

## **ABSTRACT**

YIN, YUANYUAN. Calculation of Membrane Protein Structures in Oriented-Sample NMR and Their Validation using Rosetta. (Under the direction of Dr. Alexander A. Nevzorov).

Membrane proteins encode about 30% of all eukaryotic genomes and constitute the majority of current drug targets. However, only 2% of solved protein structures in Protein Data Bank are membrane proteins. To contribute to this outstanding problem, we have developed a computational methodology for calculating membrane protein structures from oriented sample solid-state NMR. We have demonstrated the feasibility of obtaining atomic-resolution three-dimensional backbone structures of membrane proteins solely from the heteronuclear spin-spin dipolar couplings. We have further validated the calculated structures and determined their optimal immersion depth in the heterogeneous membrane-aqueous environment by combining the above structural fitting method and the Rosetta structure predicting package. Extension of this research include elucidation of the oligomerization states of homomeric protein assemblies by simultaneously utilizing solid-state NMR spectroscopy and Rosetta symmetry-docking protocols.

© Copyright 2013 by Yuanyuan Yin

All Rights Reserved

Calculation of Membrane Protein Structures in Oriented-Sample NMR and Their Validation  
using Rosetta

by  
Yuanyuan Yin

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Chemistry

Raleigh, North Carolina

2013

APPROVED BY:

---

Alexander A. Nevzorov  
Chair of Advisory Committee

---

Alex I. Smirnov

---

Jerry L. Whitten

---

Stefan Franzen

## **DEDICATION**

To my parents Weidong Yin and Yumin Zhang.

## BIOGRAPHY

Yuanyuan Yin was born on July 23, 1984, the hottest day of that year, in Tianjin, China. After six years study at Tianjin Shiyan Elementary School, she passed the specialized qualification exam and started her five years middle-high school life at Tianjin Yaohua High School in 1997. Here, she competed but also made lifetime friends with many awesome people.

In 2002, she attended University of Science and Technology of China (USTC) in Hefei, Anhui, China and chose chemical physics as the major. In 2005 summer, she began her undergraduate thesis research in Dr. Guangzhao Zhang's group at National Laboratory for Physical Sciences and Microscale. With the encouragement and support from Dr. Zhang, she obtained a large amount access to both chemical and instrument resources in the laboratory, succeeded in accomplishing her thesis and awarded Outstanding Dissertation. Tianjin Yaohua High School also received a Letter of Thanks from University of Science and Technology of China with affirmation of her performance as an undergraduate student. In August 2006, she enrolled as a graduate student in department of chemistry at North Carolina State University (NCSU) in Raleigh, North Carolina. After working on dendrimer synthesis for two years, she joined Dr. Alexander A. Nevzorov's group and began her research on development of methodology in membrane protein structure determination. With the guidance and advice of Dr. Nevzorov, she passed the preliminary oral qualification in six months. Now four years passed, she is going to defense and complete her Ph.D. degree study. This moment has appeared many times in her dream.

Although her work is about science, she always cannot help thinking about fortune. If she did not write USTC in her application form, if she switched to another major in her second year at USTC, if she picked another graduate school offer..... Maybe her life would be totally different, but if so, how would her life be?

After graduation, she would like to pursue an industrial or an academic career in Houston, Texas, where driving in snow would not be a concern anymore. Though this would not be an easy journey, she will keep her motto *Never Say Never* in mind. Please bless her.☺

## ACKNOWLEDGMENTS

I would like to thank, first and foremost, my advisor Alexander A. Nevzorov for giving me the great opportunity to pursue my doctoral dissertation under his direction. His enthusiasm about science has inspired me and shaped my career as a scientist. Alex has a beautiful mind and a great sense of scientific insight and intuition. He can always find a way to overcome the challenge that seemed to hinder my research. When I encountered obstacles, he was very patient, understanding, and willing to provide guidance when needed. He also gave me a great degree of freedom to try my own ideas and learn new programming languages with which I learned to think and work independently. I am grateful to all these years working in his group and none of my projects could have been accomplished without his advice. I would like to thank colleagues at the Nevzorov group, with whom I shared challenges, experiences, and successes. I would like to thank my dissertation committee members Alex I. Smirnov, Mike H. Whangbo, Jerry L. Whitten and Stefan Franzen for their great advice and suggestions. Without the great resource provided by North Carolina State University, Department of Chemistry and High Performance Computing Center, work in this dissertation would not be possible. I am also grateful to the great friends in NCSU who have made graduate school a memorable and enjoyable journey. Last but definitely not least, I would like to thank my parents, Weidong Yin and Yumin Zhang, for believing in me and loving me for whoever I am. They have always been there, supporting and encouraging me. I would like to thank my boyfriend, Xian Wang, for his constant support, urge and understanding.

## TABLE OF CONTENTS

LIST OF TABLES .....	viii
LIST OF FIGURES .....	ix
Chapter 1 Introduction .....	1
1.1 Introduction to Membrane Proteins .....	1
1.2 Protein Structure and Folding .....	4
1.2.1 Geometry of Polypeptide Chains .....	4
1.2.2 Hydrophobicity of Amino Acid Residues .....	7
1.2.3 Molecular Modeling using Rosetta Software Package .....	7
1.2.4 Rosetta Sampling Algorithm .....	9
1.2.5 Rosetta Scoring Function .....	10
1.2.6 Assessment of Model Quality .....	13
1.3 Solid-State NMR Studies of Membrane Proteins .....	14
1.3.1 Basic Principles of NMR .....	14
1.3.2 Magic Angle Spinning NMR .....	16
1.3.3 Oriented Solid-State NMR .....	17
1.3.4 Structural Fitting using PISEMA .....	20
1.3.5 Analytical Framework for Solid-State NMR Observables in Torsion-Angle Space .....	21
1.4 Prediction of Helical Membrane Proteins Structures using Rosetta .....	26
1.4.1 De Novo Prediction using Membrane Ab Initio in Rosetta .....	26
1.4.2 High-Resolution Prediction using Adapted All-Atom Energy Function .....	28
1.5 Research Summary .....	29
Chapter 2 Structure Determination in “Shiftless” Solid State NMR of Oriented Protein Samples .....	48
2.1 Introduction .....	48
2.2 NMR Observables in the Spherical Basis and the Algorithm .....	52
2.3 Results and Discussion .....	59
2.3.1 Structural Fitting of Simulated Three-Dimensional Spectra with $^{15}\text{N}$ CSA, and $^1\text{H}$ - $^{15}\text{N}$ and $^1\text{H}$ - $^{13}\text{C}_\alpha$ Dipolar Couplings .....	59
2.3.2 Structural Fitting of the Three-Dimensional “Shiftless” Spectra with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ Dipolar Couplings .....	60
2.3.3 Structural Fitting of Four-Dimensional Data with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ , $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ and $^{13}\text{C}'$ - $^{15}\text{N}$ Dipolar Couplings .....	61
2.3.4 Structural Fitting of a Helical Hairpin Derived from Bacteriorhodopsin .....	63
2.4 Conclusions .....	65
Chapter 3 Validation of SSNMR Structures using Rosetta .....	75
3.1 Introduction .....	75
3.2 Application of Membrane Ab Initio Protocol in Rosetta .....	75
3.2.1 Membrane Ab Initio Calculation of Bacteriorhodopsin (1C3W, 1PY6) .....	76

