the parameter 'temperature' has no physical meaning, but is simply a measure of the probability of the macromolecule to cross an energy barrier (i.e. its kinetic energy). Temperature control is performed by coupling the system with a thermal bath [Berendsen, Postma 1984]. As a result, temperature coupling will cause "heat" (kinetic energy) to be added or removed from the system, as it is needed to maintain the temperature.

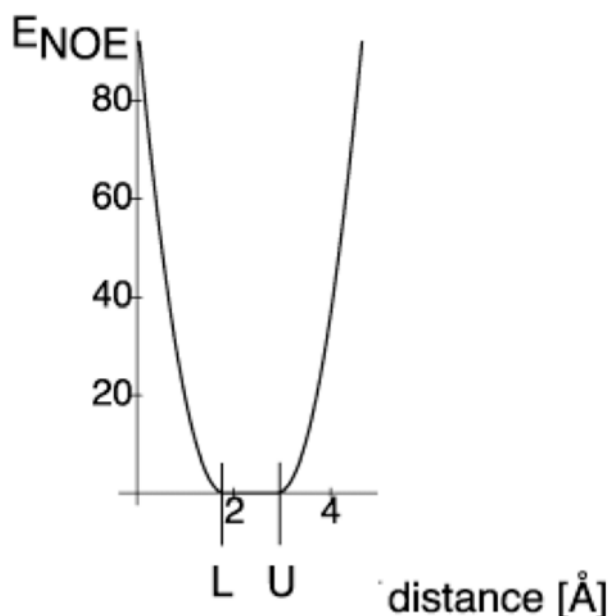


**Figure 1.12:** Compared to standard minimisation methods, SA allows the system to overcome local energy barriers, reducing the risk of the molecule to get trapped in local minima of the target function.

Energy terms for distance- and angle restraints are usually provided by specifying a lower and upper limit, L and U. During simulated annealing, a violation is generated whenever the distance between the two atoms is not contained between the boundaries, leading to an increase in energy of the system. Typically, when restraints from manually assigned peaks are provided as input, energy terms for NOE-based distance restraints and dihedral angle restraints are present in the form of flat-bottomed parabolic functions (Figure 1.13). The flat bottom is delimited by L and U:

$$E_{\text{dis\_res}} = \begin{cases} (L-d)^2 & \text{if } d < L \\ 0 & \text{if } L < d < U \\ (d-U)^2 & \text{if } U < d \end{cases} \quad (1.17)$$

$$E_{\text{dih\_res}} = \begin{cases} (L-\phi)^2 & \text{if } \phi < L \\ 0 & \text{if } L < \phi < U \\ (\phi-U)^2 & \text{if } U < \phi \end{cases}$$

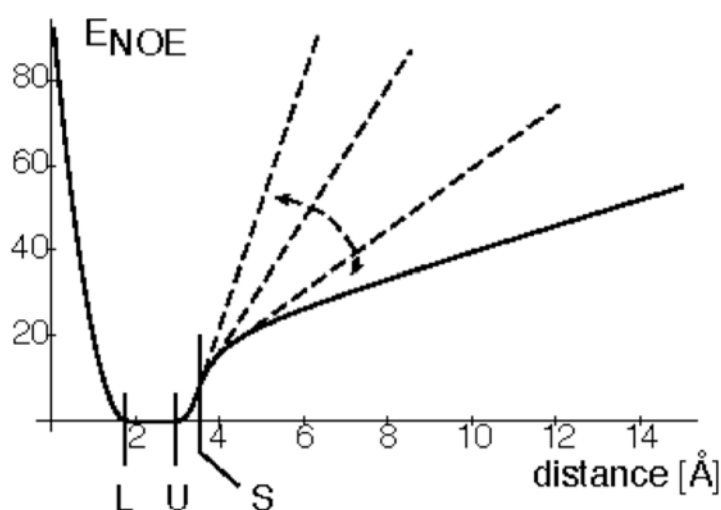


**Figure 1.13:** Standard “flat-bottom” potential used for NOE-derived distance restraints.

Structure calculations using iterative methods for automated cross-peak assignment [Linge, Habeck 2003;Linge, O'donoghue 2001] are characterized by large violations

of experimental distance restraints in the atomic model during the early iterations. Hence, a linear function is substituted in Equation 1.14 for large violations (Equation 1.15), in order to avoid numeric instabilities arising from the high penalty for large violations (Figure 1.14).

$$E_{\text{dis\_res}} = \begin{cases} (L - d)^2 & \text{if } d < L \\ 0 & \text{if } L < d < U \\ (d - U)^2 & \text{if } U < d < S \\ A(d - U)^{-1} + B(d - U) + C & \text{if } d > S \end{cases} \quad (1.18)$$



**Figure 1.14:** Soft potential for NOE-derived restraints used in iterative methods for automated cross-peak assignment.

Similar minimization procedure is used for RDCs and other experimental restraints during the structure calculation. Without the experimentally determined distance, torsion angles constraints and RDCs from the NMR spectra, the protein molecule can adopt a huge number of conformations due to the free rotation around its chemical bonds (except for the peptide bond, the N - C $^{\alpha}$  bond and C $^{\alpha}$  - C' bond). All these possible conformations are summed up in the so-called conformational space. Therefore, it is important to identify as many constraints as possible from the NMR



spectra to restrict the conformational space as much as possible, thus getting close to the true structure of the protein. In fact, the number of constraints employed is more important than the accuracy of proton-proton distance. The precision with which a structure can be calculated is directly related to the number of experimental restraint used to generate it. Structure of low resolution may be obtained with as few as five restraints per residue, whereas the most precise structures obtained from NOE constraints alone may have up to 15 restraints per residue.

The result of simulation is a minimum energy protein structure, but it cannot be excluded that this structure is stuck in a local minimum without ever reaching global minimum, which is marginally lower in energy. Therefore, about twenty different starting structures with random folds are used, which reach their final structure via different paths in energyhyperspace. These resulting structures are iteratively re-used as starting structures for another SA with slightly changed input protocols, until no further reduction in global energy is observed and the structures converge in conformational space.

After the structural calculations a family of structures is obtained instead of an exactly defined structure. This family spans out a relatively narrow conformational space. Therefore, the quality of a NMR structure can be defined by the mean deviation of each structure from this family (RMSD) from an energy minimized mean structure, which has to be calculated previously. The smaller the deviations from this mean structure, the narrower the conformational space. Another definition of RMSD is to compare pair wise the structures of a family and calculate the mean of these deviations.

The RMSD is different for different parts of the protein structure: Regions with flexible structure or without secondary structure (loops) show a larger deviation than those with rigid and well defined secondary structure. This higher RMSD in

loops results in first instance from the smaller number of distance constraints for these parts of the protein structure. Additionally it can originate from real flexibility, but this diagnosis can only be confirmed by measuring the relaxation times for the protein.

### **1.3 Solid-State NMR**

Solid-state NMR has long been the poor cousin of liquid state NMR in structural biology so it is important to know why one might be interested in using such a technically difficult method. Considerable advances have meant that solid state NMR is now readily applicable to biological systems and a wide variety of information can be elucidated with solid-state NMR. How far the technique can be pushed, its limitations and how they might be removed is an area of intense discussion and has seen rapid progress in recent years.

Solution-state NMR has constantly developed methods since the first protein-structure determination in 1985 by K. Wüthrich and currently a large toolbox of multidimensional NMR experiments, labeling schemes, automated or semi-automated programs exist for sequential assignment and structure calculation. Proteins up to 30kDa can be routinely studied by solution-state NMR and successful applications have been demonstrated for proteins with molecular weights up to 100kDa. This limitation in solution-NMR appears due to the increase of the correlation time with the molecular weight ('correlation time problem') that shortens the transverse relaxation time and degrades spectral resolution. The problem becomes even more apparent in the case of membrane proteins where the size of the lipid-protein assemblies (micelles, bicelles, liposomes) can easily reach or exceed the above limits even for small proteins.

The situation is different in solid-state NMR. As implied by the name, solid-

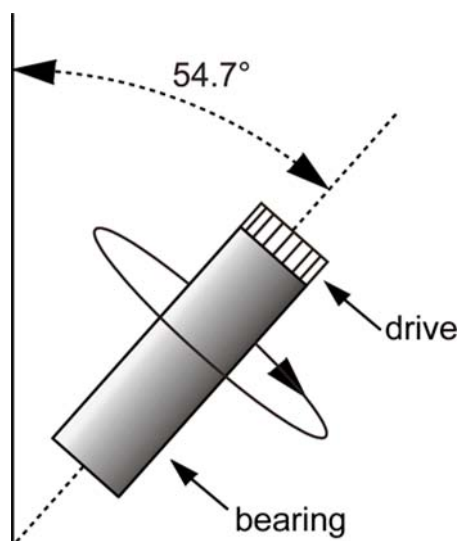
state NMR is used for solids, nearly solids, or strongly anisotropic or immobilized proteins, that is, samples where the reorientation of a molecule is very slow or non-existent, as is the case for microcrystalline powders of soluble proteins or membrane proteins reconstituted in lipid bilayers. Therefore, solid-state NMR is a technique to investigate proteins whenever the dissolution of the sample is not possible or desirable and the X-ray approach is also not feasible. This may involve: 1) insoluble proteins, 2) proteins which aggregate in solution, 3) membrane-bound proteins, often insoluble or structurally altered in aqueous solution, in their synthetic or natural membrane environment. Compared to solution NMR, solid-state NMR has the disadvantage of a much lower resolution in terms of achievable line-widths relative to the chemical shift range. Conversely, molecular tumbling is not a band-narrowing mechanism in solid-state NMR; therefore, the size of the protein does not influence the line-width and, thus, is not an intrinsic limitation on resolution. In addition, fast internal dynamics may improve resolution.

The chemical shift tensor and the dipolar interaction tensor all contain an orientation dependent factor:

$$3\cos^2\theta - 1 \quad (1.19)$$

Most of the experimental techniques used in solid-state NMR have been developed to achieve (1) line-narrowing or (2) signal enhancement. Intrinsic broad lines due to the anisotropy of NMR interactions and short T2 relaxation times characterize solid-state NMR. Anisotropically broadened lineshapes affect mainly randomly oriented (sometimes called ‘powder’) samples that are also the most interesting for biological applications. There are line-narrowing methods that average-out anisotropic interaction either in real space by magic angle spinning (MAS) [Mehring 1979; Maricq 1979] or in spin-space [Lee and Goldburg 1965]. Motional averaging, an intrinsic property of liquids, can be simulated by rapid sample spinning

at the magic angle at  $\theta_m = 54.733^\circ$  (Figure 1.15) relative to the magnetic field  $B_0$ , which leads to the vanishing of the anisotropy factor contained in Equation 1.19.



**Figure 1.15:** Principle of the magic-angle spinning (MAS) technique. The rotor containing the sample is tilted at  $54.7^\circ$  with respect to the external magnetic field

Anisotropic interactions, like dipolar couplings and the anisotropic part of the chemical shift can be averaged out if the spinning is sufficiently rapid. Thus, also for solid-state samples it is possible to obtain, with some limitations, an isotropic Hamiltonian, giving rise to relatively narrow signals with absorption line shapes. In practice, to completely average out the anisotropies, the spinning rate has to be significantly larger than the interaction strength (in Hz). This can be easily achieved with interactions such as homonuclear and heteronuclear dipolar couplings not involving protons ( $< 10$  kHz).

In biological solid-state NMR, homonuclear and heteronuclear correlations between  $^{13}\text{C}$  and  $^{15}\text{N}$  are most commonly measured. Such correlations are mostly mediated by dipolar interactions. The current strategy for uniformly labeled proteins in MAS solid-state NMR relies heavily in  $^{13}\text{C}$  (detected) and  $^{15}\text{N}$  nuclei, making isotope labeling mandatory [Kainosho 1997]. While resonance assignment methods in the solution-state make extensive use of the J-coupling to direct polarization along the

polypeptide chain [Ikura, Kay and Bax 1990;Ikura, Kay, Tschudin 1990], solid-state NMR mixing schemes can employ in principle both through-bond or through-space transfer mechanism to achieve sequential resonance assignments under MAS conditions. The residue type are identified in ( $^{13}\text{C}$ ,  $^{13}\text{C}$ ) homonuclear correlation spectra based on unique spin connectivities and distinct  $^{13}\text{C}$  chemical shifts of each residue, while sequential assignment is obtained from combination of heteronuclear NCACX and NCOCX (where X denotes second spin involved in magnetization transfer step) spectra that link neighbor residues via the common amide  $^{15}\text{N}$  nucleus [Baldus 2002;Tycko 1996]. Sequential assignments can be probed also in  $^{13}\text{C}$ - $^{13}\text{C}$  correlation spectra under specific, so called '*weak coupling conditions*' [Seidel, Lange 2004], or in  $^{15}\text{N}$ - $^{15}\text{N}$  correlations when possible [Van Rossum, Castellani 2003]. To improve resolution and obtain long range constraints in spin diffusion [Bloembergen 1949;Suter and Ernst 1985] spectra, special  $^{13}\text{C}$  labeling schemes have been proposed [Castellani, Van Rossum 2002]. Finally, sequential resonance assignments can also be obtained from proton-proton [Lange, Luca 2002] or  $^1\text{H}$ - $^{13}\text{C}$  mediated [Etzkorn, Bockmann 2004] correlation spectroscopy.

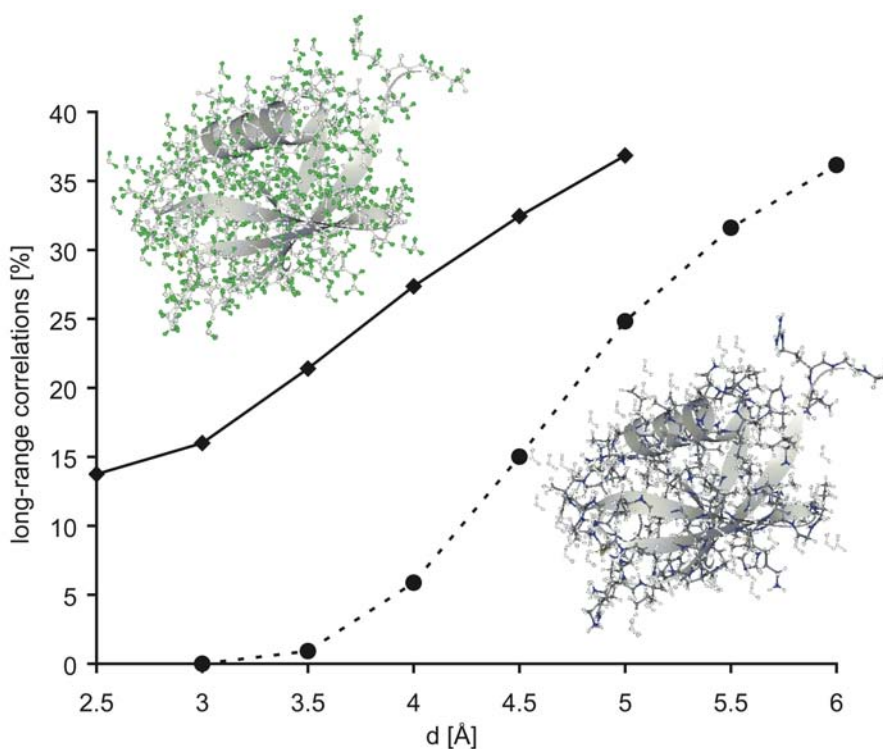
### **1.3.1 Protein structure determination *via* solid-state NMR**

Recently, a simple methodology was developed to investigate protein structures from solid-state MAS NMR data (§ Chapter 3). This approach bears many parallels with the strategy used for structure determination of proteins *via* solution NMR (§ 1.2.1).

#### **1.3.1.1 Collection of structural restraints**

For structure determination of proteins by MAS NMR it is first necessary to perform a resonance assignment. Subsequently, distance measurements are done, for which

there are different approaches in solid-state NMR. In liquid state NMR, the collection of distance restraints for structure calculation protocol relies on the NOE between pairs of protons. In solid-state NMR, for uniformly  $^{13}\text{C}$  labeled samples, homonuclear  $^{13}\text{C}$ - $^{13}\text{C}$  and heteronuclear  $^{13}\text{C}$ - $^{15}\text{N}$  correlations are more commonly used for extracting distance information, from proton-driven-spin-diffusion (PDS) [Castellani, Van Rossum 2002] types of spectra, as the large size of dipolar proton-proton interactions makes it challenging to measure  $^1\text{H}$ - $^1\text{H}$  correlations. The cross-peak volumes in the PDS spectra show either an  $r^{-3}$  or  $r^{-6}$  dependency. By assigning



**Figure 1.16:** Statistical analysis of all proton-proton and carbon-carbon distance constraints up to distance  $d$  in ubiquitin. The fraction of long-range distances (residue difference  $> 4$ ) is shown.

cross-peaks, a large number of inaccurate  $^{13}\text{C}$ - $^{13}\text{C}$  distance restraints in the range of 2-7.5 Å are derived. As an alternative to the measurement of carbon or nitrogen distances, a set of multidimensional NMR experiments for structure elucidation of proteins under MAS conditions has been proposed, which relies on the measurement of proton-proton distances [Heise, Seidel 2005] (used in this thesis). It uses proton

homonuclear transfers, bracketed by  $^{13}\text{C}$  and/or  $^{15}\text{N}$  evolution times for the sake of spectral resolution. This scheme is able to produce high-resolution spectra connecting carbon or nitrogen spins through space [Lange, Luca 2002; Lange, Seidel 2003]. Compared to  $^{13}\text{C}$ - $^{13}\text{C}$  or  $^{13}\text{C}$ - $^{15}\text{N}$  correlations in proteins, proton-proton interactions are more abundant and, because of their peripheral nature, contain a larger fraction of long-range contacts for the shortest internuclear distances [Baldus 2007].

Finally, from the conformational dependent backbone chemical shifts ( $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$ ), a set of dihedral angle restraints can be obtained with the TALOS program (§1.2.4.2).

### **1.3.1.2 Structure calculation *via* Simulated Annealing**

Structure calculation approach is similar as described in (§ 1.2.5).

## Chapter 2

### Fast High-resolution Protein Structure Determination by Using Unassigned NMR Data

#### 2.1 Introduction

Structural studies of proteins are critical for understanding biological processes at the molecular level. NMR spectroscopy provides high-resolution structural information of biomolecules in near-physiological conditions. Although significant improvements were achieved in NMR spectroscopy in the last 20 years [Wuthrich 2003], the increase in genome sequencing data has created a need for rapid and efficient methods of NMR-based structure determination [Abbott 2005;Montelione, Zheng 2000]. In contrast to X-ray crystallography, growing of suitable crystals is not required for NMR spectroscopy, and the time-consuming steps are data acquisition and data analysis. NMR data acquisition can be accelerated significantly, when highly sensitive spectrometers equipped with cryogenic probes are combined with newly developed methods for sampling chemical shifts in multidimensional NMR experiments. For approximately 1mM protein solutions, the NMR data, which are required for three-dimensional (3D) structure determination, may be obtained in one to nine days, depending on the size and properties of the protein [Liu, Shen 2005]. Therefore, data analysis and in particular the requirement to assign chemical shifts and inter-residue correlations to specific atoms is the major bottleneck of rapid NMR-based structure determination.

Good progress has been made towards prediction of high-resolution 3D protein structures from amino acid sequences. Using improved conformational sampling methods, structural models were predicted that had a  $C_{\alpha}$ -root mean square



deviation (rmsd) to the native structures of less than 1.5 Å [Bradley, Misura 2005]. However for many proteins the quality of predictions was insufficient and prediction of the detailed structures of functionally relevant parts remains a formidable challenge.

These limitations of *ab initio* structure prediction may be overcome, when a limited number of easily accessible NMR data is combined with *ab initio* methods. For example, medium-resolution folds can be determined rapidly from chemical shifts and residual dipolar couplings (RDCs) [Delaglio, Kontaxis 2000], especially when combined with the *de novo* structure prediction algorithm Rosetta [Jung, Sharma 2004;Meiler and Baker 2003;Rohl and Baker 2002]. To obtain high-resolution protein structures, experimental distance information is required. The distance information can be extracted from Nuclear Overhauser effect (NOE) spectra and can be obtained from NOESY (NOE spectroscopy) spectra.

NOESY spectra provide large number of distances between protons that are close in space (not farther than 5 Å, approximately). Unfortunately, the interpretation of NOESY spectra is never straightforward. Signal degeneracy leads to several assignment options for many peaks. With larger proteins, these effects are usually so dramatic, that a complete manual interpretation of the spectra is hampered. Consequently, the automation of this step, towards high-resolution structure, has become an issue of great interest in the last years. The distance restraints can be extracted from NOESY spectra with little manual intervention, when assignment of NOE peaks and structure calculation are performed iteratively. Today, several software packages exist for a more or less fully automated NOE assignment. ARIA [Linge, Habeck 2003;Linge, O'donoghue 2001], CANDID [Herrmann, Guntert 2002], DYANA [Guntert, Mumenthaler 1997], KNOWNOE [Gronwald, Moussa 2002], NOAH [Mumenthaler and Braun 1995;Mumenthaler, Guntert 1997],

AUTOSTRUCTURE [Moseley, Monleon 2001;Moseley and Montelione 1999] and have recently been reviewed [Guntert 2003].

All these approaches require the sequence-specific resonance assignment and lists with cross-peaks from NOESY spectra as input. These automated procedures follow the same general scheme of iteratively assigning the multi-dimensional NOESY data and structure calculation. A key-step towards complete automation was the introduction of Ambiguous Distance Restraints (ADR) [Nilges 1993;Nilges 1995] into structure calculation strategies based on simulated annealing. ARIA uses such an ambiguous constraint strategy, involving multiple ambiguous distance constraints for each ambiguous NOESY peak. The program CANDID, combined with DYANA, also uses ambiguous constraint strategies but, in addition, employs network anchoring and constraint-combination methods, minimizing deleterious effects when this correctness assumption is not satisfied. Initial structures are first built using ambiguous constraint strategies and then iteratively refined. Therefore, it is important to obtain a well-converged initial fold for the rest of the cycles to achieve correct structures. To converge to the correct structure, however, a nearly complete and error-free manual assignment of backbone and side chain chemical shifts is essential [Jee and Guntert 2003].

Alternatively, it may be possible to obtain a 3D protein structure in the absence of any chemical shift assignments from the distance information provided in NOESY spectra. Several attempts have been made to devise a strategy for NMR protein structure determination that circumvents the tedious chemical shifts assignment step. The underlying idea of assignment-free NMR structure calculation methods is to exploit the fact that NOESY spectra provide distance information even in the absence of any chemical shift assignments. This proton-proton distance information can be exploited to calculate a spatial proton distribution. Since there is

no association with the covalent structure at this point, the protons of the protein are treated as a gas of unconnected particles. Provided that the emerging proton distribution is sufficiently clear, a model can then be built into the proton density in a manner analogous to X-ray crystallography in which structural model is constructed into the electron density. The most recent approach to NMR structure determination without chemical shift assignment is the CLOUDS [Grishaev and Llinas 2002a; Grishaev and Llinas 2002b] approach. For the first time, the feasibility of the method has been demonstrated using experimental data rather than simulated data sets. The CLOUDS method relies on precise and abundant inter-proton distance constraints calculated via a relaxation matrix analysis of set of experimental NOESY cross peaks. It showed that assignment-free NMR structure calculation could successfully generate 3D protein structures from experimental data. Nevertheless, in the course of a de novo structure determination it may not be straightforward to produce a NOESY peak list of the completeness and quality used for these test calculations. Unfortunately, it was assumed that the NOEs could be unambiguously identified, as it is always not feasible for the automation. Up to now, no robust method to determine high-resolution protein structure from unassigned NMR data has been reported.

Herein, we present a method termed, FastNMR (**F**ast **S**tructure determination by **NMR**) [Korukottu, Bayrhuber 2007], which enables automatic, high-resolution NMR structure determination of domain-sized proteins from unassigned NMR data. We combined MARS (automatic backbone assignment program), RosettaNMR (*ab initio* structure prediction program from limited number of NMR data) and CYANA (automatic NOE assignment and structure calculation program) to facilitate the high-resolution structure determination. Simultaneous backbone assignment and fold determination was followed by automatic NOE assignment and structure calculation

using the initial fold. FastNMR differs from other approaches in that it starts from unassigned chemical shifts, NOEs and residual dipolar couplings (RDCs), avoids wrong structures by cross-validation, works for real experimental data, requires only a limited number of NMR spectra, requires no manual assignment of chemical shifts or inter-residue correlations and produces high-resolution ( $< 1 \text{ \AA}$ ) structures.

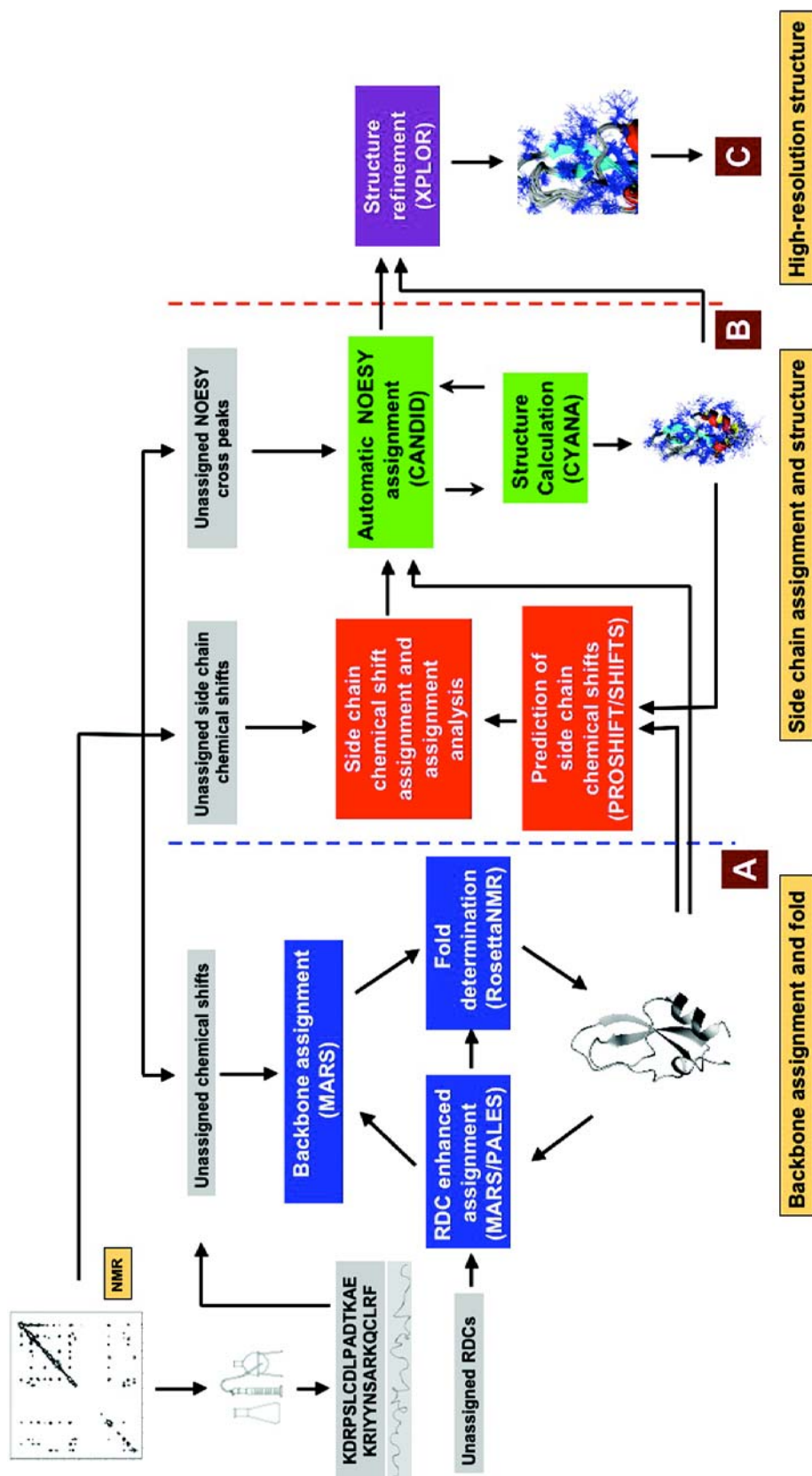
This new method was tested on two proteins and for the first time using an automatic structure determination approach, the de novo structure of the 65-residue protein cone snail neurotoxin Conkunitzin-S2 (Conk-S2) was determined automatically.

## 2.2 Materials and Methods

A simplified flowchart of FastNMR is shown in Figure 2.1. In short, FastNMR determines the backbone resonance assignment and backbone fold of the protein by iterative assignment and structure calculation, in stage A, starting from unassigned backbone chemical shifts and residual dipolar couplings (RDCs). In stage B, experimental side chain chemical shifts are assigned by comparison with values that are back-calculated from the protein backbone fold obtained in stage A. In combination with the stage A backbone conformation, the assigned side chain chemical shifts enables an automatic NOE assignment that is performed iteratively with structure calculation. In stage C, the NOE-based structure is refined in the presence of all the assigned restraints including RDCs.

### 2.2.1 Input of unassigned NMR data

Experiments and the structural constraints that we obtain from the observed experimental NMR parameters are listed in Table 2.1. To obtain lists of unassigned experimental data, NMR spectra can be analyzed manually or in an automated fashion. Here, all NMR spectra were processed using NMRPipe [Delaglio, Grzesiek 1995]. Referencing of spectra, peak picking and peak grouping were performed using the public domain graphics program Sparky 3 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). All chemical shifts and RDCs were grouped according to the common  $^1\text{H}^N$ ,  $^{15}\text{N}$  chemical shifts, which can be observed in the spectra used here.



**Figure 2.1:** FastNMR flow chart. FastNMR starts from unassigned chemical shifts, RDCs and NOE peak lists. It consists of three major phases: (A) Backbone assignment and fold determination. (B) Side chain assignment, NOE analysis and NOE-based structure calculation. (C) Refinement to a high-resolution structure. All steps are done automatically by FastNMR. Interactive work is confined to the processing of NMR spectra and the preparation of input lists.

<b>NMR experiment</b> (for review see [Bax and Grzesiek 1993])	<b>NMR parameters</b>	<b>Structural constraints</b>
3D HNCACB [Wittekind and Mueller 1993] 3D CBCA(CO)NH [Grzesiek and Bax 1992]	$^{13}\text{C}_\alpha$ , $^{13}\text{C}_\beta$ , $^1\text{H}_\text{N}$ , $^{15}\text{N}$ chemical shifts	$\phi/\psi$ Backbone dihedral angles
3D H(CCO)NH-TOCSY [Montelione, Lyons 1992] 3D C(CO)NH-TOCSY [Grzesiek, Anglister 1993]	$^{13}\text{C}$ , $^1\text{H}$ chemical shifts of aliphatic side chain	
2D $^{15}\text{N}$ - $^{13}\text{C}^\gamma$ , $^{13}\text{C}$ - $^{13}\text{C}^\gamma$ difference experiments [Hu and Bax 1997; Hu, Grzesiek 1997]	$^3\text{J}_{\text{CC}^\gamma}$ , $^3\text{J}_{\text{NC}^\gamma}$ scalar couplings	$\chi_1$ angles
3D RDC-TROSY-HNCO [Chou, Delaglio 2000; Vijayan and Zweckstetter 2005] 3D RDC-CBCA(CO)NH [Chou and Bax 2001; Grzesiek and Bax 1992; Vijayan and Zweckstetter 2005]	$^1\text{D}_{\text{N-H}}$ , $^1\text{D}_{\text{C-N}}$ , $^1\text{D}_{\text{C}\alpha\text{-H}\alpha}$ , $^1\text{D}_{\text{C}\alpha\text{-C}}$ RDCs	Orientation of one-bond vectors
3D $^{15}\text{N}$ -edited NOESY 3D $^{13}\text{C}$ -edited NOESY	NOE peak volumes	Distance restraints

**Table 2.1:** List of NMR experiments and structural constraints obtained from respective measured NMR parameters.

### **2.2.2 Backbone fold and assignment**

Starting from lists of unassigned chemical shifts, FastNMR obtains an initial backbone assignment of the target protein by using the MARS assignment method [Jung and Zweckstetter 2004b]. Based on the assigned residues and the corresponding chemical shifts and RDCs, molecular fragments are selected from a database of high-resolution protein crystal structures. The fragments are subsequently assembled to a compact structural model using the RosettaNMR algorithm [Meiler and Baker 2003]. In the next step, the initial backbone assignment is improved by comparison of experimental RDCs with values back-calculated from the assembled structural models [Jung and Zweckstetter 2004a]. Due to the improved assignment, more chemical shifts and RDCs are available for a second round of fragment selection and assembly using RosettaNMR. The iterative process of automatic assignment and structure calculation is continued until backbone assignment and conformation have converged. 20 lowest-energy folds were used for further analysis [Jung, Sharma 2004].

The next step is to get from a protein backbone to a 3D structure including side chains. For this aim, side chains were built onto the protein backbone using the program SCWRL [Canutescu, Shelenkov 2003] to the 20 lowest-energy backbone fold obtained above.

### **2.2.3 Side chain assignment and structure**

In stage B, experimental side chain chemical shifts are assigned by comparison of experimental values with those computed from the stage A protein fold, which now has all the side chains. FastNMR uses the program SHIFTS [Osapay and Case 1994] to compute proton chemical shifts using density functional theory. Carbon chemical shifts are predicted using the program PROSHIFT [Meiler 2003], which relies on



artificial neural networks. Predicted chemical shifts from all of the 20 lowest-energy conformers, which were transferred from stage A to stage B, were averaged. Experimental chemical shifts are then matched to the predicted values based on the minimal chemical shift difference.

Proton-proton distance constraints are derived from unassigned NOESY cross peaks detected in 3D  $^{13}\text{C}$ - and  $^{15}\text{N}$ -edited NOESY experiments. The structure of the protein backbone and assignment of the backbone and side-chain chemical shifts are subsequently used for NOE assignment. Using the Caliba macro in the program CYANA version 2.1 [Guntert 2003]. NOESY cross peaks are assigned automatically with the NOE assignment algorithm CANDID [Herrmann, Guntert 2002] as implemented in CYANA. In case of Conk-S1 and Conk-S2, the chemical shift tolerance window of protons and non-hydrogen atoms was set to 0.04 ppm (0.02 ppm in the direct dimension) and 0.5 ppm, respectively. The corresponding values for ubiquitin were 0.05 ppm for protons and 0.5 ppm for non-hydrogen atoms.

The automatically assigned distance restraints, the protein fold obtained from stage A of FastNMR, backbone dihedral angles obtained from the program TALOS [Cornilescu, Delaglio 1999] on the basis of backbone chemical shifts, and  $\chi_1$  torsion angles are used as input for the structure calculation using the CYANA torsion angle dynamics algorithm [Herrmann, Guntert 2002]. CYANA, however, is not only used for NOE assignment. All distance constraints involving proton chemical shifts, which could not be unambiguously assigned by comparison with predicted values, are treated as ambiguous NOEs. In this way, the assignment of side chain chemical shifts is partly done as part of the automated NOE assignment. Using the NOE-based 3D structure the prediction of chemical shifts is improved and a second round of automated NOE assignment is performed.

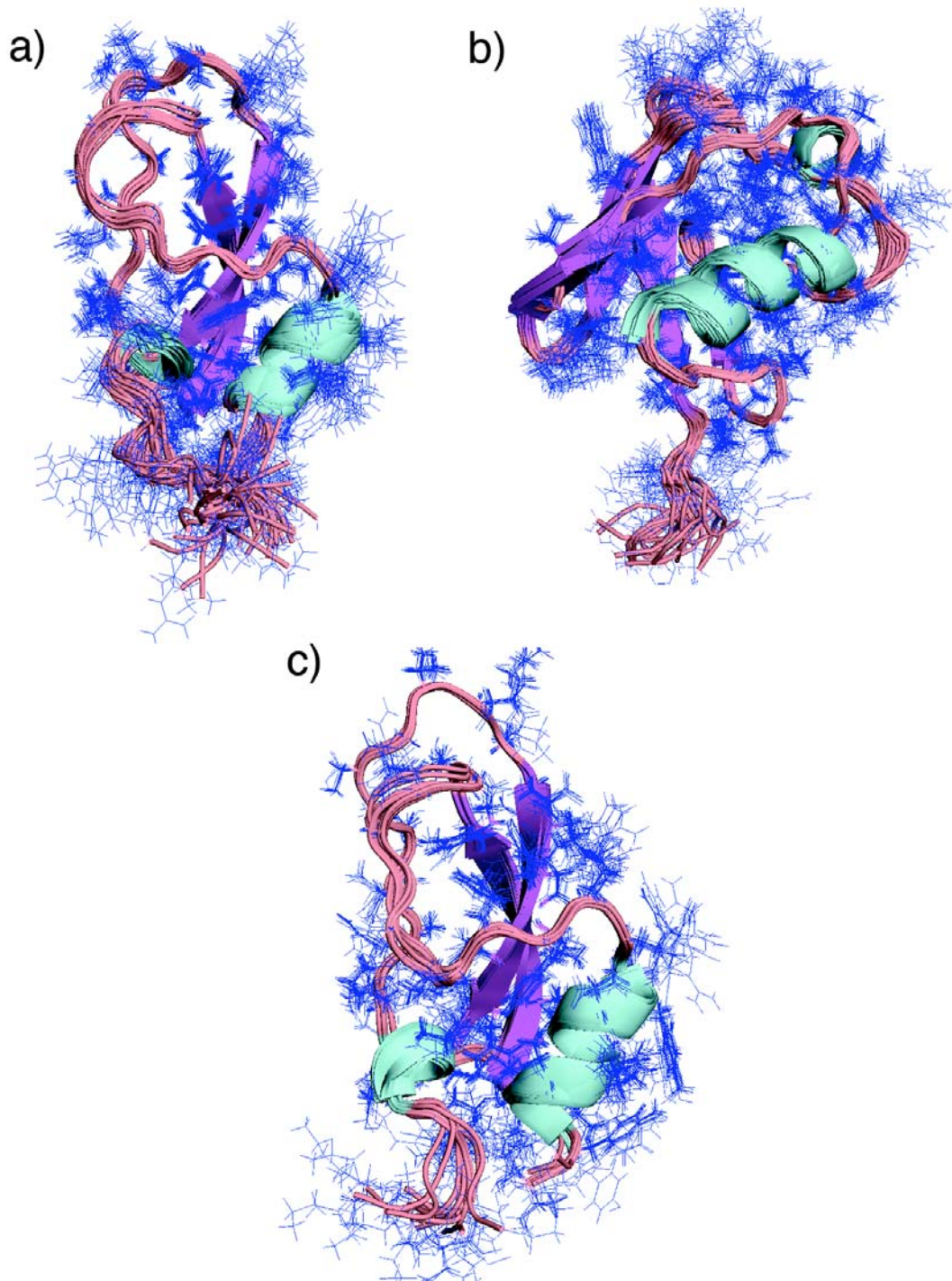
#### **2.2.4 Refinement of 3D structure to high-resolution**

Finally, in stage C of FastNMR, NOE distance restraints, which were unambiguously assigned in stage B, are evaluated. All of the NOE cross peaks assigned to protons with experimental chemical shifts are accepted. In addition, NOE distance restraints assigned to protons with predicted chemical shifts are used in the final structure refinement only if the same proton (predicted chemical shift) is assigned to two or more NOE peaks and the experimental chemical shift of the two NOE peaks differ by less than 0.1 ppm. In combination with the backbone conformation that is already established prior to the NOE analysis, this allows the determination of high-resolution protein structures. Finally, in combination with all other available experimental NMR information, the refined list of unambiguously assigned NOEs is used to refine the stage B structure in the presence of explicit water using Xplor-NIH [Schwieters, Kuszewski 2003]. The refinement follows a previously developed protocol [Nederveen, Doreleijers 2005], which was supplemented by empirical backbone-backbone hydrogen bonding potential [Grishaev and Bax 2004]. After the refinement, the 20 lowest-energy structures are selected and the automatic FastNMR structure determination is completed.

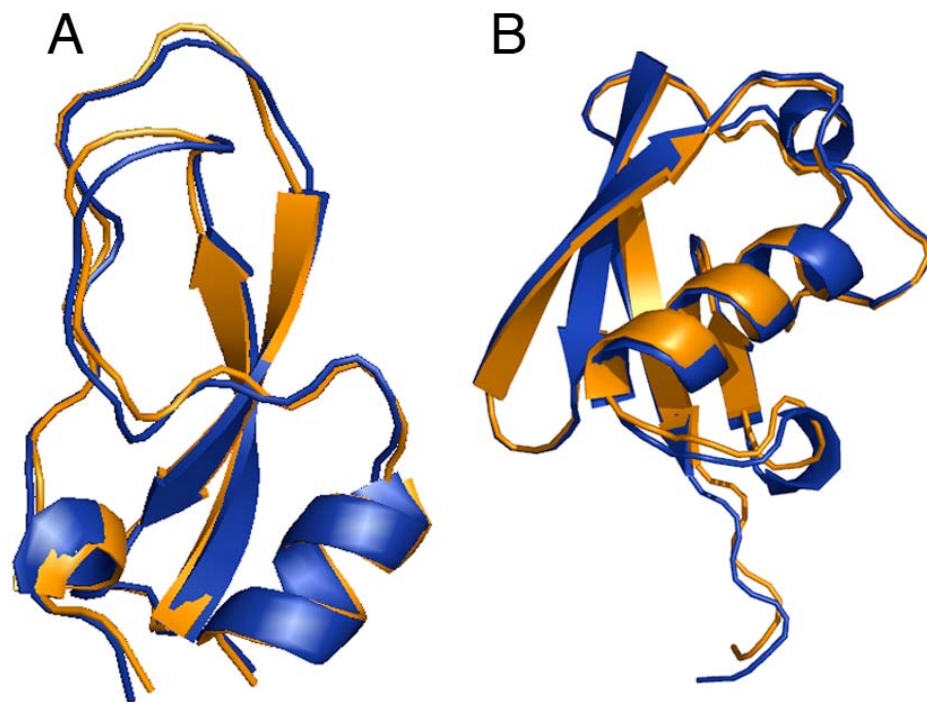
## 2.3 Results and Discussions

FastNMR was tested on the 60-residue neurotoxin Conkunitzin-S1 (Conk-S1) and the 76-residue protein Ubiquitin. For both proteins, 3D structures as well as chemical shift assignments are known allowing best evaluation and validation for FastNMR determined structures [Bayrhuber, Graf 2006; Bayrhuber, Vijayan 2005; Cornilescu, Marquardt 1998]. In addition, the high-resolution structure of the 65-residue toxin Conkunitzin-S2, which has 72% sequence identity to Conk-S1, was determined *de novo* by FastNMR. Neither NMR data nor a 3D structure were previously available for Conk-S2.