3.2.2 Membrane Ab Initio Calculation of Mercuric Ion Transport Proteins (MerF, MerFt) .....	78
3.2.2.1 Membrane Ab Initio Calculation of MerF (2LJ2).....	79
3.2.2.2 Membrane Ab Initio Calculation of MerFt (2H3O).....	80
3.2.2.3 Comparison of Structure Prediction from Different OCTOPUS Topology Files .....	81
3.2.2.4 Effect of Hydrogen Bonding on Conformation Prediction of Transmembrane Helices .....	83
3.3 Structure Validation by PISEMA Back-Calculation .....	84
3.4 Determination of Transmembrane Helices Embedding Positions in Membrane....	87
3.5 Conclusions.....	91
Chapter 4 Future Directions.....	114
4.1 Homomeric Ion Channels .....	114
4.2 Application of Symmetry Docking to M2 (2L0J).....	115
Chapter 5 Conclusions .....	124
REFERENCES .....	126
APPENDICES .....	141
Appendix A.....	142
Appendix B.....	153
Appendix C.....	188

## LIST OF TABLES

Table 3.1 Modified OCTOPUS topologies for MerFt (residue 27-70). .....	111
Table 3.2 $^{15}\text{N}$ CSA and $^1\text{H}$ - $^{15}\text{N}$ dipolar coupling frequencies of MerFt (residue 27-68)..	112
Table 3.3 Distance constraints used in calculated MerFt structures screening.....	113
Table 4.1 Summary of symmetry docking results for 2L0J.....	121
Table 4.2 Summary of symmetry docking results for helical part of 2L0J. ....	122
Table 4.3 Summary of distances measured in native structure, simulated whole structure and simulated helical part of M2 (2L0J).....	123

## LIST OF FIGURES

Figure 1.1 Schematic representation of the structure types of transmembrane proteins .	30
Figure 1.2 Yearly growth of all protein structures and unique membrane protein structures. .....	31
Figure 1.3 Distribution of all experimental methods used in protein structural studies. .	32
Figure 1.4 Representation of the unit of peptide backbone. ....	33
Figure 1.5 Resonance forms of the peptide group. ....	34
Figure 1.6 <i>Trans</i> and <i>cis</i> isomers of the peptide plane. ....	35
Figure 1.7 Representation of torsion angles $\Phi$ and $\Psi$ , respectively. ....	36
Figure 1.8 Torsion angles $\phi/\psi$ definition. ....	37
Figure 1.9 Experimentally determined hydrophobicity scale for amino acids. ....	38
Figure 1.10 Rosetta general sampling algorithm. ....	39
Figure 1.11 Splitting of nuclei spin states in an external magnetic field $\mathbf{B}_0$ when the angular momentum quantum number $I$ equals $\frac{1}{2}$ (A) and $I$ equals 1 (B). ....	40
Figure 1.12 Nuclear energy levels in an external magnetic field $\mathbf{B}_0$ when the angular momentum quantum number $I$ equals $\frac{1}{2}$ . ....	41
Figure 1.13 Schematic Diagram of magic angle spinning solid state NMR. ....	42
Figure 1.14 Representation of a uniaxially oriented membrane protein sample in lipid bilayer. .....	43
Figure 1.15 PISA wheel patterns as a function of tilt angle. ....	44
Figure 1.16 Description of the molecular frame (MF) associated with the peptide plane.	45

Figure 1.17 Membrane definition in OCTOPUS method. ....	46
Figure 1.18 Definition of membrane layers in Rosetta membrane ab initio protocol. ....	47
Figure 2.1 Local molecular frame associated with a peptide plane. ....	66
Figure 2.2 Normalized Ramachandran plots for different residue types. ....	67
Figure 2.3 A flowchart for the structural fitting algorithm. ....	68
Figure 2.4 Structural fitting of the simulated three-dimensional solid-state NMR spectrum of protein G obtained from PDB coordinates 2GB1. ....	69
Figure 2.5 (A) Two-dimensional projection of the simulated spectra (circles) with the same dimensions as in Fig. 2.4 with $^{15}\text{N}$ CSA input data randomly varied within 1 ppm for each simulation. (B) Back-calculated structures deviate substantially for each simulation. ...	70
Figure 2.6 (A) Structural fitting of simulated three-dimensional “shiftless” solid-state NMR spectra and histogram of the RMSDs. (B) The root-mean square deviations (in Hz) at each residue. ....	71
Figure 2.7 (A) Structural fitting of simulated four-dimensional data for protein G. (B) The root-mean square deviations (in Hz) at each residue for 1000 fits. ....	72
Figure 2.8 (A) and (B): Histograms of the RMSDs of 1000 back-calculated structures obtained from three-dimensional “shiftless” solid state NMR spectra of protein G (2GB1) with a tolerance of 25 Hz (A) and 50 Hz (B). (C) and (D): Histograms of the RMSDs for the structures satisfying the distance restraints between the Ca carbons of residues Ile6 and Thr53, and of residues Glu15 and Thr44 for the 25 Hz tolerance (C) and 50 Hz tolerance (D) .....	73

Figure 2.9 (A) Structural fitting of simulated three-dimensional “shiftless” spectra of two transmembrane $\alpha$ -helices (residues 104 to 155) of bacteriorhodopsin. (B) The mean of the deviation (in Hz) at each residue for 100 calculated structures. (C) When the initial structure is rotated by about 90 degrees, with the tolerance of 25Hz, about 60% of the RMSDs for the back-calculated structures are more than 5Å. (D) The rms deviation (in Hz) at each residue for 100 calculated structures. ....	74
Figure 3.1 Structure prediction. (A) 1C3WD (B) 1C3WE (C) 1PY6.....	93
Figure 3.2 The products of the mer operon(s). ....	94
Figure 3.3 The membrane ab initio energy versus RMSD for MerF.....	95
Figure 3.4 The membrane ab initio energies calculated with original OCTOPUS topology versus RMSD for MerFt. ....	96
Figure 3.5 The membrane ab initio energies calculated with modified OCOTPUS topology versus RMSD for MerFt. ....	97
Figure 3.6 (Left) The membrane ab initio energy calculated with a series of modified OCOTPUS predicted topology versus RMSD plot for MerFt. (Right) The lowest energy model predicted with each modified OCTOPUS topology .....	98
Figure 3.7 Tendency of RMSDs of MerFt structures predicted by membrane ab initio protocol in Rosetta3.4 along the change of the weight for the hydrogen bonding potential.	101
Figure 3.8 Schematic diagram of membrane rotation.....	102
Figure 3.9 Comparison of experimental data from solid state NMR experiments and back-calculated frequencies for structures predicted by membrane ab initio protocol in Rosetta.	103

Figure 3.10 Schematic diagram for translation of an NMR-derived structure across lipid bilayer along the orientation of membrane normal.....	104
Figure 3.11 Change in all-atom energy of 5-helix 1C3WD along z-axis.....	105
Figure 3.12 Change in all-atom energy of 2-helix 1C3WE along z-axis.....	106
Figure 3.13 Change in all-atom energy of bacteriophage Pf1 (2KSJ) along z-axis. ....	107
Figure 3.14 Change in all-atom energy of calculated MerFt along z-axis.....	108
Figure 3.15 Comparison between experimental data and back-calculated frequencies for <sup>15</sup> N CSA and <sup>1</sup> H- <sup>15</sup> N dipolar coupling.....	109
Figure 3.16 Summary of MolProbity results for MerFt (2H3O) (top), MerF (2LJ2) (middle), and the partial structure in 2LJ2 with the same sequence as 2H3O (bottom).....	110
Figure 4.1 Symmetry docking results for 2L0J. ....	117
Figure 4.2 Symmetry docking results for helical part of 2L0J. ....	118
Figure 4.3 Illustration of h1A30_h2I35 and h1I35_h2A30 distances measured in PyMOL.	119
Figure 4.4 (A) and (B): Histogram of distances between Ala30 on helix 1 and Ile35 on helix 2. (C) and (D): Histogram of distances between Ile35 on helix 1 and Ala30 on helix 2.	120
Figure A.1 Effect of the variations in the principal components of the <sup>15</sup> N CSA tensor on the observed CSA values for protein G. ....	147
Figure A.2 Effect of the orientation of the <sup>15</sup> N CSA tensor on the observed CSA values for protein G. ....	148
Figure A.3 Combined effect of the variations in both the principal components of the <sup>15</sup> N CSA tensor and its orientation on the individual observed CSA values for protein G....	149

Figure A.4 Histograms of the RMSDs of 1000 back-calculated structures by simulating four-dimensional “shiftless” solid state NMR spectra of protein G (2GB1) with the tolerance of (A) 25 Hz (B) 50 Hz. .... 150

Figure A.5 Histograms of the deviations of the back-calculated distances from the “true” distances (from 2GB1) corresponding to the C $\alpha$  atoms of residues Ile6 and Thr53, and residues Glu15 and Thr44. (A) 25 Hz tolerance, Ile6-Thr53 distances; (B) 25 Hz tolerance, Glu15-Thr44 distances; (C) 50 Hz tolerance, Ile6-Thr53 distances; (D) 50 Hz tolerance, Glu15-Thr44 distances..... 151

Figure A.6 Histograms of the RMSDs for the back-calculated structures obtained from fitting three-dimensional “shiftless” spectra for a helical hairpin of bacteriorhodopsin (1C3W) with the tolerances of (A) 200 Hz (B) 250 Hz (C) 300 Hz (D) 350 Hz..... 152

# Chapter 1

## Introduction

### 1.1 Introduction to Membrane Protein

Membrane proteins are associated with biological membranes and play a significant role in many dynamic processes in cells, such as substance transportation, signal transduction, cell recognition, cell-mediated immunity, neurotransmission, and metabolism modulation. They can be divided into two main categories according to their topology: peripheral membrane proteins and integral membrane proteins. Peripheral membrane proteins cling on the biological membrane temporarily, while integral membrane proteins attach to the biological membrane permanently. Transmembrane proteins are the most common type of integral membrane proteins and span the entire biological membrane, connecting the two bilayer leaflets. (Markus Sällman Almén 2009) Only two kinds of structural types have yet been observed in transmembrane proteins:  $\alpha$ -helices in majority (single  $\alpha$ -helix and bundles of  $\alpha$ -helices), mostly present in the inner membrane of bacterial cells and the plasma membrane of eukaryotes, and  $\beta$ -barrel in minority, present in outer membranes of Gram-negative bacteria, cell wall of Gram-positive bacteria, and outer membranes of mitochondria and chloroplasts. (HEIJNE 1998; Wimley 2003) (Figure 1.1)

Statistically, about 30% of all proteins in all organisms are membrane proteins. The majority of prescription pharmaceuticals on the market also target membrane proteins due to their physiological importance, in particular ion channels and G-protein coupled receptors

(GPCRs), like Eli Lilly's Zyprexa, Schering-Plough's Clarinex, GlaxoSmithKline's Zantac, and Novartis's Zelnorm. (Filmore 2004) In 2012, Nobel Prize in Chemistry was awarded to Brian K. Kobilka and Robert J. Lefkowitz for studies of GPCRs, indicating the increasing worldwide appreciation of the studies on membrane proteins.

In the past decade, the number of solved protein structures has increased by about seven fold with the improvement in both experimental instruments and methodologies, but only 2% of protein structures in protein data bank are membrane proteins. There are about 1000 distinct folds for integral membrane proteins, but only about 10% have been determined. (Figure 1.2) Hence, there is yet a large unknown world of membrane protein to explore in order to obtain a more profound insight into functions and mechanisms of human organelles. (Yalini Arinaminpathy 2009)

The technical challenges in elucidating high-resolution membrane protein structures come from three main aspects. First, though some membrane proteins can be expressed in large quantities in specific environments, like rhodopsin in the retina (Palczewski 2006) Ca-ATPase in muscle sarcoplasmic reticulum (Toyoshima C 2000), and bacteriorhodopsin in *Escherichia coli* (Kazumi Shimono 2009), the really interesting mammalian membrane proteins involved in intercellular communication and regulation of transmembrane ion and metabolite concentrations, exist in extremely low levels and usually in multiple isoforms in human cells, therefore are difficult to over-express in sufficient quantities for in-vitro studies. (Grisshammer R. 1995; Grisshammer 1996; Tate 2001) Secondly, it is difficult to extract membrane proteins from their native cellular membranes with an appropriate detergent and purify them in their functional form for full characterization. Thirdly, the crystallization of

membrane proteins can be destroyed by the aggregation of detergent-solubilized protein, and crystals often do not diffract well due to high solvent contents. (Caffrey 2003; Guidotti 2009) X-ray crystallography and solution nuclear magnetic resonance spectroscopy are the most widely used experimental methods in probing atomic resolution three-dimensional structures of biological macromolecules. (Figure 1.3)

X-ray crystallography determines the arrangement of atoms in a protein from a three-dimensional electron density map obtained by scattering X-ray beams onto the crystal sample of the protein, so the prerequisite in this methodology is to prepare a high-quality single crystal sample. Due to the challenges mentioned above, though nearly 50000 crystal structures of biological molecules have been determined by X-ray crystallography, very few of them are membrane proteins.