Figure 2.2 shows the 20 lowest energy structures of the three proteins calculated by FastNMR and Table 2.2 show that for all three proteins FastNMR calculated high-resolution 3D structures from unassigned NMR data. The spread (root-mean square deviation within the ensemble) in the ensemble of 20 lowest-energy structures was below 0.7 Å for the backbone (N, C<sup>α</sup> and C) and below 1.4 Å for all heavy atoms. Side chains are well defined. It is evident that the ensemble of structures satisfies the experimental constraints and exhibits only small deviations from the idealized covalent geometry. Majority of the dihedral angles appear on the most-favored region of the Ramachandran plot. Table 2.2 also compares the FastNMR structure with conventionally determined structure. The manually and automatically determined structures were of similar energy. The FastNMR structures of Conk-S1 and Ubiquitin deviate by 0.4 Å and 0.6 Å from the conventionally determined NMR structures [Bayrhuber, Graf 2006; Bayrhuber, Vijayan 2005; Cornilescu, Marquardt 1998]. Figure 2.3 shows the direct comparison of the two FastNMR determined structure with the manually determined structures. The FastNMR calculation of each protein was completed in less than 24 hours.



**Figure 2.2:** FastNMR 3D structure of a) Conk-S2, b) Ubiquitin, and c) Conk-S1.



**Figure 2.3:** Comparison of FastNMR structures with structures deposited in the Protein Data Bank (PDB). (A) conk-S1. (B) Ubiquitin. FastNMR structures are shown in orange, structures deposited in the PDB (PDB codes: 1D3Z and 2CA7 for ubiquitin and Conk-S1, respectively) are shown in blue.

**Table 2.2.** Structural statistical data of the investigated peptides. <sup>[a]</sup>

	Conk-S1		Ubiquitin		Conk-S2
	2CA7	FastNMR	1D3Z <sup>[b]</sup>	FastNMR	FastNMR
Number of NOEs <sup>[c]</sup>	551	464	1744 <sup>[d]</sup>	635	570
Long-range	113	72	731	119	160
Medium-range	73	57	291	106	79
Short-range	365	335	722	410	331
Number of dihedral angles	126	126	98	127	167
Violations > 5°	3 ± 1	3 ± 1	0	1 ± 1	2 ± 1
Number of RDCs	190 <sup>1</sup>	190 <sup>1</sup>	200 <sup>2</sup>	200 <sup>2</sup>	138 <sup>3</sup>
RDC types <sup>[e]</sup>	1,2,3,4	1,2,3,4	1,2,3	1,2,3	1,2,3
Energy [kcal mol <sup>-1</sup> ]	-1267.2	-1387.9	-2767.9	-2247.6	-1025.0
	Ramachandran plot [%]				
Most-favored	88.2	87.5	95.0	95.5	84.5
Disallowed	2.0	1.0	0.0	0.0	1.9
	Coordinate precision <sup>[f]</sup> [Å]				
Backbone atoms	0.6	0.7	0.3	0.4	0.6
All heavy atoms	1.2	1.4	0.9	1.1	1.2

[a] Statistics for ensembles of 20 structures. [b] Structure recalculated based on experimental restraints of 1D3Z. [c] None of the structures exhibited distance violations greater than 0.5 Å. [d] Only 58% of the long-range NOEs are non-redundant. [e] 1, 2, 3, 4 refer to <sup>1</sup>D<sub>N-H</sub>, <sup>1</sup>D<sub>C-N</sub>, <sup>1</sup>D<sub>C $\alpha$ -C</sub>, <sup>1</sup>D<sub>C $\alpha$ -H $\alpha$</sub> , RDCs, respectively. [1] 46, 50, 47 and 47 RDCs of types 1, 2, 3 and 4, respectively. [2] 69, 67 and 64 RDCs of types 1, 2 and 3, respectively. [3] 54, 54 and 30 RDCs of types 1, 2 and 3, respectively. [f] Defined as the average RMS difference between the final 20 FastNMR structures and the mean coordinates for residues 2-72 (ubiquitin), 3-60 (Conk-S1) and 5-60 (Conk-S2).

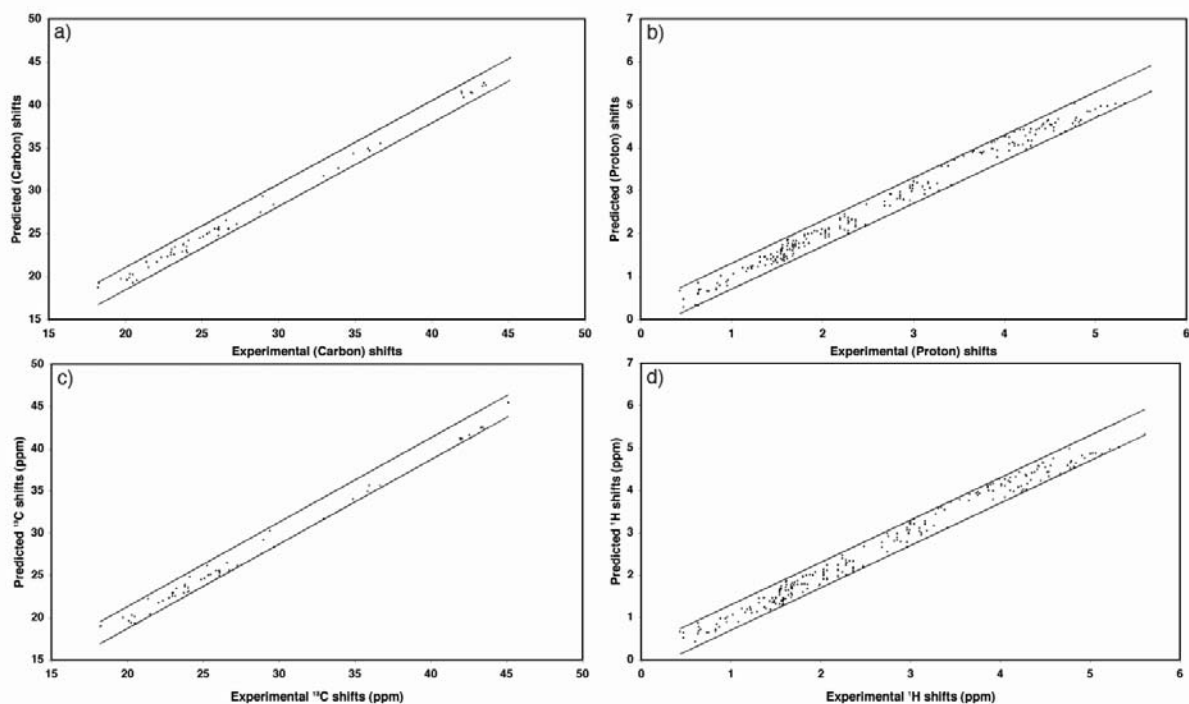
### **2.3.1 Backbone assignment and fold determination by FastNMR**

In case of Conk-S1, all 55 residues, for which signals in 2D  $^1\text{H}/^{15}\text{N}$  correlation spectra were detected, had been correctly assigned at the end of stage A of FastNMR and nearly complete backbone assignment was obtained for Ubiquitin. With nearly complete backbone assignment and a large number of RDCs (190 and 200, respectively, for Conk-S1, Ubiquitin), low-resolution folds of both the proteins were obtained at the end of stage A.

### **2.3.2 Side chain assignment and assignment analysis**

In FastNMR, the assignment of experimental side-chain resonances is performed automatically by comparison with values predicted from protein folds established at stage A in the FastNMR calculation. Assignments are considered only when the difference between predicted and measured chemical shift is less than 0.3 ppm for protons and 1.3 ppm for carbons. To justify the parameters, tests were done, on Ubiquitin, by predicting the chemical shifts from the protein fold obtained at the end of stage A of FastNMR, that are deviating by 2.4 and 3.5 Å from the native structure, 1D3Z. The predicted values were compared to the experimental chemical shifts obtained from 3D HCCONH- and CCONH-TOCSY experiments recorded on the protein. These tests show that the root-mean-square deviation (rmsd) between predicted and experimental chemical shifts is 0.19 ppm for protons and 1.1 ppm for carbons visible in HCCONH- and CCONH-TOCSY spectra [Meiler 2003; Osapay and Case 1994] (Figure 2.4).

In addition, when two experimentally observed  $^1\text{H}$  chemical shifts, which belong to the same residue (as established by HCCONH and CCONH-TOCSY spectra), differ by less than 0.3 ppm, any NOE to either of the two shifts is considered



**Figure 2.4:** Comparison of predicted chemical shifts with experimental values obtained from 3D HC(CO)NH- and C(CO)NH-TOCSY experiments recorded on ubiquitin. (A)  $^{13}\text{C}$  chemical shifts were predicted using an artificial neural network as implemented in the software PROSHIFT [Meiler 2003] from a structure that deviates  $2.4 \text{ \AA}$  from the native structure (PDB code: 1d3z [Cornilescu, Marquardt 1998]). Note that the comparison does not include  $\text{C}_\alpha$  and  $\text{C}_\beta$  chemical shifts as these are assigned based on the chemical shift information obtained from backbone triple-resonance NMR experiments. (B)  $^1\text{H}$  chemical shifts were predicted using empirical rules as implemented in the software SHIFTS [Osapay and Case 1994] from the same structure as in A). (C) Same as A) but using a structure that deviates  $3.5 \text{ \AA}$  from the 1d3z structure. (D)  $^1\text{H}$  chemical shifts were predicted using empirical rules as implemented in the software SHIFTS from the same structure as in C). The straight lines bracket the range  $\pm 0.3 \text{ ppm}$  between measured and predicted  $^1\text{H}$  chemical shifts and  $\pm 1.3 \text{ ppm}$  between measured and predicted  $^{13}\text{C}$  chemical shifts.



**Table 2.3 NMR chemical shift assignments obtained by FastNMR**

Protein	# of residues	Chemical shifts <sup>a</sup>	Backbone assignment (%) <sup>b</sup>	Side chain assignment (%) <sup>c</sup>			
				Start <sup>d</sup>		Final <sup>e</sup>	
				<sup>13</sup> C	<sup>1</sup> H	<sup>13</sup> C	<sup>1</sup> H
Conk-S1	60	C <sub>α</sub> /C <sub>β</sub>	100 (97)	100.0/ 0 (70.0/ 0)	89.7/ 10.3 62.8/ 7.2	100.0/ 0 70.0/ 0	93.7/ 6.3 65.6/ 4.4)
Ubiquitin	76	C <sub>α</sub> /C <sub>β</sub>	95 (92)	90.0/ 0 (50.0/ 0)	88.5/ 11.5 57.4/ 7.4	90.0/ 0 50.0/ 0	96.7/ 3.3 62.8/ 2.1)

<sup>a</sup> Chemical shifts used to establish sequential connectivity.

<sup>b</sup> Percentage of assigned experimental spin systems. Assignment scores with respect to the total number of assignable residues are given in brackets.

<sup>c</sup> Percentage of measured side chain chemical shifts that were assigned by FastNMR. Unambiguous (“numerator”) and ambiguous (“denominator”) assignment scores are listed. Assignment scores with respect to the total number of side chain atoms are given in brackets.

<sup>d</sup> Start: prior to starting the NOE analysis in stage B of FastNMR.

<sup>e</sup> Final: upon completion of FastNMR. As part of the iterative NOE assignment and structure calculation, the NOE assignment is selected that produces the best-fit between the experimental NOE distance restraints and those back-calculated from the structure. Thus, during stage B the number of unambiguously assigned proton chemical shifts was increased from 89.7 to 93.7% for Conk-S1 and from 88.5 to 96.7% for ubiquitin. The assignments are highly useful for further NMR studies, such as relaxation time measurements to probe backbone and side chain motions.

as ambiguous (indistinguishable assignment) during the automated NOE assignment.

By using this approach, all experimentally observed carbon chemical shifts of Conk-S1, Conk-S2 and ubiquitin were assigned unambiguously. <sup>1</sup>H chemical shifts, however, are often degenerate and about 10% of the measured side chain <sup>1</sup>H chemical shifts could not be assigned unambiguously.

Previously, it was suggested that for successful automated NOE assignment at least 90% of all proton chemical shifts have to be assigned [Jee and Guntert 2003]. FastNMR in its current implementation, however, only uses 3D CCONH- and HCCONH-TOCSY NMR experiments and only ~60 % of all protons were assigned by FastNMR prior to starting the NOE analysis, as shown in Table 2.3. Hence, in an attempt to increase the number of chemical shifts, we used theoretical chemical shifts. So, for all protons (including side chains) for which no experimental chemical shifts

were available in 3D HCCONH-TOCSY spectra, FastNMR uses predicted chemical shifts, obtained from SHIFTS, for automated NOE assignment. As the predicted chemical shifts are not very accurate, the window size that is used for matching NOEs to  $^1\text{H}$  chemical shifts was increased from 0.05 ppm to 0.3 ppm. In case of carbons, missing chemical shifts are not substituted by theoretical values to reduce the risk of wrong NOE assignments. In addition, NOE distance restraints assigned to protons with predicted chemical shifts are used in the final structure refinement only if the same proton (predicted chemical shift) is assigned to two or more NOE peaks and the experimental chemical shift of the two NOE peaks differ by less than 0.1 ppm. In combination with the backbone conformation that is already established prior to the NOE analysis, this allows the determination of high-resolution protein structures.

FastNMR was applied to NMR data of three proteins with all the inherent difficulties of peak overlap, missing backbone resonances, noise peaks and multiple conformations. NOE peaks with multiple chemical shift assignments are fully taken into account by the use of ambiguous distance constraints. Many different tests were performed to highlight the reliability of FastNMR and to show that FastNMR works for real experimental data. In addition, we tested the impact of lower data quality and wrong backbone assignments. These tests are explained in the following section in detail.

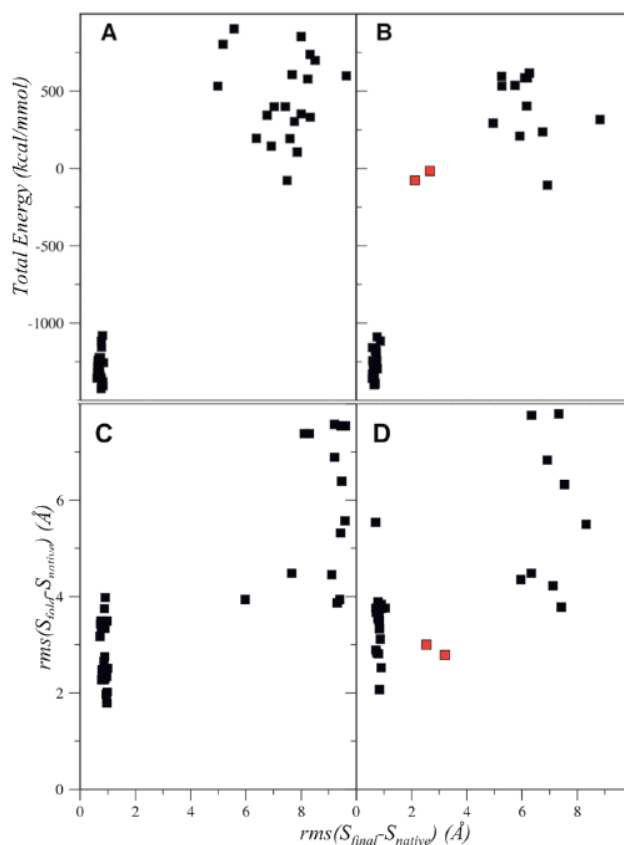
### **2.3.3 FastNMR produces correct high-resolution structures**

For testing the robustness of FastNMR, the FastNMR structure determination of the 60-residue neurotoxin Conk-S1 was interrupted prior to entering the NOE analysis step (Figure 2.1) and protein folds,  $S_{fold}$ , were selected that, on purpose, covered a wide range of deviations ( $\sim 2 - 7 \text{ \AA}$ ) from the native structure,  $S_{native}$ . These  $S_{fold}$  conformations were fed into the following stages of the FastNMR calculation

resulting in a large number of different ensembles,  $S_{final}$ , each containing 20 structures. The energy of these structures was compared to the rms (root-mean-square) deviation from the manually determined solution structure  $S_{native}$  of Conk-S1 (Figure 2.5A). Two groups of structures were observed. In one group, the 20 lowest-energy structures didn't converge to the same fold or, despite they had the same fold, the conformation differed by more than 5 Å from the native structure, i.e.  $rms(S_{final} - S_{native}) > 5$  Å. The conformations belonging to this group did not properly fulfill all experimental restraints, as indicated by the very high energies. Structures in the second group had very low total energies and were very close to the manually determined NMR structure of Conk-S1.

### 2.3.4 FastNMR tolerates missing and wrong backbone assignments

With respect to backbone chemical shifts, the FastNMR structure of Conk-S1 (Figure 2.2C and Table 2.2) was determined using  $^1\text{H}_\text{N}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}_\alpha(\text{i})$ ,  $^{13}\text{C}_\alpha(\text{i-1})$ ,  $^{13}\text{C}_\beta(\text{i})$  and  $^{13}\text{C}_\beta(\text{i-1})$  chemical shifts. These chemical shifts had been obtained from 3D HNCACB and a 3D CBCA(CO)NH experiment. To test the influence of lower data quality, the  $^{13}\text{C}_\beta(\text{i})$  and  $^{13}\text{C}_\beta(\text{i-1})$  chemical shifts were removed from the input lists and FastNMR was rerun. This corresponds to the situation in which connectivity information is only available from a HNCA experiment. Actually, the situation is even more challenging, as  $^{13}\text{C}_\alpha$  chemical shifts obtained from HNCACB and CBCA(CO)NH spectra are less accurate due to the lower resolution of these spectra compared to HNCA and HN(CO)CA spectra. Due to the lower quality of the input data, only 93% of all backbone signals (54 residues) were assigned prior to entering the NOE analysis,



**Figure 2.5:** Stability tests of FastNMR performed on Conk-S1. In panels A and C, 97% of backbone resonances were assigned automatically by FastNMR (using  $C_{\alpha}$  and  $C_{\beta}$  connectivity information and four types of RDCs) prior to entering the NOE analysis. After removing  $C_{\beta}$  connectivity information from the input to FastNMR, panels B and D, FastNMR obtained backbone assignments for only 93 % of residues. In addition, we exchanged the assignment of two out of the 93% assigned residues on purpose, to probe the influence of wrong backbone assignments.

**(A) and (B):** Comparison of the energy of  $S_{final}$  structures with the rms deviation of these structures from the conventionally determined solution structure (PDB code: 2CA7),  $rms(S_{final} - S_{native})$ .

**(C) and (D):** Evaluation of the required accuracy of the protein fold,  $S_{fold}$ . The deviation of  $S_{final}$  ensembles from the native structure (PDB code: 2CA7),  $rms(S_{final} - S_{native})$ , was compared to the rms deviation of the native structure from the selected fold,  $rms(S_{fold} - S_{native})$ . These data show that the deviation of the backbone-only conformation from the native structure,  $rms(S_{fold} - S_{native})$ , has to be less than 3.5 Å to achieve convergence to a high-resolution structure using FastNMR. Note that in two cases (marked red in panel D) FastNMR did not converge to a high-resolution structure despite the fact that the intermediate fold deviated by only  $\sim 3$  Å from the native structure. Both of these problematic cases, however, are unambiguously identified as not converged by their high final energy (marked red in panel B).

whereas 97 % had been assigned when  $C_{\beta}$  connectivity information was included.

Despite the missing backbone assignments, FastNMR produced a structure that was

very close to the manually determined structure, with an rms deviation of 0.8 Å for the backbone heavy atoms. With the reduced lists of backbone chemical shifts (without  $^{13}\text{C}_\beta(i)$  and  $^{13}\text{C}_\beta(i-1)$  chemical shifts) the robustness tests described above were repeated, i.e. the FastNMR structure determination of the 60-residue neurotoxin Conk-S1 was interrupted prior to entering the NOE analysis step, protein folds were selected that covered a wide range of deviations from the native structure, and the selected conformations were fed into the following stages of FastNMR. To make the tests even more demanding, the backbone assignment of two residues was interchanged, i.e. two residues were wrongly assigned on purpose. Nevertheless, all structures, which fulfilled the experimental restraints and had very low energies, deviated by less than 1.0 Å from the native conformation (Figure 2.5B) (also refer text below Table 2.4).

### **2.3.5 Influence of intermediate backbone conformation on convergence to a high-resolution structure**

How much is the intermediate backbone conformation (obtained at the end of stage A) allowed to deviate from the correct structure, such that FastNMR converges to a high-resolution structure? To address this question, we compared the deviation of the backbone conformations, selected in our robustness tests, from the native structure,  $rms(S_{fold} - S_{native})$ , with the deviation of the structures, obtained from these backbone conformations after completion of FastNMR, again from the native structure,  $rms(S_{final} - S_{native})$ . In all cases, FastNMR produced a high-resolution structure of Conk-S1 when the backbone conformation, prior to entering side chain chemical shift assignment, deviated by less than ~3.5 Å from the native structure (Figure 2.5C). Whereas for *ab initio* prediction this cannot generally be achieved, backbone conformations of domain-sized proteins obtained from NMR-based iterative backbone

assignment and fold determination in most cases fulfill this requirement [Jung, Sharma 2004;Rohl and Baker 2002]. Note that no information about homology to proteins with known 3D structures is used in FastNMR, but could easily be incorporated.

Figure 2.5D shows the comparison of  $rms(S_{fold} - S_{native})$  with  $rms(S_{final} - S_{native})$  for the robustness tests, in which no  $C_{\beta}$  connectivity information was used. Despite the missing and two wrong backbone assignments, the convergence properties are very similar. When  $rms(S_{fold} - S_{native})$  is smaller  $\sim 3.5$  Å, FastNMR converges to a high-resolution structure of Conk-S1. Only in two cases (marked in red in Figure 2.5D) FastNMR did not converge to a high-resolution structure despite the fact that the intermediate fold deviated by only  $\sim 3$  Å from the native structure. However, both these structures have very high energies (Figure 2.5B) clearly identifying them as not converged. Noteworthy in one case, FastNMR converged to a high-resolution structure (backbone rmsd of 0.7 Å to the native structure) although the backbone conformation prior to side chain chemical shift assignment deviated by 5.8 Å from the native structure (Figure 2.5D).

### **2.3.6 FastNMR tolerates spurious peaks and multiple conformations**

In the various 3D experiments, in which the magnetization is detected on the  $^1H_N$  nuclei, one pair of  $^1H_N$  and  $^{15}N$  chemical shifts is expected for each amino acid in the protein except for prolines and the C-terminal residue. However, when parts of the protein exchange between different conformations, backbone signals can be missing if the exchange is intermediate on the NMR time scale, or additional signals can be present if the exchange is slow on the NMR time scale. Additional signals can also be due to spectral artifacts.

The influence of multiple conformations and spurious peaks on the performance of FastNMR was tested for the 76-aa protein ubiquitin. The ten most C-terminal residues were removed from the primary sequence of ubiquitin, while keeping all 70 experimentally observed spin systems. Despite the presence of four spin systems, which did not correspond to any position in the primary sequence, FastNMR calculated a high-resolution structure of ubiquitin.

The tests on ubiquitin are strongly supported by the results obtained from the *de novo* FastNMR structure determination of the 65-aa toxin Conk-S2. Initially, Conk-S2 NMR samples were prepared in 50 mM sodium acetate buffer, pH 5.2. In a first 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC recorded on Conk-S2, however, about 95 peaks potentially corresponding to  $^1\text{H}/^{15}\text{N}$  backbone nuclei were present. Therefore, the pH was raised to pH 6.3 (50 mM sodium phosphate buffer). At pH 6.3, 60 backbone signals were visible in the 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum and in the triple-resonance spectra, roughly fitting to the expected 59 backbone signals (Conk-S2 contains 5 prolines). The high-resolution structure of Conk-S2 was then determined at pH 6.3 (see Fig. 1A and Table 1). FastNMR not only produced the high-resolution structure of Conk-S2, but also the assignment of backbone and side chain resonances. Surprisingly, only for 53 residues the backbone signals (and therefore also the side chain resonances) had been assigned by FastNMR, i.e. seven backbone signals remained unassigned. Therefore, additional triple-resonance experiments (3D HNC(O), 3D HN(CA)CO, 3D HNCA and 3D HN(CO)CA) were measured to see if the remaining seven  $^1\text{H}/^{15}\text{N}$  backbone signals could be assigned. Manual analysis of the six 3D triple resonance spectra finally proved that two of these signals correspond to V64 and G65. The others five spin systems, however, were multiple conformations of residues G65 (2x), V64 (1x), R6 (1x). One spin system could not be assigned. In addition, after close inspection of some weaker spin systems (not counted in the seven) were assigned to residues Q58,

Q61 and Y62. No signals were observed for residues S8, G42, T53 and N54 due to chemical exchange intermediate on the NMR time scale. Thus, FastNMR was able to determine a high-resolution *de novo* structure of Conk-S2 despite significant complications due to chemical exchange. Despite these complications, FastNMR produced high-resolution structures, including the *de novo* structure of Conk-S2. These results demonstrate that FastNMR is highly robust.

**Table 2.4 Deviation between different 3D structures of ubiquitin and cross-validation by  $^1D(C^\alpha H^\alpha)$  and  $^1D(C^\alpha, C^\beta)$  RDCs.<sup>a</sup>**

Structure	1ubq <sup>b</sup>	1d3z <sup>c</sup>	1d3z refined <sup>d</sup>	FastNMR
1ubq		0.35	0.52	0.55
1d3z			0.45	0.56 <sup>g</sup>
1d3z refined	0.52	0.45		0.41 <sup>g</sup>
FastNMR	0.55	0.56	0.41	
RDC cross-validation <sup>e</sup>	0.97	0.99 <sup>f</sup>	0.98	0.97

<sup>a</sup> Values given are rmsd values in Å and calculated for residues 2-72 between the mean structure of the NMR ensembles or the 1.8 Å X-ray structure.

<sup>b</sup> 1.8 Å X-ray structure of ubiquitin (PDB code: 1ubq).

<sup>c</sup> NMR solution structure of ubiquitin, which was determined with an extremely large number of experimental restraints, including six different types of RDCs (PDB code: 1d3z).

<sup>d</sup> Structure of ubiquitin that was recalculated from the distance restraints, dihedral angles and three types of RDCs ( $^1D_{N-H}$ ,  $^1D_{C-N}$ ,  $^1D_{C^\alpha-C}$ ), which are available from the PDB (PDB code: 1d3z.mr).

<sup>e</sup> Pearson's correlation coefficient for the comparison of 62  $^1D_{C^\alpha-H^\alpha}$  and 39  $^1D_{C^\alpha-C^\beta}$  RDCs with values back-calculated from the various structures using singular-value decomposition.

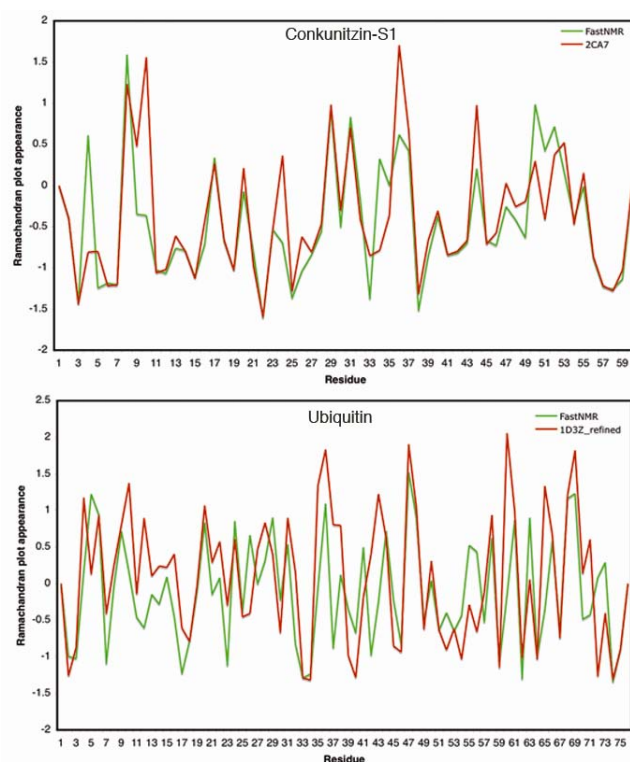
<sup>f</sup> Note that the very high RDC correlation is due to the fact that the  $^1D_{C^\alpha-H^\alpha}$  and  $^1D_{C^\alpha-C^\beta}$  RDCs were used in the structure calculation of 1d3z.

<sup>g</sup> Note that the rms deviation of the 20 lowest-energy structures of the FastNMR ensemble with respect to the average structure of this ensemble is 0.42 Å, whereas the maximum deviation of any of the 20 structures of the ensemble with respect to the average structure is 0.58 Å.

Cross-validation ensures that no wrong structures are produced by FastNMR: For backbone assignment and fold determination only RDCs and chemical shifts are used, whereas during automated NOE assignment RDCs are not used. Thus, in case the initial fold is incorrect it is unlikely that a sufficient number of NOEs is assigned during automated, structure-based NOE assignment. Even if a large enough number of NOEs is assigned, the NOE-based structure will likely differ significantly from the



initial fold and disagree with the RDCs. Therefore, in the final stage of FastNMR when all experimental data are combined convergence to a low energy structure is not possible. This is clearly visible from Figure. 2.5D and the additional stability tests: Only for correct, high-resolution structures a low total energy is obtained by FastNMR. In addition, FastNMR structures have to pass the following check points: (i) at each stage during FastNMR structures must have converged to a unique conformation, (ii) structural changes during FastNMR must be less than 3.5 Å from



**Figure 2.6:** Comparison of per-residue Ramachandran plot quality Z-score between FastNMR structures and conventionally determined structures. (A) PDB (2CA7) structure (red) and FastNMR structure (green) of Conk-S1. (B) FastNMR structure of ubiquitin (green) and structure of ubiquitin recalculated from manually evaluated NMR data deposited in PDB (1D3Z.mr) (red). Average values for the twenty lowest-energy structures are shown. Additional structural statistics can be found in Table 1.1.

the initial fold to the high-resolution structure, (iii) more than 85% of the backbone resonances must have been assigned prior to starting the automated NOE assignment,

and (iv) FastNMR structures have to pass the standard quality criteria such as violations of experimental restraints (Table 2.2).

## 2.4 Conclusions

The strategy of FastNMR is based on the approach that has proven itself as robust in manual structure determination. This includes usage of information from triple-resonance experiments for sequential backbone assignment, use of iterative NOE assignment and structure calculation, and structure refinement using RDCs. Key to the success of FastNMR is, however, the simultaneous determination of the backbone assignment and the protein fold prior to analysis of NOE data.

FastNMR in its current implementation is limited to domain-sized proteins. This is mainly due to the fact that the only experiments which are used for extraction of side chain chemical shifts are CCONH- and HCCONH-TOCSY experiments. The performance of these experiments decreases with increasing molecular weight of the protein and they also do not allow access to chemical shifts of aromatic groups. A larger number of chemical shifts will be available, when 3D (H)CCH-COSY and 3D H(C)CH- TOCSY spectra [Kay, Xu 1993] are incorporated into FastNMR. In addition, aromatic chemical shifts can be obtained from two-dimensional  $(H_{\beta})C_{\beta}(C_{\gamma}C_{\delta})H_{\delta}$  and  $(H_{\beta})C_{\beta}(C_{\gamma}C_{\delta}C_{\epsilon})H_{\epsilon}$  spectra [Yamazaki, Forman-Kay 1993]. The incorporation of these experiments into FastNMR is in progress.

In conclusion, we have demonstrated that it is possible to determine high-resolution structures of domain-sized proteins within 24 hours starting from unassigned chemical shifts, RDCs and NOE peak lists and have used this approach to determine the *de novo* structure of Conk-S2. The calculation time can be reduced even further, when FastNMR is executed in parallel on several computers or when faster computers are available. FastNMR runs automatic, avoids wrong structures by cross-validation, works for experimental data, only requires a limited number of NMR spectra and produces high-resolution structures. No manual assignment of chemical

shifts or inter-residue correlations is required. Interactive work is confined to the processing of NMR spectra and to the preparation of input lists, which contain the unassigned, experimental chemical shifts, residual dipolar couplings and NOE peaks. As FastNMR is highly robust with respect to missing or wrong chemical shift assignments, we expect that it will be possible to perform peak picking and grouping also fully automatic.

FastNMR is a method for *de novo* structure determination, i.e. no prior structure or fold is assumed. However, with the rapid increase in the number of available 3D structures, it becomes more and more likely to find a close homologue in the Protein Data Bank (PDB) ([www.rcsb.org](http://www.rcsb.org)). The structure of this homologue can be supplied as input to FastNMR. This will be important for larger proteins, for which the fold cannot be determined reliably using only RDCs and chemical shifts. When FastNMR is combined with methods for fast data acquisition, such as G-Matrix Fourier Transform NMR spectroscopy [Shen, Atreya 2005], it appears possible to obtain high-resolution NMR structures in less than one week after preparation of suitable NMR samples. Fast and efficient determination of high-resolution structures in solution will make biomolecular NMR a more efficient tool for Structural Biology.

## Chapter 3

### High-resolution 3D Structure Determination of Kalitoxin by Solid-state NMR Spectroscopy

#### 3.1 Introduction

Structural characterization of membrane proteins and many other biological systems by X-ray crystallography or solution NMR spectroscopy is difficult because of problems with crystallization, solubility or molecular size. Significant advances, however, have been made to construct three-dimensional (3D) molecular structure from solid-state NMR data obtained under Magic Angle Spinning (MAS) [Andrew, Bradbury 1958] conditions [Griffin 1998;Luca, Heise 2003;Tycko 2001]. These efforts resulted in high-resolution 3D conformations for small peptides [Jaroniec, Macphee 2004;Nomura, Takegoshi 1999;Petkova, Ishii 2002;Rienstra, Tucker-Kellogg 2002] and the determination of medium-resolution backbone structures for a few solid-phase proteins [Castellani, Van Rossum 2002;Franks, Zhou 2005;Lange, Becker 2005;Zech, Wand 2005]. No high-resolution structure of a protein in the solid-phase has been reported so far.

Recently, progress has been made to determine three-dimensional molecular structures existing in solid phase by magic angle spinning [Andronesi, Pfeifer 2004;Glaubitx and Watts 1998]. Unless sample orientation provides a direct route to monitor molecular structure under MAS conditions, the collection of medium and long-range distance constraints is most crucial. Two strategies have been developed in this direction: (i) measurement of  $^{13}\text{C}$ - $^{13}\text{C}$  distances on  $^{13}\text{C}$  block-labeled protein microcrystals [Castellani, Van Rossum 2002] and (ii) extraction of  $^1\text{H}$ - $^1\text{H}$ -distance restraints from  $^{13}\text{C}$ ,  $^{13}\text{C}$ - and  $^{15}\text{N}$ ,  $^{13}\text{C}$ -encoded  $^1\text{H}/^1\text{H}$  mixing experiments on a

uniformly  $^{13}\text{C}/^{15}\text{N}$ -labeled sample [Lange, Becker 2005]. In the first approach, it uses proton-driven spin diffusion (PDS) pulse sequence or dipolar assisted rotational resonance (DARR) [Takegoshi, Nakamura 2001] acquired  $^{13}\text{C}$ - $^{13}\text{C}$  distances and involves site directed isotope spin labeling, which leads to attenuation of the dipolar effects. Such approaches still depend on the extent of the labeling and require a high spectral dispersion to detect large number of long range  $^{13}\text{C}$ - $^{13}\text{C}$  correlations and the lack of it, leads to cross-peak overlaps. When spectral overlaps are present, manual assignment of the distance correlations becomes tedious, hence, requiring automatic cross-peak assignment. A recent approach, SOLARIA [Fossi, Castellani 2005], used solution-state method to overcome assignment of ambiguous PDS cross-peaks and speed up structure calculation for solid phase proteins, thus proposing a method to automate the whole structure determination procedure. This approach though, is sensitive to the intermolecular peaks present in the peak list leading to local distortions in the structure.

Crucial to 3D structure determination is the detection of long-range distance correlations. Relayed magnetization transfer leads to the attenuation in the precision of correlations at long distances. Hence, additional difficulties in extraction of distance restraints from PDS or DARR approach could be overcome by using single uniformly isotope labeled sample and from  $^{13}\text{C}$ ,  $^{13}\text{C}$  encoded  $^1\text{H}$ ,  $^1\text{H}$  mixing experiments (CHHC), an indirect detection of  $^1\text{H}$ - $^1\text{H}$  distances. In contrast to carbon correlations, all interactions in this spectra carries valuable structural information. This experiments allowed identifying a sufficient number of unambiguous long-range distance correlations, which defined the global 3D fold of KTX [Lange, Becker 2005]. Identifying an unambiguous cross peak for larger and complicated systems would be challenging in the presence of intermolecular peaks and wrong assignments. Here we combine this approach with a probabilistic assignment algorithm originally

developed for the automatic assignment of  $^1\text{H}$ - $^1\text{H}$  correlations in Nuclear Overhauser Effect spectra recorded on proteins in solution [Kuszewski, Schwieters 2004]. We determine the high-resolution structure of the 38-residue scorpion toxin kaliotoxin (KTX) and show that the structure of KTX in the solid phase deviates from the one observed in solution.

## 3.2 Material and Methods

We developed a strategy for rapid high-resolution structure determination from solid-state NMR (ssNMR) data obtained under the conditions of magic-angle spinning (MAS). In addition to the 3D structure of the protein, the method also yields cross-peak assignments of all spectra automatically. The strategy starts from a list of chemical shifts obtained from a series of two-dimensional  $^{13}\text{C}$ ,  $^{13}\text{C}$ - and  $^{15}\text{N}$ ,  $^{13}\text{C}$ -correlation experiments and peak lists from proton-proton correlated 2D CHHC/NHHC spectra. The first step in this procedure is to assign the peak positions in the 2D N/CHHC spectra [Lange, Becker 2005] obtained at different mixing times. The probabilistic assignment algorithm for automated structure determination (PASD) [Kuszewski, Schwieters 2004] automatically performs this task by matching the assigned chemical shifts with the peak positions in 2D N/CHHC spectra. Earlier, the backbone fold of KTX (Protein Databank (PDB) code: 1XSW) [Lange, Becker 2005] in the solid phase was deduced from 28 manually assigned interresidue CHHC correlations using a conventional simulated annealing protocol. Building on the backbone fold and the few manual assignments of cross peaks, most of the peaks observed in 2D N/CHHC spectra were here assigned using PASD, a probabilistic assignment algorithm for automated structure determination. Subsequently, structure calculation was carried out in torsion angle space using the assigned distance restraints. Distance restraints were supplemented by torsion angle restraints derived from  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  chemical shifts using the program TALOS [Cornilescu, Delaglio 1999]. Three successive PASD passes of cross peak assignment and simulated annealing were performed and each pass was started from a set of randomly generated coordinates. After completion of the PASD calculations, cross peak assignments were selected that had a likelihood of 1.0 (as defined by PASD). Selected assignments were



then carefully verified by manual inspection of 2D CHHC and 2D NHHC spectra. In the final stage, manually verified cross peak assignments were used for calculation of a high-resolution structure using an optimized simulated annealing protocol [Linge, Williams 2003]. The calculations were started from random initial coordinates and all verified distance restraints were active during the course of calculation. In addition, torsion angle restraints derived from backbone chemical shifts were included.

PASD was applied largely following published procedures [Kuszewski, Schwieters 2004]. In short, three successive passes of simulated annealing calculations in torsion angle space were carried out. 500 independent structures were calculated for each pass. Each pass was started from a set of randomly generated coordinates. The target function comprised a potential function for experimental distance restraints (e.g. obtained from NOEs or CHHC correlations), a quadratic van der Waals repulsion term, a square-well potential for torsion angles and a torsion angle database potential of mean force. Pass 1 and 2 protocol comprised two high-temperature phases (4000 K) and a slow cooling phase (from 4000 to 100 K) with a linear NOE potential. Pass 3 comprised a single high-temperature phase (4000 K) followed by a cooling phase with a quadratic NOE potential. Final assignment likelihoods were determined at the end of pass 3 calculations. Calculations were carried out on a Linux cluster of 32 processors and took about two days for each structure.

The method shares the same overall architecture of cross-peak assignment and structure calculation as in PASD. However, there are also significant differences. First, we have not used the ensemble of structures present at the end of pass 1 and 2 for calculation of likelihood estimates, but the high-resolution structure of KTX for solution-state structure, obtained using the same procedure. Second, since the final structure at the end of pass 3 do not represent fully-refined NMR structures,

assignments of cross-peaks with a final likelihood of 1.0 were taken and verified by manual inspection of 2D CHHC and NHHC spectra. Manually verified cross-peaks were finally used for further refinement of the structures using an optimized simulated annealing protocol [Linge, Williams 2003;Schwieters, Kuszewski 2003]. The calculations were started from random initial coordinates and distance restraints were not randomly switched on and off during the calculation, but all verified restraints were active during the course of calculation.