Solution NMR spectroscopy has acted as a powerful tool in exploring the structure of organic molecules for many decades before scientists were capable of resolving the structure of proteins in solution by the well-established  $^1\text{H}$  NMR spectroscopy around 1985. (Bax 1994) Sharp resonance bands and intense resonance signals usually predicate a high resolution of the NMR spectrum. The water-soluble proteins tumble very rapidly in the aqueous solution and, therefore, yield short correlation time, narrow resonance bands and good intensity. As for the membrane proteins, the hydrophobicity caused by the transmembrane regions spanning the phospholipid bilayer leads to the demand for lipids or detergents to get solubilized as well. The resulting protein-detergent complexes have a larger size than the protein itself, which slows down the tumbling in the solution, prolongs the correlation time, and thus broadens the line-widths and attenuates the signal intensity, becoming the main

limitation in resolving the structure of membrane proteins by solution NMR spectroscopy. Only a small number of membrane protein structures with a relative molecular mass of less than 35kDa solubilized in micelles have been determined by solution NMR. (Almeida FC 1997; Kevin R. MacKenzie 1997; Yu L 2005; Schnell JR 2008) Due to the above difficulties and limitations in X-ray crystallography and solution NMR spectroscopy, solid-state NMR tools have been developed as an alternative to study membrane proteins.

In solid state NMR of oriented samples, membrane protein samples are prepared by either mechanically aligned on glass plates or magnetically aligned in large bicelles or lipid bilayers. In principle, the bilayers stacked between glass plates should be the most suitable substrate for investigating the structures and functions of the membrane proteins because of its similarity to the biological membrane and, thus, the capability of mimicking the real environment for the membrane proteins. (De Angelis and Opella 2007) The membrane protein samples are immobilized via the interaction with the surrounding lipids, resulting in a much longer correlation time comparing to the timescale of experiments. Radio frequency, magic angle spinning and sample alignment rather than tumbling act as the mechanism of line narrowing. (Jaume Torres 2003; Marassi 2004)

## **1.2 Protein Structure and Folding**

### **1.2.1 Geometry of Polypeptide Chains**

Proteins are long chains of amino acids connected to each other by peptide bonds. The basic unit is formed by three atoms: the amide nitrogen (N), the alpha carbon ( $C_\alpha$ ), and the carbonyl carbon ( $C'$ ). (Figure 1.4) The unit starts from the amide nitrogen, and ends at the

carbonyl carbon, labelled as  $N_i$ ,  $C_{\alpha i}$ , and  $C'_i$ , respectively, where  $i$  is the number of the residue.

Usually, the peptide group  $-C(=O)NH-$  has two resonance forms, one of which has a double bond between carbonyl carbon atom and amide nitrogen atom. (Figure 1.5) This partial (~40%) double bond character renders the peptide group planar, ensures the rigidity of the peptide plane and consequently the rigidity of the whole polypeptide molecule.

The partial double bond character also leads to the existence of *trans* and *cis* isomers. (Figure 1.6) In the folded polypeptide, the *trans* isomer is always sterically and energetically favorable, so  $C_{\alpha i}$  and  $C_{\alpha i+1}$  are mostly in *trans* position, with the dihedral angle  $\omega$  defined by  $C_{\alpha i}$ ,  $C'_i$ ,  $N_{i+1}$ , and  $C_{\alpha i+1}$ , four atoms in the same peptide plane, equal to  $180^\circ$ . The *cis* isomer, where the dihedral angle  $\omega$  is equal to  $0^\circ$ , rarely appears except in the case of proline. (Creighton 1984)

The rigidity of the peptide plane results in only two bonds,  $N_i-C_{\alpha i}$  and  $C_{\alpha i}-C'_i$ , rotating freely.  $\Phi$ , the torsion angle along  $N_i-C_{\alpha i}$  bond axis, is the dihedral angle defined by  $C'_{i-1}$ ,  $N_i$ ,  $C_{\alpha i}$  and  $C'_i$ .  $\Psi$ , the torsion angle along  $C_{\alpha i}-C'_i$  bond axis, is the dihedral angle defined by  $N_i$ ,  $C_{\alpha i}$ ,  $C'_i$  and  $N_{i+1}$ . (Figure 1.7) Hence, if we assume that the protein is a biological macromolecule consists of a number of identical peptide planes, we can obtain the secondary and tertiary backbone structure of the protein as long as the torsion angles are determined.

The common secondary-structural types, like  $\alpha$ -helices,  $\beta$ -strands, and  $\gamma$ -turns have their own characteristic combination of torsion angles  $\phi$  and  $\psi$  (Lovell, Davis et al. 2003), which can be visualized respectively in a Ramachandran plot. Ramachandran plot was developed early in 1963 by G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan by computing hard

sphere models. (G.N. Ramachandran 1963) In their experiments, atoms were mimicked by hard spheres with their van der Waals radii, and torsion angles  $\phi$  and  $\psi$  were varied systematically to find stable conformations, in which no collision among atoms occurred. Figure 1.8 is a classic Ramachandran plot defining torsion angles. (Lovell, Davis et al. 2003) The conformations with  $\phi$  and  $\psi$  falling into dark gray regions, i.e.  $\beta$ -sheet and right handed  $\alpha$ -helix, have no steric conflict, so they are strictly allowed. If slightly shorter van der Waals radii were used in the calculation,  $\phi$  and  $\psi$  in medium gray regions are allowed as well. At the same time, the left-handed  $\alpha$ -helix region showed up. White regions are sterically disallowed for almost all amino acids, because with  $\phi$  and  $\psi$  in this part the distance between atoms is smaller than the sum of their van der Waals radii, atoms collide with each other, resulting in unstable conformations. However, glycine has no side chains, so it becomes the only amino acid structure allowed in white regions. (Figure 1.8) With the increasing availability of high-resolution protein structures obtained in the past few decades, refined Ramachandran plots can be used to estimate the backbone conformation and evaluate the quality of the structure refinement simultaneously.

Specifically, the protein helices are mainly categorized into alpha-helices,  $3_{10}$  helices and  $\pi$ -helices based on the number of residues and atoms per turn. Alpha helices have 3.6 residues and 13 atoms per turn,  $3_{10}$  helices have 3.2 residues and 12 atoms per turn, and  $\pi$ -helices have 4.4 residues and 16 atoms per turn. Torsion angles phi and psi are accordingly different in these three cases. We need to modify the values of torsion angles in structural fitting according to different objects. For alpha-helices,  $\phi = -65^\circ$ ,  $\psi = -40^\circ$ ; for  $3_{10}$  helices,  $\phi = -68^\circ$ ,  $\psi = -17^\circ$ ; for  $\pi$ -helices,  $\phi = -57^\circ$ ,  $\psi = -70^\circ$ . (Sanguk Kim 2004)

### **1.2.2 Hydrophobicity of Amino Acid Residues**

Hydrophobicity is very useful in prediction and evaluation of transmembrane helices region of membrane proteins as well. The more positive the hydrophobicity scale, the more likely are the amino acids to reside in the hydrophobic core of biological membrane.(Jack Kyte 1982; White 1996; Tara Hessa 2005) (Figure 1.9) kdHydrophobicity, wwHydrophobicity and hhHydrophobicity are the hydrophobicity scales most commonly assigned to amino acid residues. Among them, wwHydrophobicity, i.e. the Wimley-White whole residue hydrophobicity scale, is based on transfer free energies of polypeptides directly determined in experiments and considers not only the backbone peptide bonds but also the contribution from side chain. It is significant in determination of the positions of amino acid residues across the membrane.

### **1.2.3 Molecular Modeling using Rosetta Software Package**

To get an idea about the complexity of protein folding, let us consider the backbone structure by sequentially connecting the amino acid residues and looking for all possible torsion angles associated with each peptide plane, assuming the value of torsion angles  $\Phi$  and  $\Psi$  are chosen from one of the three most stable regions, i.e. the dark grey regions shown in Figure1.8. For a small polypeptide containing only 50 amino acid residues and thus 49 peptide bonds and 98 torsion angles, in order to calculate all of its possible structures it would take  $3^{98}$  multiples of the time steps that are needed to find one torsion angle conformation. Therefore, the total time to find the native structure could be even longer than the age of the universe, as originally described by Levinthal's paradox in 1969. (E.Munck 1969)

Molecular dynamics is commonly used in the study of protein folding using force fields to simulate the physical movements and the interactions between atoms by solving Newton's equation of motion. However, in order to simulate the real dynamic processes occurred in the object system, molecular dynamics modeling needs a large amount of computational time even with the latest CPU. Some well-known packages, such as CHARMM (Bernard R. Brooks 1983; Alexander D. MacKerell Jr. 2000; B. R. Brooks 2009), AMBER (Jay W. Ponder 2003; David A. Case 2005), GROMACS (H.J.C. Berendsen 1995; David Van Der Spoel 2005) have been well established and widely applied in research of both soluble proteins and membrane proteins.

Early in 1961, Anfinsen (C. B. Anfinsen 1961) and coworkers had demonstrated that the three-dimensional protein structures are completely determined by their amino acid sequences. Starting from development of *de novo* prediction methodology, Rosetta has been expanding rapidly since its first launch in May of 2004. Besides the initial *de novo* folding protocol, comparative modeling, protein-protein docking, protein-ligand docking, loop modeling, protein design and enzyme design protocols etc. have been developed in the past few years. (Baker 2000; Brian Kuhlman 2003; Jeffrey J. Gray 2003; Daniel J Mandell 2009; James Thompson 2011) Applications such as RosettaNMR (Yang Shen 2008; Yang Shen 2009) and RosettaEPR (Stephanie J. Hirst 2011) are also developed to build protein models from sparse experimental data. In RosettaNMR, local distance restraints are derived from NOEs and long range angular restraints are obtained from residual dipolar couplings.

Combining Monte Carlo structural perturbation with energy minimization, thousands of local minima are sampled to find a global minimum. Basically, the protocols in Rosetta can be divided into two steps: sampling and scoring.

#### **1.2.4 Rosetta Sampling Algorithm**

Noticing the fact that the folding of local segments of a polypeptide is not dependent on the folding of the whole protein, the Baker group has developed a molecular modeling package, Rosetta, to predict the conformation of biological macromolecules based on the Bayesian statistical analysis of available conformations of sequence segments in known protein structures available in protein data bank. (Kim T. Simons 1997; Bonneau, Tsai et al. 2001; Richard Bonneau 2002) The protein folding starts from an extended polypeptide chain. The 3-residue and 9-residue fragments with similar sequences found in the fragment library are inserted into the protein backbone to build simulated models. Usually about 30000 9-residues fragments and 10000 3-residue fragments are searched and combined in each model prediction. (Carol A. Rohl 2004) Then the low-resolution energies for all the predicted models are calculated to produce a broad range of energy minimum, and with Metropolis Monte Carlo simulated annealing optimization algorithm the global energy minimum can be determined. In the optimization procedure, the initial conformation is selected randomly, then a local perturbation is conducted by modifying torsion angles or replacing a rotamer at side chains. If the energy of the new conformation is lower than that of the old model, the new conformation will be accepted. Otherwise, the value of  $e^{-(E_{new}-E_{old})/\kappa T}$  will be calculated and compared to a random probability P, where  $0 \leq P \leq 1$ : if  $e^{-(E_{new}-E_{old})/\kappa T} > P$ , the new conformation will still be accepted, and will act as the starting point of the next searching

step; if  $e^{-(E_{new}-E_{old})/\kappa T} < P$ , the old conformation will be kept, and a new perturbation will be conducted. Consequently, the searching will be able to hop the local energy minima and achieve the global energy minimum. Experimental data and constraints derived from NMR and EPR can also assist in improving the resolution of predicted protein structure. (Philip Bradley 2005; Kira M. S. Misura 2006; Bin Qian 2007)

### **1.2.5 Rosetta Scoring Function**

Rosetta scoring functions, i.e. energy function, can be categorized into two major classes: low resolution and high resolution. The former utilizes reduced atom representation and describes side-chain with “super atoms”. This centroid-mode scoring function is mainly used for de novo folding protocols. The high-resolution scoring function describes side-chain with rotamers developed by Dunbrack et al (Roland L. Dunbrack Jr 1993; Karplus 1994; Roland L. Dunbrack Jr. 1997), i.e. all-atom representation, and uses the sum of a series of weighted energy terms as the total energy for the model. In the present project, low resolution centroid-mode scoring function is used in de novo prediction of membrane proteins, and high resolution all-atom scoring function is used in validation of membrane protein structures calculated from solid state NMR experimental data.