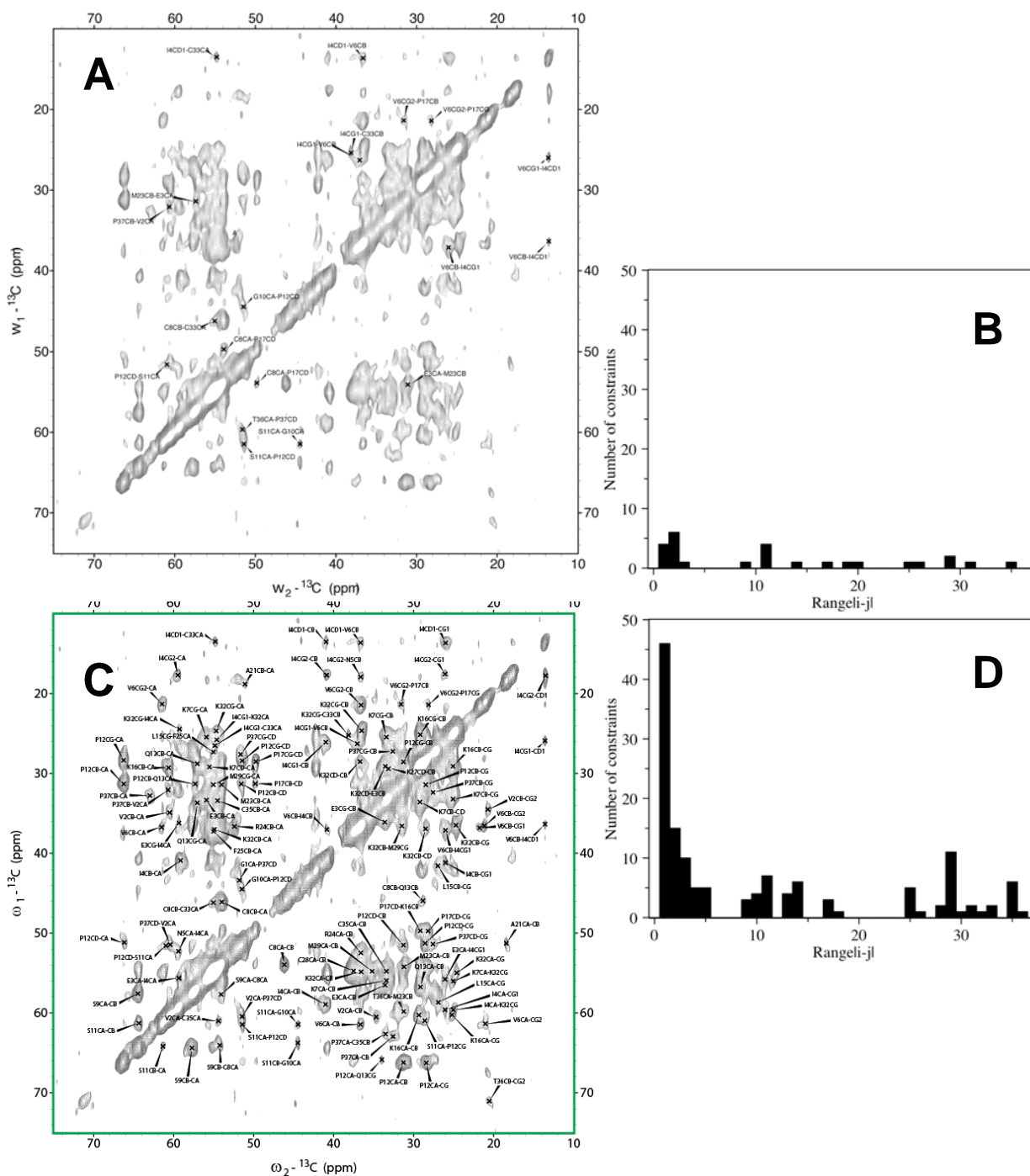
At the end of pass 1 and 2 the PASD algorithm calculates likelihood estimates that each particular assignment associated with a cross-peak is correct. The likelihoods are calculated using the ensemble of structures present at the end of the corresponding pass. Thus, they are a metric of how consistent a given assignment is with the ensemble of structures at the end of each calculation pass. Here we have not used the ensemble of structures present at the end of pass 1 and 2 for calculation of likelihood estimates, but the high-resolution structures of KTX obtained under different conditions or the KTX fold determined previously. The structure calculation was done twice, in which the high-resolution structures of the solution state (obtained from the similar procedure from solution NMR data) and the solid-state was used for calculation of the likelihood estimates. For example, the structure calculations of KTX (solid phase) was done once by using KTX (solution state) for calculation of the likelihood estimates at the end of pass 1 and 2. Then a second calculation was done, in which the likelihood estimates were determined using KTX (solid phase). This approach was done iteratively for both the states. Thus, the structure calculations were biased on purpose toward the structure of the other state. In both the cases, the structures obtained for a specific state from the two different calculations were indistinguishable. Note, that identical structure calculation protocols were used in all cases.

### 3.3 Results and Discussions

#### 3.3.1 High-resolution structure of KTX

The approach was used to solve the high-resolution 3D structure of the 38-residue potassium channel blocker toxin KTX obtained after lyophilization and rehydration [Lange, Becker 2005].  $^1\text{H}$ - $^1\text{H}$  correlations were detected in three CHHC spectra with mixing times of 250, 325 (Figure 3.1A) and 400  $\mu\text{s}$ , and one NHHC spectrum (MAS rate of 12.5 kHz;  $B_0=18.8$  T) [Lange, Becker 2005]. Proton-proton correlation 2D CHHC and 2D NHHC Spectra of the solid-state and 2D NOESY solution-state spectra was automatically peak picked using SPARKY 3 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Diagonal peaks were manually removed. Peak intensities obtained from the 2D CHHC and NHHC spectra were classified into four ranges and converted into distance ranges of 1.8-2.7, 1.8-3.3, 1.8-5.0 and 1.8-6.0 Å, respectively. The classification was done independently for the four proton-proton correlation spectra. For analysis of the solid-state spectra by PASD,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts were labeled as if they were proton chemical shifts. The tolerances for matching chemical shifts to cross-peaks were set to 0.38 ppm in the acquisition dimension and to 0.60 ppm in the indirect dimension. PASD structures do not represent fully-refined NMR structures [Kuszewski, Schwieters 2004]. Therefore, we selected cross-peaks with a final likelihood assignment (as defined by PASD) of 1.0. For KTX(solution) and KTX(solid), 31% and 28%, respectively, of all long-range restraints, 83% and 80%, respectively, of the medium-range restraints, and 99% and 100%, respectively, of the sequential restraints had final restraint likelihoods of 1.0. All other restraints had final restraint likelihoods of 0. Assignments obtained for these cross-peaks by PASD were verified by manual inspection of the 2D CHHC spectra and the 2D NHHC spectrum.

Previously, 15 long-range, 7 medium-range and 6 short-range correlations could be assigned manually using the same set of CHHC and NHHC spectra [Lange, Becker 2005]. Using the above-described semi-automated approach a total of 254  $^1\text{H}$ - $^1\text{H}$  distance correlations could be assigned unambiguously (Figure 3.1A, 3.1B and Table 3.1). 64 of these were long-range, 37 medium-range and 153 sequential. The 3D solid-state structure of KTX that was calculated from the 254 distance restraints and 58 dihedral angle restraints is shown in Figure 3.2B (PDB code: 2UVS). The resulting ensemble of KTX structures tightly converged with a coordinate precision of 0.5 Å and 1.4 Å for backbone and side chain heavy atoms, respectively (Figure 3.3A). Backbone and most side chains had a well-defined orientation except the N- and C-terminal residue and Asn30 located in the loop connecting the second and third  $\beta$ -strand of KTX. Weak correlations between RMSD of backbone and side chains may be arising from the side chain flexibility in free form of the structure. Overall, the structure showed a backbone fold consisting of a  $\alpha$ -helix affixed by disulphide bridges to a three-stranded  $\beta$ -sheet that is typical for the charybdotoxin family of K<sup>+</sup> channel-blocking scorpion toxins [Grissmer, Nguyen 1994]. 87% of residues were found in the most favored region of the Ramachandran plot, whereas 3% of residues were found in the disallowed region. The high-resolution solid-state structure of KTX deviates by 2.4 Å from the backbone conformation (PDB code: 1XSW) obtained on the basis of 28 manually assigned distance restraints [Lange, Becker 2005], which deviates by 2.7 Å from the solution structure. The most pronounced deviation between 1XSW and the high-resolution structure was observed at the N-terminus, where four residues were rotated by about 50°, such that the first beta-strand was straight and not bent as seen in the high-resolution structure (in the solid phase or in solution).



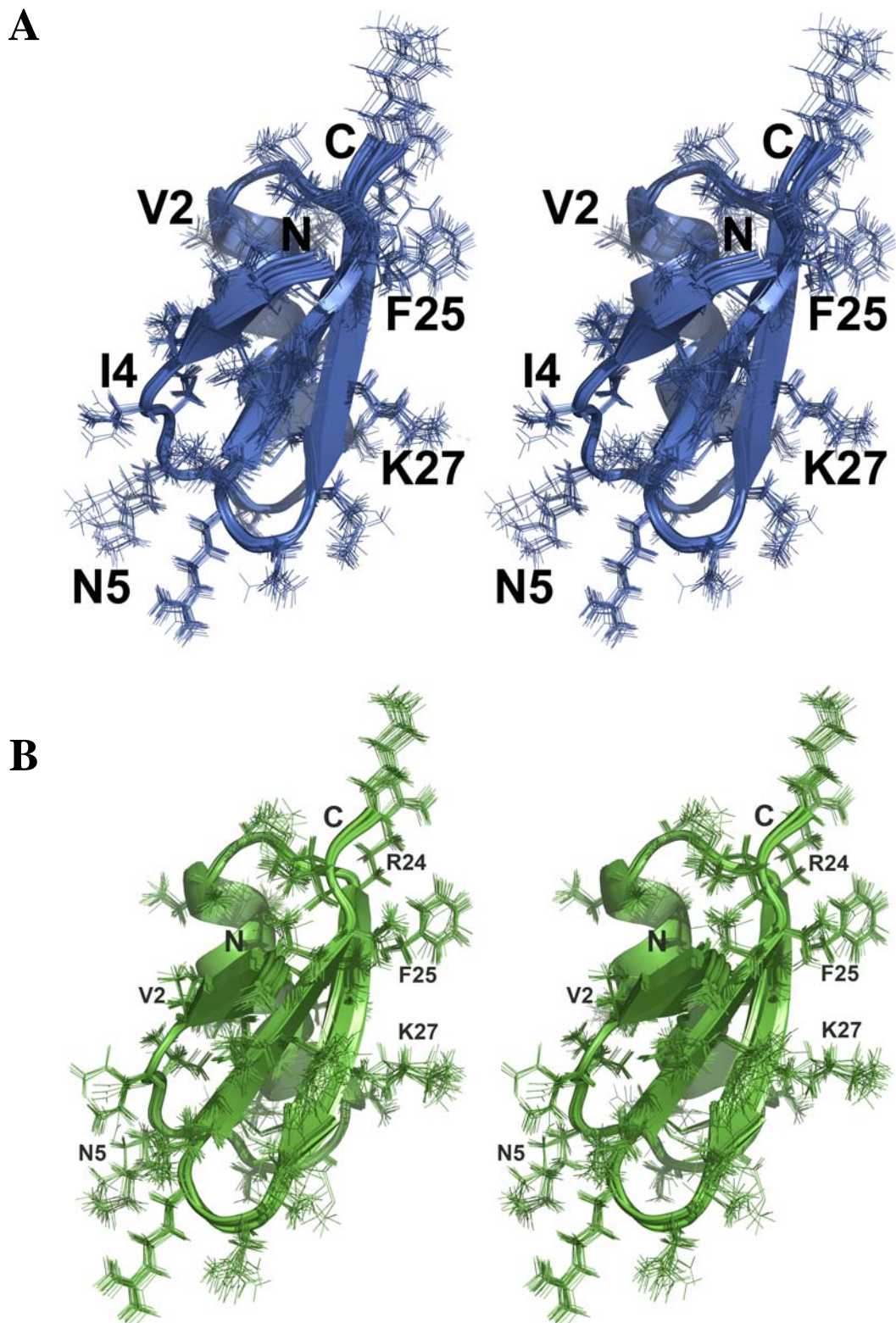
**Figure 3.1:** Comparison of interresidue correlations assigned manually (**A** and **B**) and assigned unambiguously by PASD and verified manually (**C** and **D**) for KTX in the solid phase. Signals assigned in the 2D CHHC spectrum of U- $[{}^{13}\text{C}, {}^{15}\text{N}]$ -KTX recorded with a mixing time of 325  $\mu\text{s}$  are labeled. Spectra in (A) and (C) are identical. (B) and (D) show the number of unambiguously assigned distance constraints as a function of residue difference  $i$  and  $j$ .

**Table 3.1:** Structural statistics for the 20 lowest-energy structures of KTX in solution and in the solid phase.

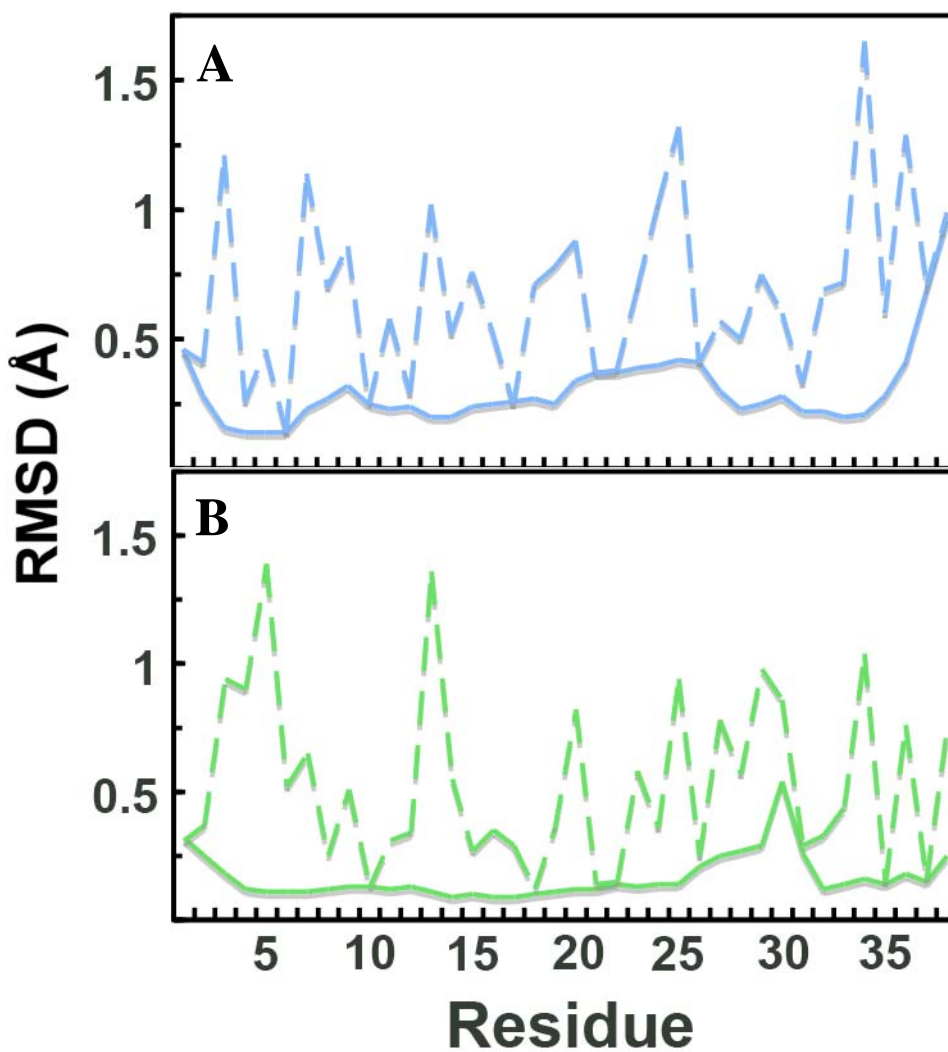
	Solution	Free
<b><u>Proton-Proton distance correlations</u></b>		
Total	314	254
Short range	199	153
Medium range	45	37
Long range	70	64
Distance violations (>0.5 Å)	0	2 <sup>b</sup>
<b><u>Dihedral angles</u></b>		
	64	58
<b><u>Energy (kcal/mol)</u></b>		
Total	-1203.32	-1307.05
Bond	34.02	32.49
Angle	123.23	130.48
Improper	36.77	35.04
Dihedral	133.40	109.58
NOE/CHHC	166.82	156.58
<b><u>RMSD</u></b>		
Bond	0.0076	0.0074
Angles	0.7832	0.8973
Improper	0.8727	0.9013
Dihedral	3.2033	4.7092
NOE/CHHC	0.0782	0.0791
<b><u>Coordinate precision</u></b> <sup>a</sup>		
Backbone atoms (Å)	0.7	0.5
All heavy atoms (Å)	1.6	1.4
<b><u>Ramachandran statistics</u></b>		
Most favored region (%)	86.7	84.0
Disallowed region (%)	3.3	3.7

<sup>a</sup> Defined as the average rmsd difference between the 20 structures and the mean coordinates.

<sup>b</sup> The two distance restraints E3(H $\alpha$ )-C33(H $\beta$ 1) and C8(H $\beta$ 1)-C33(H $\alpha$ ) were violated by 0.65 and 0.52 Å, respectively. The two restraints came from weak cross peak in the spectra and were assigned a distance range of 1.8 - 6.0 Å in the calculations.

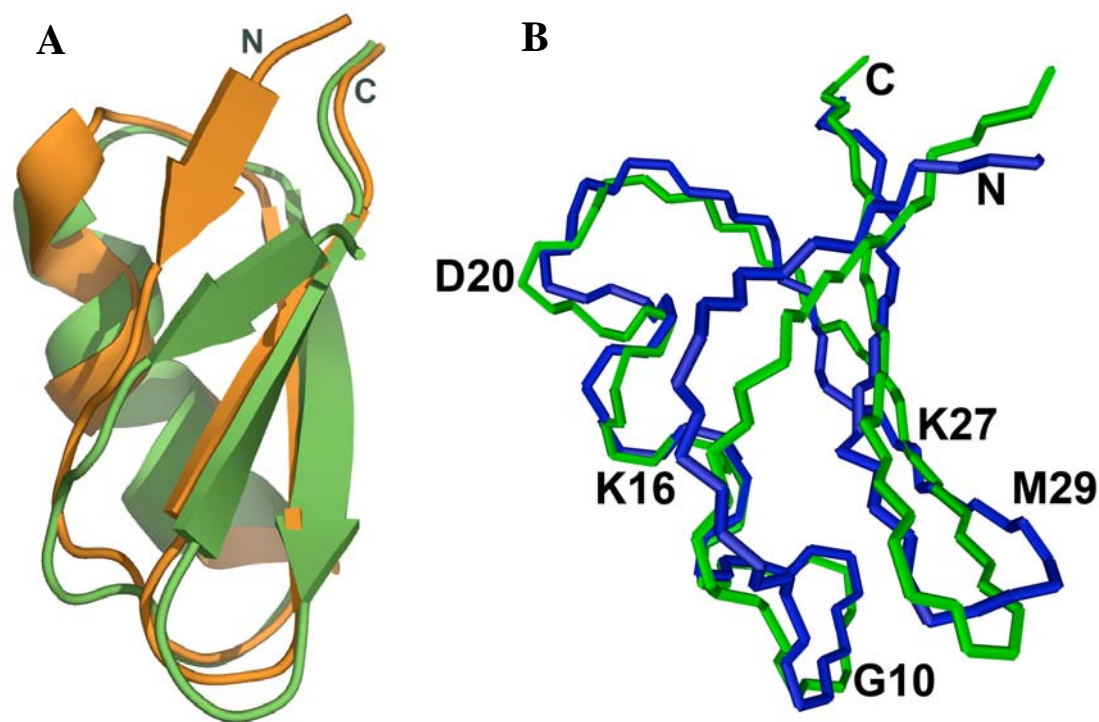


**Figure 3.2:** Stereo view of 20 lowest energy ensemble of high-resolution structures of Kaliotoxin in solution-state (A) (upper blue), and solid-state (B) (lower green)



**Figure 3.3:** Coordinate precision of KTX in solution and solid-phase. Residue-based rms deviations of the atomic coordinates within the ensemble of 20 lowest energy structure of KTX in solution (A), in the solid phase (B). Deviations for the backbone atoms and for the non-hydrogen side chain atoms are shown as solid and dashed line, respectively.





**Figure 3.4:** High-resolution solid-state structure of Kaliotoxin (green; determined in this study) compared to the backbone fold obtained previously from 28 manually assigned CHHC correlations (orange; PDB code: 1XSW) (Left). High-resolution solid-state structure of Kaliotoxin (green) compared to solution structure (blue; PDB code: 2KTX) (right).

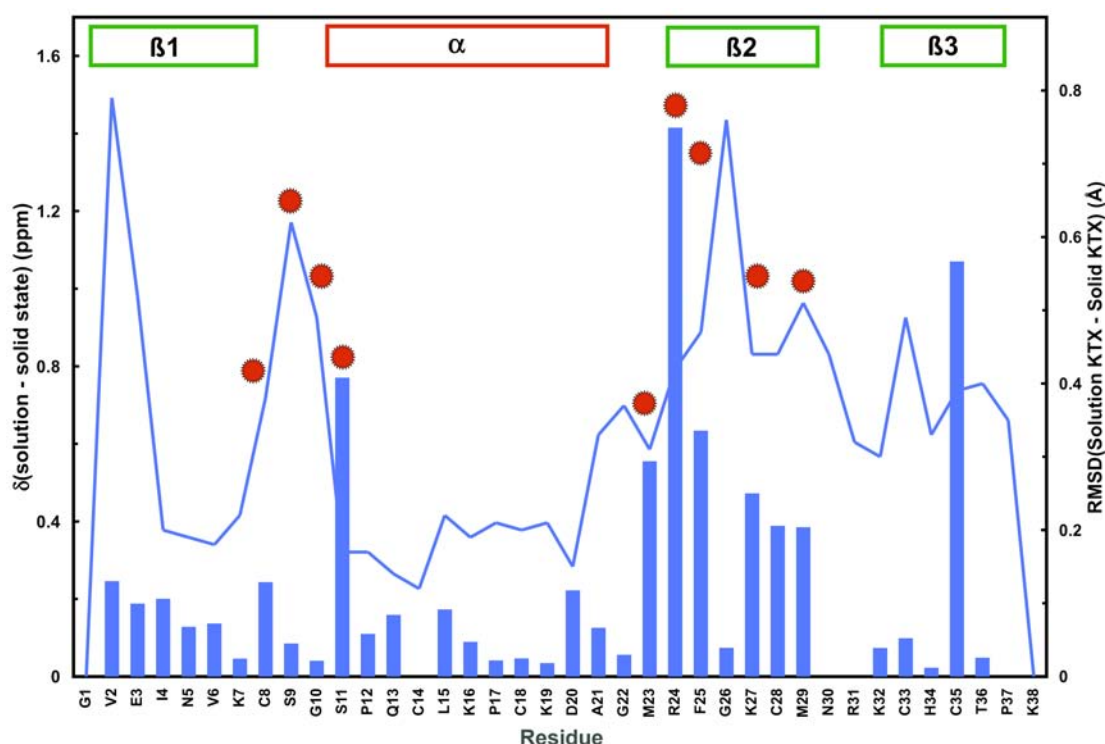
To obtain convergence to a unique conformation, we fixed the 28 manually obtained CHHC cross peak assignments [Lange, Becker 2005]. Although some of the 28 distance restraints could be removed, a minimum of five long-range and two medium-range restraints were necessary to obtain convergence. In addition, the 1XSW backbone fold had to be used for calculation of PASD likelihood estimates to converge to a unique structure. Various additional tests were performed to probe the convergence of the structure calculations and support the accuracy of the high-resolution solid-state structure (discussed below): (i) use of different conformations for calculating likelihood estimates in PASD; (ii) influence of chemical shift tolerances; (iii) sensitivity towards distance ranges used for interresidue correlations; (iv) dependence on the number of CHHC spectra; (v) influence of disulphide bond

restraints. In all cases, the backbone of the calculated structures deviated by less than 0.7 Å from the backbone of the structure shown in Figure 3.2B.

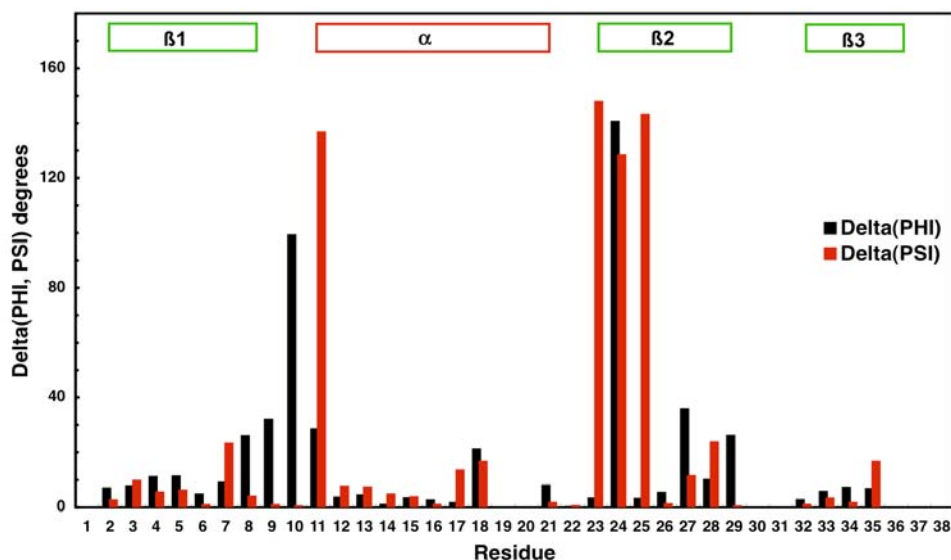
Recently, a method for automatic assignment of cross peaks in  $^{13}\text{C}$ - $^{13}\text{C}$  correlation spectra was developed [Fossi, Castellani 2005]. The approach called SOLARIA was used to analyze proton-driven spin diffusion (PDS) spectra recorded on  $^{13}\text{C}$ - block-labeled, microcrystalline preparations of the  $\alpha$ -spectrin SH3 domain. To attenuate dipolar truncation effects, the method requires different selectively  $^{13}\text{C}$ -enriched protein preparations. In addition, as cross peak volumes in PDS spectra do not depend exclusively on the distance, a large distance boundaries had to be used in structure calculation. In that study, only a modest improvement in the 3D backbone structure was observed, and was also reported to be sensitive to the intermolecular peaks present in the peak list that lead to local distortions in the structure. Also, in such iterative approaches, if the initial assignment contains false constraints, the erroneous global fold of structures generated in the first calculation could bias the result of successive cycles. In this respect, the probabilistic algorithm, PASD [Kuszewski, Schwieters 2004], recently implemented in the structure determination package Xplor-NIH [Schwieters, Kuszewski 2003], provided a highly error tolerant approach for automated constraint identification and structure calculation cycles. Thus, no high-resolution structure of the backbone and the side chain of SH3 domain was reported. In contrast, our strategy based on C/NHHC correlations leads to an atomic resolution definition of both the backbone and the side-chain structure of KTX. We attribute these improvements to the higher fraction of long-range contacts in initial-rate N/CHHC spectra that allows for the same small distances boundaries [Baldus 2007; Lange, Seidel 2003] during structure calculation as used in liquid-state NMR.

### 3.3.2 Comparison with high-resolution solution-state NMR structure of KTX

The 3D structure of KTX in solution had been previously determined by liquid state NMR spectroscopy. To enable a direct comparison with the structure of KTX in solid phase, we determined the solution structure of KTX employing the identical strategy as used for KTX in the solid phase. Tolerances for matching chemical shifts to cross-peaks were 0.015 ppm in direct and indirect dimensions of the 2D NOESY spectrum. After completion of PASD calculations, cross peak assignments were selected that had a likelihood of 1.0. Selected assignments were carefully verified by manual inspection of the 2D NOESY spectrum. 70 long-range, 45 medium-range and 199 sequential NOEs could be assigned unambiguously, closely resembling the amount and distribution of distance restraints obtained from 2D N/CHHC spectra for



**Figure 3.5:** Comparison of averaged C $\alpha$ /C $\beta$  chemical shift differences (blue bars; calculated according to  $0.256 * [\Delta\delta C_{\alpha}^2 + \Delta\delta C_{\beta}^2]^{1/2}$ ) with rms deviation between the mean structures (blue line) of KTX in solution and in the solid-phase. Red dots mark residues, for which the backbone dihedral angles predicted by TALOS differ in solution and the solid state.



**Figure 3.6:** Differences between the backbone dihedral angles predicted by TALOS from the solution-state and solid-state secondary chemical shifts. The average uncertainty of TALOS prediction as stated on the web page is  $\sim 12^\circ$ . In Figure 3.5, residues were marked with a red dot, for which the difference in either *phi* or *psi* exceeds twice this value.

KTX in the solid phase (Table 3.1). The newly determined solution structure deviates by  $0.6 \text{ \AA}$  from a previously determined solution structure of KTX (PDB code: 2KTX) [Gairi, Romi 1997]. The backbone of the high-resolution solid-state structure of KTX deviates by  $1.3 \text{ \AA}$  from that observed in solution (Figure 3.4B). Structural differences between the KTX (as seen from the pairwise root mean square deviation between the high-resolution structures of KTX), existing in solution and solid phase, were observed for the two N-terminal residues, the loop between the first beta-strand and the  $\alpha$ -helix, and the C-terminal beta-sheet in particular next to G26 (Figure 3.5).

The averaged  $C\alpha/C\beta$  chemical shift differences between the solution and solid phase, in the above-mentioned regions in the structure, also supports this finding. The structural differences are due to a combination of changes in interresidue cross peaks and in backbone dihedral angles (Figure 3.6). For residues 8-11, 23-25, K27 and M29, backbone dihedral angles predicted by TALOS from the solid-state secondary

chemical shifts clearly deviated from those predicted by TALOS from the solution-state secondary chemical shifts (Figure 3.6).

### **3.3.3 Reliability of High-resolution solid-state structure of KTX**

To check the reliability of the high-resolution solid-state structure and to account for the structural difference we performed several tests.

(i) Use of different structures for calculating likelihood estimates in PASD: biasing on purpose the calculation towards the solution-state structure.

At the end of pass 1 and 2 the PASD algorithm calculates likelihood estimates that each particular assignment associated with a cross-peak is correct. The likelihoods are calculated using the ensemble of structures present at the end of the corresponding pass. Thus, they are a metric of how consistent a given assignment is with the ensemble of structures at the end of each calculation pass [Kuszewski, Schwieters 2004]. Here we have not used the ensemble of structures present at the end of pass 1 and 2 for calculation of likelihood estimates, but either the high-resolution structure of KTX obtained under different conditions or a medium-resolution backbone fold. This improved convergence in the structure calculations and was justified as we previously established that the fold of KTX in solution and in the solid phase is the same [Lange, Becker 2005].

For example, the PASD calculations of KTX(solid) were done once by using KTX(solution) (PDB CODE: 2KTX) for calculation of the likelihood estimates at the end of pass 1 and 2. Thus, we biased on purpose the calculation towards the solution-state structure. Then a second PASD calculation was done, in which the likelihood estimates were determined using the medium resolution backbone fold obtained previously for KTX in the solid phase (PDB CODE: 1XSW) [Lange, Becker 2005]. In all cases, the structures obtained from the two different PASD calculations were

indistinguishable. This supports the relevance of the differences between the solution and solid-state structure. Note, that identical structure calculation protocols were used in all cases.

(ii) Influence of chemical shift tolerances.

For the calculations reported in the main part of the manuscript, tolerances for matching chemical shifts to cross-peaks were set to 0.38 ppm and 0.60 ppm in the acquisition and indirect dimension, respectively. We repeated the structure calculations with chemical shift tolerances of 0.38 ppm and 0.4 ppm in the acquisition and indirect dimension, respectively. The resulting structure deviated by less than 0.7 Å (rms value for all N, C $\alpha$ , CO backbone atoms) from the structure shown in Figure 3.2B.

(iii) Influence of distance ranges: using only a single distance range (1.8 – 6.0 Å).

Peak intensities obtained from the 2D CHHC and NHHC spectra were classified into four ranges and converted into distance ranges of 1.8-2.7, 1.8-3.3, 1.8-5.0 and 1.8-6.0 Å, respectively. The classification was done independently for the four proton-proton correlation spectra (see Results and Discussion). To test the sensitivity of the solid-state structure to the used distance ranges, we repeated the structure calculations assigning to all N/CHHC correlations a distance range of 2.4-6.0 Å. The resulting structure deviated by less than 0.3 Å (rms value for all N, C $\alpha$ , CO backbone atoms) from the structure shown in Figure 3.2B.

(iv) Influence of the available spectra: usage of only two CHHC spectra instead of all four N/CHHC spectra.

The structure reported in the manuscript was calculated from distance restraints that were extracted from three CHHC spectra (mixing times of 250 $\mu$ s,

325 $\mu$ s and 400 $\mu$ s) and one NHHC spectrum. When we only used the peak lists from the 250 and 400  $\mu$ s CHHC spectrum, the number of long-range, medium-range and short-range correlations was reduced to 59, 33 and 102, respectively. Nevertheless, the 3D structure of KTX(solid) was highly similar to the one obtained with all four proton-proton correlation spectra (backbone rms deviation of 0.65 Å).

(v) Influence of disulphide bond restraints.

For both KTX(solution) and KTX(solid), structure calculations were performed without and with restraints for the three-disulphide bonds. The resulting structures did not differ (backbone rms deviation below 0.5 Å) and only the results of calculations, in which the disulphide bonds were not enforced, were reported.

(vi) Combining solid-state distance restraints with solution-state dihedral angles (and vice versa).

Are the structural differences due to an uncertainty in the analysis of N/CHHC spectra? To address this question, we recalculated the structure (using XPLOR-NIH and starting from an extended strand) using the same solid-state N/CHHC distance restraints, but supplementing them with the dihedral angles obtained by TALOS from the solution-state chemical shifts (instead of those obtained from the solid-state chemical shifts). The backbone of the resulting structure deviated by 0.5 Å from the high-resolution solid-state structure. The coordinate precision for backbone and all heavy atoms was 0.7 Å and 1.7 Å, respectively (Table 3.2). However, two dihedral angle violations were introduced (for residues 2 and 24) and residue 24 moved into the disallowed region of the Ramachandran plot. In addition, the total energy increased from  $-1307\pm 54$  kcal/mol to  $-1032\pm 48$  kcal/mol, the dihedral angle energy from  $-110\pm 6$  kcal/mol to  $-16\pm 36$  kcal/mol and the distance restraint energy from -

157±28 kcal/mol to -18±67 kcal/mol (when compared to the pure solid-state structure calculation).

**Table 3.2:** Structural statistics for the 20-lowest energy structures of KTX in solution and solid-phase, for test case *vi*, when combining the solid-state distance with solution-state backbone dihedral angles (and vice versa).

	Solution	Free
<b><u>Proton-Proton distance correlations</u></b>		
Total	314	254
Short range	199	153
Medium range	45	37
Long range	70	64
Distance violations (>0.5 Å)	1	1
<b><u>Dihedral angles</u></b>	64	58
Dihedral violations (>5°)	1	2
<b><u>Energy (kcal/mol)</u></b>		
Total	-1154.5 ± 65.3	-1031.6 ± 48.5
Bond	-34.6 ± 5.8	-17.7 ± 5.5
Angle	-105.4 ± 35.1	-75.1 ± 36.2
Improper	-36.77 ± 11.2	-35.04 ± 14.6
Dihedral	-54.6 ± 23.2	-16.3 ± 36.4
NOE/CHHC	-30.6 ± 52.9	-17.9 ± 66.7
<b><u>Coordinate precision</u></b> <sup>a</sup>		
Backbone atoms (Å)	0.7	0.5
All heavy atoms (Å)	1.6	1.4
<b><u>Ramachandran statistics</u></b>		
Most favored region (%)	86.7	84.0
Disallowed region (%)	3.3	3.7

<sup>a</sup> Defined as the average rmsd difference between the 20 structures and the mean coordinates.

Similarly, when the solution-state distance restraints were combined with the solid-state dihedral angles, one dihedral angle violation (for S9) was introduced, the total energy was increased from 1203±60 kcal/mol to -1154±65 kcal/mol, the dihedral angle energy from -133±6 kcal/mol to -55±32 kcal/mol and the distance restraint energy from -167±34 kcal/mol to -31±53 kcal/mol (when compared to the pure solution-state structure calculation). The backbone of the resulting structure deviated



by 0.6 Å from the high-resolution solution-state structure. The coordinate precision for backbone and all heavy atoms was 0.8 Å and 1.9 Å, respectively. These data demonstrate that the solid-state distance restraints are only in agreement with the solid-state backbone chemical shifts, and the solution-state distance restraints are only in agreement with the solution-state backbone chemical shifts.

### 3.4 Conclusions

This study demonstrates that high-resolution 3D structures of globular proteins can be obtained from solid-state NMR data. The combination of  $^{15}\text{N}$ ,  $^{13}\text{C}$ -encoded  $^1\text{H}/^1\text{H}$  mixing experiments with a probabilistic cross peak assignment algorithm is particularly powerful, as short distances between protons provide the principal source of long-range structural information. Depending on the molecule under investigation, the presented approach can be combined with other solid-state NMR spectroscopic methods. Applications to larger proteins may benefit from the use of block [Castellani, Van Rossum 2002], modular [Pickford and Campbell 2004] or stereo-array [Kainosho, Torizawa 2006] isotope labeling schemes, and/or acquisition of higher-dimensional spectra and will therefore extend the applicability of this method in the determination of high-resolution structures of amyloid or membrane proteins and more complex systems.

## Chapter 4

### **High-resolution 3D Structure of Kaliotoxin bound to the Potassium channel and toxin-channel complex structural model determined by Solid-state NMR Spectroscopy**

#### **4.1 Introduction**

Interactions between voltage-dependent potassium (Kv) channels and peptide toxins from scorpion venom are among the strongest and most specific protein-peptide complex formations known, rivaling those of antibody-antigen interactions [De La Vega, Merino 2003;Gouaux and Mackinnon 2005;Ranganathan, Lewis 1996]. Kaliotoxin (KTX), a peptide neurotoxin from scorpion *Androctonus mauretanicus* venom, belongs to the charybdotoxin (CTX) family of K<sup>+</sup> channel-blocking scorpion toxins [Grissmer, Nguyen 1994;Lange, Becker 2005]. KTX inhibits the eukaryotic Kv1.3 channel both with high specificity and high affinity by binding with 1:1 stoichiometry to the outer pore region and blocking K<sup>+</sup> conduction [Miller 1995]. While several three-dimensional structures of unbound toxins and unbound K<sup>+</sup> channels were reported [Aiyar, Withka 1995;Doyle, Cabral 1998;Gouaux and Mackinnon 2005;Jiang, Lee 2002;Kuo, Gulbis 2003;Long, Campbell 2005;Miller 1995], a detailed view of the structural changes and motional degrees of freedom important for the formation of toxin-channel complexes is currently lacking. Recently, we have analyzed by solid-state nuclear magnetic resonance (ssNMR) spectroscopy the interaction of KTX with the chimaeric K<sup>+</sup> channel KcsA-Kv1.3 [Lange, Giller 2006]. Recognition of KTX by the channel was accompanied by conformational changes in both the KcsA-Kv1.3 outer pore region and in KTX. This observation was

difficult to understand considering the principle, referred to as ‘the induced fit’ hypothesis [Koshland 1958], that conformational changes produce a high specificity in protein-ligand interactions at the expense of ligand affinity.

A more detailed insight into the molecular mechanism of KTX – KcsA-Kv1.3 and other ligand – membrane protein interactions requires the determination of amplitudes and time scales of motions related to complex formation. Such information is difficult to obtain from X-ray crystallography or mutant cycle analysis. NMR has provided much insight into the dynamical aspects of complex formation in solution [Kern and Zuiderweg 2003] but requires the use of ssNMR methods to study proteins in lipid bilayers. While good progress in the determination of 3D molecular structures from ssNMR data has been made [Castellani, Van Rossum 2002;Fossi, Castellani 2005;Lange, Becker 2005;Rienstra, Tucker-Kellogg 2002], no high-resolution protein structure has been reported to date [Castellani, Van Rossum 2002;Fossi, Castellani 2005;Rienstra, Tucker-Kellogg 2002;Zech, Wand 2005].

To have a detailed understanding and to elucidate the importance of structural dynamics observed in KTX upon binding to the potassium channel, high-resolution structure of the kaliotoxin bound to the potassium channel is necessary. From the method developed in Chapter 3, in this chapter, we determine the high-resolution structure of KTX bound to a membrane-embedded potassium channel and an improved docking model for the KTX - KcsA-Kv1.3 complex is also proposed based on mutagenesis studies and chemical shift differences of KTX observed upon binding to the potassium channel.

## 4.2 Materials and Methods

### 4.2.1 Solid-state NMR data

The strategy starts from a list of chemical shifts obtained from a series of two-dimensional  $^{13}\text{C}$ ,  $^{13}\text{C}$ - and  $^{15}\text{N}$ ,  $^{13}\text{C}$ - correlation experiments conducted on a uniformly [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ] labeled sample of KTX bound to KcsA-Kv1.3 under MAS conditions. Cross-peak lists of indirectly detected ( $^1\text{H}$ ,  $^1\text{H}$ ) correlations obtained from CHHC [Lange, Becker 2005] spectrum with a ( $^1\text{H}$ ,  $^1\text{H}$ ) mixing time of 325 $\mu\text{s}$  was used as the input for the probabilistic assignment approach [Kuszewski, Schwieters 2004] described in Chapter 3. TALOS derived backbone dihedral angles [Cornilescu, Delaglio 1999] obtained from solid-state secondary chemical shifts ( $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$  and  $^{15}\text{N}$ ) for KTX bound to potassium channel, also served as input for the structure calculation.