The energy terms are physical potentials calculated by Newton’s law, van der Waals forces as well as Coulomb’s law, and knowledge based potentials obtained statistically from the physical characteristics of experimental protein structures available in PDB (Frances C. Bernstein 1977; H. M. Berman 2002; Gautam Dantas 2003), including 6-12 Lennard-Jones potential, solvation potential, electrostatic interactions, side chain conformation-dependent

term, torsion angle preference term and an energy term based on the probability of a certain amino acid at a given pair of phi/psi angles. (Baker 2000; Brian Kuhlman 2003)

The 6-12 Lennard-Jones potential is used to describe van der Waals interactions and expressed by an attractive energy term  $E_{ljatr}$  and a repulsive energy term  $E_{ljrep}$ . (Eyal Neria 1996; Baker 2000)

$$E_{ljatr} = \sum_i^{natom} \sum_{j>i}^{natom} \left[ \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}}{d_{ij}} \right)^6 \right] e_{ij} \quad \text{if} \quad \frac{r_{ij}}{d_{ij}} < 1.12 \quad (1.1)$$

$$E_{ljrep} = \sum_i^{natom} \sum_{j>i}^{natom} \left[ 10.0 - 11.2 \left( \frac{d_{ij}}{r_{ij}} \right) \right] \quad \text{if} \quad \frac{r_{ij}}{d_{ij}} > 1.12 \quad (1.2)$$

where i and j are indices of different atoms,  $d_{ij}$  is the distance between the two atoms,  $r_{ij}$  is the sum of the van der Waals radii, and  $e_{ij}$  is the square root of the product of the well depths by adapting the CHARMM19 parameter set via using 5% shorter van der Waals radii. When  $d_{ij}$  equals  $r_{ij}$ , the value of the attractive energy is the sum of  $-e_{ij}$ . When the distance between two atoms is 0 Å, the repulsive energy reaches its maximum 10.0 kcal/mol.

The solvation potential  $E_{solv}$  is derived from semi-empirical implicit solvent model parameterized with experimental data, which was initially promoted by Lazaridis and Karplus. (Themis Lazaridis 1999) Burial of polar atoms is also penalized as part of the solvation potential.  $E_{solv}$  penalizes surface exposure of hydrophobic residues and favors exposure of hydrophilic residues.

$$E_{solv} = \sum_i^{natom} \left( \Delta G_i^{ref} - \sum_{j>i}^{natom} \left\{ \frac{2\Delta G_i^{free}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-d_{ij}^2) \mathcal{V}_j + \frac{2\Delta G_j^{free}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-d_{ij}^2) \mathcal{V}_i \right\} \right) \quad (1.3)$$

$d_{ij}$  and  $r_{ij}$  have the same definition as mentioned above.  $\Delta G^{ref}$  is the solvation free energy of the fully solvated atom.  $\lambda_{ij}$  is a correlation length, and  $V$  is the atomic volume.

The electrostatic interactions are demonstrated by residue pair potential  $E_{pair}$  (Kim T. Simons 1999), and hydrogen bonding potential  $E_{hbond}$ . The residue pair potential is knowledge based and obtained from the probability of a pair of polar residues found in the vicinity of each other in PDB database.

$$E_{pair} = \sum_i^{nres} \sum_{j>i}^{nres} \frac{P(aa_i, aa_j | d_{ij}, env_i, env_j)}{P(aa_i | d_{ij}, env_i)P(aa_j | d_{ij}, env_j)} \quad (1.4)$$

In Eq. (1.4),  $i$  and  $j$  are indices for different amino acid residues,  $P(aa_i | d_{ij}, env_i)$  is the probability of finding amino acid type  $aa_i$  within the distance  $d_{ij}$  in local environment type  $i$ , and  $P(aa_j | d_{ij}, env_j)$  is the probability of finding amino acid type  $aa_j$  within the distance  $d_{ij}$  in local environment type  $j$ .

The hydrogen bonding potential is orientation-dependent and consists of long-range and short-range items, i.e.  $hbond\_lr\_bb$ ,  $hbond\_sr\_bb$ ,  $hbond\_bb\_sc$  and  $hbond\_sc$  (D Benjamin Gordon 1999; William J. Wedemeyer 2003).

$$E_{hbond} = \left[ 5 \left( \frac{r_{ij}}{d_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}}{d_{ij}} \right)^{10} \right] F(q) \quad (1.5)$$

$d_{ij}$  is the distance between acceptor and donor of the hydrogen bond, and  $r_{ij}$  is the optimal hydrogen bonding distance.  $F(q)$  is the angular dependence of the hydrogen bond calculated from vector-angle potential, vector-tensor potential and vector-displacement potentials developed by Wedemeyer et.al. (William J. Wedemeyer 2003)

Rotamer self-energy  $E_{rot}$  is calculated from the probability of finding a certain rotamer at a given pair of phi/psi torsion angles in protein data bank, and is based on Dunbrack rotamer library. (Roland L. Dunbrack Jr 1993; Karplus 1994; Roland L. Dunbrack Jr. 1997)

$$E_{rot} = \sum_i^{residue} -\ln(prob(rot(i) | phi(i), psi(i))) \quad (1.6)$$

The torsion potential  $E_{rama}$  is based on the Ramachandran torsion angle preferences and is calculated from the probability of finding a certain amino acid in a secondary protein structure at a given pair of phi/psi torsion angles in protein data bank. (G.N. Ramachandran 1963; William J. Wedemeyer 2003; Carol A. Rohl 2004)

$$E_{rama} = \sum_i^{residue} -\ln[prob(phi(i), psi(i) | aa_i, ss_i)] \quad (1.7)$$

, where  $aa$  = amino acid type and  $ss$  = secondary structure type.

### 1.2.6 Assessment of Model Quality

The accuracy of predicted models are evaluated mainly by three factors, i.e. RMSD, GDTMM and MaxSub. RMSD is the root mean square deviation of the simulated structure to the native structure if a crystal structure is available. Otherwise, the lowest energy structure is usually used as the baseline to calculate RMSD. GDTMM (James Thompson 2011) measures the superposition between the simulated structures and the native structure, and with a given distance cutoff, the closer is GDTMM to 1.0, the more similar is the predicted model to the native structure. MaxSub (Naomi Siew 2000) describes the largest subset of  $C_\alpha$  atoms of a simulated structure that superimpose well over the native structure with a given RMSD threshold.