### 4.2.2 Cross-peak assignment and Structure calculation

The first step in this procedure is to assign the peak positions in the 2D CHHC spectrum obtained at 325 $\mu\text{s}$  mixing times. The probabilistic assignment algorithm for automated structure determination (PASD) [Kuszewski, Schwieters 2004] automatically performs this task by matching the assigned chemical shifts with the peak positions in 2D CHHC spectrum. Earlier, the backbone fold of KTX (Protein Databank (PDB) code: 1XSW) [Lange, Becker 2005] in the solid phase was deduced from 28 manually assigned interresidue CHHC correlations using a conventional simulated annealing protocol. Building on the backbone fold and the unassigned cross peaks lists, most of the peaks observed in 2D CHHC spectrum was here assigned using PASD. Subsequently, structure calculation was carried out in torsion angle space using the assigned distance restraints. Distance restraints were supplemented by

torsion angle restraints derived from  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  chemical shifts using the program TALOS [Cornilescu, Delaglio 1999]. After completion of the PASD calculations, cross peak assignments were selected that had a likelihood of 1.0 (as defined by PASD). Selected assignments were then carefully verified by manual inspection of 2D CHHC and 2D NHHC spectra. In the final stage, manually verified cross peak assignments were used for calculation of a high-resolution structure using an optimized simulated annealing protocol [Linge, Williams 2003; Schwieters, Kuszewski 2003]. The calculations were started from random initial coordinates and all verified distance restraints were active during the course of calculation. In addition, torsion angle restraints derived from backbone chemical shifts were included.

### **4.2.3 Modeling of KcsA-Kv1.3**

Homology modeling of the chimaeric KcsA-Kv1.3 potassium channel was calculated using the crystal structure of conductive KcsA solved at 2Å resolution (PDB code 1K4C) [Zhou, Morais-Cabral 2001]. Positions R52, G53, A54, G56, A57, Q58, L59, I60, T61, Y62 and R64 in 1K4C were mutated to A52, D53, D54, T56, S57, G58, F59, S60, S61, I62 and D64, respectively in the model channel. Using KcsA as a template structure and the sequence alignment of KcsA with the model, generated with CLUSTALW [Thompson, Higgins 1994], 100 models were calculated using the program MODELLER [Marti-Renom, Stuart 2000]. The 20 best models were then energy minimized to remove steric overlaps and clashes of the side chains. The r.m.s. deviation between KcsA (1K4C) and the lowest-energy model was 0.334 Å. The average r.m.s deviation of the 20 lowest-energy models from the mean structure was 0.15 and 0.9 Å for the backbone and heavy atoms, respectively.

#### 4.2.4 Docking

A model of the KTX-KcsA-Kv1.3 complex was calculated using the docking program HADDOCK version 1.3 [Dominguez, Boelens 2003]. The active and passive residues used for the docking in HADDOCK were chosen based on the chemical shift perturbation observed for KTX upon binding to the potassium (K<sup>+</sup>) channel and relative residue solvent accessibility as calculated by the program NACCESS [Hubbard and Thornton 1993]. Active residues were those residues, for which the chemical shift differences were larger than the mean value (1.675 ppm for KTX 0.509 ppm for KcsA-Kv1.3) and the relative residue solvent accessibility (as calculated by NACCESS) exceeded 50% for either side chain or backbone atoms. Passive residues were defined as the neighbors of the active residues whose relative residue solvent accessibility was larger than 50%. Thus, five amino acids (M23, R24, K27, M29 and T36) of KTX and three amino acids (D64, E71 and D80) from each subunit of the tetrameric KcsA-Kv1.3 were defined as active residues. Besides, we defined five passive amino acids (F25, G26, C28, R31, K32) in KTX and three (I62, P63, L81) in each subunit of the tetrameric structure of KcsA-Kv1.3. In HADDOCK, the docking is driven by ambiguous intermolecular restraints (AIRs) [Nilges 1993]. An AIR is defined as an ambiguous intermolecular distance with a maximum value of 3.0Å between any atom of an active residue of KTX and any atom of both active and passive residues of KcsA-Kv1.3 and inversely for KcsA-Kv1.3 with the toxin KTX. The ensemble of 20 structures of the bound KTX, i.e. the 3D structure of KTX determined in this study using solid-state NMR spectra recorded on uniformly labeled [<sup>13</sup>C, <sup>15</sup>N] KTX in complex with KcsA-Kv1.3 reconstituted into proteoliposomes, was used as a starting conformer for the toxin and kept rigid throughout the docking protocol. For KcsA-kv1.3, the lowest-energy model obtained by homology modeling

was used. Restraints for backbone torsion angles obtained for residues within the selectivity filter of KcsA-Kv1.3 from solid-state secondary chemical shifts [Lange, Giller 2006] by the program TALOS [Cornilescu, Delaglio 1999] were used and the respective residues (T75 to G79) were kept rigid during the docking. All other active and passive residues were defined as semi-flexible segments during the docking.

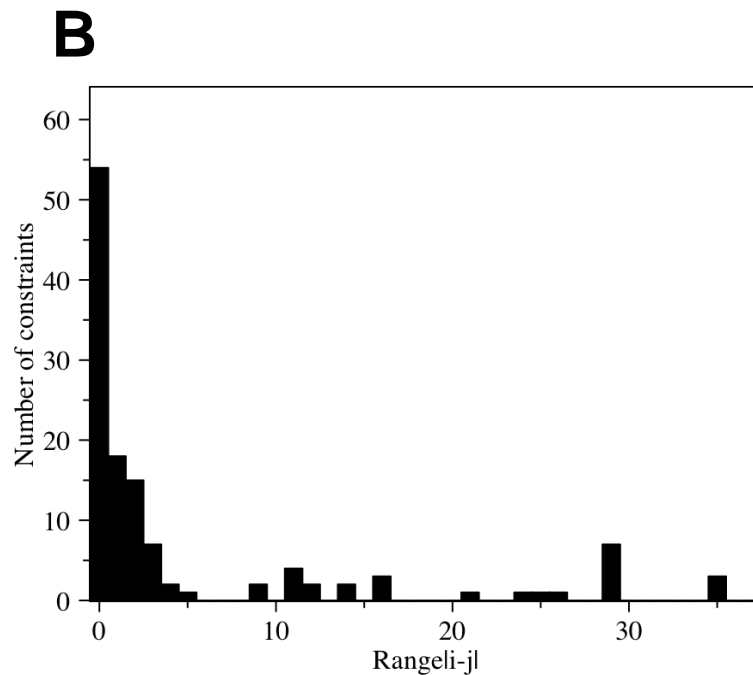
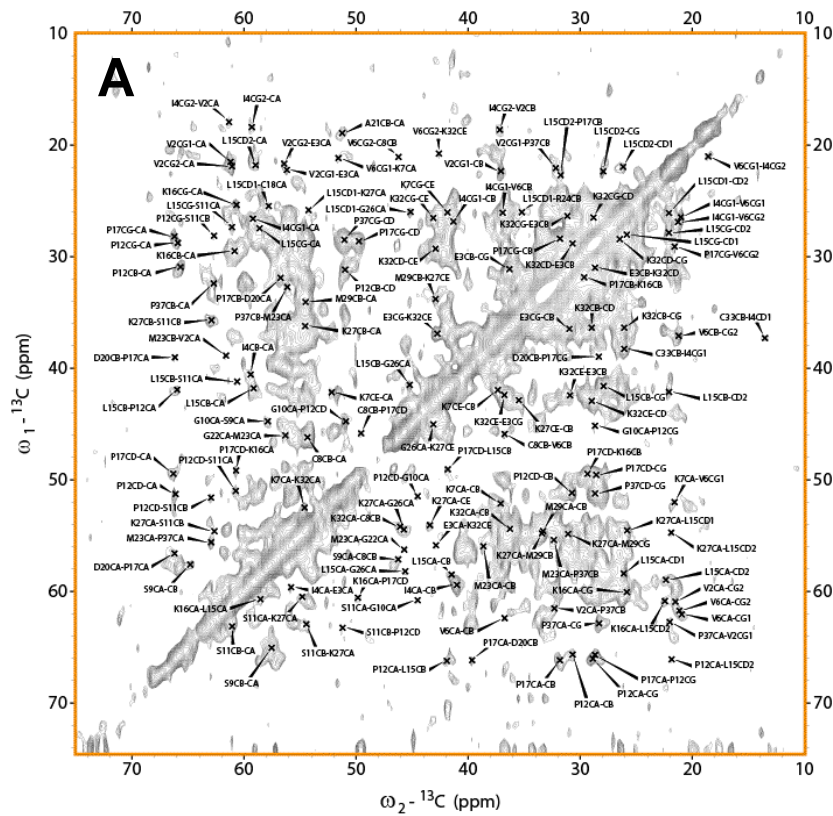


## 4.3 Results and Discussions

### 4.3.1 Solid-state NMR structure of KTX in complex with KcsA-Kv1.3

The approach was used to solve the high-resolution 3D structure of the 38-residue potassium channel blocker toxin KTX existing in complex with membrane-embedded KcsA-Kv1.3.  $^1\text{H}$ - $^1\text{H}$  correlations were detected in 2D CHHC spectrum with mixing times of 325 $\mu\text{s}$ . Proton-proton correlation 2D CHHC spectrum of the solid-state was automatically peak picked using SPARKY 3 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Diagonal peaks were manually removed. Peak intensities obtained from the 2D CHHC spectrum was classified into four ranges and converted into distance ranges of 1.8-2.7, 1.8-3.3, 1.8-5.0 and 1.8-6.0 Å, respectively. For analysis of the solid-state spectra by PASD,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts were labeled as if they were proton chemical shifts. The tolerances for matching chemical shifts to cross-peaks were set to 0.38 ppm in the acquisition dimension and to 0.60 ppm in the indirect dimension. PASD structures do not represent fully-refined NMR structures [Kuszewski, Schwieters 2004]. Therefore, we selected cross-peaks with a final likelihood assignment (as defined by PASD) of 1.0. For KTX bound to the potassium channel, 19% of all long-range restraints, 76% of the medium-range restraints, and 100% of the sequential restraints had final restraint likelihoods of 1.0. All other restraints had final restraint likelihoods of 0. Assignments obtained for these cross-peaks by PASD were verified by manual inspection of the 2D CHHC spectrum.

A large number of distance restraints were unambiguously assigned in solid-state NMR spectra recorded on proteoliposomes containing uniformly labeled [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ] KTX bound to purified KcsA-Kv1.3 (Table 4.1). The structures based on distance and dihedral angle restraints shows very well defined secondary structure



**Figure 4.1:** Interresidue correlations observed in U-[ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]-KTX in complex with unlabeled KcsA-Kv1.3. (A) 2D CHHC spectrum recorded with a mixing time of  $325\mu\text{s}$ . Unambiguously assigned signals are labeled. (B) Number of unambiguously assigned CHHC distance constraints as a function of residue difference  $i$  and  $j$ .

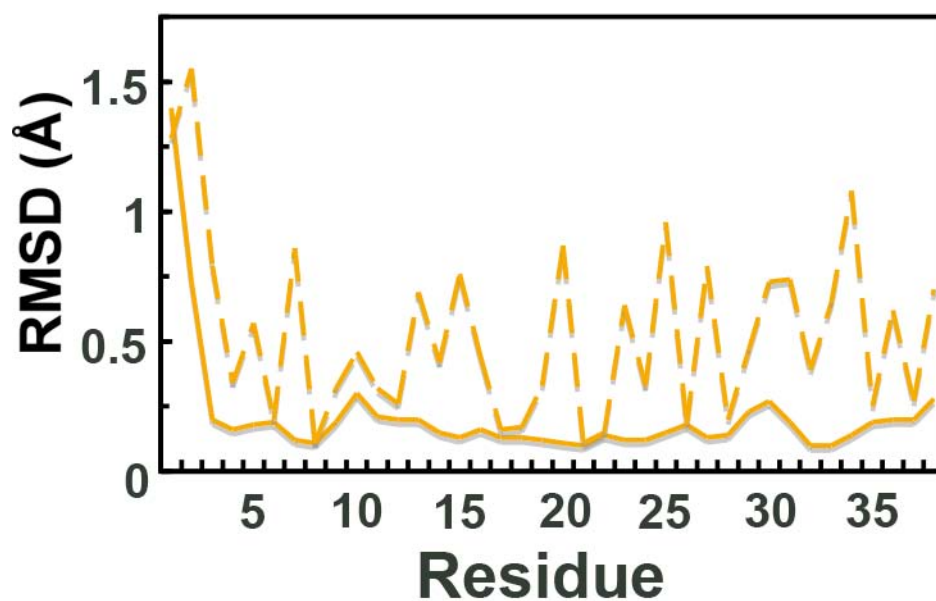
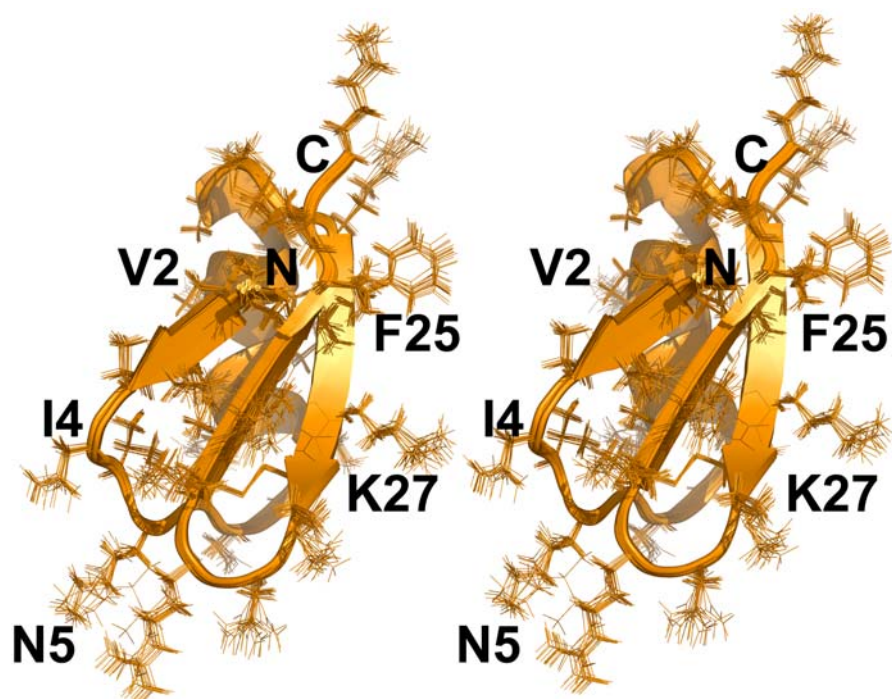
**Table 4.1:** Structural statistics for the 20 lowest-energy structures of KTX in complex with KcsA-Kv1.3.

---

<b><u>Proton-Proton distance correlations</u></b>	
Total	113
Short range	70
Medium range	18
Long range	25
Distance violations (>0.5 Å)	1
<b><u>Dihedral angles</u></b>	
Total	54
Angle violations	3
<b><u>Energy(kcal/mol)</u></b>	
Total	-1241.42
Bond	28.35
Angle	165.25
Improper	46.78
Dihedral	128.31
NOE	93.22
<b><u>Coordinate precision</u></b> <sup>a</sup>	
Backbone RMSD (Å)	0.5
Heavy side chain RMSD (Å)	1.2
<b><u>Ramachandran statistics</u></b>	
Most favored region (%)	83.0
Additionally allowed region (%)	3.7
Generously allowed region (%)	3.3
Disallowed region (%)	0.0

---

<sup>a</sup> Defined as the average rmsd difference between the 20 structures and the mean coordinates.

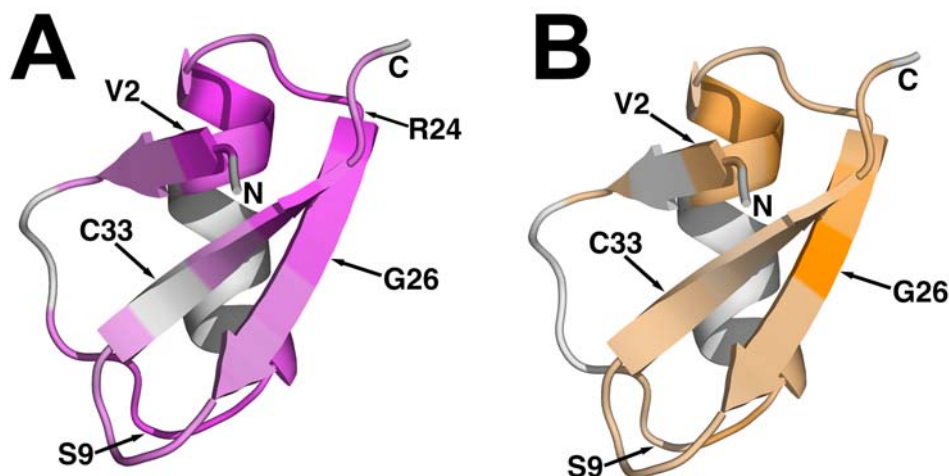


**Figure 4.2:** Stereo view of 20 lowest energy ensemble of high-resolution solid-structure of Kaliotoxin in complex with KcsA-Kv1.3 (upper). Residue based rms deviation of the atomic coordinates within the ensemble of 20 lowest energy structures (lower). Deviations for the backbone atoms and for the non-hydrogen side chain are shown as solid and dashed line, respectively.

elements consisting of  $\alpha$ -helix affixed by disulphide bridges to a three-stranded  $\beta$ -sheet, that is typical for the charybdotoxin family of  $K^+$  channel-blocking scorpion toxins [Grissmer, Nguyen 1994], with rms deviations in the backbone and heavy atoms of 0.5 and 1.2 Å, respectively, for all residues. The backbone and side chains in the resulting structures of KTX in complex with the  $K^+$  channel were well defined as seen from Figure 4.2 and Table 4.1. 83% of residues were found in the most favored region of Ramachandran plot and none of the 38-residues was found in the disallowed region.

At the end of pass 1 and 2 the PASD algorithm calculates likelihood estimates that each particular assignment associated with a cross-peak is correct. The likelihoods are calculated using the ensemble of structures present at the end of the corresponding pass. Thus, they are a metric of how consistent a given assignment is with the ensemble of structures at the end of each calculation pass. Here we have not used the ensemble of structures present at the end of pass 1 and 2 for calculation of likelihood estimates, but the high-resolution structures of KTX obtained under different conditions (solution-state or solid-state free KTX). This improved convergence in the structure calculations and was justified as we previously established that the fold of KTX in solution and in the solid phase is the same. For each state of KTX, in complex with the KcsA-Kv1.3, the structure calculation was done twice, in which the high-resolution structures of the two other states were used for calculation of the likelihood estimates. The PASD calculations of KTX(solid/bound) were done once by using KTX(solution) for calculation of the likelihood estimates at the end of pass 1 and 2. Then a second PASD calculation was done, in which the likelihood estimates were determined using KTX(solid/free). Thus, the PASD calculations were biased on purpose toward the structure of the other state.

In both cases, the structures obtained for KTX in complex with the channel were indistinguishable. Note, that identical structure calculation protocols were used in all cases.



**Figure 4.3:** (A) Structural differences between KTX in solution and the KTX in complex with the potassium channel in solid phase mapped onto the solution-state structure (pink scale). (B) Structural differences between free KTX in the solid phase and in complex with the potassium channel mapped onto the solution-state structure (orange scale).

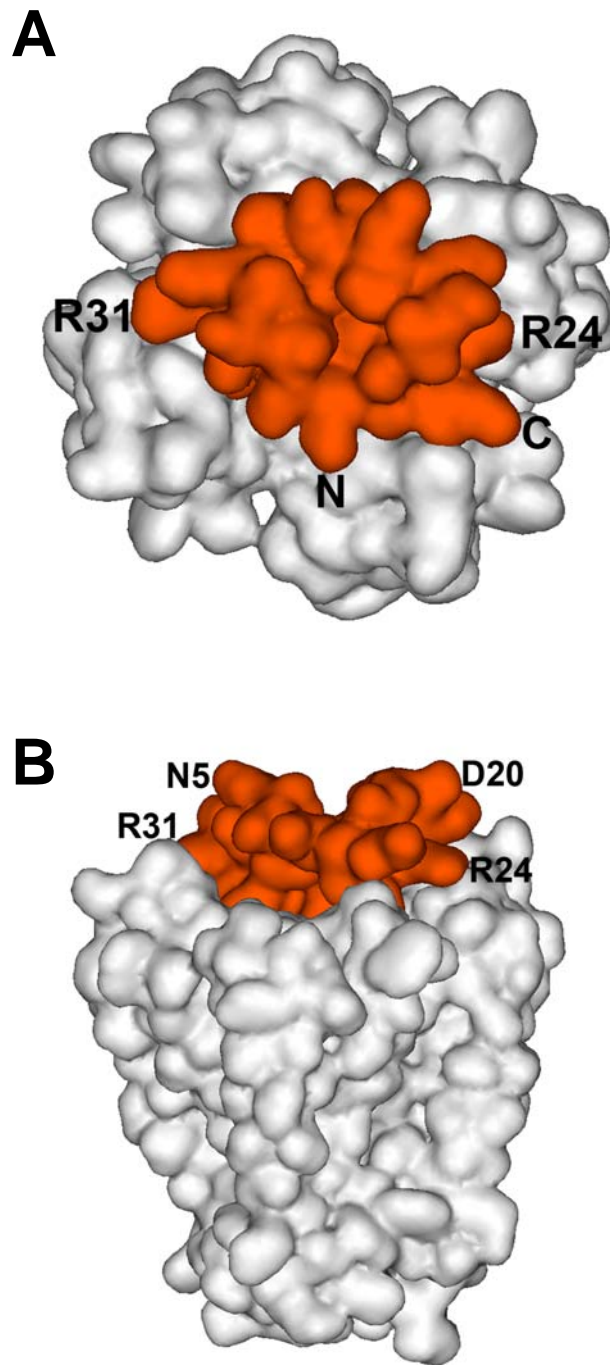
### 4.3.2 Comparison with high-resolution solution-state and solid-state (free) NMR structures of KTX

Differences in the backbone conformation between KTX in solution and KTX in complex with the potassium channel in the solid phase is shown in Figure 4.3A. Residue-based root mean square deviations (one-residue-window) were mapped onto the solution-state structure using a continuous pink scale. Also in Figure 4.3B, differences in the backbone conformation between free KTX in the solid phase and in complex with the potassium channel is shown. Structural differences are mapped onto the solution-state structure using a continuous orange scale. Mapping the structural differences onto the structure of KTX indicated that they mainly occurred in one particular half of the KTX molecule including the C-terminal end of the  $\alpha$ -helix, the

N-terminal half of the second  $\beta$ -strand and the Cys18-Cys35 disulfide bridge connecting  $\alpha$ -helix and third  $\beta$ -strand. The differences between solution and solid-state structures of KTX correlate well with those that were observed upon binding of KTX to the KcsA-Kv1.3 channel. This correlation indicates that transitions between different conformational states of KTX involve flexibility at specific points in the  $\beta$ -sheet of the KTX polypeptide.

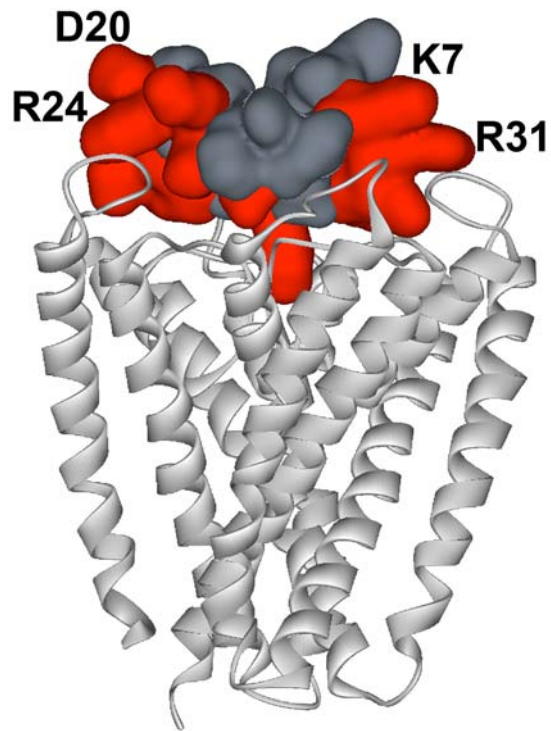
### **4.3.3 Improved model of KTX-KcsA-Kv1.3**

We used the high-resolution 3D structure of bound KTX, the backbone conformation of the selectivity filter of KcsA-Kv1.3 in the complex and functional data to derive an improved model for the KTX – KcsA-Kv1.3 complex (Figure 4.4). Docking of KTX with the model of the KcsA-kv1.3 was performed with HADDOCK version 1.3 in combination with CNS version 1.1 [Brunger, Adams 1998;Dominguez, Boelens 2003]. The starting structures for the docking were the 20 solid-state NMR structures of the bound KTX and the model of the KcsA-Kv1.3 obtained above. In total, 2000 rigid-body docking solutions for the ensemble of 20 NMR structures of bound KTX were first generated by energy minimization. In the initial stages, and the docking is mainly based on the AIRs. Once the structures are within the non-bonded cut-off of 8.5Å, van der Waals and electrostatic energy terms become important. The 200 best solutions, based on the intermolecular energies, were subjected to semi-flexible simulated annealing in torsion angle space followed by a final refinement in explicit water. During simulated annealing and water refinement, residues at the interface, both side chains and backbone, were allowed to move to avoid steric overlaps and clashes.

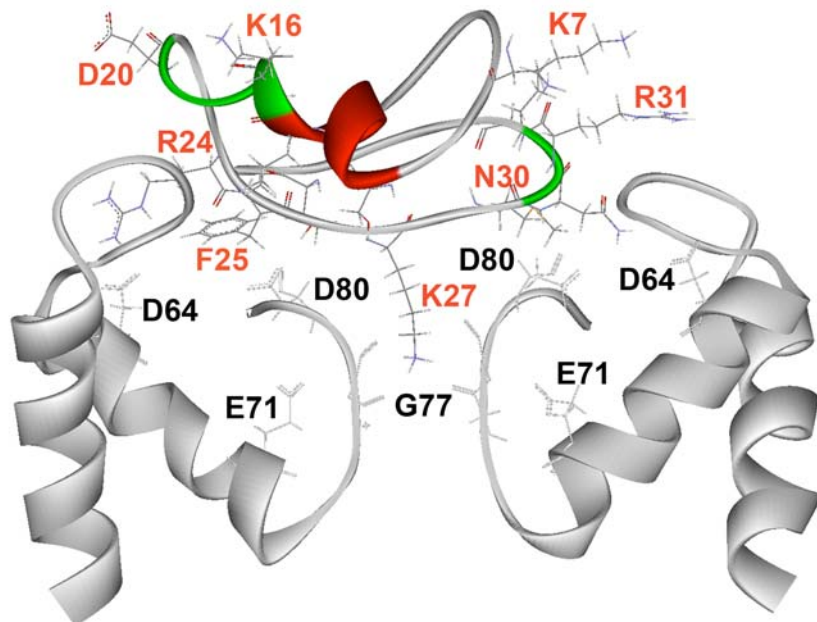


**Figure 4.4:** Model of the KTX – KcsA-Kv1.3 complex shown in different views. In (B), residues 149-160 of KcsA-Kv1.3 were removed to allow a better view of bound KTX.

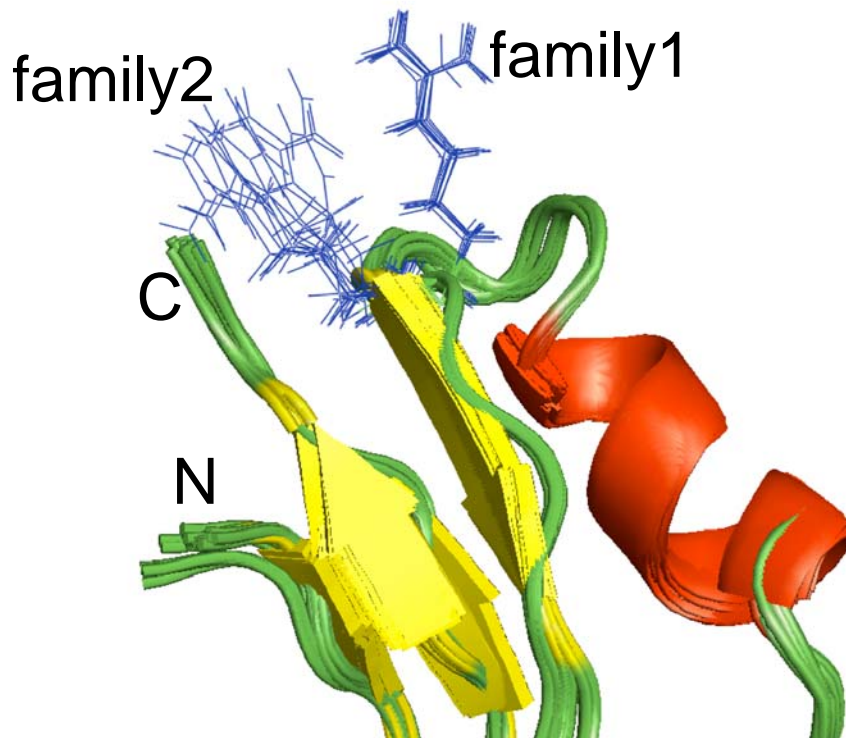




**Figure 4.5:** Model of the complex between KTX and KcsA-Kv1.3 (Protein Data Bank 1K4C). KTX residues that are important for binding to the channel according to mutagenesis studies are highlighted in orange.



**Figure 4.6:** Lys 27 of KTX deeply inserts into the pore region the KcsA-Kv1.3



**Figure 4.7:** Cartoon representation of two families of structures. Two families differed in the orientation of the side chain of R24.

The final 200 structures were clustered into two families based on pairwise backbone r.m.s.d. 20 lowest energy structures were selected in each family. The average r.m.s.d values to the mean structure were 0.45 and 0.46 Å for the backbone atoms of cluster 1 and cluster 2, respectively and the average r.m.s.d to the mean structure were 0.75 and 0.73 Å for the side chain atoms of cluster1 and cluster 2, respectively. The Lowest-energy complex from cluster 1 is shown in the Figure 4.5. Structures in cluster 2 differed from those in cluster 1 by a 180° rotation relative to the structure of KcsA-Kv1.3. We used the high-resolution 3D structure of bound KTX, the backbone conformation of the selectivity filter of KcsA-Kv1.3 in the complex and functional data to derive an improved model for the KTX – KcsA-Kv1.3 complex (Figure 4.5). The positively charged side chain of Lys27, which is highly conserved in scorpion toxins, protrudes from the interaction surface and plugs the K<sup>+</sup> channel pore and comfortably fits into the selectivity filter near G77 (Figure 4.6).

Structure calculation for KTX(solid/bound) resulted in two families that had very similar backbone conformations (maximum r.m.s. deviation of 0.3 Å), but differed in the orientation of the side chain of R24 (Figure 4.7). Out of 50 calculated structures, 30 belonged to family 1 and 20 to family 2.  $^{13}\text{C}_\alpha$ ,  $^{13}\text{C}_\beta$  and  $^{15}\text{N}$  chemical shifts back-calculated from each family using SHIFTX did fit to a comparable degree to experimental values. However, the  $^{13}\text{C}_\alpha$  chemical shift predicted for R24 from family 2 structures differed by 0.9 ppm from the experimental value, whereas the value predicted for family 1 conformations differed by only 0.1 ppm. In addition, the  $^{15}\text{N}$  chemical shift predicted for M23 on the basis of family 1 was much closer to the experimental value (0.6 ppm difference instead of 5.4 ppm difference in case of family 2). When compared to the solid-state structure of free KTX, a smaller structural change is present in family 1 conformations. Both the family 1 and family 2 ensembles were used for docking of KTX to KcsA-Kv1.3. Only in docking solutions obtained with family 1 structures of KTX(solid/bound), R24 was able to form a hydrogen bond with D64. The R24-D64 hydrogen bond is thought to be important for binding of KTX to K<sup>+</sup> channels according to mutagenesis studies. These data suggest that the orientation of the side chain of R24 seen in the family 1 ensemble is the conformation R24 adopts in complex with KcsA-Kv1.3. For further analysis, therefore, only family 1 of KTX(solid/bound) was used.

## 4.4 Conclusions

In this study, the previously developed approach (Chapter 3) allowed the determination of high-resolution structure of KTX bound to a membrane-embedded potassium channel, KcsA-Kv1.3. An improved model of KTX-KcsA-Kv1.3 was presented based on the functional data and chemical shift differences observed for KTX upon binding to the tetrameric KcsA-Kv1.3. In consistent with the previous results and observed data, Lys 27 protrudes from the interaction surface of the KTX and inserts deeply into the selectivity filter of the chimaeric channel. Though, the structure of the bound KTX could be improved by measuring spectra at different mixing times, as it was done for KTX in free form, the overall quality of the structure and the data presented would still hold true. High-resolution structural data along with dynamics studies at the interfacial regions would provide the basis for the formation of a tight complex with the active site of the K<sup>+</sup> channel.

## Chapter 5

### Understanding and Prediction of alignment and residual dipolar couplings of a protein from its known three-dimensional structure

#### 5.1 Introduction

##### 5.1.1 Residual Dipolar Couplings (RDCs)

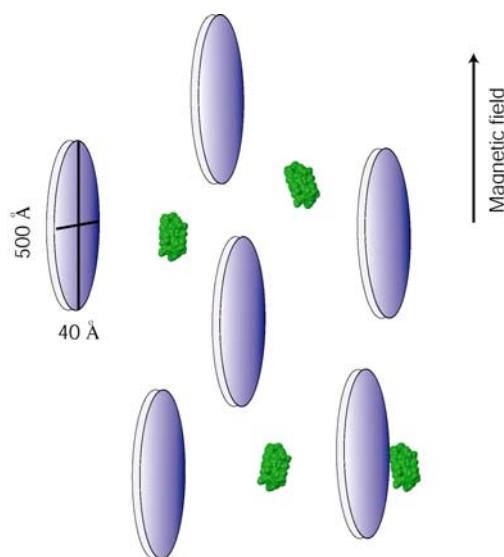
Residual dipolar coupling (RDCs) (§ 1.2.4.3 Residual Dipolar Couplings) offers new opportunities in NMR studies of macromolecules existing in solution [Bax 2003; Prestegard, Al-Hashimi 2000]. This anisotropic interaction contains valuable structural information; hence it is used to address a number of different problems in structural biology. These range from structure validation, to high-resolution structural applications to proteins, to the determination of inter domain orientation in multi domain or multi subunit systems, to characterization of structure and dynamics of proteins and non-protein systems [Clore and Garrett 1999; Cornilescu, Marquardt 1998; Delaglio, Kontaxis 2000; Fischer, Losonczi 1999; Hus, Marion 2001; Jung, Sharma 2004; Markus, Gerstner 1999; Meiler, Prompers 2001; Skrynnikov, Goto 2000; Tolman 2002]. This anisotropic interaction is not normally observable in high-resolution NMR because, it effectively average to zero with molecular tumbling [Prestegard 1998]. However, they can be observed, if small degrees of molecular order are induced. The desired very weak ordering (on the order of  $10^{-3}$ ) or in other words the partial alignment of the biomolecule in solution is usually accomplished by means of a dilute liquid crystalline medium (Figure 5.1) or an anisotropically compressed hydrogels [Barrientos, Dolan 2000; Barrientos, Louis 2001; Clore, Starich

1998;Fleming, Gray 2000;Hansen, Mueller 1998;Prestegard and Kishore 2001;Ruckert and Otting 2000;Tjandra and Bax 1997]. The resulting residual dipolar interactions are scaled down by three orders of magnitude relative to their static values, but nevertheless and because of that can be measured at high accuracy.

It is important to introduce a controllable, but small amount of alignment. Under these conditions, the rotational diffusion of the protein is more or less unhindered, thus maintaining the narrow linewidth typical of liquid state NMR. The partial alignment results also in a net alignment, which allows the measurement of RDCs only for those nuclei that are very strongly coupled. So, the small tunable degree of alignment provides the structural information and narrow lines. RDCs can also be measured in the absence of any liquid crystalline medium when the molecule of interest has a sufficiently large intrinsic anisotropy of its magnetic susceptibility to achieve a preferred orientation with respect to the external magnetic field [Tolman, Flanagan 1995]. This has been shown for both paramagnetic and diamagnetic biomolecules [Contreras, Ubach 1999]. Even for paramagnetic proteins, the natural alignment with a high magnetic field is usually one order of magnitude smaller than that achieved by the use of an orienting medium.

It has been found that not only the degree of alignment, but also its characteristics can be modified by changing the medium. This property turns out to be useful because, if the structure of the macromolecule is assumed to be independent of the medium, the availability of RDCs for different alignment provides a set of independent data [Al-Hashimi, Valafar 2000;Peti, Meiler 2002]. A clear understanding of the dependence of the orientational behavior of a macromolecule on the structure of the aligning medium has not been reached, yet. Alignment would be the result of steric repulsions, which are likely to be similar in the various systems, and electrostatic interactions, which on the contrary, are expected to depend on the

features of the colloidal suspensions. It was previously demonstrated that in a bicelle medium, consisting of large diameter, disk shaped zwitterionic phospholipid micelles; the principal axes of the molecular alignment tensor closely coincide with those of the rotational diffusion tensor [De Alba, Baber 1999;Tjandra and Bax 1997;Tjandra, Omichinski 1997]. This shows that in this nearly neutral medium, alignment is solely defined by the solute's shape. There also have been attempts to provide theoretical predictions of macromolecular alignment. Previously, in a computational approach implemented in



**Figure 5.1:** Alignment of biomolecules in dilute liquid crystalline media.

the program called Prediction of ALignment from Structure (PALES) [Zweckstetter and Bax 2000], it was shown that the magnitude and orientation of the steric component of the molecular alignment could be predicted accurately from the solute's three-dimensional shape. This was done, under the assumption that there are no attractive or long-range repulsive interactions between the solute and the disk or rod shaped liquid crystal particles; steric obstruction effects dominate solute alignment. In this model, the alignment tensor is calculated by averaging over all non-overlapping configurations of a liquid crystal particle and macromolecule. The macromolecular

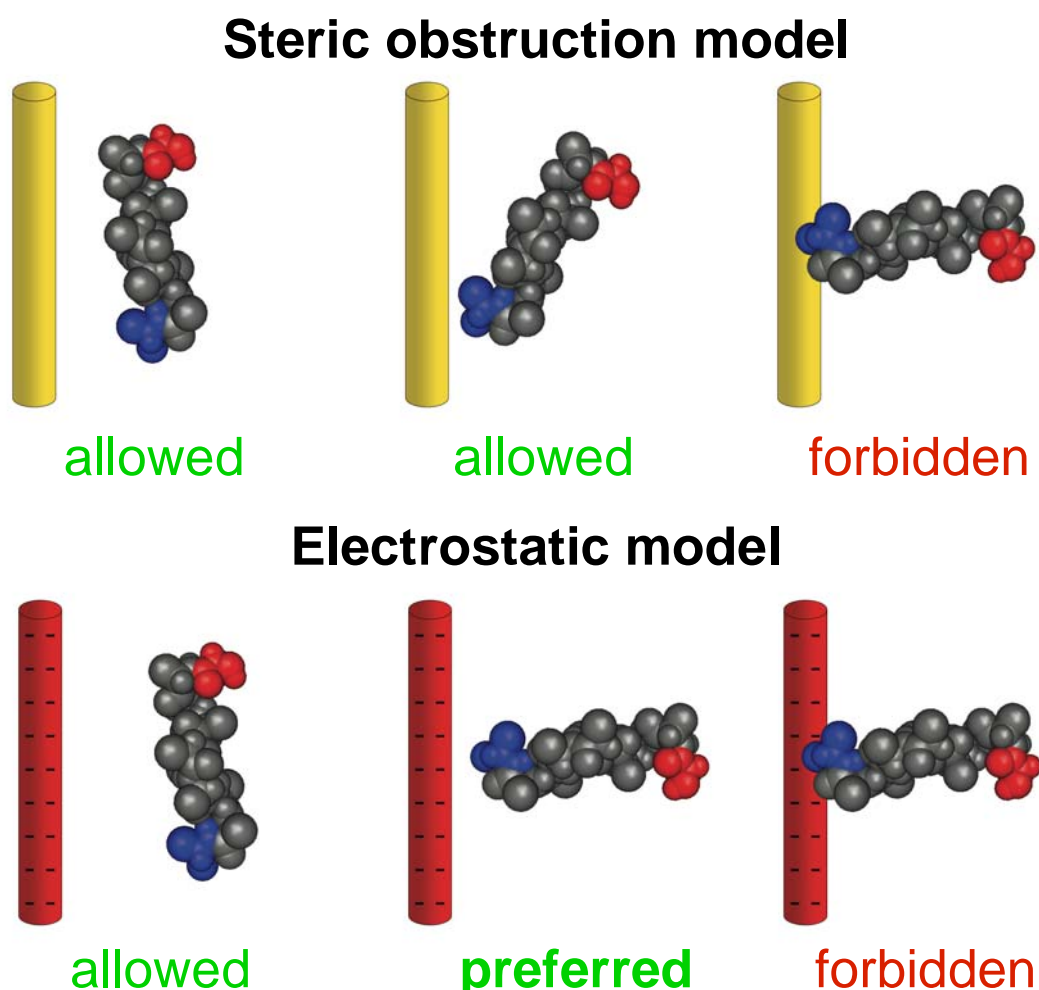
shape is represented with atomistic detail, whereas bicelle was modeled as an infinite wall. This method's ability to predict a protein alignment tensor and hence the dipolar coupling allowed an opportunity in validating structural models [Bewley 2001; Clore and Garrett 1999], determining relative orientation of protein domains [Bewley and Clore 2000], refining nucleic acid structure [Warren and Moore 2001], differentiating monomeric from homodimeric states [Zweckstetter and Bax 2000]. The ability to predict dipolar couplings for a given protein structure also provides unique opportunities when attempting to classify protein fold families on the basis of unassigned NMR data, potentially increasing data throughput in structural genomics [Valafar and Prestegard 2003].

However, when a net charge was added to the bicelles [Ramirez and Bax 1998], the alignment tensor of the protein showed a modification, demonstrating that electrostatic interactions could play a role. As a further support to this view, it has been observed that alignment in virus suspensions can be tuned by changing the ionic strength: the differences with respect to bicellar suspensions are gradually reduced when the ionic strength is increased [Zweckstetter and Bax 2001]. As a direct consequence, the prediction of the macromolecular alignment tensor was less accurate based on the steric obstruction alone for negatively charged liquid-crystalline media. Analysis of solute alignment observed in apolar organic liquid-crystalline media shows that steric interactions frequently dominate, but electrostatic forces could be important [Dingemans, Photinos 2003; Terzis, Poon 1996]. There is still debate as to which electrostatic interactions are most important. Some studies argue that the dipole moment plays a major role [Photinos, Poon 1992; Photinos and Samulski 1993]. Other studies indicate that in apolar media it is primarily quadrupole moment of the solute that dominates its alignment [Syvitski and Burnell 1997; Syvitski and Burnell 2000]. Full calculation of electrostatic forces between charged macromolecules in aqueous



solution is very complex and computationally demanding [Davis and Mccammon 1990;Honig, Sharp 1993].