## **1.3 Solid-State NMR Studies of Membrane Proteins**

### **1.3.1 Basic Principles of NMR**

In NMR spectroscopy, Nuclear Overhauser Effect (NOE), chemical shifts, J-coupling, dipole-dipole coupling and quadrupolar couplings are the nuclear spin interactions normally utilized for structure determination. NOE (Overhauser 1953; Kaiser 1963) describes the magnetization transfer between dipole-dipole coupled nuclear spins and mainly provides information about distance between these nuclei pairs. J-couplings (Maxwell 1952) describe the inter-nuclear spin interaction through chemical bonds and provide both distance and angular information. Both NOE's and J-couplings are usually probed in solution NMR spectroscopy. By contrast, chemical shift anisotropy, dipolar couplings and quadrupolar couplings are all anisotropic in solid samples, and, therefore, represent a useful source of information for membrane protein structure determination from solid-state NMR spectroscopy. Chemical shift anisotropy demonstrates the orientation dependence of the chemical shift and arises from that fact that the charge distribution on a nuclei is not spherically symmetric. The orientation of the atoms affect the orientation of the electron cloud over the nuclei, and thus affect the chemical shielding. Dipolar couplings, different from J-couplings which describe through-space dipole-dipole interaction, indicate direct interactions between dipole-dipole coupled nuclear spins. Distances and bond orientations between the nuclei can be obtained directly from NMR spectroscopic data. Quadrupolar couplings (Lucio Frydman 1995) appear in nuclei with a spin larger than  $\frac{1}{2}$  and depend on the electric field gradient. Both dynamic and orientational information can be obtained in solid-state NMR by analyzing these spectroscopic observables.

In most NMR experiments, backbone atoms (amide nitrogen, alpha carbon and carbonyl carbon) in polypeptide chains are isotopically labeled for frequency measurement.

Nuclear magnetism is originated from nuclear spin angular momentum,  $I$ , which is an intrinsic property of nucleus. When  $I=0$ , usually found in atoms with both even mass number and even atomic number like  $^{12}_6\text{C}$ , the nuclei have no magnetic moment and no NMR signal can be detected. Both  $^{15}\text{N}$  and  $^{13}\text{C}$  used to label a protein sample in solid state NMR experiments have a nuclear spin with  $I$  equal to  $\frac{1}{2}$ . Quadrupolar couplings can be measured in nuclei with  $I > \frac{1}{2}$  like  $^{14}\text{N}$ .

The magnetic moment  $\mu$  can be expressed by the product of nuclear spin angular momentum  $S$  and gyromagnetic ratio  $\gamma$ , i.e.  $\mu = \gamma S$ . If we represent the spin angular momentum with  $S$ , then the square of its observable value is given by:

$$S^2 = I(I+1)\hbar^2 \quad (1.8)$$

where  $I$  stands for the angular momentum quantum number and  $\hbar$  is the reduced Planck's constant.

For a static magnetic field  $\mathbf{B}_0$ , its orientation is defined along the z-axis, and the projection of the angular momentum  $S$  on the z axis can be described by  $S_z = m\hbar$ , where  $m$  is the magnetic quantum number, and  $m$  ranges from  $I, I-1, \dots$  to  $-(I-1), -I$ . Then the z-component of the magnetic moment  $\mu_z = \gamma m\hbar$ . Therefore, in the magnetic field  $\mathbf{B}_0$ , the energy of a nuclear spin state can be calculated using  $E = -\mu_z B_0 = -\gamma m\hbar B_0$ , and the difference of energy between two nuclear spin states with spin  $\frac{1}{2}$  as shown in Figure 1.11 satisfies  $\Delta E = \gamma\hbar B_0$ . Similarly, for a

nucleus with spin  $I=1$ , the energy difference is  $\Delta E=2\gamma\hbar B_0$ . The angular momentum operators are also used in the calculation of Wigner rotation matrix in structural fitting. (see Eq. (2.4)) The ratio of the populations of two spin states,  $N_{(-1/2)}$  and  $N_{(+1/2)}$  at temperature  $T$  is related to the Boltzmann constant  $\kappa$  and the energy difference between the two states.

$$N_{(-1/2)} / N_{(+1/2)} = N_- / N_+ = e^{-\Delta E/\kappa T} = e^{-h\nu/\kappa T} \quad (1.9)$$

By absorbing energy from radio frequency,  $N_-$  will increase and  $N_+$  will decrease, thus the ratio  $N_- / N_+$  will increase. Based on this phenomenon, the sensitivity of sparse nuclear spins in NMR experiments, such as  $^{15}\text{N}$  and  $^{13}\text{C}$ , can be increased via their interaction with a collection of abundant spins like  $^1\text{H}$ . The abundant spin system is prepared under an assigned low temperature, then the sparse spin system is brought into the cold system of the abundant spins via thermal contact. As the energy is transferred from the sparse spin system to the abundant spin system, the population between the higher energy and lower energy spin states will increase. Therefore, the sensitivity of the NMR experiments can be improved.

The resonant condition for absorbing radio frequency radiation is given by:  $E=h\nu$ ,  $\nu=\gamma B_0/2\pi$ . Consequently, the resonance frequencies observed for different nuclei are proportional to their corresponding gyromagnetic ratio.

### 1.3.2 Magic Angle Spinning NMR

In magic angle spinning (MAS) solid-state NMR, the orientations of the protein molecules immobilized in randomly oriented lipid bilayers with respect to the external magnetic field is averaged out by the sample rotation about  $54.74^\circ$ , i.e. the magic angle. By mechanically spinning membrane protein samples at the magic angle at a high frequency of 1-70 kHz, nuclear dipole-dipole interactions are averaged out almost to zero, and the chemical shift

anisotropy is averaged to the isotropic component of the CSA tensor, and the quadrupolar coupling is partially averaged to a residual second moment. This results in a high resolution spectrum with narrower lines containing only isotropic shifts and J couplings, such as those obtained from solution NMR, thus providing distance and torsion angle information. Figure 1.13 shows a schematic diagram of a typical MAS experiment. (E. R. ANDREW 1958; Lowe 1959; Andrew 1981; Jaume Torres 2003)

### **1.3.3 Oriented Solid-State NMR**

In oriented solid state NMR, bicelles and lipid bilayers, which encompass the membrane protein samples and serve as a native-like environment (Isabelle Marcotte 2005), are magnetically aligned either perpendicular or parallel (with the addition of lanthanide ions) to the external magnetic field. (Figure 1.14) Orientation-dependent chemical shift anisotropy and heteronuclear spin-spin coupling are the conformational constraints used to predict membrane protein structures. Many experimental methods, such as MREV8 (W-K. Rhim 1973), TMREV (M. Hohwy 2000), PISEMA (Wu, Ramamoorthy et al. 1994) and SAMMY (Alexander A Nevzorov 2003), have been developed to measure the dipolar nuclear spin interactions since the establishment of solid-state NMR spectroscopy. These pulse sequences belong to the so-called class of high-resolution separated-local field experiments. For instance, PISEMA (Polarization Inversion Spin Exchange at Magic Angle) is a useful tool to analyze the conformations of helical membrane proteins, while recent publications demonstrated that  $\beta$ -barrel fractions in membrane proteins can be resolved as well. (Ana Carolina Zeri 2003; Nevzorov and Opella 2003; David S. Thiriot 2004; De Angelis, Howell et al. 2006; Sang Ho Park 2010; Bibhuti B. Das 2012)