Recently, a simple extension of the steric obstruction model, which combined the short-range steric and long-range electrostatic interactions, allowed improved predictions of the magnitude and orientation of the alignment tensor of protein and DNA solutes in charged nematic media, such as widely used medium of filamentous phage Pf1 [Zweckstetter, Hummer 2004] (Figure 5.2). Filamentous phage was modeled as an infinitely long cylinder with uniform charge distribution on its surface. The extended model approximates the electrostatic interaction between a solute and the alignment media as that between the solute's surface charges and electric field



**Figure 5.2:** Schematic representation of the steric and electrostatic PALES [Zweckstetter and Bax 2000;Zweckstetter, Hummer 2004]. For the allowed and preferred orientation of the biomolecules, individual alignment tensors were weighed by the relative electrostatic contribution through Boltzmann probability.

strength generated by the medium. Hence the model is not limited to neutral media such as bicells and alkyl poly-(ethylene glycol)-based liquid crystals, but extends to charged liquid-crystalline media. Recently, a similar approach to predict alignment in charged nematic media [Ferrarini 2003], was also reported.

Though, the overall quality was good considering the simplified electrostatic model, there were few notable problems. Due to the continuum level treatment of molecular solvents, which excludes the tight ionic interactions at low salt concentration and ion packing contributions which plays a role at high salt concentrations, the magnitude of the predicted alignment tensors varied considerably from the observed values, for salt concentrations between 0.1 M and 0.3 M and the predictions varied for different proteins. Also the uniform charge distribution on the liquid crystal particle rules out the possibility of the specific interactions with the interfacial charge distribution, when the protein is close to the interface. Rigid approximation of the protein structure and charge distribution on the protein and interface may lead to errors for proteins with flexible charged side chains that could adjust their conformations to optimize electrostatic interactions with interface. Because minor rearrangement of side chains of charged surface residue dramatically affected the prediction. This may be either due to the quality of the structural model or due to less detailed charged model. Hence demanding for a detailed atomistic treatment of the solute and solvents.

Here in this study we focus on the incorporation of improved electrostatic model in PALES to predict the molecular alignment and residual dipolar couplings of a protein from its known three-dimensional structure. Detailed electrostatic calculations were performed, by solving nonlinear Poisson-Boltzmann equations for the protein/bilayer system, where protein and bilayer were represented in atomic details. Nonlinear Poisson-Boltzmann equation was solved for different orientations and

translational positions of a protein in front of a liquid crystal particle. Electrostatic potentials and free energies were then used within PALES to calculate the alignment tensor and the residual dipolar couplings of the protein. Dipolar couplings calculated from these electrostatic free energies will be compared to experimental values, which had been measured in the Rec A binding protein DinI dissolved in Pf1 bacteriophage. Purple membranes, like bacteriophage Pf1, are highly negatively charged, and the alignments in both cases are dominated by electrostatic interactions. It was shown previously that, observed alignment tensors for proteins dissolved in purple membrane or bacteriophage Pf1 are similar [Koenig, Hu 1999;Lakomek, Carlomagno 2006;Lakomek, Griesinger (personal communications)], depending on salt concentrations of the solute medium. These findings provide experimental evidence supporting our case of using RDCs measured in bacteriophage Pf1 medium for the calculation of alignment tensors in membrane system (here lipid bilayers). Although short-range interactions between the liquid crystal particle and the protein are not taken correctly into account in this manner, the influence of the protein on the overall electrostatic potential can be taken into account explicitly. Thus, we can overcome one of the major limitations of the simple electrostatic model currently used in PALES. Solving a more detailed electrostatic model expected to improve the quality of the charge/shape prediction of the alignment.

### **5.1.2 Electrostatics**

Electrostatic interactions play an important role in both the structure and function of molecules [Honig, Sharp 1993]. The complex shape of the macromolecules dramatically affects their electrostatic field, and can be crucial to their function. The finite difference Poisson-Boltzmann (FDPB) method [Davis and Mccammon

1990;Honig and Nicholls 1995;K. A. Sharp and B. Honig 1990] has been widely applied to describe the electrostatic properties of proteins, nucleic acids, and membranes [Baker and Mccammon 2003]. In a series of experimental/theoretical studies over the past decade, several research groups have used this method to describe the binding of charged peptides and proteins to membrane surfaces. These works demonstrated that the FDPB method is remarkably accurate in treating electrostatic properties associated with non-specific binding. This is because many of the relevant interactions are long range and depend only on the charge distribution of the system. The computational methodology is based on finite difference solutions to the full (nonlinear) Poisson-Boltzmann (PB) equation, an equation of classical electrostatics [Gallagher and Sharp 1998;Honig and Nicholls 1995;Holst 1993;Gilson 1988;Kim A. Sharp and Barry Honig 1990;K. A. Sharp and B. Honig 1990]. The finite difference method is a numerical procedure that allows the solution of the PB equation for biological molecules that often have complex shapes. The PB equation is as follows:

$$\nabla[\epsilon(r)\nabla\varphi(r)] - \epsilon_r\kappa(r)^2 \sinh[\varphi(r)] + \frac{e^2}{(\epsilon_0 k_B T)} \rho^f(r) = 0 \quad (5.1)$$

where  $\epsilon(r)$  is the dielectric constant,  $\varphi(r)$  is the electrostatic potential,  $\kappa(r)$  is the Debye-Hückel parameter and  $\rho^f(r)$  is the fixed charge density of the protein and lipids; all four variables are functions of space,  $r = (x, y, z)$ . A protein/membrane system is represented atomistically and is mapped onto a three-dimensional lattice of points; each lattice point represents a small region of the protein or membrane ( $\epsilon = 2$ ; to approximate electronic polarizability) or the aqueous phase ( $\epsilon = 80$ ; to account for the highly polar nature of water molecules).

The solutions to the PB equation are the electrostatic potentials,  $\varphi(r)$ , which are

used to calculate the electrostatic free energy of a system,  $G_{el}$  [Kim A. Sharp and Barry Honig 1990]. The electrostatic free energy of interaction between protein and membrane,  $\Delta G_{el}$ , is calculated as the difference between the electrostatic free energy of the protein and membrane when they are close together,  $G_{el}(P.M)$ , and when they are far apart,  $G_{el}(P)$  and  $G_{el}(M)$  [Ben-Tal, Honig 1996]:

$$\Delta G_{el} = G_{el}(P.M) - [G_{el}(P) + G_{el}(M)] \quad (5.2)$$

When the protein is far from the membrane surface, it experiences an electrostatic attraction that drives it toward the negatively charged membrane. This ‘‘Coulombic’’ attraction increases as the protein approaches the membrane. Close to the membrane surface, charged and polar groups on the both the protein and membrane are desolvated, or stripped of water molecules, which is energetically unfavorable and results in a repulsion at small distances. The minimum electrostatic free energy of interaction of protein is predicted to occur where the van der Waals surface of the peptide and membrane are separated by a distance of about the thickness of a layer of water. At this distance the Coulombic attractive force is balanced by the repulsive desolvation force. The components of the electrostatic free energy of interaction, namely the Coulombic attraction and the desolvation repulsion, can be calculated explicitly. The desolvation penalty  $\Delta G_{desolv}$  is determined at each R by calculating to what extent the presence of the protein and the membrane shield the others from favorable interactions with solvent [Misra and Honig 1995; Misra, Sharp 1994]. This is done by discharging the protein and membrane in turn and calculating the electrostatic free energy of the system:

$$\Delta G_{desolv} = \Delta G_{el}(P(Q=0).M) + \Delta G_{el}(P.M(Q=0)) \quad (5.3)$$

The Coulombic component  $\Delta G_{Coul}$  is simply the difference between the total electrostatic free energy of interaction and the desolvation penalty:

$$\Delta G_{Coul} = \Delta G_{el} - \Delta G_{desolv} \quad (5.4)$$

The membrane/protein association could generally be given as combination of nonspecific electrostatic and hydrophobic interactions. The non-polar contribution could be given in simple expression:

$$\Delta G_{np} = \gamma \Delta A, \quad (5.5)$$

where  $\gamma$  is the surface tension coefficient and  $\Delta A$  is the change in solvent accessible surface area upon association. An approximation to the total nonspecific free energy of membrane association of a peripheral protein can be taken as the sum of  $\Delta G_{el}$  and  $\Delta G_{np}$ , so that

$$\Delta G = \Delta G_{el} + \Delta G_{np} \quad (5.6)$$

The additivity of electrostatic and non-polar interactions qualitatively explains membrane binding behavior with protein. Long-range electrostatic attraction increases probability of a protein/membrane interaction and helps orient the protein. As the protein approaches the membrane surface, there are two competing short-range interactions: (1) the (repulsive) desolvation of both the protein and membrane as both, which contain charged and polar groups, begin to lose their favorable interactions with the polar aqueous solvent; and (2) the favorable non-polar partitioning of hydrophobic groups into the membrane interface as described by the interfacial hydrophobicity scale. The FDPB method, allows us to examine quantitatively how the combination of favorable electrostatic attraction, repulsive desolvation, and favorable non-polar interactions leads to different membrane binding behaviors. Each case must be considered separately since the relative contribution of each of these energetic terms will depend on the structure and composition of the membrane/protein system.

## 5.2 Materials and Methods

### 5.2.1 Experimental residual dipolar couplings

Dipolar coupling was previously measured for the Rec-A-inactivating protein DinI dissolved in dilute suspension of Pf1 bacteriophage on 300 mM salt concentrated  $^{15}\text{N}$ -enriched sample [Zweckstetter, Hummer 2004].  $^{15}\text{N}$ - $^1\text{H}$  RDCs were derived from in-phase/anti-phase (IPAP) [ $^{15}\text{N}$ ,  $^1\text{H}$ ]-heteronuclear single quantum coherence (HSQC) experiments [Ottiger, Delaglio 1998]. Alignment tensors in liquid-crystalline or paranematic Pf1 were obtained by fitting  $^{15}\text{N}$ - $^1\text{H}$  RDCs to high-resolution NMR DinI (PDB code: 1GHH) [Ramirez, Voloshin 2000] structures, using singular value decomposition (SVD) as implemented in the program PALES [Zweckstetter and Bax 2000].

### 5.2.2 Coordinates and Structural Models

In our calculations high-resolution NMR structure for 81-residue RecA-inactivating protein DinI was used from the protein data bank (PDB) (PDB code: 1GHH) [Ramirez, Voloshin 2000]. Structural models for phospholipid bilayers were built as described [Peitzsch, Eisenberg 1995]. We used all-atom molecular model bilayer containing 33% acidic lipid (2:1 PC/PS). The zwitterionic PC and negatively charged PS lipids in each leaflet, containing 192 lipids, were uniformly distributed in a hexagonal lattice. Each lipid headgroup occupies an area of  $68 \text{ \AA}^2$  in the plane of the membrane and the lipid head group regions from the two opposing membrane leaflets encompass about half the thickness of the bilayer. It is assumed that the lipids change neither structure nor position upon interaction with the protein. These approximations are not unreasonable for the protein DinI which is charged and do not penetrate the membrane interface to a large degree. Previous work has shown that the membrane

partitioning of basic peptides that reside outside the polar envelope of the membrane is independent of whether the membrane is in the liquid crystalline or gel phase [Ben-Tal, Honig 1996], suggesting that the use of static bilayer models is appropriate for calculating the alignment of peripheral proteins.

### 5.2.3 Model of Protein/bilayer

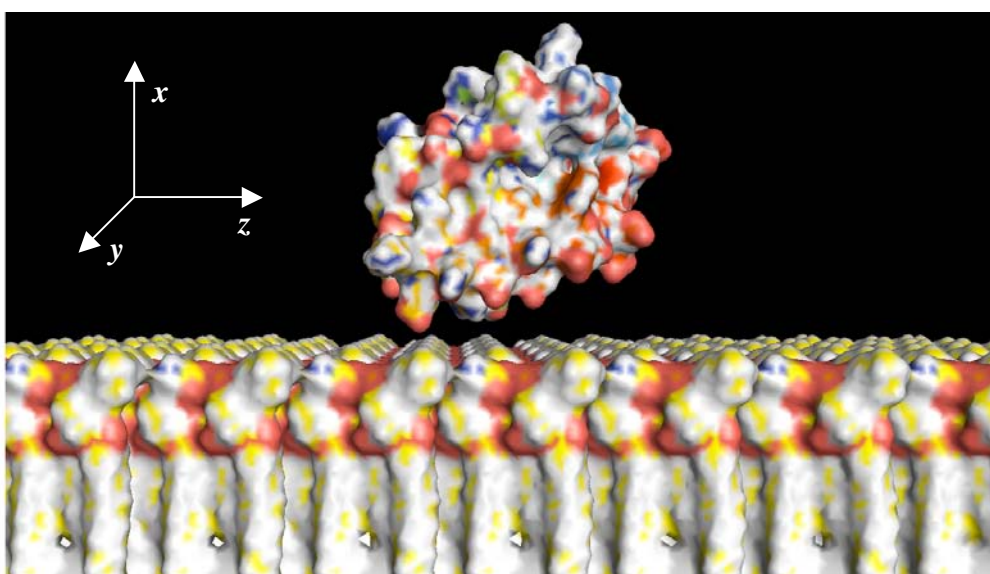
In the calculations described in this work, we represented the protein DinI and the lipid bilayer in atomic detail, as opposed to simplified electrostatic model in PALES, and the solvent as homogenous medium of constant dielectric. Both the Debye length and the dimension of DinI are much smaller than the radius of the vesicles so we can assume the protein DinI interact with a planar surface. Each atom of a protein/bilayer system is assigned a radius and partial charge that is located at its nucleus. The charges and radii used for the amino acids were taken from a CHARMM22 parameter set [Brooks 1983]; those used for lipids are the ones described and used in previous studies [Ben-Tal, Honig 1997; Ben-Tal, Honig 1996; Murray, Hermida-Matsumoto 1998]. The lipid bilayers are negatively charged with a total surface charge density of  $1 e/204 \text{ \AA}^2$  (33% PS), which closely mimics the surface charge density, of  $-0.475 e/\text{nm}^2$ , on bacteriophage Pf1.

We built atomic models with the DinI molecules in different configurations relative to the lipid bilayers; one of them is shown in Figure 5.3. For each configuration, the protein was docked in the aqueous solution above the membrane and steric clash was avoided by manual inspection. Each configuration is determined by six coordinates: three Cartesian coordinates ( $x, y, z$ ) define location of the geometrical center of the protein relative to the center of the membrane section, and three angular coordinates ( $\theta, \alpha, \eta$ ) define its orientation with respect to the bilayer.



The rotational angles were measured relative to the orientation of Figure 5.3.  $\theta$  denotes rotations of the protein around x-axis, the normal to the membrane surface,  $\alpha$  denotes rotation of the protein around z, an axis parallel to the membrane surface in the plane of the figure, and  $\eta$  denotes rotation of the protein around y, an axis parallel to the membrane surface and perpendicular to the plane of the figure.

We define  $R$  as the minimal distance between the van der Waals surfaces of the protein and the lipid bilayer along the x-axis, with  $R = 0$  at van der Waals contact.



**Figure 5.3:** Molecular model of a portion of a PC/PS (2:1) bilayer membrane and protein DinI. The van der Waals surface of the protein and membrane are shown. The molecular surface created by rolling a sphere with the radius of water molecule, 1.4 Å, on the atomic model above.

Denoting the distance and orientation dependent electrostatic free energy of interaction between a protein molecule and the membrane at a given configuration by  $E_{el}(x, y, R, \alpha, \eta, \theta)$  and choosing a reference state where the energy of interaction is zero when the protein is at  $R = \infty$ , i.e.,  $E_{el}(x, y, \infty, \alpha, \eta, \theta) = 0$ , the total electrostatic free energy of interaction,  $\Delta E_{el}$ , would be calculated as given by Equation 5.1.

## 5.2.4 Alignment tensor prediction

The Boltzmann probability  $p_B$  could be calculated from the electrostatic free energy of interaction as follows,

$$p_B = \exp[-\Delta E_{el}(x, y, R, \alpha, \eta, \theta) / k_B T] \quad (5.7)$$

Alignment tensor for each distance and orientation,  $\mathbf{A}_{protein}$ , was then calculated using the equation below,

$$\mathbf{A}_{protein} = \int \mathbf{A} p_B(x, y, R, \alpha, \eta, \theta) dV d\Omega / \int p_B(x, y, R, \alpha, \eta, \theta) dV d\Omega \quad (5.8)$$

where  $V$  and  $\Omega$  denotes the Cartesian and angular volume, respectively and the alignment matrix,  $\mathbf{A}$ , obtained from the steric effect has its components,

$$A_{ij} = 1 / 2(3 \cos \theta_i \cos \theta_j - \delta_{ij}) \quad (i, j = x, y, R). \quad (5.9)$$

## 5.2.5 Electrostatic calculations

We calculated electrostatic potentials,  $\phi(r)$ , and electrostatic free energies of interaction,  $E_{el}(x, y, R, \alpha, \eta, \theta)$ , for each of the 122 configurations from a modified version of the DELPHI program [Gallagher and Sharp 1998], which is adapted to solve the nonlinear Poisson-Boltzmann equation for protein/membrane systems [Ben-Tal, Honig 1996; Peitzsch, Eisenberg 1995]. DELPHI produces finite-difference solutions to the Poisson-Boltzmann equation on a cubical lattice for a system where the solvent is described in terms of a bulk dielectric constant and mean concentrations of mobile ions/atoms, while the protein and the lipids comprising the bilayer are described in terms of the coordinates of the individual atoms as well as their atomic radii and partial charges.

The protein DinI and membrane model were then mapped onto a three-dimensional cubic lattice of  $257^3$  points, each of which represents a small region of

the peptide, membrane or solvent. The molecular surfaces of the DinI and the membrane are defined as the point of contact between a spherical probe with the radius of a water molecule (1.4 Å) and the van der Waals surface. The spaces enclosed by the molecular surfaces, or interior regions, were assigned a low dielectric constant of 2. The space outside the molecular surfaces, or the exterior region, was assigned a high dielectric constant of 80 [Ben-Tal, Honig 1996]. We solved the nonlinear Poisson-Boltzmann equation in the finite-difference approximation, for each distance and orientation of the protein DinI and the membrane and calculated  $E_{el}(x, y, R, \alpha, \eta, \theta)$  at this configuration from the spatial distribution of the charges and the electrostatic potential. Electrostatic free energies are obtained from the calculated potentials, and the electrostatic free energy of interaction is determined as the difference between the electrostatic free energy of the protein DinI in a specific orientation with respect to the membrane surface and the electrostatic free energies of the DinI and membrane infinitely far apart, i.e., taken separately, as given by Equation 5.2. The numerical calculation of the potential is iterated to convergence, which is defined as the point at which the potential changes less than  $10^{-5}$  kT/e between successive iterations.

A sequence of focusing runs of increasing resolution was employed to calculate the electrostatic potentials (e.g., 0.375, 0.75, 1.5 and 3.0 grid/Å). In the initial calculation, the DinI/membrane model encompassed a small percentage of the lattice (10%), and the potentials at the boundary points of the lattice are approximately zero; this procedure ensures that the system is electro-neutral. The calculations with atomic models of membranes were performed in a final resolution of 3 grid/Å and lattice sizes of  $257^3$  were used. We also tested the convergence of the results with respect to the lattice size and scale: increasing the grid box from  $257^3$  to  $321^3$  (at a constant scale

of 3 grids/Å) alters the electrostatic free energy of interaction of DinI with a 2:1 PC/PS membrane in 300mM monovalent salt by <0.1 kcal/mol (Data not shown). The convergence with respect to lattice scale depends on the distance between the protein and membrane and on their relative orientation. The results obtained using a grid box of  $257^3$  and scales of 0.5, 1.0 and 2.0 grids/Å differed by only <0.3 kcal/mol (Data not shown) and the depth of the minimum in the electrostatic free energy, which dominates the electrostatic component of the binding free energy, changed by only <0.2 kcal/mol (Data not shown). The precision in the electrostatic free energies of interaction, determined as the difference between the results obtained at the two highest resolution scales, is 0.3 kcal/mol for all calculations.

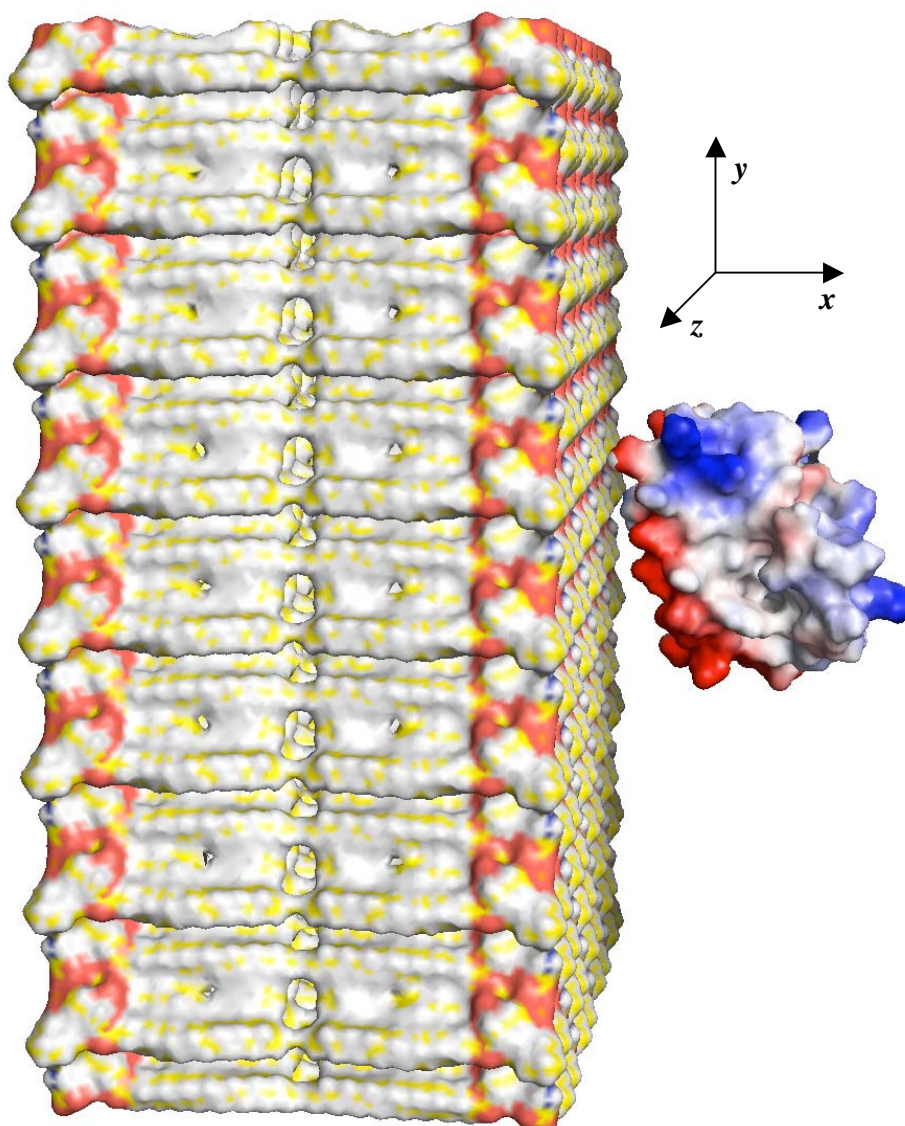
## 5.3 Results and Discussions

### 5.3.1 Calculation of electrostatic free energies

In this study, the simplistic electrostatic model as described previously [Zweckstetter, Hummer 2004], was replaced by an atomistic detailed electrostatic effect to predict the alignment tensor of the protein DinI. 2:1 PC/PS Lipid bilayer was used to mimic the charge density of infinite cylinder (Pfl bacteriophage),  $-0.5 \text{ e/nm}^2$ . The center of gravity of the van der Waals surface of the protein DinI is moved on a one-dimensional grid (along the x-axis, the normal to the bilayer surface), with a grid spacing between grid points of  $0.5 \text{ \AA}$  and  $1 \text{ \AA}$ , away from the surface of the lipid bilayer. The molecule was rotated only in the y-z plane to sample 122 orientations on a unit sphere, as determined by a double cubic lattice method, as it provides a highly uniform sampling [Eisenhaber 1995] and the  $20^\circ$  rotation around the x-axis was not performed, as that would mean solving non-linear Poisson-Boltzmann equation for 2196 configurations, which is computationally very expensive. Therefore, for all results presented in this study, only 122 orientations were used.

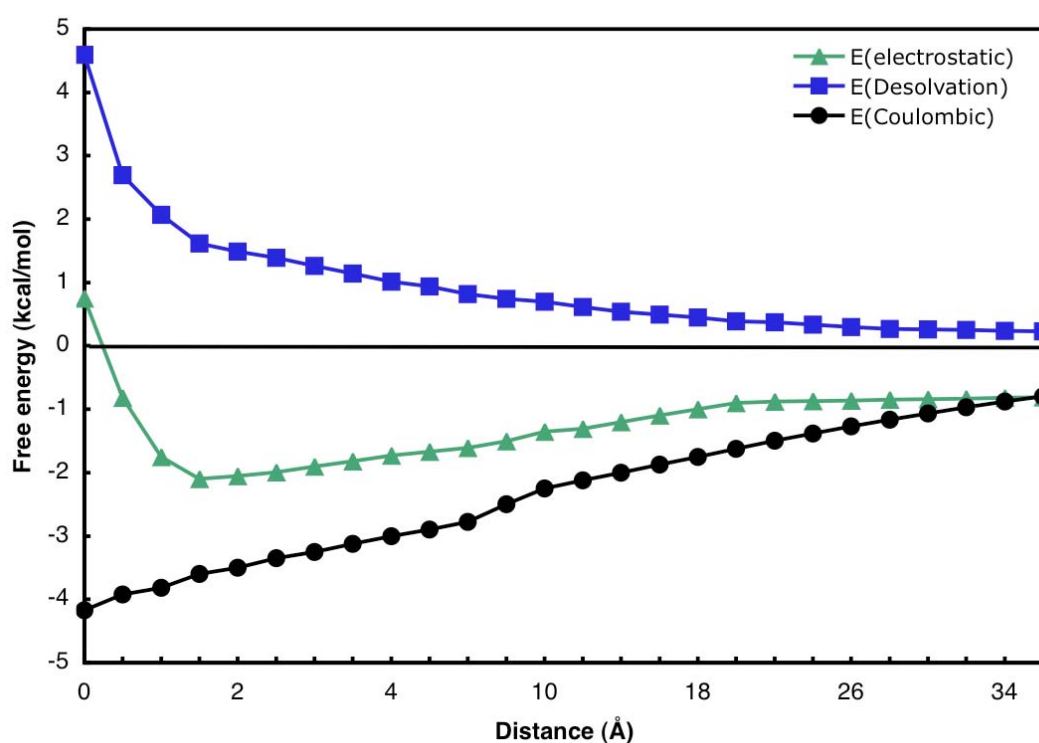
Figure 5.4 shows the one (mean) of the 122 orientations of the DinI in front of the phospholipid bilayer. Phospholipid bilayer lies in the y-z plane and protein is translated along the x-axis, normal to the bilayer surface. In the simplified electrostatic model, for each orientation of the protein, with respect to the membrane surface, nonexcluded alignment matrix was calculated according to Equation 5.9. Each nonexcluded matrix was weighed according to its Boltzmann probability after calculating the electrostatic potential from the simplified electrostatic model [Zweckstetter, Hummer 2004]. In this model, the electrostatic free energy in Equation 5.7 was replaced by the solution to the nonlinear Poisson-Boltzmann equation calculated with full atomistic detail.

The distance and orientation dependent electrostatic free energy was calculated as discussed in Materials and Methods. Figure 5.4 shows one (mean) of the 122 orientations of the DinI in front of the phospholipid bilayer. The most attractive orientation should have the maximum number of positive charges facing the lipid bilayer because binding is driven mainly by electrostatic interactions. The positive



**Figure 5.4:** Model of the alignment of protein. One of the 122 orientations of protein DinI at the surface of a 2:1 PC/PS bilayer in 300mM salt is shown. Lipid bilayer mimics the Pf1 surface's charge density of  $-0.5 \text{ e/nm}^2$ . Van der Waals surface of the bilayer and the electrostatic surface of the protein are shown. Lipid bilayer is oriented with the normal to its surface along the x-axis. The minimal distance between the van der Waals surfaces of the protein and the membrane in this configuration is  $R = 2 \text{ \AA}$ .

potential due to the protein extends to the surface of the bilayer, accounting for the strong favorable electrostatic interaction. Some of the acidic groups on the protein are clearly repelled by the bilayer (e.g., regions of negative potential in the protein). Though suggestive, the figure clearly indicates the limitations of qualitative analysis. There is no simple way, for example, to know the relative magnitudes of the contribution of different regions of the protein. Also, there would be different contributions to the electrostatic free energy of interactions such as Coulombic, non-polar etc [Ben-Tal, Honig 1997; Murray, Arbuzova 2002]. This would be due to the fact that, when the protein approaches the membrane surface, the charges on the protein and the membrane are transferred to a low dielectric region, which would give rise to Born repulsion (desolvation effects).



**Figure 5.5:** Electrostatic free energy of interaction for the orientation, shown in Figure 5.4, between DinI and a 2:1 PC/PS bilayer in 300mM salt. Free energy of interaction is plotted as a function of distance between van der Waals surfaces of DinI and the membrane; Black circles represent the long-range Coulombic attraction. Blue squares represent the unfavorable short-range desolvation repulsion, and the balance between the long-range Coulombic and short-range desolvation repulsion results in an electrostatic free energy minimum (green triangles) when the protein and membrane separated by a layer of water (refer Equation 5.4).

Figure 5.5 plots the electrostatic free energies of interaction for DinI with a 2:1 PC/PS bilayer in 300mM salt as a function of  $R$ , the distance between the surfaces of the protein and membrane. As depicted by the curves, the electrostatic free energy becomes increasingly attractive as the protein approaches the membrane surface but is highly repulsive at short distances. The long-range attraction is due to Coulombic interactions between negatively charged membrane and the positively charged surface of the protein, while the repulsion is due principally to the removal of water from the charged surfaces of the protein and membrane at short distances (desolvation effects) but also repulsive Coulombic interactions between negatively charged groups on the protein and membrane. The combination of these opposing effects - Coulombic attraction versus Coulomb and desolvation repulsions - results in a minimum in the electrostatic free energy at  $R \sim 2.0 \text{ \AA}$ , the diameter of a water molecule. The orientation corresponding to the minimum in the free energy is depicted in Figure 5.4.

### 5.3.2 Prediction of molecular alignment

To compare the detailed electrostatic calculation with the simplified electrostatic model [Zweckstetter, Hummer 2004], we predicted the alignment tensor from the electrostatic free energy of interaction, as obtained by solving full nonlinear Poisson-Boltzmann equation, and compared with the alignment tensor obtained from simplified model. Different contributions (as from Figure 5.5) were also taken into account to predict the alignment tensor.

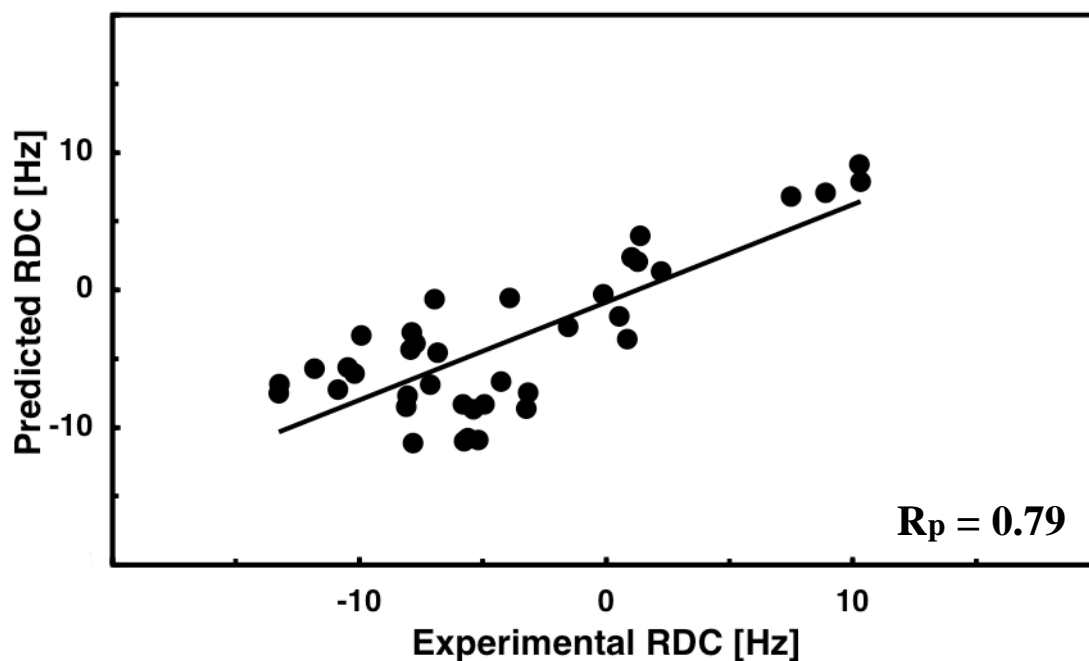
Experimental RDCs have been measured for protein DinI and the RDCs fit very well with the high-resolution NMR structure of DinI molecule, when evaluated by SVD, with a Pearson's linear correlation coefficient,  $R_p$ , of 0.985 [Zweckstetter, Hummer 2004].  $^{15}\text{N}$ - $^1\text{H}$  dipolar couplings ( $^1D_{\text{NH}}$ ) predicted on the basis of a purely



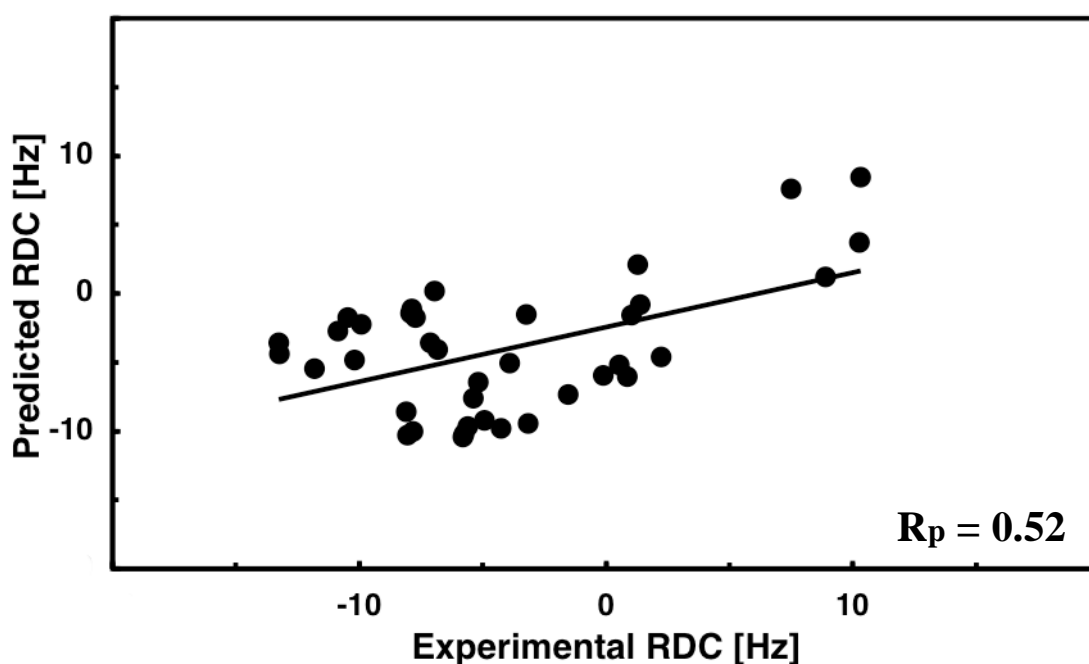
steric obstruction effect and simplified electrostatic model correlate with the experimentally observed RDCs, with the correlation coefficient of 0.34 and 0.96, respectively, with DinI molecule. This already indicates that in highly charged liquid-crystalline media, such as bacteriophage Pf1, interactions other than steric obstruction are important for molecular alignment and inclusion of electrostatic interactions, treated in the highly simplified manner improved the prediction of RDCs and the magnitude of alignment, dramatically. Pearson's correlation coefficient of 0.96 was obtained with large number of (2196) orientations. Since the atomistic models require more computation time, much less, namely 122 could be calculated. When the number of orientations 122 and 1 was considered, the simplified model, as implemented in PALES [Zweckstetter, Hummer 2004], yielded only a correlation coefficient of 0.78 and 0.57, respectively.

### **Prediction of alignment from the minimum energy orientation**

Detailed electrostatic calculations were done to predict the residual dipolar couplings from the minimum energy orientation (mean orientation). Figure 5.5 represents the minimum electrostatic free energy of interaction between protein DinI and lipid bilayer. In that particular (mean) orientation, for each of the distance along the x-axis in front of the lipid bilayer, the nonexcluded matrix A, which was obtained within PALES, was weighted according to its Boltzmann probability, obtained from the detailed electrostatic calculations (the total electrostatic free energy of interaction (Figure 5.5) was used to calculate the alignment tensor and the RDCs as by Equation 5.8). Saupe order matrix was obtained from the weighted nonexcluded alignment matrix. For the 81-residue rec-A-binding protein DinI, experimental  $^1D_{NH}$  values correlated with the values predicted from the Saupe ordered matrix (calculated from the electrostatic free energies for the mean orientation by atomistic model) with a



**Figure 5.6:** Comparison of experimental  $^1D_{NH}$  values measured at 300mM salt with values predicted from the total minimum electrostatic free energy configuration as in Figure 5.4.

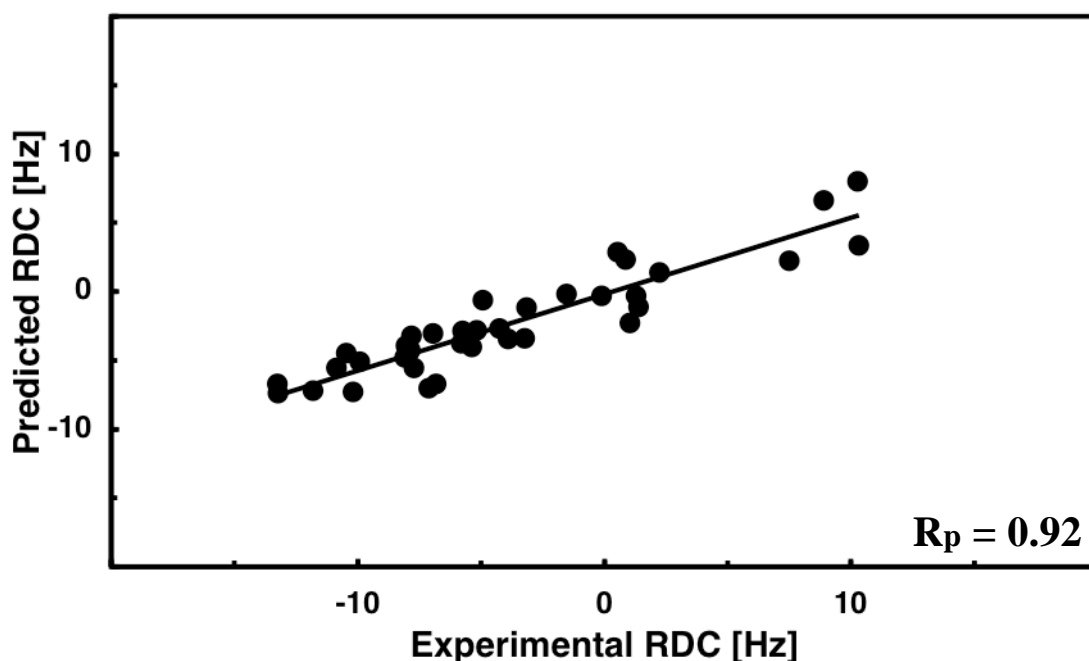


**Figure 5.7:** Comparison of experimental  $^1D_{NH}$  values measured at 300mM salt with values predicted, based only on the Coulombic term, from the total minimum electrostatic free energy configuration as in Figure 5.4.

Pearson's correlation coefficient  $R_p = 0.79$  (Figure 5.6), whereas, the correlation coefficient obtained from the simplistic electrostatic model was 0.57 (by sampling only one orientation), as compared to the 2196 orientations that obtained correlation factor of 0.96 [Zweckstetter, Hummer 2004]. To take into account only the contributions from the electrostatic (Coulombic) effect, the desolvation term was removed from the total electrostatic free energy of interaction as by Equation 5.3 and Equation 5.4. Boltzmann probability obtained only from the Coulombic energy was used to predict the dipolar couplings. The correlation between the experimental and predicted RDCs decreased to 0.52 (Figure 5.7), suggesting that the Coulombic term needs the compensation by the water exclusion term.

### **Prediction of alignment from 122 Orientations**

Accuracy of the predictions depend also on the total number of conformations sampled adjacent to the membrane surface. Hence detailed electrostatic calculations were also done by increasing the number of conformational samplings to 122, opposed to restricting to only for the mean orientation as discussed above. Here, for each of the distance along the x-axis and for each of the 122 orientation of the protein with respect to the lipid bilayer in y-z plane, the nonexcluded matrix A, which was obtained within PALES, was weighted according to its Boltzmann probability, obtained from the detailed electrostatic calculations for 122 orientations. Saupe order matrix was obtained from the weighted nonexcluded alignment matrix. For the 81-residue rec-A-binding protein DinI, experimental  $^1D_{NH}$  values correlated with the values predicted from the Saupe ordered matrix with a Pearson's correlation coefficient  $R_p = 0.92$  (Figure 5.8). Clearly, when we compare to the results above (0.79 for mean orientation) the large sampling yielded better correlation. In addition, the predicted magnitude of alignment,  $G_{mag}$ , was approximately five times less than

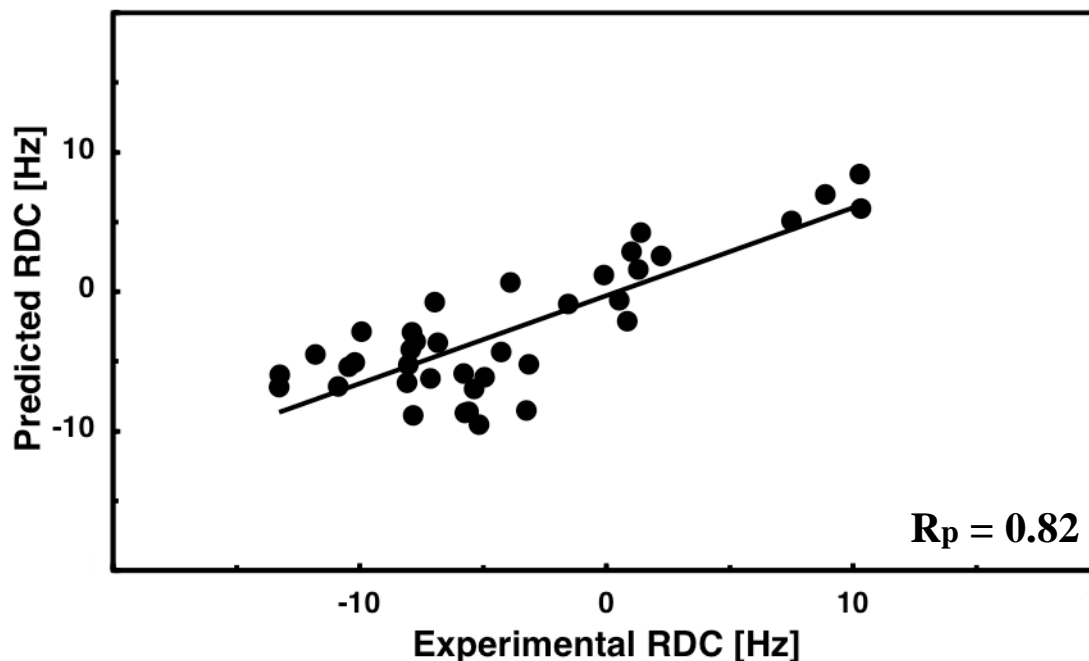


**Figure 5.8:** Comparison of experimental  $^1D_{NH}$  values measured at 300mM salt with values predicted on the basis of the molecule's three-dimensional shape and atomistic electrostatic effects for protein DinI, considering 122 different orientations.

the experimental value of  $4.1 \times 10^{-3}$ .

To take into account only the contributions from the electrostatic (Coulombic) effect, the desolvation term was removed from the total electrostatic free energy of interaction as by Equation 5.3 and Equation 5.4. Boltzmann probability obtained from the Coulombic energy was used to predict the residual dipolar couplings. The correlation between the experimental and predicted RDCs decreased to 0.82 (shown in Figure 5.9). Thus supporting the fact that the short-range repulsion term contributes to the accuracy in the prediction of residual dipolar couplings. It is expected that, by sampling 2196 orientations in the detailed electrostatic model, the predictions should improve when compared to the simplified model. In addition, the predicted magnitude of alignment,  $G_{mag}$ , was  $1.25 \times 10^{-3}$ , which was three times less than the experimental magnitude of alignment, whereas the alignment magnitude predicted by electrostatic PALES was  $7.97 \times 10^{-3}$ . Inaccuracies in the prediction of magnitude of alignment

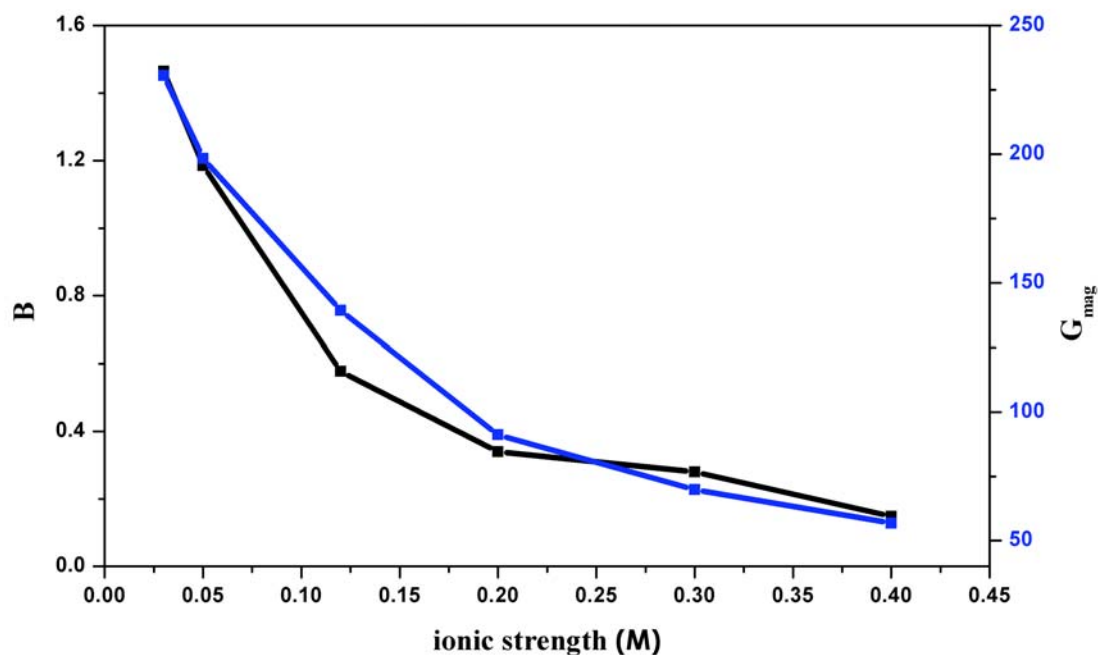
could be arising from the fact, that the translational grid was under sampled, that is only up to 30 Å.



**Figure 5.9:** Comparison of experimental  $^1D_{NH}$  values measured at 300mM salt with values predicted, considering only the Coulombic contribution from the total electrostatic free energy, on the basis of the molecule's three-dimensional shape and atomistic electrostatic effects for protein DinI, considering 122 different orientations.

### Prediction of magnitude of alignment

For the simplified, model, the predicted alignment magnitudes deviated significantly for salt concentrations below 0.01 M and above 0.3 M from experimental values [Zweckstetter, Hummer 2004]. Hence, to investigate the accuracy of the detailed model, the magnitude of alignment was predicted by repeating the detailed electrostatic calculations for the protein in the mean orientation at salt concentrations ranging between 0.03 - 0.4 M. Minimum electrostatic free energy obtained at various salt concentrations were converted to Boltzmann factor (according to equation 5.7) to obtain the binding factor, B. Figure 5.10 shows the overlay of  $G_{mag}$  as obtained from experimental dipolar couplings by SVD (blue lines) and B values (black lines) as



**Figure 5.10:** Comparison of experimental alignment magnitude ( $G_{\text{mag}}$ ), obtained by SVD, (blue lines) with B values predicted from atomistic model (black lines).

predicted from detailed electrostatic model. The overall profile obtained from the detailed model is comparable to the experimental values, as opposed to the simplified model. Thus, the preliminary results suggests that the presence of desolvation term in the detailed model already improves the accuracy of prediction at different salt concentrations and future works are planned to perform the similar calculations for different systems reported in Zweckstetter, Hummer 2004.

The study presented here introduced a possible extension to the existing simplistic electrostatic model used in PALES, by adding more detailed electrostatic effects. Previous model approximated the electrostatic interaction between a solute and an ordered phage particle as that between the solute's surface charges and the electric field of the phage [Zweckstetter, Hummer 2004]. This simplified geometry allowed one to obtain an analytical expression for the energy of interaction at each protein-phage distance in terms of Gouy-Chapman-Stern theory [Chapman 1913;Gouy 1910]. Here, in this method the alignment medium (here lipid bilayer) and

the protein (DinI) were represented in atomic detail. The reciprocal effects of the protein and the membrane on each other's electrostatic potentials, which were ignored in the previous model, are taken into account and give rise to a Born repulsion between the protein and the membrane at a short distance. The short-range repulsion, when added to the long-range coulombic attraction yields a minimum in the electrostatic free energy of attraction between the membrane and the protein.

**Table 5.1:** Table representing the correlations obtained using two different electrostatic models.

Electrostatic Model	Correlation ( $R_p$ )	
	Mean Orientation	122 Orientations
Simplistic electrostatic model (PALES)	0.57 <sup>§</sup>	0.78 <sup>§</sup>
Detailed electrostatic model		
Only Coulombic	0.52	0.82
Coulombic + short range Born-repulsion	0.79	0.92

[§] When performed for 2196 orientations, correlation coefficient was 0.96 [Zweckstetter, Hummer 2004].

The simplified electrostatic model predicted correctly not only the RDCs but also the magnitude of alignment for protein DinI [Zweckstetter, Hummer 2004], which implies other attractive nonpolar (e.g. hydrophobic) interactions between the protein and the membrane may compensate for Born-repulsion. The prediction of the detailed atomic model, which considers the Born-repulsion effects and the influence of protein on the bilayer, was comparable to the simplified model and reasonably accurate when the same number of orientations was assumed (Table 5.1). However, since the simplified model is extremely fast, more orientations could be considered so that it still represents the standard for accuracy given a certain amount of time. When we use an atomistic model that assumes purely electrostatic (Coulombic) effect, the predicted alignment tensors for DinI was further away from the observed values (0.52

and 0.82, respectively for the mean and 122 orientations), but was similar to the values obtained from simplified model (0.57 and 0.78, respectively), for the given number of orientations (Table 5.1). The quality of the predictions obtained by considering full electrostatic free energy (Coulombic and desolvation effects) was better compared to the predictions obtained from the simplistic electrostatic model, for the given number of orientations (Table 5.1). Sampling large number of orientations of proteins adjacent to the membrane surface is expected to further improve the accuracy of the predictions of alignment tensor and residual dipolar couplings. Also, the calculations here were restricted up to a translation grid of  $\sim 30$  Å, unlike in electrostatic PALES, where it was sampled until the potential reduces to the range of  $1e-06$  (which is nearly  $\sim 85$  Å). Increasing the translations grid would help in improving the accuracy in predictions of magnitude of alignment.

In addition, addressing the following concerns in the detailed model could lead to better accuracy in the predictions.

- 1) The detailed model ignores some attractive interactions. This attractive interaction could arise from nonpolar contributions to the binding of protein/membrane system. Assuming the nonpolar interactions are proportional to the water-accessible surface area [Ben-Tal, Honig 1997], the water-accessible surface area would decrease as the protein approaches the bilayer, which will give rise to short-range attraction.

- 2) The lower correlations in the detailed model could also arise from either structural changes in the bilayer that occur on binding or some other deficiency in the model, as reported previously [Ben-Tal, Honig 1997; Zhou, Schulten 1995; Woolf, Roux 1996]. The “soft” or nonrigid nature of the bilayer surface may strongly affect the calculations of the short-range Born repulsions and nonpolar attractions, but



should have a much smaller effect on the calculation of the long-range Coulomb attraction.

3) Deficiencies could also relate to the accuracy of the electrostatic potential defined by the membrane bilayer model used here compared to the electrostatic potential experienced by a protein in front of the bacteriophage Pf1.

4) It is also well known that the Coulomb energy does not depend on the set of atomic partial or real charges and radii used for the calculations [Ben-Tal, Honig 1996]. But the Born repulsion does depend strongly on parameters used for the calculations, i.e., the set of atomic partial charges and radii used in the Born calculations.

## 5.4 Conclusions

In this study, a more detailed electrostatic model was used in order to improve the quality of prediction of the alignment tensor of a protein from its known three-dimensional structure. To this aim, Non-linear Poisson-Boltzmann equation was solved for the protein/bilayer system, where protein and lipid bilayer were represented in atomistic detail. Calculated electrostatic free energies were used to compute molecular alignment tensors. Residual dipolar couplings predicted from the total electrostatic free energies correlated reasonably well with the experimental values, for the Rec A binding protein DinI dissolved in Pf1 bacteriophage. Preliminary result suggests that, the prediction of charge-induced molecular alignment could be improved by employing detailed electrostatic model. Contributions from the short-range Born repulsion to the total electrostatic free energy seemed crucial when the protein approaches the bilayer surface, indicating that apart from the steric interactions, short-range interactions arising from desolvation effect should also be taken into account to increase the accuracy of the prediction of alignment tensor and residual dipolar couplings for a protein. Influence of the protein on the bilayer is taken into these calculations, through both short-range interactions and long-range electrostatic effects, thus, overcoming one of the major limitations of the simple electrostatic model currently used in PALES. Though tests should be done for all the 2196 orientations of DinI molecule and for different systems at various ionic strengths. Also, from these studies, it will be possible to determine how much charge (positive/negative) is required to improve the molecular alignment of a protein for a given medium and also the correlation between the predicted and observed residual dipolar couplings over the simplified electrostatic model.

## 5.5 Outlook

- 1) It was reported previously that, the simplistic electrostatic model under or over estimates the magnitude of alignment tensor at low ionic strengths in few cases. Hence, to evaluate the scaling of the magnitude of the alignment as a function of ionic strength, calculations must be done for different members of the ensemble of different protein structures (such as Ubiquitin, GB1, GB $\delta$ , Dickerson dodecamer) and for many different configurations (all 2196 orientations, i.e., 122 orientations for every 18° around the x-axis) of the protein adjacent to the membrane surface. This test is also expected to address the concern related to dependence of quality of the available input structure on evaluating the molecular alignment in the simplified electrostatic model.
- 2) Instead of lipid bilayers, a more realistic model of the bacteriophage could also be used for the electrostatic calculations. A cylindrical surface with atomistic detail could be modeled with DELPHI to mimic the shape and surface charge density of the Pf1 Bacteriophage. Mimicking also the non-uniform electrostatic potential experienced by the protein along the z-axis of the Pf1.

## Bibliography

- Abbott, A., (2005) Protein structures hint at the shape of things to come. *Nature*. 7042, **435**, 547.
- Aiyar, J., Withka, J. M., Rizzi, J. P., Singleton, D. H., Andrews, G. C., Lin, W., Boyd, J., Hanson, D. C., Simon, M., Dethlefs, B., Lee, C. L., Hall, J. E., Gutman, G. A., Chandy, K. G., (1995) Topology of the Pore-Region of a K<sup>+</sup> Channel Revealed by the Nmr-Derived Structures of Scorpion Toxins. *Neuron*. 5, **15**, 1169-1181.
- Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K., Prestegard, J. H., (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *Journal of Magnetic Resonance*. 2, **143**, 402-406.
- Andrew, E. R., Bradbury, A., Eades, R. G., (1958) Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed. *Nature*. 4650, **182**, 1659-1659.
- Andronesi, O. C., Pfeifer, J. R., Al-Momani, L., Ozdirekcan, S., Rijkers, D. T., Angerstein, B., Luca, S., Koert, U., Killian, J. A., Baldus, M., (2004) Probing membrane protein orientation and structure using fast magic-angle-spinning solid-state NMR. *Journal of Biomolecular NMR*. 3, **30**, 253-265.
- Baker, N. A., Mccammon, J. A., (2003) Electrostatic interactions. *Methods Biochem Anal*. **44**, 427-440.
- Baldus, M., (2002) Correlation experiments for assignment and structure elucidation of immobilized polypeptides under magic angle spinning. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 1-2, **41**, 1-47.
- Baldus, M., (2007) ICMRBS founder's medal 2006: Biological solid-state NMR, methods and applications. *Journal of Biomolecular NMR*. 1, **39**, 73-86.
- Barrientos, L. G., Dolan, C., Gronenborn, A. M., (2000) Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *Journal of Biomolecular NMR*. 4, **16**, 329-337.
- Barrientos, L. G., Louis, J. M., Gronenborn, A. M., (2001) Characterization of the cholesteric phase of filamentous bacteriophage fd for molecular alignment. *Journal of Magnetic Resonance*. 1, **149**, 154-158.
- Bax, A., (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Science*. 1, **12**, 1-16.
- Bax, A., Grzesiek, S., (1993) Methodological Advances in Protein Nmr. *Accounts of Chemical Research*. 4, **26**, 131-138.

- Bayrhuber, M., Graf, R., Ferber, M., Zweckstetter, M., Imperial, J., Garrett, J. E., Olivera, B. M., Terlau, H., Becker, S., (2006) Production of recombinant Conkunitzin-S1 in *Escherichia coli*. *Protein Expr Purif.* 2, **47**, 640-644.
- Bayrhuber, M., Vijayan, V., Ferber, M., Graf, R., Korukottu, J., Imperial, J., Garrett, J. E., Olivera, B. M., Terlau, H., Zweckstetter, M., Becker, S., (2005) Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. *Journal of Biological Chemistry.* 25, **280**, 23766-23770.
- Ben-Tal, N., Honig, B., Miller, C., Mclaughlin, S., (1997) Electrostatic binding of proteins to membranes. Theoretical predictions and experimental results with charybdotoxin and phospholipid vesicles. *Biophys Journal.* 4, **73**, 1717-1727.
- Ben-Tal, N., Honig, B., Peitzsch, R. M., Denisov, G., Mclaughlin, S., (1996) Binding of small basic peptides to membranes containing acidic lipids: theoretical models and experimental results. *Biophys Journal.* 2, **71**, 561-575.
- Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R., (1984) Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics.* 8, **81**, 3684-3690.
- Bewley, C. A., (2001) Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-HIV protein cyanovirin-N using residual dipolar couplings. *Journal of American Chemical Society.* 5, **123**, 1014-1015.
- Bewley, C. A., Clore, G. M., (2000) Determination of the Relative Orientation of the Two Halves of the Domain-Swapped Dimer of Cyanovirin-N in Solution Using Dipolar Couplings and Rigid Body Minimization. *Journal of American Chemical Society.* 25, **122**, 6009-6016.
- Bloembergen, N., (1949) On the Interaction of Nuclear Spins in a Crystalline Lattice. *Physica.* 3-4, **15**, 386-426.
- Bradley, P., Misura, K. M., Baker, D., (2005) Toward high-resolution de novo structure prediction for small proteins. *Science.* 5742, **309**, 1868-1871.
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus., (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry.* 2, **4**, 187-217.
- Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L., (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D.* **54**, 905-921.

- Brunger, A. T., Adams, P. D., Rice, L. M., (1997) New applications of simulated annealing in X-ray crystallography and solution NMR. *Structure*. 3, **5**, 325-336.
- Brunger, A. T., Nilges, M., (1993) Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy. *Q Rev Biophys*. 1, **26**, 49-125.
- Brunner, E., (2001) Residual dipolar couplings in protein NMR. *Concepts in Magnetic Resonance*. **13**, 238-259.
- Canutescu, A. A., Shelenkov, A. A., Dunbrack, R. L., (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Science*. 9, **12**, 2001-2014.
- Castellani, F., Van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., Oschkinat, H., (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*. 6911, **420**, 98-102.
- Chapman, D. L., (1913) A contribution to the theory of electrocapillarity. *Philosoph. Magaz.* **25**, 475-481.
- Chou, J. J., Bax, A., (2001) Protein side-chain rotamers from dipolar couplings in a liquid crystalline phase. *Journal of American Chemical Society*. 16, **123**, 3844-3845.
- Chou, J. J., Delaglio, F., Bax, A., (2000) Measurement of one-bond <sup>15</sup>N-<sup>13</sup>C' dipolar couplings in medium sized proteins. *Journal of Biomolecular NMR*. 2, **18**, 101-105.
- Clore, G. M., Garrett, D. S., (1999) R-factor, Free R, and Complete Cross-Validation for Dipolar Coupling Refinement of NMR Structures. *Journal of American Chemical Society*. 39, **121**, 9008-9012.
- Clore, G. M., Starich, M. R., Gronenborn, A. M., (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *Journal of American Chemical Society*. 40, **120**, 10571-10572.
- Contreras, M. A., Ubach, J., Millet, O., Rizo, J., Pons, M., (1999) Measurement of one bond dipolar couplings through lanthanide-induced orientation of a calcium-binding protein. *Journal of American Chemical Society*. 38, **121**, 8947-8948.
- Cornilescu, G., Delaglio, F., Bax, A., (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*. 3, **13**, 289-302.
- Cornilescu, G., Marquardt, J. L., Ottiger, M., Bax, A., (1998) Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid

- Crystalline Phase. *Journal of American Chemical Society*. 27, **120**, 6836-6837.
- Davis, M. E., Mccammon, J. A., (1990) Electrostatics in biomolecular structure and dynamics. *Chem. Rev.* 3, **90**, 509-521.
- De Alba, E., Baber, J. L., Tjandra, N., (1999) The Use of Residual Dipolar Coupling in Concert with Backbone Relaxation Rates to Identify Conformational Exchange by NMR. *Journal of American Chemical Society*. 17, **121**, 4282-4283.
- De Alba, E., Tjandra, N., (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2, **40**, 175-197.
- De La Vega, R. C. R., Merino, E., Becerril, B., Possani, L. D., (2003) Novel interactions between K<sup>+</sup> channels and scorpion toxins. *Trends in Pharmacological Sciences*. 5, **24**, 222-227.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A., (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular NMR*. 3, **6**, 277-293.
- Delaglio, F., Kontaxis, G., Bax, A., (2000) Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings. *Journal of American Chemical Society*. 9, **122**, 2142-2143.
- Dingemans, T., Photinos, D. J., Samulski, E. T., Terzis, A. F., Wutz, C., (2003) Ordering of apolar and polar solutes in nematic solvents. *Journal of Chemical Physics*. 15, **118**, 7046-7061.
- Dominguez, C., Boelens, R., Bonvin, A. M., (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of American Chemical Society*. 7, **125**, 1731-1737.
- Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A. L., Gulbis, J. M., Cohen, S. L., Chait, B. T., Mackinnon, R., (1998) The structure of the potassium channel: Molecular basis of K<sup>+</sup> conduction and selectivity. *Science*. 5360, **280**, 69-77.
- Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., Scharf M., (1995) The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*. **16**, 273-284.
- Etzkorn, M., Bockmann, A., Lange, A., Baldus, M., (2004) Probing molecular interfaces using 2D magic-angle-spinning NMR on protein mixtures with different uniform labeling. *Journal of American Chemical Society*. 45, **126**, 14746-14751.

- Ferrarini, A., (2003) Modeling of Macromolecular Alignment in Nematic Virus Suspensions. Application to the Prediction of NMR Residual Dipolar Couplings. *Journal of Physical Chemistry. B.* 31, **107**, 7923-7931.
- Fischer, M. W., Losonczi, J. A., Weaver, J. L., Prestegard, J. H., (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry.* 28, **38**, 9013-9022.
- Fleming, K., Gray, D., Prasanna, S., Matthews, S., (2000) Cellulose Crystallites: A New and Robust Liquid Crystalline Medium for the Measurement of Residual Dipolar Couplings. *Journal of American Chemical Society.* 21, **122**, 5224-5225.
- Fossi, M., Castellani, F., Nilges, M., Oschkinat, H., Van Rossum, B. J., (2005) SOLARIA: a protocol for automated cross-peak assignment and structure calculation for solid-state magic-angle spinning NMR spectroscopy. *Angewandte Chemie International Edition Engl.* 38, **44**, 6151-6154.
- Franks, W. T., Zhou, D. H., Wylie, B. J., Money, B. G., Graesser, D. T., Frericks, H. L., Sahota, G., Rienstra, C. M., (2005) Magic-angle spinning solid-state NMR spectroscopy of the beta 1 immunoglobulin binding domain of protein G (GB1): N-15 and C-13 chemical shift assignments and conformational analysis. *Journal of American Chemical Society.* 35, **127**, 12291-12305.
- Gaemers, S., Bax, A., (2001) Morphology of three lyotropic liquid crystalline biological NMR media studied by translational diffusion anisotropy. *Journal of American Chemical Society.* 49, **123**, 12343-12352.
- Gairi, M., Romi, R., Fernandez, I., Rochat, H., Martin-Eauclaire, M. F., Van Rietschoten, J., Pons, M., Giralt, E., (1997) 3D structure of kaliotoxin: is residue 34 a key for channel selectivity? *Journal of Peptide Science.* 4, **3**, 314-319.
- Gallagher, K., Sharp, K., (1998) Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophysical Journal.* 2, **75**, 769-776.
- Gilson, M. K., Sharp, K. A., Honig, B., (1988) Calculating the electrostatic potential of molecules in solution: Method and error assessment. *Journal of Computational Chemistry.* 4, **9**, 327-335.
- Glaubitx, C., Watts, A., (1998) Magic angle-oriented sample spinning (MAOSS): A new approach toward biomembrane studies. *Journal of Magnetic Resonance.* 2, **130**, 305-316.
- Gouaux, E., Mackinnon, R., (2005) Principles of selective ion transport in channels and pumps. *Science.* 5753, **310**, 1461-1465.
- Gouy, D. L., (1910) Sur la constitution de la charge électrique à la surface d'un électrolyte. *Ann. Phys.* **9**, 457-468.



- Griffin, R. G., (1998) Dipolar recoupling in MAS spectra of biological solids. *Nature Structural Biology*. **5 Suppl**, 508-512.
- Grishaev, A., Bax, A., (2004) An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *Journal of American Chemical Society*. **126**, 7281-7292.
- Grishaev, A., Llinas, M., (2002a) CLOUDS, a protocol for deriving a molecular proton density via NMR. *Proceedings of the National Academy of Sciences U S A*. **99**, 6707-6712.
- Grishaev, A., Llinas, M., (2002b) Protein structure elucidation from NMR proton densities. *Proceedings of the National Academy of Sciences U S A*. **99**, 6713-6718.
- Grissmer, S., Nguyen, A. N., Aiyar, J., Hanson, D. C., Mather, R. J., Gutman, G. A., Karmilowicz, M. J., Auperin, D. D., Chandy, K. G., (1994) Pharmacological characterization of five cloned voltage-gated K<sup>+</sup> channels, types Kv1.1, 1.2, 1.3, 1.5, and 3.1, stably expressed in mammalian cell lines. *Mol Pharmacol*. **6**, 1227-1234.
- Gronwald, W., Moussa, S., Elsner, R., Jung, A., Ganslmeier, B., Trenner, J., Kremer, W., Neidig, K. P., Kalbitzer, H. R., (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). *Journal of Biomolecular NMR*. **4**, 23, 271-287.
- Grzesiek, S., Anglister, J., Bax, A., (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins by Isotropic Mixing of C-13 Magnetization. *Journal of Magnetic Resonance Series B*. **1**, **101**, 114-119.
- Grzesiek, S., Bax, A., (1992) Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr. *Journal of American Chemical Society*. **114**, 6291-6293.
- Guntert, P., (2003) Automated NMR protein structure calculation. *Progress in Nuclear Magnetic Resonance Spectroscopy*. **3-4**, **43**, 105-125.
- Guntert, P., Mumenthaler, C., Wuthrich, K., (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology*. **1**, **273**, 283-298.
- Hansen, M. R., Mueller, L., Pardi, A., (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*. **12**, **5**, 1065-1074.
- Heise, H., Seidel, K., Etzkorn, M., Becker, S., Baldus, M., (2005) 3D NMR spectroscopy for resonance assignment and structure elucidation of proteins under MAS: novel pulse schemes and sensitivity considerations. *Journal of Magnetic Resonance*. **1**, **173**, 64-74.

- Herrmann, T., Guntert, P., Wuthrich, K., (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *Journal of Molecular Biology*. 1, **319**, 209-227.
- Holst, M., Saied, F., (1993) Multigrid solution of the Poisson-Boltzmann equation. *Journal of Computational Chemistry*. 1, **14**, 105-113.
- Honig, B., Nicholls, A., (1995) Classical electrostatics in biology and chemistry. *Science*. 5214, **268**, 1144-1149.
- Honig, B., Sharp, K., Yang, A. S., (1993) Macroscopic models of aqueous solutions: biological and chemical applications. *Journal of Physical Chemistry*. 6, **97**, 1101-1109.
- Hu, J. S., Bax, A., (1997)  $\chi_1$  angle information from a simple two-dimensional NMR experiment that identifies trans  $(^3J_{NC})_{\gamma}$  couplings in isotopically enriched proteins. *Journal of Biomolecular NMR*. 3, **9**, 323-328.
- Hu, J. S., Grzesiek, S., Bax, A., (1997) Two-Dimensional NMR Methods for Determining  $\chi_1$  Angles of Aromatic Residues in Proteins from Three-Bond  $J_{CC\gamma}$ ; and  $J_{NC\gamma}$ ; Couplings. *Journal of American Chemical Society*. 7, **119**, 1803-1804.
- Hubbard, S. J., Thornton, J. M. (Department of Biochemistry and Molecular Biology, University College London, 1993), vol. 54.
- Hus, J. C., Marion, D., Blackledge, M., (2001) Determination of protein backbone structure using only residual dipolar couplings. *Journal of American Chemical Society*. 7, **123**, 1541-1542.
- Ikura, M., Kay, L. E., Bax, A., (1990) A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins - Heteronuclear Triple-Resonance 3-Dimensional Nmr-Spectroscopy - Application to Calmodulin. *Biochemistry*. 19, **29**, 4659-4667.
- Ikura, M., Kay, L. E., Tschudin, R., Bax, A., (1990) 3-Dimensional Noesy-Hmqc Spectroscopy of a C-13-Labeled Protein. *Journal of Magnetic Resonance*. 1, **86**, 204-209.
- Jaroniec, C. P., Macphee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M., Griffin, R. G., (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences U S A*. 3, **101**, 711-716.
- Jee, J., Guntert, P., (2003) Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J Struct Funct Genomics*. 2-3, **4**, 179-189.

- Jeener, J., Meier, B. H., Bachmann, P., Ernst, R. R., (1979) Investigation of Exchange Processes by 2-Dimensional Nmr-Spectroscopy. *Journal of Chemical Physics*. 11, **71**, 4546-4553.
- Jiang, Y. X., Lee, A., Chen, J. Y., Cadene, M., Chait, B. T., Mackinnon, R., (2002) The open pore conformation of potassium channels. *Nature*. 6888, **417**, 523-526.
- Jung, Y. S., Sharma, M., Zweckstetter, M., (2004) Simultaneous assignment and structure determination of protein backbones by using NMR dipolar couplings. *Angewandte Chemie International Edition Engl.* 26, **43**, 3479-3481.
- Jung, Y. S., Zweckstetter, M., (2004a) Backbone assignment of proteins with known structure using residual dipolar couplings. *Journal of Biomolecular NMR*. 1, **30**, 25-35.
- Jung, Y. S., Zweckstetter, M., (2004b) Mars -- robust automatic backbone assignment of proteins. *Journal of Biomolecular NMR*. 1, **30**, 11-23.
- Kainosho, M., (1997) Isotope labelling of macromolecules for structural determinations. *Nature Structural Biology*. **4**, 858-861.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. M., Guntert, P., (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature*. 7080, **440**, 52-57.
- Karplus, M., (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. *Journal of American Chemical Society*. 18, **85**, 2870.
- Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., Formankay, J. D., (1993) A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain <sup>1</sup>H and <sup>13</sup>C Correlations in H<sub>2</sub>O Samples of Proteins. *Journal of Magnetic Resonance, Series B*. 3, **101**, 333-337.
- Kern, D., Zuiderweg, E. R. P., (2003) The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology*. 6, **13**, 748-757.
- Kirkpatrick, S., Gelatt, C. D., Jr., Vecchi, M. P., (1983) Optimization by Simulated Annealing. *Science*. 4598, **220**, 671-680.
- Koenig, B. W., Hu, J. S., Ottiger, M., Bose, S., Hendler, R. W., Bax, A., (1999) NMR measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments. *Journal of American Chemical Society*. **121**, 1385-1386.
- Korukottu, J., Bayrhuber, M., Montaville, P., Vijayan, V., Jung, Y. S., Becker, S., Zweckstetter, M., (2007) Fast high-resolution protein structure determination by using unassigned NMR data. *Angewandte Chemie International Edition Engl.* 7, **46**, 1176-1179.

- Koshland, D. E., (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences USA*. 2, **44**, 98-104.
- Kuo, A. L., Gulbis, J. M., Antcliff, J. F., Rahman, T., Lowe, E. D., Zimmer, J., Cuthbertson, J., Ashcroft, F. M., Ezaki, T., Doyle, D. A., (2003) Crystal structure of the potassium channel KirBac1.1 in the closed state. *Science*. 5627, **300**, 1922-1926.
- Kuszewski, J., Schwieters, C. D., Garrett, D. S., Byrd, R. A., Tjandra, N., Clore, G. M., (2004) Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments. *Journal of American Chemical Society*. 20, **126**, 6258-6273.
- Lakomek, N. A., Carlomagno, T., Becker, S., Griesinger, C., Meiler, J., (2006) A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *Journal of Biomolecular NMR*. **34**, 101-115.
- Lange, A., Becker, S., Seidel, K., Giller, K., Pongs, O., Baldus, M., (2005) A concept for rapid protein-structure determination by solid-state NMR spectroscopy. *Angewandte Chemie-International Edition*. 14, **44**, 2089-2092.
- Lange, A., Giller, K., Hornig, S., Martin-Eauclaire, M. F., Pongs, O., Becker, S., Baldus, M., (2006) Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. *Nature*. 7086, **440**, 959-962.
- Lange, A., Luca, S., Baldus, M., (2002) Structural constraints from proton-mediated rare-spin correlation spectroscopy in rotating solids. *Journal of American Chemical Society*. 33, **124**, 9704-9705.
- Lange, A., Seidel, K., Verdier, L., Luca, S., Baldus, M., (2003) Analysis of proton-proton transfer dynamics in rotating solids and their use for 3D structure determination. *Journal of American Chemical Society*. 41, **125**, 12640-12648.
- Lee, M., Goldberg, W. I., (1965) Nuclear-magnetic-resonance line narrowing by a rotating rf field. *Physical Review*. **140**, 1261-1271.
- Linge, J. P., Habeck, M., Rieping, W., Nilges, M., (2003) ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*. 2, **19**, 315-316.
- Linge, J. P., O'donoghue, S. I., Nilges, M., (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol*. **339**, 71-90.
- Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., Nilges, M., (2003) Refinement of protein structures in explicit solvent. *Proteins*. 3, **50**, 496-506.

- Liu, G., Shen, Y., Atreya, H. S., Parish, D., Shao, Y., Sukumaran, D. K., Xiao, R., Yee, A., Lemak, A., Bhattacharya, A., Acton, T. A., Arrowsmith, C. H., Montelione, G. T., Szyperski, T., (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. *Proceedings of the National Academy of Sciences U S A.* 30, **102**, 10487-10492.
- Long, S. B., Campbell, E. B., Mackinnon, R., (2005) Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science.* 5736, **309**, 897-903.
- Luca, S., Heise, H., Baldus, M., (2003) High-resolution solid-state NMR applied to polypeptides and membrane proteins. *Accounts of Chemical Research.* 11, **36**, 858-865.
- Macura, S., Ernst, R. R., (1980) Elucidation of Cross Relaxation in Liquids by Two-Dimensional Nmr-Spectroscopy. *Molecular Physics.* 1, **41**, 95-117.
- Maricq, M. M., Waugh, J. S., (1979) NMR in rotating solids. *Journal of Chemical Physics.* 7, **70**, 3300-3316.
- Markus, M. A., Gerstner, R. B., Draper, D. E., Torchia, D. A., (1999) Refining the overall structure and subdomain orientation of ribosomal protein S4 delta41 with dipolar couplings measured by NMR in uniaxial liquid crystalline phases. *Journal of Molecular Biology.* 2, **292**, 375-387.
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., Sali, A., (2000) Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure.* **29**, 291-325.
- Mehring, M., Waugh, J. S., (1972) Magic-angle NMR experiments in solids. *Physical Review B.* 5, **9**, 3459-3471.
- Meiler, J., (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. *Journal of Biomolecular NMR.* 1, **26**, 25-37.
- Meiler, J., Baker, D., (2003) Rapid protein fold determination using unassigned NMR data. *Proceedings of the National Academy of Sciences U S A.* 26, **100**, 15404-15409.
- Meiler, J., Peti, W., Griesinger, C., (2000) DipoCoup: A versatile program for 3D-structure homology comparison based on residual dipolar couplings and pseudocontact shifts. *Journal of Biomolecular NMR,* **17**, 283-294.
- Meiler, J., Prompers, J. J., Peti, W., Griesinger, C., Bruschweiler, R., (2001) Model-Free Approach to the Dynamic Interpretation of Residual Dipolar Couplings in Globular Proteins. *Journal of American Chemical Society* 25, **123**, 6098-6107.
- Miller, C., (1995) The Charybdotoxin Family of K<sup>+</sup> Channel-Blocking Peptides. *Neuron.* 1, **15**, 5-10.

- Misra, V. K., Honig, B., (1995) On the magnitude of the electrostatic contribution to ligand-DNA interactions. *Proceedings of the National Academy of Sciences U S A.* 10, **92**, 4691-4695.
- Misra, V. K., Sharp, K. A., Friedman, R. A., Honig, B., (1994) Salt effects on ligand-DNA binding. Minor groove binding antibiotics. *Journal of Molecular Biology.* 2, **238**, 245-263.
- Montelione, G. T., Lyons, B. A., Emerson, S. D., Tashiro, M., (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *Journal of American Chemical Society.* 27, **114**, 10974-10975.
- Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C., Szyperski, T., (2000) Protein NMR spectroscopy in structural genomics. *Nature Structural Biology.* **7 Suppl**, 982-985.
- Moseley, H. N., Monleon, D., Montelione, G. T., (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Methods Enzymol.* **339**, 91-108.
- Moseley, H. N., Montelione, G. T., (1999) Automated analysis of NMR assignments and structures for proteins. *Current Opinions in Structural Biology.* 5, **9**, 635-642.
- Mumenthaler, C., Braun, W., (1995) Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. *Journal of Molecular Biology.* 3, **254**, 465-480.
- Mumenthaler, C., Guntert, P., Braun, W., Wuthrich, K., (1997) Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *Journal of Biomolecular NMR.* 4, **10**, 351-362.
- Murray, D., Arbusova, A., Honig, B., Mclaughlin, S., (2002) The role of electrostatic and nonpolar interactions in the association of peripheral proteins with membranes. *Peptide-Lipid Interactions.* **52**, 277-307.
- Murray, D., Hermida-Matsumoto, L., Buser, C. A., Tsang, J., Sigal, C. T., Ben-Tal, N., Honig, B., Resh, M. D., Mclaughlin, S., (1998) Electrostatics and the membrane association of Src: theory and experiment. *Biochemistry.* 8, **37**, 2145-2159.
- Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., Bonvin, A. M., (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins.* 4, **59**, 662-672.

- Nilges, M., (1993) A calculation strategy for the structure determination of symmetric dimers by  $^1\text{H}$  NMR. *Proteins*. 3, **17**, 297-309.
- Nilges, M., (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *Journal of Molecular Biology*. 5, **245**, 645-660.
- Nomura, K., Takegoshi, K., Terao, T., Uchida, K., Kainosho, M., (1999) Determination of the complete structure of a uniformly labeled molecule by rotational resonance solid-state NMR in the tilted rotating frame. *Journal of American Chemical Society*. 16, **121**, 4064-4065.
- Osapay, K., Case, D. A., (1994) Analysis of proton chemical shifts in regular secondary structure of proteins. *Journal of Biomolecular NMR*. 2, **4**, 215-230.
- Ottiger, M., Delaglio, F., Bax, A., (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance*. 2, **131**, 373-378.
- Peitzsch, R. M., Eisenberg, M., Sharp, K. A., Mclaughlin, S., (1995) Calculations of the electrostatic potential adjacent to model phospholipid bilayers. *Biophysical Journal*. 3, **68**, 729-738.
- Permi, P., Annala, A., (2001) A new approach for obtaining sequential assignment of large proteins. *Journal of Biomolecular NMR*. **20**, 127-133.
- Peti, W., Meiler, J., Bruschweiler, R., Griesinger, C., (2002) Model-free analysis of protein backbone motion from residual dipolar couplings. *Journal of American Chemical Society*. 20, **124**, 5822-5833.
- Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., Tycko, R., (2002) A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences U S A*. 26, **99**, 16742-16747.
- Photinos, D. J., Poon, C. D., Samulski, E. T., Toriumi, H., (1992) Nmr-Study of the Effects of Electric-Dipole Interactions on the Ordering of Chain Solutes in the Nematic Phase. *Journal of Physical Chemistry*. 20, **96**, 8176-8180.
- Photinos, D. J., Samulski, E. T., (1993) Electric-Dipole Interactions of Chain Molecules in Nematics - the Analysis of Segmental Ordering in Alpha, Omega-Dibromoalkanes. *Journal of Chemical Physics*. 12, **98**, 10009-10016.
- Pickford, A. R., Campbell, I. D., (2004) NMR studies of modular protein structures and their interactions. *Chemical Reviews*. 8, **104**, 3557-3565.
- Prestegard, J. H., (1998) New techniques in structural NMR - anisotropic interactions. *Nature Structural Biology*. **5**, 517-522.

- Prestegard, J. H., Al-Hashimi, H. M., Tolman, J. R., (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys.* 4, **33**, 371-424.
- Prestegard, J. H., Kishore, A. I., (2001) Partial alignment of biomolecules: an aid to NMR characterization. *Curr Opin Chem Biol.* 5, **5**, 584-590.
- Ramirez, B. E., Bax, A., (1998) Modulation of the Alignment Tensor of Macromolecules Dissolved in a Dilute Liquid Crystalline Medium. *Journal of American Chemical Society.* 35, **120**, 9106-9107.
- Ramirez, B. E., Voloshin, O. N., Camerini-Otero, R. D., Bax, A., (2000) Solution structure of DinI provides insight into its mode of RecA inactivation. *Protein Sci.* 11, **9**, 2161-2169.
- Ranganathan, R., Lewis, J. H., Mackinnon, R., (1996) Spatial localization of the K<sup>+</sup> channel selectivity filter by mutant cycle-based structure analysis. *Neuron.* 1, **16**, 131-139.
- Rienstra, C. M., Tucker-Kellogg, L., Jaroniec, C. P., Hohwy, M., Reif, B., McMahon, M. T., Tidor, B., Lozano-Perez, T., Griffin, R. G., (2002) De novo determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences U S A.* 16, **99**, 10260-10265.
- Rohl, C. A., Baker, D., (2002) De novo determination of protein backbone structure from residual dipolar couplings using Rosetta. *Journal of American Chemical Society.* 11, **124**, 2723-2729.
- Ruckert, M., Otting, G., (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *Journal of American Chemical Society* 32, **122**, 7793-7797.
- Sanders, C. R., Schaff, J. E., Prestegard, J. H., (1993) Orientational behavior of phosphatidylcholine bilayers in the presence of aromatic amphiphiles and a magnetic field. *Biophysical Journal.* **64**,1069–1080.
- Saupe, A., Englert, G., (1963) High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules. *Physical Review Letters.* 10, **11**, 462.
- Schwieters, C. D., Kuszewski, J. J., Tjandra, N., Clore, G. M., (2003) The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance.* 1, **160**, 65-73.
- Seidel, K., Lange, A., Becker, S., Hughes, C. E., Heise, H., Baldus, M., (2004) Protein solid-state NMR resonance assignments from (C-13, C-13) correlation spectroscopy. *Physical Chemistry Chemical Physics.* 22, **6**, 5090-5093.



- Sharp, K. A., Honig, B., (1990) Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *Journal of Physical Chemistry*. 19, **94**, 7684-7692.
- Sharp, K. A., Honig, B., (1990) Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem*. **19**, 301-332.
- Shen, Y., Atreya, H. S., Liu, G. H., Szyperski, T., (2005) G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. *Journal of American Chemical Society*. 25, **127**, 9085-9099.
- Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W. Y., Tolman, J. R., Mueller, G. A., Kay, L. E., (2000) Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *Journal of Molecular Biology*. 5, **295**, 1265-1273.
- Solomon, I., (1955) Relaxation Processes in a System of 2 Spins. *Physical Review*. 2, **99**, 559-565.
- Stein, E. G., Rice, L. M., Brunger, A. T., (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *Journal of Magnetic Resonance*. 1, **124**, 154-164.
- Suter, D., Ernst, R. R., (1985) Spin Diffusion in Resolved Solid-State Nmr-Spectra. *Physical Review B*. 9, **32**, 5608-5627.
- Syvitski, R. T., Burnell, E. E., (1997) Dipole-induced ordering in nematic liquid crystals fact or fiction? *Chemical Physics Letters*. 1-3, **281**, 199-206.
- Syvitski, R. T., Burnell, E. E., (2000) Dipole-induced ordering in nematic liquid crystals. II. The elusive holy grail. *Journal of Chemical Physics*. 8, **113**, 3452-3465.
- Takegoshi, K., Nakamura, S., Terao, T., (2001) C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters*. 5-6, **344**, 631-637.
- Terzis, A. F., Poon, C. D., Samulski, E. T., Luz, Z., Poupko, R., Zimmermann, H., Muller, K., Toriumi, H., Photinos, D. J., (1996) Shape-Dominated Ordering in Nematic Solvents. A Deuterium NMR Study of Cycloalkane Solutes. *Journal of American Chemical Society*. 9, **118**, 2226-2234.
- Thompson, J. D., Higgins, D. G., Gibson, T. J., (1994) Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research*. 22, **22**, 4673-4680.

- Tjandra, N., Bax, A., (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*. 5340, **278**, 1111-1114.
- Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., Bax, A., (1997) Use of dipolar  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}$ - $^{13}\text{C}$  couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Structural Biology*. 9, **4**, 732-738.
- Tolman, J. R., (2002) A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *Journal of American Chemical Society*. 40, **124**, 12020-12030.
- Tolman, J. R., Flanagan, J. M., Kennedy, M. A., Prestegard, J. H., (1995) Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *Proceedings of the National Academy of Sciences U S A*. 20, **92**, 9279-9283.
- Tycko, R., (1996) Prospects for resonance assignments in multidimensional solid-state NMR spectra of uniformly labeled proteins. *Journal of Biomolecular NMR*. 3, **8**, 239-251.
- Tycko, R., (2001) Biomolecular solid state NMR: Advances in structural methodology and applications to peptide and protein fibrils. *Annual Review of Physical Chemistry*. **52**, 575-606.
- Valafar, H., Prestegard, J. H., (2003) Rapid classification of a protein fold family using a statistical analysis of dipolar couplings. *Bioinformatics*. 12, **19**, 1549-1555.
- Van Rossum, B. J., Castellani, F., Pauli, J., Rehbein, K., Hollander, J., De Groot, H. J. M., Oschkinat, H., (2003) Assignment of amide proton signals by combined evaluation of HN, NN and HNCA MAS-NMR correlation spectra. *Journal of Biomolecular NMR*. 3, **25**, 217-223.
- Vijayan, V., Zweckstetter, M., (2005) Simultaneous measurement of protein one-bond residual dipolar couplings without increased resonance overlap. *Journal of Magnetic Resonance*. 2, **174**, 245-253.
- Warren, J. J., Moore, P. B., (2001) Application of dipolar coupling data to the refinement of the solution structure of the sarcin-ricin loop RNA. *Journal of Biomolecular NMR*. 4, **20**, 311-323.
- Wishart, D. S., Sykes, B. D., (1994) Chemical shifts as a tool for structure determination in Nuclear Magnetic Resonance, Pt C, Volume **239**, *Methods in Enzymology*. 363-392.
- Wittekind, M., Mueller, L., (1993) Hncacb, a High-Sensitivity 3d NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon

- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance Series B*. 2, **101**, 201-205.
- Woolf, T. B., Roux, B., (1996) Structure, energetics, and dynamics of lipid-protein interactions: a molecular dynamics study of the gramicidin A channel in a DMPC bilayer. *Proteins*. **24**, 92-114.
- Wuthrich, K., (1986) NMR of proteins and nucleic acids. *Wiley, New York*.
- Wuthrich, K., (2003) NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew Chem Int Ed Engl*. 29, **42**, 3340-3363.
- Wuthrich, K., Wider, G., Wagner, G., Braun, W., (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high-resolution proton nuclear magnetic resonance. *Journal of Molecular Biology*. 3, **155**, 311-319.
- Yamazaki, T., Forman-Kay, J. D., Kay, L. E., (1993) Two-dimensional NMR experiments for correlating carbon-13.β and proton.δ/ε chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *Journal of American Chemical Society*. 23, **115**, 11054-11055.
- Zech, S. G., Wand, A. J., McDermott, A. E., (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *Journal of American Chemical Society*. 24, **127**, 8618-8626.
- Zhou, F., Schulten, K., (1995) Molecular dynamics study of a membrane-water interface. *Journal of Physical Chemistry*. **99**, 2195-2207.
- Zhou, Y. F., Morais-Cabral, J. H., Kaufman, A., Mackinnon, R., (2001) Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel-Fab complex at 2.0 angstrom resolution. *Nature*. 6859, **414**, 43-48.
- Zweckstetter, M., Bax, A., (2000) Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *Journal of American Chemical Society*. 15, **122**, 3791-3792.
- Zweckstetter, M., Bax, A., (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *Journal of Biomolecular NMR*. 4, **20**, 365-377.
- Zweckstetter, M., Hummer, G., Bax, A., (2004) Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys J*. 6, **86**, 3444-3460.

# Appendix A

## Chemical shift assignment of KTX in solution-state

1	43.502	CA	1	45	1.689	HB3	7
2	3.988	QA	1	46	1.53	QD	7
3	61.434	CA	2	47	1.448	QG	7
4	34.274	CB	2	48	8.446	H	7
5	4.303	HA	2	49	125.793	N	7
6	1.91	HB	2	50	53.855	CA	8
7	0.882	QQG	2	51	47.115	CB	8
8	8.846	H	2	52	4.87	HA	8
9	117.577	N	2	53	3.778	HB3	8
10	55.85	CA	3	54	3.162	HB2	8
11	34.069	CB	3	55	2.854	HG	8
12	4.296	HA	3	56	7.935	H	8
13	2.077	HB2	3	57	114.088	N	8
14	1.865	HB3	3	58	57.997	CA	9
15	1.537	QG	3	59	64.589	CB	9
16	8.693	H	3	61	4.619	HA	9
17	125.316	N	3	62	3.997	HB2	9
18	59.821	CA	4	63	3.784	HB3	9
19	40.435	CB	4	64	9.342	H	9
20	4.584	HA	4	65	112.099	N	9
21	1.865	HB	4	66	44.364	CA	10
22	0.942	QG1	4	68	4.444	QA	10
23	0.748	QG2	4	69	7.696	H	10
24	8.24	H	4	70	108.998	N	10
25	118.472	N	4	71	61.458	CA	11
26	52.797	CA	5	72	61.323	CB	11
27	36.966	CB	5	73	4.191	HA	11
28	4.805	HA	5	74	4.118	HB3	11
29	2.912	HB2	5	75	4.12	HB2	11
30	2.596	HB3	5	76	9.385	H	11
31	8.699	H	5	77	118.787	N	11
32	119.879	N	5	78	57.427	CA	13
33	61.748	CA	6	79	29.223	CB	13
34	36.251	CB	6	80	34.321	CG	13
35	4.134	HA	6	81	4.194	HA	13
36	1.386	HB	6	82	2.227	QB	13
37	0.931	QG1	6	83	2.551	HG2	13
38	0.851	QG2	6	84	2.386	HG3	13
39	8.045	H	6	85	7.312	H	13
40	121.024	N	6	86	111.535	N	13
41	56.061	CA	7	87	4.582	HA	14
42	33.283	CB	7	88	3.096	QB	14
43	4.826	HA	7	89	2.568	HG	14
44	1.769	HB2	7	90	8.256	H	14

91	111.534	N	14	143	9.068	H	20
92	59.041	CA	15	144	120.47	N	20
93	41.45	CB	15	145	51.697	CA	21
94	4.095	HA	15	146	18.65	CB	21
95	1.813	HB2	15	147	4.49	HA	21
96	1.691	HB3	15	148	1.489	HB	21
97	1.005	HD1	15	149	7.355	H	21
98	0.882	HD2	15	150	119.295	N	21
99	7.19	H	15	151	45.701	CA	22
100	122.213	N	15	153	3.845	QA	22
101	60.766	CA	16	154	8.028	H	22
102	29.49	CB	16	155	105.76	N	22
103	4.22	HA	16	156	54.312	CA	23
104	1.727	HB2	16	157	33.31	CB	23
105	1.651	HB3	16	158	4.607	HA	23
106	1.94	HE2	16	159	2.052	HB2	23
107	1.824	HE3	16	160	1.755	HB3	23
108	1.457	HG2	16	161	2.413	HG2	23
109	1.37	HG3	16	162	2.317	HG3	23
110	9.028	H	16	163	8.099	H	23
111	118.647	N	16	164	117.91	N	23
113	66.436	CA	17	165	57.604	CA	24
114	31.432	CB	17	166	38.402	CB	24
115	28.368	CG	17	167	4.425	HA	24
116	4.284	HA	17	168	1.354	HB2	24
117	2.334	HB2	17	169	1.269	HB3	24
118	2.094	HB3	17	170	2.847	HD2	24
119	3.647	HD2	17	171	2.8	HD3	24
120	3.537	HD3	17	172	1.15	HG2	24
121	1.658	QG	17	173	1.048	HG3	24
122	36.325	CB	18	174	8.374	H	24
123	4.424	HA	18	175	115.225	N	24
124	3.034	QB	18	176	56.882	CA	25
125	2.382	HG	18	177	38.009	CB	25
126	7.719	H	18	178	4.459	HA	25
127	112.583	N	18	179	3.255	HB2	25
128	59.608	CA	19	180	3.163	HB3	25
129	32.139	CB	19	181	8.177	H	25
130	4.283	HA	19	182	113.537	N	25
131	1.966	HB2	19	183	46.159	CA	26
132	1.915	HB3	19	184	4.726	HA2	26
133	1.668	HD2	19	185	3.782	HA3	26
134	1.55	HD3	19	186	7.935	H	26
135	1.428	QG	19	187	106.22	N	26
136	8.057	H	19	188	35.151	CB	27
137	122.728	N	19	189	4.634	HA	27
138	56.049	CA	20	190	1.688	HB2	27
139	40.472	CB	20	191	1.615	HB3	27
140	4.414	HA	20	192	1.352	HG2	27
141	2.822	HB3	20	193	1.265	HG3	27
142	2.829	HB2	20	194	8.873	H	27

195	121.393	N	27	233	5.364	HA	32
196	55.519	CA	28	234	1.79	QB	32
197	38.209	CB	28	235	1.586	QD	32
198	4.884	HA	28	236	1.939	QE	32
199	2.823	HB2	28	237	1.445	QG	32
200	2.495	HB3	28	238	7.761	H	32
201	8.865	H	28	239	117.452	N	32
202	125.044	N	28	240	55.451	CA	33
203	54.868	CA	29	241	38.209	CB	33
204	36.879	CB	29	242	5.064	HA	33
205	31.115	CG	29	243	3.947	HB2	33
207	4.657	HA	29	244	2.665	HB3	33
208	2.141	HB3	29	245	2.503	HG	33
209	1.691	HB2	29	246	8.434	H	33
210	1.33	QE	29	247	122.075	N	33
211	2.247	HG2	29	248	52.271	CA	34
212	2.127	HG3	29	249	30.939	CB	34
213	8.716	H	29	250	5.332	HA	34
214	127.77	N	29	251	3.154	HB2	34
215	54.606	CA	30	252	3.023	HB3	34
216	37.296	CB	30	253	9.502	H	34
217	4.331	HA	30	254	119.271	N	34
218	3.023	HB2	30	255	54.196	CA	35
219	2.748	HB3	30	256	37.546	CB	35
220	9.485	H	30	257	5.329	HA	35
221	124.232	N	30	258	3.097	HB3	35
222	58.079	CA	31	259	3.095	HB2	35
223	4.035	HA	31	260	2.676	HG	35
224	2.264	HB2	31	261	9.245	H	35
225	2.172	HB3	31	262	119.717	N	35
226	3.255	HD2	31	263	60.073	CA	36
227	3.2	HD3	31	264	71.108	CB	36
228	1.595	QG	31	265	4.941	HA	36
229	8.6	H	31	266	4.004	HB	36
230	105.574	N	31	267	1.402	HG2	36
231	54.653	CA	32	268	8.766	H	36
232	36.702	CB	32	269	117.142	N	36

# Appendix B

## Distance restraints of KTX in solid-state (free form)

12	PRO	HA	15	LEU	HG	32	LYS	HA	32	LYS	QD
9	SER	QB	10	GLY	QA	4	ILE	QG1	35	CYS	HA
4	ILE	HB	6	VAL	HA	3	GLU	QB	4	ILE	QG1
6	VAL	HA	7	LYS	HA	4	ILE	QG1	33	CYS	QB
36	THR	HA	37	PRO	QD	12	PRO	QB	16	LYS	QG
1	GLY	QA	2	VAL	HA	15	LEU	QD1	25	PHE	HA
15	LEU	HA	15	LEU	QB	29	MET	QB	32	LYS	QG
18	CYS	HA	21	ALA	QB	4	ILE	QD1	6	VAL	QG1
13	GLN	HA	16	LYS	QB	6	VAL	QG2	7	LYS	QD
6	VAL	HA	7	LYS	HA	18	CYS	HA	21	ALA	QB
7	LYS	QG	32	LYS	HA	18	CYS	HA	21	ALA	QB
11	SER	HA	28	CYS	HA	21	ALA	QB	35	CYS	QB
23	MET	HA	37	PRO	QB	4	ILE	QG2	6	VAL	HB
8	CYS	HA	17	PRO	QD	4	ILE	QD1	17	PRO	QG
5	ASN	HA	5	ASN	QB	4	ILE	QD1	17	PRO	QB
33	CYS	HA	34	HIS	HA	4	ILE	QD1	6	VAL	QG1
36	THR	HA	37	PRO	QD	4	ILE	QD1	21	ALA	QB
35	CYS	HA	37	PRO	QD	2	VAL	HB	34	HIS	HA
16	LYS	HA	17	PRO	QD	12	PRO	HA	15	LEU	HG
8	CYS	HA	17	PRO	QD	12	PRO	QB	13	GLN	QG
12	PRO	QB	17	PRO	QD	15	LEU	HA	25	PHE	HA
8	CYS	QB	10	GLY	QA	3	GLU	HA	34	HIS	HA
1	GLY	QA	3	GLU	HA	16	LYS	HA	17	PRO	QG
1	GLY	QA	2	VAL	HA	6	VAL	HA	6	VAL	QG1
15	LEU	QB	16	LYS	QG	3	GLU	HA	33	CYS	QB
15	LEU	HA	15	LEU	QB	3	GLU	QG	32	LYS	QG
4	ILE	HB	33	CYS	QB	2	VAL	QG1	3	GLU	QG
33	CYS	QB	34	HIS	HA	32	LYS	QG	32	LYS	QD
4	ILE	QG1	33	CYS	QB	7	LYS	QG	32	LYS	QB
10	GLY	QA	28	CYS	QB	3	GLU	HA	4	ILE	HA
33	CYS	HA	33	CYS	QB	2	VAL	HA	35	CYS	HA
6	VAL	HB	7	LYS	QB	8	CYS	HA	9	SER	QB
6	VAL	HB	17	PRO	QB	8	CYS	HA	9	SER	HA
4	ILE	HA	6	VAL	HB	8	CYS	QB	33	CYS	HA
29	MET	QB	29	MET	QG	2	VAL	HA	2	VAL	HB
2	VAL	HB	2	VAL	QG1	3	GLU	QG	4	ILE	HA
3	GLU	QB	32	LYS	QB	12	PRO	QB	13	GLN	HA
2	VAL	HB	37	PRO	QB	32	LYS	QG	33	CYS	QB
29	MET	QB	29	MET	QG	32	LYS	QB	32	LYS	QD
16	LYS	QG	17	PRO	QB	8	CYS	QB	13	GLN	QB
6	VAL	HB	17	PRO	QB	7	LYS	HA	32	LYS	QG
23	MET	QB	37	PRO	QD	10	GLY	QA	11	SER	HA
7	LYS	QD	32	LYS	HA	23	MET	QB	36	THR	HA
27	LYS	QG	27	LYS	QD	4	ILE	HA	32	LYS	QG
3	GLU	HA	32	LYS	QD	3	GLU	HA	4	ILE	QG1
6	VAL	HB	17	PRO	QG	12	PRO	HA	13	GLN	QG

36	THR	HB	36	THR	QG2	17	PRO	QB	17	PRO	QD
12	PRO	HA	12	PRO	QB	27	LYS	QB	27	LYS	QD
12	PRO	HA	12	PRO	QG	3	GLU	QB	32	LYS	QD
9	SER	HA	9	SER	QB	16	LYS	QB	16	LYS	QG
11	SER	HA	11	SER	QB	37	PRO	QG	37	PRO	QB
10	GLY	QA	11	SER	QB	15	LEU	HG	25	PHE	HA
37	PRO	HA	37	PRO	QB	4	ILE	QG1	33	CYS	HA
35	CYS	QB	37	PRO	HA	4	ILE	QG1	4	ILE	QD1
6	VAL	HA	6	VAL	HB	4	ILE	QG1	32	LYS	HA
6	VAL	HA	6	VAL	QG2	7	LYS	HA	7	LYS	QG
2	VAL	HA	37	PRO	QD	6	VAL	QG2	17	PRO	QG
16	LYS	HA	16	LYS	QB	6	VAL	QG2	17	PRO	QB
16	LYS	HA	16	LYS	QG	4	ILE	QG2	5	ASN	QB
4	ILE	HA	4	ILE	QG1	4	ILE	QG2	4	ILE	HB
4	ILE	HA	4	ILE	HB	4	ILE	QG2	4	ILE	QD1
13	GLN	HA	13	GLN	QB	4	ILE	HA	4	ILE	QG2
3	GLU	HA	3	GLU	QB	4	ILE	QG2	4	ILE	QG1
7	LYS	HA	7	LYS	QB	4	ILE	QD1	33	CYS	HA
32	LYS	HA	32	LYS	QG	4	ILE	HB	4	ILE	QD1
29	MET	HA	29	MET	QB	7	LYS	HA	7	LYS	QD
28	CYS	HA	28	CYS	QB	37	PRO	QG	37	PRO	QD
35	CYS	HA	35	CYS	QB	3	GLU	QB	3	GLU	QG
23	MET	HA	23	MET	QB	11	SER	HA	12	PRO	QG
8	CYS	HA	8	CYS	QB	15	LEU	HA	15	LEU	HG
24	ARG	HA	24	ARG	QB	17	PRO	HA	18	CYS	HA
4	ILE	HA	5	ASN	HA	11	SER	HA	28	CYS	HA
11	SER	HA	12	PRO	QD	17	PRO	QD	18	CYS	HA
12	PRO	QD	12	PRO	QB	8	CYS	QB	9	SER	HA
12	PRO	QD	12	PRO	QG	8	CYS	QB	13	GLN	HA
21	ALA	HA	21	ALA	QB	1	GLY	QA	4	ILE	HA
12	PRO	QD	12	PRO	HA	11	SER	HA	15	LEU	QB
17	PRO	QD	17	PRO	QG	37	PRO	QB	37	PRO	QD
16	LYS	QB	17	PRO	QD	2	VAL	QG2	3	GLU	HA
10	GLY	QA	12	PRO	QD	6	VAL	QG1	32	LYS	HA
1	GLY	QA	37	PRO	QD	21	ALA	QB	35	CYS	QB
15	LEU	QB	15	LEU	HG	2	VAL	QG2	35	CYS	QB
4	ILE	HB	4	ILE	QG1	37	PRO	QD	37	PRO	QB
25	PHE	HA	25	PHE	QB	37	PRO	QD	37	PRO	QB
4	ILE	QG1	6	VAL	HB	4	ILE	HA	5	ASN	QB
32	LYS	HA	32	LYS	QB	37	PRO	HA	37	PRO	QG
4	ILE	HB	6	VAL	HB	2	VAL	HA	37	PRO	QG
6	VAL	HB	6	VAL	QG1	25	PHE	QB	27	LYS	QD
6	VAL	HB	6	VAL	QG2	21	ALA	QB	35	CYS	QB
29	MET	QG	32	LYS	QB	15	LEU	QD2	15	LEU	HG
32	LYS	QB	32	LYS	QG	2	VAL	QG2	37	PRO	QG
4	ILE	QD1	6	VAL	HB	26	GLY	QA	35	CYS	QB
2	VAL	HB	2	VAL	QG2	34	HIS	HA	35	CYS	QB
13	GLN	HA	13	GLN	QG	36	THR	HA	37	PRO	QD
7	LYS	QB	7	LYS	QD	35	CYS	QB	36	THR	HB
7	LYS	QB	7	LYS	QG	17	PRO	HA	17	PRO	QG
2	VAL	HA	37	PRO	QB	17	PRO	HA	17	PRO	QB
29	MET	HA	29	MET	QG	17	PRO	HA	21	ALA	HA
12	PRO	QB	12	PRO	QG	12	PRO	HA	15	LEU	HG



9	SER	QB	10	GLY	QA	2	VAL	HB	37	PRO	QB
15	LEU	QD1	16	LYS	HA	23	MET	QB	37	PRO	QD
29	MET	QB	33	CYS	HA	13	GLN	QB	13	GLN	QG
28	CYS	HA	33	CYS	HA	32	LYS	HA	32	LYS	QD
12	PRO	QB	17	PRO	QD	6	VAL	HB	17	PRO	QG
15	LEU	QB	25	PHE	QB	16	LYS	QG	17	PRO	QB
15	LEU	HA	15	LEU	QB	15	LEU	QD1	15	LEU	HG
11	SER	QB	15	LEU	QB	4	ILE	QD1	6	VAL	QG1
5	ASN	HA	5	ASN	QB	6	VAL	QG1	17	PRO	QB
3	GLU	QB	32	LYS	QB	4	ILE	QD1	5	ASN	QB
3	GLU	QB	32	LYS	QG	4	ILE	QD1	17	PRO	QB
36	THR	HA	37	PRO	QB						

# Appendix C

## Chemical shift assignment of Conk-S2

1	1.390	HB	1	47	1.896	HB2	6
2	3.688	HA	1	48	3.148	HD2	6
3	7.876	H	1	49	3.175	HD3	6
4	1.526	HG3	2	50	4.513	HA	6
5	1.640	HG2	2	51	8.280	H	6
6	1.978	HB3	2	52	53.840	CA	6
7	1.998	HB2	2	53	122.000	N	6
8	3.164	HD2	2	54	1.951	HB3	7
9	3.193	HD3	2	55	1.979	HG2	7
10	4.714	HA	2	56	2.022	HG3	7
11	8.200	H	2	57	2.352	HB2	7
12	54.280	CA	2	58	3.807	HD3	7
13	118.900	N	2	59	3.825	HD2	7
14	2.017	HB3	3	60	4.534	HA	7
15	2.051	HG2	3	61	3.959	HB2	8
16	2.102	HG3	3	62	3.989	HB3	8
17	2.309	HB2	3	63	4.345	HA	8
18	4.087	HD3	3	64	8.136	H	8
19	4.185	HD2	3	65	3.048	HB3	9
20	4.449	HA	3	66	3.081	HB2	9
21	1.394	HG3	4	67	4.470	HA	9
22	1.394	HG2	4	68	6.824	HE2	9
23	1.720	HB3	4	69	6.980	H	9
24	1.720	HD2	4	70	7.200	HD2	9
25	1.720	HD3	4	71	36.420	CB	9
26	1.720	HB2	4	72	57.840	CA	9
27	2.964	HE3	4	73	116.300	N	9
28	2.964	HE2	4	74	2.671	QG	10
29	4.230	HA	4	75	2.808	HB2	10
30	8.270	H	4	76	2.808	HB3	10
31	24.710	CG	4	77	4.387	HA	10
32	29.230	CD	4	78	7.410	H	10
33	42.160	CE	4	79	38.640	CB	10
34	56.260	CA	4	80	57.760	CA	10
35	33.240	CB	4	81	119.500	N	10
36	121.200	N	4	82	2.700	HB3	11
37	2.460	HB2	5	83	3.050	HB2	11
38	2.573	HB3	5	84	4.950	HA	11
39	4.556	HA	5	85	7.400	H	11
40	8.040	H	5	86	52.860	CA	11
41	41.860	CB	5	87	39.510	CB	11
42	53.820	CA	5	88	44.160	CG	11
43	120.400	N	5	89	111.000	N	11
44	1.384	HG3	6	90	0.825	HD2	12
45	1.431	HG2	6	91	1.557	HB2	12
46	1.640	HB3	6	92	1.671	HG	12

93	1.718	HB3	12	146	8.080	H	20
94	4.292	HA	12	147	21.420	CG2	20
95	7.510	H	12	148	69.230	CB	20
96	53.250	CA	12	149	60.350	CA	20
97	121.800	N	12	150	106.000	N	20
98	1.973	HB3	13	151	1.266	HG3	21
99	2.006	HG2	13	152	1.448	HG2	21
100	2.056	HG3	13	153	1.684	HD3	21
101	2.288	HB2	13	154	1.732	HD2	21
102	3.840	HD2	13	155	1.837	HB3	21
103	3.961	HD3	13	156	2.193	HB2	21
104	4.271	HA	13	157	2.920	HE3	21
105	0.561	HB	14	158	2.975	HE2	21
106	3.422	HA	14	159	4.478	HA	21
107	7.040	H	14	160	7.740	H	21
108	51.960	CA	14	161	53.070	CA	21
109	15.360	CB	14	162	122.900	N	21
110	120.700	N	14	163	1.914	HB3	22
111	2.609	HB2	15	164	1.940	HG3	22
112	2.303	HB3	15	165	1.982	HG2	22
113	4.844	HA	15	166	2.198	HB2	22
114	7.640	H	15	167	3.436	HD3	22
115	52.070	CA	15	168	3.844	HD2	22
116	43.630	CB	15	169	4.082	HA	22
117	39.780	CG	15	170	2.116	HB2	23
118	121.600	N	15	171	2.288	HG3	23
119	4.680	HA	16	172	2.288	HG2	23
120	4.083	HB2	16	173	2.116	HB3	23
121	3.941	HB3	16	174	4.606	HA	23
122	9.080	H	16	175	9.070	H	23
123	64.300	CB	16	176	41.340	CD	23
124	61.770	CA	16	177	36.340	CG	23
125	122.100	N	16	178	34.440	CB	23
126	3.763	HA2	17	179	54.590	CA	23
127	4.268	HA3	17	180	120.900	N	23
128	8.890	H	17	181	1.993	HB2	24
129	45.510	CA	17	182	1.993	HB3	24
130	108.400	N	17	183	2.195	HG2	24
131	3.695	HB2	18	184	2.334	HG3	24
132	3.695	HB3	18	185	4.658	HA	24
133	4.773	HA	18	186	8.670	H	24
134	8.740	H	18	187	29.940	CB	24
135	65.610	CB	18	188	48.280	CD	24
136	56.630	CA	18	189	56.240	CA	24
137	117.900	N	18	190	121.800	N	24
138	3.617	HA3	19	191	1.802	HB2	25
139	4.598	HA2	19	192	1.240	HG2	25
140	7.920	H	19	193	1.802	HB3	25
141	43.350	CA	19	194	1.187	HG3	25
142	105.300	N	19	195	3.307	HD2	25
143	1.029	HG2	20	196	2.817	HD3	25
144	4.544	HA	20	197	4.654	HA	25
145	4.544	HB	20	198	8.720	H	25

199	34.600	CB	25	252	7.920	H	31
200	29.140	CG	25	253	18.400	CB	31
201	42.610	CD	25	254	54.700	CA	31
202	52.550	CA	25	255	124.900	N	31
203	127.500	N	25	256	1.158	HG3	32
204	0.941	HG2	26	257	1.296	HG2	32
205	1.267	HG12	26	258	1.741	HB2	32
206	1.140	HG13	26	259	1.741	HB3	32
207	0.755	QD1	26	260	1.488	HD2	32
208	1.700	HB	26	261	1.488	HD3	32
209	5.315	HA	26	262	2.890	HE3	32
210	8.880	H	26	263	2.890	HE2	32
211	28.010	CG1	26	264	4.088	HA	32
212	20.780	CG2	26	265	7.540	H	32
213	12.170	CD1	26	266	41.970	CE	32
214	40.640	CB	26	267	29.100	CD	32
215	58.630	CA	26	268	25.290	CG	32
216	119.700	N	26	269	33.760	CB	32
217	2.715	HB2	27	270	55.880	CA	32
218	2.715	HB3	27	271	115.200	N	32
219	5.220	HA	27	272	1.256	HG3	33
220	6.638	HE2	27	273	1.301	HG2	33
221	7.165	HD2	27	274	1.652	HD2	33
222	9.870	H	27	275	1.652	HD3	33
223	54.810	CA	27	276	1.704	HB3	33
224	41.420	CB	27	277	1.855	HB2	33
225	126.500	N	27	278	2.998	HE3	33
226	2.696	HB2	28	279	2.056	HE2	33
227	3.428	HB3	28	280	3.544	HA	33
228	4.304	HA	28	281	7.870	H	33
229	6.674	HE2	28	282	24.940	CD	33
230	6.878	HD2	28	283	29.090	CB	33
231	10.370	H	28	284	42.610	CE	33
232	58.130	CA	28	285	56.940	CA	33
233	39.080	CB	28	286	116.400	N	33
234	123.200	N	28	287	1.375	HB2	34
235	1.912	HB3	29	288	1.375	HB3	34
236	2.992	HB2	29	289	2.173	HG2	34
237	4.395	HA	29	290	1.828	HG3	34
238	8.030	H	29	291	4.520	HA	34
239	44.180	CG	29	292	6.780	H	34
240	39.050	CB	29	293	34.430	CB	34
241	51.680	CA	29	294	33.160	CG	34
242	126.600	N	29	295	50.140	CD	34
243	2.824	HB2	30	296	53.810	CA	34
244	3.058	HB3	30	297	112.300	N	34
245	4.023	HA	30	298	2.643	HG	35
246	8.310	H	30	299	3.171	HB2	35
247	63.250	CB	30	300	3.171	HB3	35
248	61.100	CA	30	301	5.223	HA	35
249	121.400	N	30	302	8.500	H	35
250	1.455	HB	31	303	44.600	CB	35
251	4.199	HA	31	304	57.780	CA	35

305	122.500	N	35	358	4.470	HA3	42
306	2.127	HB	36	359	7.237	H	42
307	4.659	HA	36	360	1.246	HG3	43
308	9.170	H	36	361	1.502	HD2	43
309	0.880	HG1	36	362	1.502	HD3	43
310	0.599	HG2	36	363	1.246	HG2	43
311	22.230	CG1	36	364	1.582	HB3	43
312	20.190	CG2	36	365	1.582	HB2	43
313	35.760	CB	36	366	2.878	HE3	43
314	60.210	CA	36	367	2.878	HE2	43
315	122.700	N	36	368	4.643	HA	43
316	1.079	HG2	37	369	8.140	H	43
317	4.209	HB	37	370	24.730	CG	43
318	5.456	HA	37	371	29.390	CD	43
319	7.970	H	37	372	35.640	CB	43
320	22.180	CG2	37	373	42.280	CE	43
321	71.250	CB	37	374	55.020	CA	43
322	61.990	CA	37	375	121.500	N	43
323	112.800	N	37	376	3.775	HA2	44
324	2.961	HB3	38	377	4.128	HA3	44
325	3.216	HB2	38	378	8.590	H	44
326	4.942	HA	38	379	45.170	CA	44
327	6.527	HE2	38	380	106.500	N	44
328	7.091	HD2	38	381	3.676	HA3	45
329	9.460	H	38	382	4.532	HA2	45
330	41.190	CB	38	383	9.200	H	45
331	56.070	CA	38	384	43.700	CA	45
332	121.300	N	38	385	112.800	N	45
333	0.958	HG1	39	386	2.893	HB2	46
334	1.810	HG2	39	387	2.968	HB3	46
335	4.289	HA	39	388	4.870	HA	46
336	3.919	HB	39	389	9.050	H	46
337	8.540	H	39	390	40.750	CB	46
338	22.420	CG2	39	391	51.530	CA	46
339	62.730	CA	39	392	118.800	N	46
340	69.060	CB	39	393	3.578	HA2	47
341	117.000	N	39	394	4.210	HA3	47
342	2.454	HB2	40	395	8.270	H	47
343	2.226	HB3	40	396	45.530	CA	47
344	4.804	HA	40	397	104.000	N	47
345	6.858	HE2	40	398	1.912	HB3	48
346	6.868	HD2	40	399	3.213	HB2	48
347	8.860	H	40	400	4.691	HA	48
348	41.880	CB	40	401	8.180	H	48
349	54.580	CA	40	402	34.980	CB	48
350	129.100	N	40	403	50.410	CA	48
351	2.572	HB3	41	404	120.600	N	48
352	2.713	HB2	41	405	2.372	HB2	49
353	4.455	HA	41	406	2.460	HB3	49
354	9.140	H	41	407	4.715	HA	49
355	55.630	CA	41	408	6.650	H	49
356	124.300	N	41	409	38.690	CB	49
357	4.292	HA2	42	410	53.520	CA	49

411	118.700	N	49	464	3.731	HA	57
412	2.615	HB3	50	465	3.150	HD2	57
413	3.285	HB2	50	466	3.150	HD3	57
414	4.930	HA	50	467	8.420	H	57
415	7.014	HE2	50	468	28.220	CG	57
416	9.930	H	50	469	30.400	CB	57
417	25.503	HD2	50	470	43.720	CD	57
418	42.350	CB	50	471	59.680	CA	57
419	55.920	CA	50	472	122.100	N	57
420	123.800	N	50	473	2.082	HB3	58
421	3.880	HB2	51	474	2.082	HB2	58
422	3.949	HB3	51	475	2.447	HG2	58
423	4.607	HA	51	476	2.447	HG3	58
424	9.390	H	51	477	4.009	HA	58
425	64.070	CB	51	478	8.000	H	58
426	59.960	CA	51	479	28.960	CB	58
427	116.000	N	51	480	34.190	CG	58
428	1.353	HG2	52	481	57.970	CA	58
429	1.639	HG3	52	482	115.300	N	58
430	1.759	HB3	52	483	1.054	HG2	59
431	1.919	HB2	52	484	4.082	HA	59
432	3.256	HD2	52	485	3.954	HB	59
433	3.260	HD3	52	486	7.350	H	59
434	4.609	HA	52	487	22.330	CG2	59
435	7.430	H	52	488	65.770	CA	59
436	54.260	CA	52	489	69.370	CB	59
437	116.800	N	52	490	112.200	N	59
438	1.094	HG2	53	491	1.849	HG	60
439	4.106	HA	53	492	2.106	HB2	60
440	4.352	HB	53	493	2.106	HB3	60
441	8.128	H	53	494	4.684	HA	60
442	2.780	HB3	54	495	7.730	H	60
443	2.846	HB2	54	496	41.040	CB	60
444	4.543	HA	54	497	53.390	CA	60
445	7.967	H	54	498	112.000	N	60
446	2.884	HB3	55	499	1.965	HB3	61
447	3.073	HB2	55	500	2.119	HB2	61
448	4.425	HA	55	501	2.145	HG2	61
449	6.830	H	55	502	2.227	HG3	61
450	40.620	CB	55	503	3.810	HA	61
451	57.000	CA	55	504	7.340	H	61
452	118.000	N	55	505	30.440	CB	61
453	2.967	HB2	56	506	33.780	CG	61
454	2.511	HB3	56	507	58.280	CA	61
455	2.012	HG	56	508	120.800	N	61
456	6.710	H	56	509	2.940	HB2	62
457	43.860	CB	56	510	3.077	HB3	62
458	58.950	CA	56	511	4.764	HA	62
459	120.400	N	56	512	6.769	HE2	62
460	1.527	HG2	57	513	6.897	HD2	62
461	1.640	HG3	57	514	8.190	H	62
462	1.742	HB3	57	515	55.390	CA	62
463	1.742	HB2	57	516	119.600	N	62

517	1.904	HG2	63	528	62.020	CA	64
518	1.948	HG3	63	529	33.360	CB	64
519	1.963	HB3	63	530	21.340	CG1	64
520	2.242	HB2	63	531	20.050	CG2	64
521	3.633	HD2	63	532	119.200	N	64
522	3.787	HD3	63	533	3.963	HA2	65
523	4.365	HA	63	534	4.094	HA3	65
524	0.9327	HG2	64	535	7.938	H	65
525	2.102	HB	64	536	46.200	CA	65
526	4.187	HA	64	537	117.600	N	65
527	8.058	H	64				

# Appendix D

## NMR restraints for Conk-S2

### D.1 Distance restraints for Conk-S2

25	ARG	H	40	TYR	H	4.92
16	SER	HA	40	TYR	H	5.19
24	GLN	HA	40	TYR	H	5.46
39	THR	HA	40	TYR	H	2.91
22	PRO	HA	40	TYR	H	4.91
39	THR	HB	40	TYR	H	4.53
40	TYR	H	40	TYR	HB3	3.41
40	TYR	H	40	TYR	HB2	3.62
24	GLN	HA	25	ARG	H	2.70
25	ARG	H	39	THR	HA	4.95
24	GLN	HG2	25	ARG	H	5.20
24	GLN	HG3	25	ARG	H	4.65
24	GLN	HB3	25	ARG	H	3.95
25	ARG	H	37	THR	QG2	4.69
29	ASN	H	34	GLN	H	3.95
29	ASN	H	35	CYSS	HA	3.90
29	ASN	H	34	GLN	HA	4.82
29	ASN	H	36	VAL	QG2	3.88
27	TYR	H	37	THR	HA	5.00
26	ILE	HA	27	TYR	H	2.93
27	TYR	H	27	TYR	HB3	3.48
26	ILE	QG2	27	TYR	H	3.42
27	TYR	H	38	PHE	H	5.50
27	TYR	H	36	VAL	H	3.71
27	TYR	H	27	TYR	HD2	3.15
27	TYR	H	36	VAL	QG2	4.45
30	SER	H	31	ALA	H	3.96
31	ALA	H	31	ALA	QB	3.11
40	TYR	H	41	ASN	H	4.33
38	PHE	HE2	41	ASN	H	5.19
40	TYR	HD2	41	ASN	H	5.50
40	TYR	HA	41	ASN	H	3.02
16	SER	HA	41	ASN	H	3.39
40	TYR	HB2	41	ASN	H	4.58
26	ILE	H	50	PHE	H	3.88
50	PHE	H	50	PHE	HE2	5.50
49	ASN	HA	50	PHE	H	3.01
50	PHE	H	50	PHE	HB2	3.64
25	ARG	HB3	50	PHE	H	4.44
26	ILE	HB	50	PHE	H	4.06
28	TYR	H	50	PHE	HE2	4.18
27	TYR	HA	28	TYR	H	3.26
28	TYR	H	28	TYR	HB3	3.40
27	TYR	HB2	28	TYR	H	3.59



28	TYR	H	28	TYR	HB2	3.70
21	LYS	H	42	GLY	H	5.50
20	THR	HB	21	LYS	H	3.60
19	GLY	HA2	21	LYS	H	4.11
35	CYSS	HB2	36	VAL	H	5.50
35	CYSS	HB3	36	VAL	H	5.50
36	VAL	H	36	VAL	HB	3.90
26	ILE	QG2	36	VAL	H	3.85
36	VAL	H	36	VAL	QG2	3.52
34	GLN	HA	35	CYSS	H	3.16
35	CYSS	H	36	VAL	H	5.03
57	ARG	H	58	GLN	H	3.74
56	CYSS	H	57	ARG	H	3.62
54	ASN	HA	57	ARG	H	4.45
56	CYSS	HB3	57	ARG	H	4.11
57	ARG	H	61	GLN	HG3	4.20
57	ARG	H	57	ARG	HB2	3.18
57	ARG	H	57	ARG	HG2	3.69
16	SER	H	17	GLY	H	3.96
15	ASP	HA	16	SER	H	3.04
16	SER	H	16	SER	HB2	3.68
15	ASP	HB3	16	SER	H	5.50
5	ASP	H	6	ARG	H	4.58
5	ASP	HA	6	ARG	H	2.70
5	ASP	HB2	6	ARG	H	4.33
6	ARG	H	6	ARG	HB3	3.70
5	ASP	HB3	6	ARG	H	4.33
11	ASN	HB3	12	LEU	H	4.23
11	ASN	HB2	12	LEU	H	3.54
12	LEU	H	12	LEU	QD1	3.35
11	ASN	H	12	LEU	H	3.43
9	TYR	HD2	12	LEU	H	5.48
9	TYR	HA	12	LEU	H	3.88
12	LEU	H	48	ASN	HB2	5.50
12	LEU	H	12	LEU	HB2	3.64
12	LEU	H	12	LEU	HG	3.93
23	GLU	HA	24	GLN	H	2.73
24	GLN	H	24	GLN	HG2	5.50
24	GLN	H	24	GLN	HG3	3.36
24	GLN	H	24	GLN	HB2	2.92
14	ALA	HA	15	ASP	H	2.96
15	ASP	H	15	ASP	HB2	3.63
15	ASP	H	15	ASP	HB3	3.74
13	PRO	HA	15	ASP	H	5.50
43	LYS	H	44	GLY	H	4.81
19	GLY	H	43	LYS	H	4.87
43	LYS	H	43	LYS	HA	2.85
17	GLY	HA3	43	LYS	H	3.36
43	LYS	H	43	LYS	HG2	4.03
29	ASN	HB2	30	SER	H	4.40
29	ASN	H	30	SER	H	5.50
29	ASN	HA	30	SER	H	3.46
29	ASN	HB3	30	SER	H	3.67

4	LYS	H	4	LYS	HB2	3.76
25	ARG	H	38	PHE	H	3.92
27	TYR	HD2	38	PHE	H	3.49
24	GLN	HA	38	PHE	H	4.43
37	THR	HB	38	PHE	H	3.72
37	THR	QG2	38	PHE	H	4.11
23	GLU	H	40	TYR	H	3.61
23	GLU	H	24	GLN	H	4.48
23	GLU	H	39	THR	HA	5.13
23	GLU	H	40	TYR	HB3	4.82
23	GLU	H	23	GLU	HG2	4.71
23	GLU	H	23	GLU	HG3	4.71
23	GLU	H	23	GLU	HB2	3.90
23	GLU	H	39	THR	HG1	4.63
60	CYSS	H	61	GLN	H	3.99
61	GLN	H	61	GLN	HB2	3.13
61	GLN	H	62	TYR	HB2	4.90
13	PRO	HA	14	ALA	H	3.00
40	TYR	HE2	48	ASN	H	5.39
40	TYR	HD2	48	ASN	H	5.50
9	TYR	HA	48	ASN	H	5.50
48	ASN	H	48	ASN	HB2	3.42
12	LEU	QD1	48	ASN	H	5.26
4	LYS	H	5	ASP	H	4.25
4	LYS	HA	5	ASP	H	3.42
4	LYS	HB3	5	ASP	H	4.10
56	CYSS	H	59	THR	H	5.39
55	ASP	H	56	CYSS	H	3.83
54	ASN	HA	56	CYSS	H	4.96
50	PHE	HB3	56	CYSS	H	4.11
56	CYSS	H	56	CYSS	HB2	3.76
56	CYSS	H	56	CYSS	HB3	3.42
52	ARG	HB2	56	CYSS	H	4.37
62	TYR	H	62	TYR	HB2	4.01
62	TYR	H	62	TYR	HB3	3.64
61	GLN	HB2	62	TYR	H	3.77
26	ILE	H	27	TYR	HD2	5.15
25	ARG	HA	26	ILE	H	3.29
25	ARG	HD3	26	ILE	H	4.54
26	ILE	H	50	PHE	HB2	5.17
25	ARG	HB3	26	ILE	H	3.26
26	ILE	H	26	ILE	HB	3.25
26	ILE	H	37	THR	QG2	5.50
26	ILE	H	26	ILE	HG13	3.37
26	ILE	H	26	ILE	QD1	3.71
61	GLN	H	62	TYR	H	3.72
60	CYSS	HB3	62	TYR	H	5.43
61	GLN	HG2	62	TYR	H	5.50
9	TYR	H	10	CYSS	H	4.13
10	CYSS	H	50	PHE	HE2	4.59
10	CYSS	H	48	ASN	HA	5.34
7	PRO	HD2	10	CYSS	H	4.56
9	TYR	HB2	10	CYSS	H	4.56

6	ARG	HB3	10	CYSS	H	5.50
10	CYSS	H	12	LEU	QD1	4.60
63	PRO	HA	64	VAL	H	3.11
64	VAL	H	64	VAL	HB	4.01
64	VAL	H	64	VAL	QG1	3.76
45	GLY	HA2	46	ASN	H	3.24
15	ASP	HB2	46	ASN	H	3.88
40	TYR	HD2	49	ASN	H	5.50
49	ASN	H	50	PHE	HE2	5.50
48	ASN	HA	49	ASN	H	3.12
49	ASN	H	49	ASN	HB3	3.56
50	PHE	HB3	55	ASP	H	5.18
55	ASP	H	55	ASP	HB2	3.34
52	ARG	HB2	55	ASP	H	3.80
52	ARG	HB3	55	ASP	H	3.96
18	SER	H	43	LYS	H	4.44
17	GLY	HA3	18	SER	H	2.99
18	SER	H	44	GLY	HA3	5.10
17	GLY	HA2	18	SER	H	3.13
18	SER	H	43	LYS	HD2	5.37
18	SER	H	43	LYS	HD3	5.37
64	VAL	HA	65	GLY	H	3.53
64	VAL	HB	65	GLY	H	4.09
64	VAL	QG1	65	GLY	H	3.83
38	PHE	H	39	THR	H	5.22
38	PHE	HD2	39	THR	H	4.00
38	PHE	HA	39	THR	H	3.13
39	THR	H	39	THR	HB	3.18
38	PHE	HB3	39	THR	H	3.74
38	PHE	HB2	39	THR	H	3.34
39	THR	H	39	THR	HG1	4.27
52	ARG	H	55	ASP	H	5.07
51	SER	HB3	52	ARG	H	4.46
50	PHE	HB3	52	ARG	H	3.59
52	ARG	H	55	ASP	HB3	5.12
52	ARG	H	55	ASP	HB2	4.02
50	PHE	HB2	52	ARG	H	4.48
52	ARG	H	52	ARG	HB2	3.80
52	ARG	H	52	ARG	HB3	3.62
29	ASN	H	33	LYS	H	5.12
32	LYS	H	33	LYS	H	3.15
33	LYS	H	34	GLN	H	3.58
33	LYS	H	33	LYS	HA	2.93
29	ASN	HB3	33	LYS	H	4.69
32	LYS	HB2	33	LYS	H	4.14
9	TYR	H	9	TYR	HB2	3.46
7	PRO	HG2	9	TYR	H	5.01
9	TYR	H	12	LEU	HG	5.50
9	TYR	H	12	LEU	QD1	4.41
51	SER	H	52	ARG	H	3.95
51	SER	H	51	SER	HB3	3.67
50	PHE	HB3	51	SER	H	4.33
58	GLN	H	60	CYSS	H	4.67

54	ASN	HA	58	GLN	H	4.61
55	ASP	HA	58	GLN	H	5.13
58	GLN	H	58	GLN	HB2	3.15
57	ARG	HB2	58	GLN	H	3.16
31	ALA	H	32	LYS	H	3.09
32	LYS	H	34	GLN	H	5.21
32	LYS	H	33	LYS	HA	5.11
32	LYS	H	32	LYS	HB2	3.93
31	ALA	QB	32	LYS	H	3.43
32	LYS	H	32	LYS	HG3	3.41
32	LYS	H	32	LYS	HG2	3.26
26	ILE	HA	37	THR	H	5.50
36	VAL	HA	37	THR	H	2.99
36	VAL	HB	37	THR	H	3.17
37	THR	H	37	THR	QG2	3.57
36	VAL	QG2	37	THR	H	4.04
29	ASN	HB3	34	GLN	H	4.29
58	GLN	H	59	THR	H	3.61
55	ASP	HA	59	THR	H	4.86
59	THR	H	59	THR	HB	2.94
58	GLN	HG2	59	THR	H	5.50
58	GLN	HG3	59	THR	H	5.50
58	GLN	HB2	59	THR	H	3.58
59	THR	H	60	CYSS	H	3.08
60	CYSS	H	60	CYSS	HB2	3.75
11	ASN	H	12	LEU	HB2	4.76
11	ASN	H	28	TYR	HD2	5.50
7	PRO	HD2	11	ASN	H	5.36
11	ASN	H	11	ASN	HB2	3.15
7	PRO	HG2	11	ASN	H	4.50
11	ASN	H	12	LEU	HG	4.85
15	ASP	HA	17	GLY	H	4.38
16	SER	HB2	17	GLY	H	4.11
43	LYS	HA	44	GLY	H	2.73
43	LYS	HG2	44	GLY	H	4.96
43	LYS	HG3	44	GLY	H	5.26
20	THR	H	21	LYS	H	3.85
20	THR	H	20	THR	QG2	4.12
19	GLY	H	19	GLY	HA2	2.78
24	GLN	HA	37	THR	HB	3.80
24	GLN	HB3	37	THR	HB	4.29
24	GLN	HG2	37	THR	HB	4.60
24	GLN	HG3	37	THR	HB	5.15
16	SER	HA	39	THR	HB	4.99
59	THR	HB	61	GLN	HG2	5.50
58	GLN	HB2	59	THR	HA	4.96
58	GLN	HB3	59	THR	HA	5.49
59	THR	HA	60	CYSS	HB2	5.50
16	SER	HB3	40	TYR	HA	4.60
16	SER	HB2	40	TYR	HA	4.82
16	SER	HA	16	SER	HB2	2.93
16	SER	HB2	17	GLY	HA3	5.09
25	ARG	HD2	51	SER	HB3	5.50

24	GLN	HA	39	THR	HA	4.16
39	THR	HA	40	TYR	HB3	4.72
39	THR	HA	39	THR	HG1	3.18
64	VAL	HA	64	VAL	QG1	2.40
54	ASN	HA	57	ARG	HA	5.05
57	ARG	HA	61	GLN	HG3	3.01
57	ARG	HA	57	ARG	HG3	3.07
28	TYR	HA	61	GLN	HA	5.18
60	CYSS	HA	61	GLN	HA	4.76
61	GLN	HA	62	TYR	HA	5.18
28	TYR	HA	28	TYR	HB2	2.55
4	LYS	HA	4	LYS	HB3	2.55
31	ALA	QB	32	LYS	HA	4.22
32	LYS	HA	32	LYS	HG3	3.62
32	LYS	HA	32	LYS	HG2	3.49
14	ALA	HA	15	ASP	HA	4.78
13	PRO	HA	14	ALA	HA	4.57
14	ALA	HA	48	ASN	HB2	4.95
14	ALA	HA	38	PHE	HB2	5.50
14	ALA	HA	15	ASP	HB2	5.26
43	LYS	HA	44	GLY	HA2	4.54
17	GLY	HA3	44	GLY	HA3	5.50
15	ASP	HB2	48	ASN	HA	5.50
15	ASP	HB3	16	SER	HA	5.33
50	PHE	HB3	56	CYSS	HB3	4.20
56	CYSS	HB2	61	GLN	HG3	4.29
56	CYSS	HB2	61	GLN	HG2	4.45
26	ILE	HB	56	CYSS	HB3	5.21
15	ASP	HB2	16	SER	HA	5.22
15	ASP	HB2	45	GLY	HA2	5.24
15	ASP	HB2	45	GLY	HA3	3.91
15	ASP	HB3	45	GLY	HA3	4.26
19	GLY	HA2	43	LYS	HG2	5.39
16	SER	HB2	19	GLY	HA2	4.98
19	GLY	HA2	43	LYS	HG3	5.50
25	ARG	HA	25	ARG	HD3	3.03
25	ARG	HA	25	ARG	HD2	3.08
25	ARG	HB3	25	ARG	HD2	3.88
25	ARG	HD2	26	ILE	QD1	5.14
50	PHE	HB2	52	ARG	HB2	5.50
40	TYR	HB3	41	ASN	HA	5.50
16	SER	HA	40	TYR	HB2	4.98
25	ARG	HG3	40	TYR	HB3	3.65
27	TYR	HB3	49	ASN	HA	4.35
14	ALA	HA	27	TYR	HB2	5.50
14	ALA	HA	27	TYR	HB3	5.50
27	TYR	HB2	28	TYR	HB3	5.50
27	TYR	HB3	38	PHE	HB3	5.14
52	ARG	HB2	55	ASP	HB3	4.39
12	LEU	QD1	46	ASN	HB3	3.66
50	PHE	HB3	55	ASP	HB3	4.16
26	ILE	HB	50	PHE	HB2	3.62
52	ARG	HB2	55	ASP	HB2	5.50

12	LEU	QD1	46	ASN	HB2	3.66
26	ILE	HB	26	ILE	QD1	3.96
7	PRO	HG2	28	TYR	HB3	5.48
29	ASN	HB3	32	LYS	HG2	4.87
10	CYSS	HB3	28	TYR	HB2	3.00
7	PRO	HG2	10	CYSS	HB2	5.23
6	ARG	HB3	10	CYSS	HB2	4.39
9	TYR	HB3	48	ASN	HA	3.89
9	TYR	HB3	12	LEU	HB2	5.50
7	PRO	HG2	9	TYR	HB2	4.41
9	TYR	HB3	12	LEU	QD1	4.00
9	TYR	HB2	12	LEU	QD1	4.52
13	PRO	HA	48	ASN	HB2	5.50
12	LEU	HB2	48	ASN	HB2	4.68
34	GLN	HB2	36	VAL	QG1	4.61
34	GLN	HB3	36	VAL	QG1	4.61
23	GLU	HB2	40	TYR	HB3	3.58
61	GLN	HA	61	GLN	HG2	3.27
32	LYS	HB3	32	LYS	HG2	2.74
60	CYSS	HA	61	GLN	HG2	4.54
60	CYSS	HA	61	GLN	HG3	5.50
32	LYS	HB2	32	LYS	HG3	2.40
57	ARG	HA	61	GLN	HB2	3.80
54	ASN	HA	57	ARG	HB2	2.91
35	CYSS	HB2	61	GLN	HB2	5.50
35	CYSS	HB3	61	GLN	HB2	5.50
24	GLN	HB3	24	GLN	HG2	2.41
25	ARG	HG2	40	TYR	HB3	3.65
32	LYS	HA	32	LYS	HD3	4.05
58	GLN	HB2	59	THR	QG2	5.50
57	ARG	HB3	57	ARG	HG2	2.40
25	ARG	HA	26	ILE	HG12	5.50
25	ARG	HA	26	ILE	HG13	4.42
57	ARG	HA	57	ARG	HG2	3.76
26	ILE	QG2	26	ILE	HG13	3.65
26	ILE	HG13	37	THR	QG2	3.80
26	ILE	QG2	26	ILE	HG12	3.68
29	ASN	HB3	32	LYS	HG3	3.59
43	LYS	HA	43	LYS	HG2	3.77
29	ASN	HB2	32	LYS	HG3	4.77
18	SER	HA	43	LYS	HG3	4.93
28	TYR	HB2	59	THR	QG2	5.00
36	VAL	HA	37	THR	QG2	4.21
36	VAL	HA	36	VAL	QG1	3.97
24	GLN	HG2	37	THR	QG2	4.78
36	VAL	HB	37	THR	QG2	4.55
24	GLN	HB3	37	THR	QG2	4.31
26	ILE	HG12	37	THR	QG2	4.49
36	VAL	QG2	37	THR	QG2	4.58
20	THR	QG2	43	LYS	HA	5.38
19	GLY	HA2	20	THR	QG2	5.50
26	ILE	QG2	56	CYSS	HB2	3.72
26	ILE	QG2	50	PHE	HB2	3.62

26	ILE	QG2	56	CYSS	HB3	3.54
26	ILE	QG2	26	ILE	QD1	2.85
36	VAL	HA	36	VAL	QG2	4.03
63	PRO	HA	64	VAL	QG1	4.92
62	TYR	HB2	64	VAL	QG1	5.26
62	TYR	HB3	64	VAL	QG1	5.50
29	ASN	HB3	36	VAL	QG2	3.90
32	LYS	HB2	36	VAL	QG2	5.49
32	LYS	HD2	36	VAL	QG2	5.50
56	CYSS	HB2	57	ARG	HG2	5.50
27	TYR	HB2	48	ASN	HB2	5.50
27	TYR	HB3	48	ASN	HB2	5.50
26	ILE	QD1	37	THR	HB	5.50
38	PHE	HB2	39	THR	HB	5.50
20	THR	HA	20	THR	QG2	3.16
10	CYSS	HA	10	CYSS	HB3	2.54
9	TYR	HA	9	TYR	HB3	2.55
10	CYSS	HA	28	TYR	HB3	3.52
9	TYR	HA	12	LEU	QD1	3.09
55	ASP	HA	55	ASP	HB3	2.55
55	ASP	HA	58	GLN	HB2	3.48
38	PHE	HA	38	PHE	HB3	2.94
11	ASN	HA	30	SER	HA	4.81
59	THR	HA	60	CYSS	HA	5.12
37	THR	H	37	THR	HB	3.99
59	THR	HB	60	CYSS	H	3.32
18	SER	HB2	19	GLY	H	5.11
18	SER	HB3	19	GLY	H	5.11
16	SER	HB3	17	GLY	H	5.13
16	SER	H	16	SER	HB3	3.78
16	SER	HB2	41	ASN	H	4.30
16	SER	HB3	38	PHE	HE2	4.63
16	SER	HB2	38	PHE	HE2	5.50
28	TYR	HD2	30	SER	HA	4.88
57	ARG	HA	62	TYR	H	5.28
57	ARG	HA	60	CYSS	H	4.92
57	ARG	HA	61	GLN	H	3.83
33	LYS	HA	34	GLN	H	3.39
45	GLY	HA3	46	ASN	H	3.52
57	ARG	H	57	ARG	HD2	5.50
57	ARG	H	57	ARG	HD3	5.50
56	CYSS	HB2	57	ARG	H	3.66
15	ASP	HB3	46	ASN	H	4.98
17	GLY	H	45	GLY	HA3	4.95
50	PHE	HB2	56	CYSS	H	4.83
38	PHE	HE2	40	TYR	HB3	4.28
40	TYR	HB3	41	ASN	H	5.50
40	TYR	HB3	42	GLY	H	5.27
38	PHE	HE2	40	TYR	HB2	4.89
27	TYR	HB3	38	PHE	H	5.37
27	TYR	HB3	27	TYR	HD2	3.05
27	TYR	HB2	50	PHE	HE2	4.45
27	TYR	HB3	27	TYR	HE2	5.13

27	TYR	HE2	38	PHE	HB2	4.75
38	PHE	HB2	38	PHE	HE2	5.26
38	PHE	H	38	PHE	HB2	4.06
38	PHE	H	38	PHE	HB3	3.84
60	CYSS	HB3	61	GLN	H	3.67
38	PHE	HB2	38	PHE	HD2	3.31
27	TYR	HE2	38	PHE	HB3	3.68
46	ASN	HB2	47	GLY	H	3.36
46	ASN	HB3	47	GLY	H	3.36
26	ILE	HB	27	TYR	HD2	5.50
55	ASP	H	55	ASP	HB3	3.89
55	ASP	HB3	56	CYSS	H	4.25
55	ASP	HB2	56	CYSS	H	3.85
11	ASN	H	11	ASN	HB3	3.72
29	ASN	H	29	ASN	HB3	3.68
28	TYR	HB3	28	TYR	HD2	3.41
28	TYR	HD2	29	ASN	HB2	5.50
27	TYR	HA	28	TYR	HB2	5.50
28	TYR	HB2	35	CYSS	HA	5.50
28	TYR	HB2	28	TYR	HD2	3.65
40	TYR	HD2	49	ASN	HB3	4.91
49	ASN	H	49	ASN	HB2	3.56
27	TYR	HA	28	TYR	HB3	4.85
10	CYSS	H	10	CYSS	HB2	3.75
10	CYSS	HB3	28	TYR	HD2	2.97
40	TYR	HD2	49	ASN	HB2	4.91
9	TYR	HB3	50	PHE	HE2	3.58
9	TYR	HB2	50	PHE	HE2	3.48
27	TYR	HE2	36	VAL	HB	5.02
25	ARG	HB2	27	TYR	HD2	4.87
25	ARG	HB2	40	TYR	HD2	5.40
25	ARG	HB3	40	TYR	HD2	5.50
32	LYS	HB2	34	GLN	H	4.02
32	LYS	HB3	34	GLN	H	4.61
56	CYSS	H	57	ARG	HB2	5.40
25	ARG	HG2	40	TYR	H	4.77
25	ARG	HG3	40	TYR	H	4.77
32	LYS	H	32	LYS	HD2	5.05
32	LYS	H	32	LYS	HD3	5.47
57	ARG	HG2	58	GLN	H	5.50
26	ILE	H	26	ILE	HG12	4.75
26	ILE	HG12	37	THR	HA	5.14
26	ILE	HG13	37	THR	HA	4.37
31	ALA	H	32	LYS	HG3	4.27
19	GLY	H	43	LYS	HG3	4.28
36	VAL	H	36	VAL	QG1	5.10
27	TYR	HE2	36	VAL	QG1	5.10
35	CYSS	HA	36	VAL	QG1	4.59
20	THR	QG2	21	LYS	H	5.43
26	ILE	H	26	ILE	QG2	4.96
26	ILE	QG2	27	TYR	HD2	5.01
26	ILE	QG2	56	CYSS	H	5.50
26	ILE	QG2	37	THR	HA	5.22



26	ILE	HA	26	ILE	QG2	3.28
27	TYR	HD2	36	VAL	QG2	4.70
27	TYR	HE2	36	VAL	QG2	3.64
28	TYR	HA	29	ASN	H	2.95
28	TYR	HA	36	VAL	H	4.19
10	CYSS	HA	50	PHE	HE2	3.82
46	ASN	HA	47	GLY	H	3.08
27	TYR	HD2	38	PHE	HB3	3.92
37	THR	HA	37	THR	QG2	3.37
36	VAL	QG2	37	THR	HA	4.00
37	THR	HA	38	PHE	HA	4.74
36	VAL	HA	37	THR	HA	4.64
26	ILE	HA	37	THR	HA	3.05
27	TYR	HD2	37	THR	HA	4.40
25	ARG	H	37	THR	HA	5.32
37	THR	HA	38	PHE	H	2.92
26	ILE	HA	26	ILE	QD1	3.90
26	ILE	HA	37	THR	QG2	4.51
26	ILE	HA	26	ILE	HG13	4.05
26	ILE	HA	27	TYR	HB3	5.26
25	ARG	HA	26	ILE	HA	4.75
26	ILE	HA	37	THR	HB	4.51
26	ILE	HA	27	TYR	HD2	4.06
26	ILE	HA	38	PHE	H	4.40
35	CYSS	HA	36	VAL	H	2.80
28	TYR	HA	35	CYSS	HA	3.01
26	ILE	QG2	35	CYSS	HA	4.41
35	CYSS	HA	36	VAL	QG2	4.30
18	SER	HA	19	GLY	H	3.26
38	PHE	HA	38	PHE	HD2	4.64
50	PHE	HA	51	SER	H	3.11
27	TYR	HA	27	TYR	HB2	2.96
26	ILE	QG2	27	TYR	HA	4.34
27	TYR	HA	50	PHE	HE2	4.82
27	TYR	HA	27	TYR	HD2	4.12
4	LYS	H	4	LYS	QG	4.83
4	LYS	HA	4	LYS	QG	3.54
5	ASP	H	5	ASP	QB	3.55
5	ASP	QB	6	ARG	H	3.75
5	ASP	QB	6	ARG	HA	5.34
8	SER	QB	9	TYR	H	4.04
8	SER	QB	11	ASN	H	5.34
9	TYR	HB3	47	GLY	QA	5.34
12	LEU	QD1	46	ASN	QB	3.20
18	SER	H	18	SER	QB	3.66
18	SER	H	43	LYS	QD	4.50
18	SER	QB	19	GLY	H	4.36
19	GLY	H	41	ASN	QB	3.11
23	GLU	H	23	GLU	QG	4.08
23	GLU	HB2	23	GLU	QG	2.58
23	GLU	HB2	25	ARG	QG	4.79
23	GLU	HB3	25	ARG	QG	5.34
23	GLU	QG	24	GLN	H	5.34

25	ARG	H	25	ARG	QG	4.03
25	ARG	HA	25	ARG	QG	3.32
25	ARG	QG	26	ILE	H	5.18
25	ARG	QG	40	TYR	H	3.97
25	ARG	QG	40	TYR	HB2	3.48
25	ARG	QG	40	TYR	HB3	3.16
26	ILE	QG2	35	CYSS	QB	4.16
32	LYS	HG2	34	GLN	QG	5.28
33	LYS	H	33	LYS	QG	3.13
33	LYS	H	33	LYS	QE	5.34
33	LYS	HA	33	LYS	QG	3.07
33	LYS	QB	33	LYS	QD	2.59
33	LYS	QB	33	LYS	QE	5.08
33	LYS	QG	33	LYS	QE	2.62
33	LYS	QD	33	LYS	QE	2.25
34	GLN	H	34	GLN	QB	3.57
34	GLN	H	34	GLN	QG	3.69
34	GLN	QB	36	VAL	QG1	3.80
34	GLN	QB	36	VAL	QG2	4.66
34	GLN	QG	35	CYSS	H	4.40
34	GLN	QG	36	VAL	QG1	5.23
35	CYSS	H	35	CYSS	QB	3.37
35	CYSS	QB	36	VAL	HA	5.34
40	TYR	HE2	49	ASN	QB	4.19
40	TYR	HD2	49	ASN	QB	4.21
41	ASN	H	41	ASN	QB	3.52
43	LYS	H	43	LYS	QB	3.68
43	LYS	HA	43	LYS	QE	4.20
43	LYS	QB	43	LYS	HG2	2.63
43	LYS	QB	44	GLY	H	3.30
43	LYS	QD	44	GLY	H	3.35
46	ASN	H	46	ASN	QB	3.37
46	ASN	HA	46	ASN	QB	2.47
46	ASN	QB	47	GLY	H	2.88
48	ASN	H	49	ASN	QB	5.34
49	ASN	H	49	ASN	QB	3.02
54	ASN	HA	57	ARG	QD	3.76
54	ASN	HA	58	GLN	QG	4.27
55	ASP	HA	58	GLN	QG	4.41
57	ARG	HA	57	ARG	QD	4.55
57	ARG	HB2	57	ARG	QD	2.53
57	ARG	HB2	58	GLN	QG	5.34
57	ARG	QD	58	GLN	H	5.34
58	GLN	H	58	GLN	QG	3.65
58	GLN	HA	58	GLN	QG	3.56
58	GLN	QG	59	THR	H	4.74
64	VAL	H	65	GLY	QA	5.34
64	VAL	HB	65	GLY	QA	5.21
64	VAL	QG1	65	GLY	QA	5.26

## D.2 Dihedral angle restraints for Conk-S2

2	ARG	PHI	-184.1	-67.3
2	ARG	PSI	121.5	161.5
4	LYS	CHI1	-65.0	-55.0
4	LYS	PHI	-125.9	-10.7
4	LYS	PSI	97.5	205.4
5	ASP	CHI1	-65.0	-55.0
5	ASP	PHI	-153.3	-65.1
5	ASP	PSI	-172.9	-69.2
6	ARG	CHI1	-65.0	-55.0
6	ARG	PHI	-106.9	-17.0
6	ARG	PSI	112.7	175.9
8	SER	CHI1	-65.0	-55.0
8	SER	PHI	36.6	103.0
8	SER	PSI	-101.9	-36.2
9	TYR	CHI1	-65.0	-55.0
9	TYR	PHI	-80.2	-40.2
9	TYR	PSI	-47.0	-3.5
10	CYSS	CHI1	-65.0	-55.0
10	CYSS	PHI	-118.2	-6.0
10	CYSS	PSI	-55.0	-10.3
11	ASN	CHI1	-65.0	-55.0
11	ASN	PHI	-150.5	6.3
11	ASN	PSI	-43.5	-3.5
12	LEU	CHI1	-65.0	-55.0
12	LEU	PHI	-118.0	-3.4
12	LEU	PSI	97.7	189.2
14	ALA	CHI1	-65.0	-55.0
14	ALA	PHI	-94.0	-54.0
14	ALA	PSI	67.6	176.5
15	ASP	CHI1	-65.0	-55.0
15	ASP	PHI	-140.0	-90.6
15	ASP	PSI	85.7	159.1
16	SER	CHI1	55.0	65.0
16	SER	PHI	-98.2	-58.2
16	SER	PSI	-52.4	-6.1
17	GLY	PHI	61.1	101.1
17	GLY	PSI	140.4	197.5
18	SER	CHI1	55.0	65.0
18	SER	PHI	-153.1	-95.1
18	SER	PSI	8.1	94.0
19	GLY	PHI	-277.5	98.3
19	GLY	PSI	2.3	340.3
20	THR	CHI1	-65.0	-55.0
20	THR	PHI	-144.1	-8.4
20	THR	PSI	-56.1	-14.6
21	LYS	CHI1	-65.0	-55.0
21	LYS	PHI	-164.6	-54.0
21	LYS	PSI	101.2	196.0
23	GLU	CHI1	-65.0	-55.0
23	GLU	PHI	-168.0	-127.2

23	GLU	PSI	115.6	170.8
24	GLN	CHI1	-65.0	-55.0
24	GLN	PHI	-138.7	-45.7
24	GLN	PSI	94.5	144.6
25	ARG	CHI1	-65.0	-55.0
25	ARG	PHI	-166.2	-110.0
25	ARG	PSI	133.7	211.7
26	ILE	CHI1	-65.0	-55.0
26	ILE	PHI	-158.3	-105.6
26	ILE	PSI	123.5	163.5
27	TYR	CHI1	55.0	65.0
27	TYR	PHI	-157.6	-114.1
27	TYR	PSI	124.5	205.1
28	TYR	CHI1	175.0	185.0
28	TYR	PHI	-143.9	-41.2
28	TYR	PSI	99.6	139.6
29	ASN	CHI1	-65.0	-55.0
29	ASN	PHI	-126.2	-44.0
29	ASN	PSI	73.1	165.8
30	SER	CHI1	-65.0	-55.0
30	SER	PHI	-82.4	-42.4
30	SER	PSI	-81.3	-29.5
31	ALA	CHI1	-65.0	-55.0
31	ALA	PHI	-84.6	-44.6
31	ALA	PSI	-67.0	-25.1
32	LYS	CHI1	-65.0	-55.0
32	LYS	PHI	-87.5	-47.5
32	LYS	PSI	-47.7	-7.1
33	LYS	CHI1	-65.0	-55.0
33	LYS	PHI	40.6	80.6
33	LYS	PSI	12.3	52.3
34	GLN	CHI1	-65.0	-55.0
34	GLN	PHI	-175.2	-92.9
34	GLN	PSI	110.5	187.0
35	CYSS	CHI1	175.0	185.0
35	CYSS	PHI	-133.0	-26.6
35	CYSS	PSI	104.6	165.3
36	VAL	CHI1	-65.0	-55.0
36	VAL	PHI	-156.8	-116.8
36	VAL	PSI	116.4	194.1
37	THR	CHI1	55.0	65.0
37	THR	PHI	-146.8	-25.4
37	THR	PSI	108.6	183.0
38	PHE	CHI1	55.0	65.0
38	PHE	PHI	-199.6	-85.0
38	PHE	PSI	132.9	200.4
39	THR	CHI1	-65.0	-55.0
39	THR	PHI	-138.4	-57.9
39	THR	PSI	90.1	142.9
40	TYR	CHI1	175.0	185.0
40	TYR	PHI	-138.8	-73.6
40	TYR	PSI	87.4	154.4
41	ASN	CHI1	-65.0	-55.0

41	ASN	PHI	-93.8	-53.8
41	ASN	PSI	-52.9	-12.9
42	GLY	PHI	86.7	126.7
42	GLY	PSI	-44.8	13.2
43	LYS	CHI1	-65.0	-55.0
43	LYS	PHI	-133.8	-37.9
43	LYS	PSI	-215.3	-128.1
44	GLY	PHI	-197.0	97.8
44	GLY	PSI	-245.4	83.6
45	GLY	PHI	19.5	154.0
45	GLY	PSI	-218.2	-116.8
46	ASN	CHI1	-65.0	-55.0
46	ASN	PHI	-178.0	-73.1
46	ASN	PSI	141.2	193.2
47	GLY	PHI	-163.6	30.7
47	GLY	PSI	-267.2	198.0
48	ASN	CHI1	-185.0	-175.0
48	ASN	PHI	-92.0	-52.0
48	ASN	PSI	31.7	146.4
49	ASN	CHI1	-185.0	-175.0
49	ASN	PHI	-198.7	-83.6
49	ASN	PSI	84.2	146.8
50	PHE	CHI1	-65.0	-55.0
50	PHE	PHI	-154.1	-75.7
50	PHE	PSI	101.4	166.4
51	SER	CHI1	55.0	65.0
51	SER	PHI	-115.1	-51.5
51	SER	PSI	-49.1	1.3
52	ARG	CHI1	-65.0	-55.0
52	ARG	PHI	-174.4	-85.8
52	ARG	PSI	126.3	166.6
53	THR	CHI1	-65.0	-55.0
53	THR	PHI	-122.3	-10.6
53	THR	PSI	-61.1	-18.3
54	ASN	CHI1	-65.0	-55.0
54	ASN	PHI	-86.7	-46.7
54	ASN	PSI	-55.4	-15.4
55	ASP	CHI1	-65.0	-55.0
55	ASP	PHI	-89.0	-34.1
55	ASP	PSI	-67.9	-23.5
56	CYSS	CHI1	55.0	65.0
56	CYSS	PHI	-104.2	-27.8
56	CYSS	PSI	-62.8	-21.5
57	ARG	CHI1	-65.0	-55.0
57	ARG	PHI	-84.5	-44.5
57	ARG	PSI	-63.6	-23.6
58	GLN	CHI1	-65.0	-55.0
58	GLN	PHI	-84.7	-44.7
58	GLN	PSI	-64.4	-24.4
59	THR	CHI1	-65.0	-55.0
59	THR	PHI	-100.1	-29.9
59	THR	PSI	-82.9	-28.2
60	CYSS	CHI1	-65.0	-55.0

60	CYSS	PHI	-256.7	106.3
60	CYSS	PSI	-219.0	143.5
61	GLN	CHI1	-65.0	-55.0
61	GLN	PHI	-103.4	-35.7
61	GLN	PSI	-63.2	-16.2
62	TYR	CHI1	-65.0	-55.0
62	TYR	PHI	-116.8	-76.8
62	TYR	PSI	87.4	127.4
64	VAL	CHI1	-65.0	-55.0
64	VAL	PHI	-149.5	-109.5
64	VAL	PSI	138.5	178.5

### D.3 Residual dipolar couplings for Conk-S2

2	N	2	HN	2.590
4	N	4	HN	-0.910
5	N	5	HN	-2.200
6	N	6	HN	-3.130
9	N	9	HN	3.530
10	N	10	HN	18.360
11	N	11	HN	-9.970
12	N	12	HN	-0.350
14	N	14	HN	-8.430
15	N	15	HN	7.190
16	N	16	HN	1.810
17	N	17	HN	2.830
18	N	18	HN	11.950
19	N	19	HN	6.700
20	N	20	HN	-3.740
21	N	21	HN	-0.010
23	N	23	HN	-9.940
24	N	24	HN	-8.410
25	N	25	HN	3.160
26	N	26	HN	-1.030
27	N	27	HN	-12.130
28	N	28	HN	-19.610
29	N	29	HN	-21.030
30	N	30	HN	-7.220
31	N	31	HN	12.520
32	N	32	HN	0.610
33	N	33	HN	-13.240
34	N	34	HN	-6.020
35	N	35	HN	-7.890
36	N	36	HN	-22.060
37	N	37	HN	-15.850
38	N	38	HN	-0.020
39	N	39	HN	-1.270
41	N	41	HN	-6.940
43	N	43	HN	7.060

44	N	44	HN	3.620
45	N	45	HN	-9.850
46	N	46	HN	-17.510
47	N	47	HN	-11.760
48	N	48	HN	-0.560
49	N	49	HN	11.340
50	N	50	HN	0.950
51	N	51	HN	8.370
52	N	52	HN	-5.400
55	N	55	HN	17.390
56	N	56	HN	-1.500
57	N	57	HN	-0.500
58	N	58	HN	8.200
59	N	59	HN	13.230
60	N	60	HN	-2.170
61	N	61	HN	-0.330
62	N	62	HN	3.170
64	N	64	HN	2.423
65	N	65	HN	-0.146
2	N	1	C	-0.010
4	N	3	C	0.350
5	N	4	C	-0.610
6	N	5	C	0.710
9	N	8	C	0.160
10	N	9	C	0.670
11	N	10	C	-1.930
12	N	11	C	0.530
14	N	13	C	-1.120
15	N	14	C	-0.130
16	N	15	C	2.050
17	N	16	C	-2.370
18	N	17	C	-0.300
19	N	18	C	1.930
20	N	19	C	-0.660
21	N	20	C	2.240
23	N	22	C	2.310
24	N	23	C	-1.640
25	N	24	C	-0.080
26	N	25	C	2.200
27	N	26	C	-0.140
28	N	27	C	0.840
29	N	28	C	-1.130
30	N	29	C	-1.820
31	N	30	C	1.980
32	N	31	C	-0.410
33	N	32	C	-0.840
34	N	33	C	2.140
35	N	34	C	-0.550
36	N	35	C	1.260
37	N	36	C	1.350
38	N	37	C	-0.800
39	N	38	C	1.620
40	N	39	C	-0.020

41	N	40	C	-1.560
43	N	42	C	-1.330
44	N	43	C	0.200
45	N	44	C	-0.390
46	N	45	C	-0.060
47	N	46	C	2.390
48	N	47	C	0.340
49	N	48	C	-0.220
50	N	49	C	-2.570
51	N	50	C	-0.310
52	N	51	C	2.270
55	N	54	C	-0.260
56	N	55	C	1.320
57	N	56	C	-2.530
58	N	57	C	-0.040
59	N	58	C	1.110
60	N	59	C	0.350
61	N	60	C	-2.330
62	N	61	C	0.440
65	N	64	C	0.248
39	CA	39	C	5.288
24	CA	24	C	-3.474
28	CA	28	C	-0.965
26	CA	26	C	0.871
40	CA	40	C	-1.961
27	CA	27	C	-0.23
20	CA	20	C	4.383
34	CA	34	C	1.157
56	CA	56	C	-3.51
15	CA	15	C	-2.628
5	CA	5	C	-1.326
23	CA	23	C	-0.643
11	CA	11	C	-4.836
14	CA	14	C	0.974
29	CA	29	C	-2.34
37	CA	37	C	0.359
4	CA	4	C	0.882
55	CA	55	C	1.342
9	CA	9	C	5.043
45	CA	45	C	-3.58
17	CA	17	C	-4.063
51	CA	51	C	4.442
32	CA	32	C	-2.945
50	CA	50	C	3.137
31	CA	31	C	4.115
36	CA	36	C	0.142
58	CA	58	C	2.675
59	CA	59	C	1.795
10	CA	10	C	-2.097
19	CA	19	C	0.075



# Curriculum Vitae

## Personal data

Name	Jegannath Korukottu
Date of birth	30.03.1980
Place of birth	Madurai, India
Citizenship	Indian

## Education

05/1996	Secondary School Education, Madurai, India
05/1998	Higher Secondary School Education, Madurai, India
06/1998-05/2001	Bachelor's of Science (Bsc.) Physics (spl.), Madurai Kamaraj University, Madurai, India.
08/2001-05/2003	Master's of Science (MSc.) Physics, Indian Institute of Technology Madras (IIT M), Chennai, India, Thesis titled "Quantum version of Bertrand's theorem".
05/2003-08/2003	Visiting student at the Institute of Mathematical Sciences, Chennai, India.
08/2003-11/2003	Visiting student at the Center for Bioinformatics, University of Saarbrücken, Germany Funded by EMBO
11/2003-05/2004	Praktikum in the group of Prof. Dr. Stefan Hell and Dr. Markus Zweckstetter at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany
06/2004-02/2008	Ph.D. thesis with the title "Rapid Determination of High-resolution Protein Structures by Solution and Solid-state NMR spectroscopy" in the department of NMR-based structural biology under the supervision of Prof. Dr. Christian Griesinger and Dr. Markus Zweckstetter at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany

# Lebenslauf

## Persönliche Daten

Name	Jegannath Korukottu
Geburtstag	30.03.1980
Geburtsort	Madurai, Indien
Staatsangehörigkeit	Indisch

## Ausbildung

05/1996	Abschluss der Sekundarstufe I (10. Klasse) in Madurai, Indien
05/1998	Erwerb des hochschulqualifizierenden Abschlusses in Madurai, Indien
06/1998-05/2001	Bachelor's of Science (Bsc.) in Physik an der Madurai Kamaraj Universität in Madurai, Indien.
08/2001-05/2003	Master's of Science (MSc.) in Physik am Indischen Institut für Technologie Madras (IIT M) in Chennai, Indien, Titel der Master-Arbeit: "Quantenmechanische Äquivalenz von Bertrands Theorem".
05/2003-08/2003	Gaststudent am Institut für Mathematik in Chennai, Indien
08/2003-11/2003	Gaststudent am Zentrum für Bioinformatik an der Universität Saarbrücken, unterstützt durch ein EMBO Stipendium
11/2003-05/2004	Praktikum in der Gruppe von Prof. Dr. Stefan Hell und der Gruppe von Dr. Markus Zweckstetter am Max-Planck-Institut für biophysikalische Chemie in Göttingen, Germany
06/2004-02/2008	Doktorarbeit mit dem Titel "Schnelle Bestimmung von hochaufgelösten Strukturen durch Lösungs- und Festkörper-NMR-Spektroskopie" unter Anleitung von Prof. Dr. Christian Griesinger und Dr. Markus Zweckstetter am Max Planck Institut für biophysikalische Chemie in Göttingen, Germany