PISEMA correlates  $^{15}\text{N}$  chemical shift anisotropy and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling for uniformly or selectively  $^{15}\text{N}$  labeled proteins highly aligned in the magnetic field. Mathematically, the observables in a PISEMA experiment can be described as

$$\pi\left(\vec{\mathbf{B}}_0\right) = (\sigma, \nu) \quad (1.10)$$

, where  $\sigma$  is chemical shift tensor and  $\nu$  is dipolar splitting tensor. (Denny, Wang et al. 2001)

The chemical shift tensor  $\sigma$  is asymmetric, and its principal axis frame (**PAF**) is represented

$$\text{as } \mathbf{PAF} = \left( \vec{\sigma}_{11}, \vec{\sigma}_{22}, \vec{\sigma}_{33} \right). \sigma_{11}, \sigma_{22}, \sigma_{33} \text{ are the corresponding principal values and satisfy } \sigma_{11} \leq$$

$\sigma_{22} \leq \sigma_{33}$ . The value of chemical shift tensor is given by

$$\sigma = \sigma_{11} \left( \vec{\mathbf{B}}_0 \cdot \vec{\sigma}_{11} \right)^2 + \sigma_{22} \left( \vec{\mathbf{B}}_0 \cdot \vec{\sigma}_{22} \right)^2 + \sigma_{33} \left( \vec{\mathbf{B}}_0 \cdot \vec{\sigma}_{33} \right)^2 \quad (1.11).$$

The dipolar splitting tensor  $\nu$  is traceless and axially symmetric with a unique rotation axis along the direction of covalent bond NH and can be represented as

$$\nu = \frac{\nu_{\parallel}}{2} \left( 3 \left( \vec{\mathbf{B}}_0 \cdot \vec{\mathbf{u}} \right)^2 - 1 \right) \quad (1.12)$$

, where  $\nu_{\parallel}$  is the value of dipolar splitting tensor when  $\vec{\mathbf{B}}_0 = \vec{\mathbf{u}}$ . If the coordinates of unit

vector of  $\mathbf{B}_0$  in **PAF** are (x, y, z), tensor  $\sigma$  and  $\nu$  can be expressed by

$$\sigma = \sigma_{11}x^2 + \sigma_{22}y^2 + \sigma_{33}z^2 \quad (1.13)$$

and

$$\nu = \frac{\nu_{\parallel}}{2} \left( 3(\cos\alpha\sin\beta x + \sin\alpha\sin\beta y + \cos\beta z)^2 - 1 \right) \quad (1.14)$$

The periodicity in the orientation of each peptide plane in an ideal helix is reflected into a characteristic PISEMA pattern, i.e. the PISA (Polar Index Slant Angle) wheel. By analyzing the PISA wheels, one can assign the resonances sequentially to the corresponding amino acid residues in the polypeptide chain and determine the orientation of alpha helices and beta sheets with respect to the external magnetic field. (J. Wang 2000; Marassi and Opella 2002; Nevzorov and Opella 2003; A. Ramamoorthy 2004) As the slant angle  $\tau$  relative to the magnetic field increases, the PISA wheel moves from top left to bottom right while the radius first expands and then reduces in its size. (Figure 1.15)

In the PISEMA coordinates, the formulation of the dipolar coupling and chemical shift anisotropy are both in a quadratic form, so the values with equivalent magnitudes but opposite signs can all satisfy the equation, leading to a large number of possible orientational solutions. The orientations of the peptide planes are four- to eight-fold degenerate owing to the quadratic nature of the relevant NMR interactions, as well as to the fact that only the absolute values of the dipolar coupling could be determined experimentally. If multidimensional spectra including  $^{13}\text{C}_\alpha$  chemical shift,  $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$  dipolar coupling and even  $^{13}\text{C}_\beta$ - $^1\text{H}_\alpha$  dipolar coupling are available, these degeneracies could be removed and the conformation of side chains can be unambiguously resolved. (Nevzorov and Opella 2006) Ramachandran maps could also be used as a means of removing these degeneracies by constricting only those peptide orientations that result in plausible torsion angles  $\phi$  and  $\psi$  as described in the present work.

### 1.3.4 Structural Fitting using PISEMA

Structural fitting of the PISEMA spectra has proved to be a useful method to determine the membrane protein structure, particularly irregular structures like kinks, twists and bends, from the frequencies measured in solid-state NMR spectra. (Thomas Vosegaard 2002; Bertram, Asbury et al. 2003; Nevzorov and Opella 2003; J. R. Quine 2004) The programs SIMPSON and SIMMOL (Mads Bak 2002) have been developed by Nielsen and coworkers since 2000, while CNS-SS02, a public software package developed from CNS (Crystallography & NMR System) software (A. T. Brünger 1998), is also a common tool used in structural fitting. Around 2003, a Monte Carlo simulation algorithm implemented in MATLAB was established by Nevzorov and Opella for structural fitting with partial or incomplete resonance assignment of the PISEMA spectra. (Nevzorov and Opella 2003; Anna A. De Angelis 2004; David S. Thiriot 2004; De Angelis, Howell et al. 2006; Sang Ho Park 2006) Nevzorov and Opella have applied this algorithm to the experimental data of an 18-residue ideal  $\alpha$ -helix containing only alanines, a 16-residue single  $\alpha$ -helical transmembrane domain of the channel-forming peptide from the acetylcholine receptor (AChR M2), and a 25-residue transmembrane  $\alpha$ -helix of the fd-coat protein to compare the structural fits under different conditions, such as different variation ranges in the torsion angles and incomplete or partial assignment of the spectra. For the ideal  $\alpha$ -helix fragment, there is no obvious difference between the structural fits for the  $\pm 5^\circ$  and  $\pm 10^\circ$  variations in the torsion angles  $\phi$  and  $\psi$ . For the  $\alpha$ -helical transmembrane domain in AChR M2, the structural fit is virtually unique when the variation region for the torsion angles is  $\pm 5^\circ$ . But when the variation region extended to  $\pm 10^\circ$ , the structural fit is no longer unique, but is still similar to each other with a

RMSD (root mean square deviation) of less than 2Å. In addition, if additional resonances of some specific types of residues are available, the structural fit could be more accurate (RMSD < 1.6 Å) even with a relatively large variation ( $\pm 10^\circ$ ) in the torsion angles. For the transmembrane  $\alpha$ -helices of the fd-coat protein, Nevzorov and Opella also compared the structural fits with and without specific resonance assignment of some types of residues, and concluded that additional resonance assignment of some residues can help achieve a convergent set of solutions with higher accuracy and precision.

### 1.3.5 Analytical Framework for Solid-State NMR Observables in Torsion-Angle Space

To implement the structural fitting of multidimensional solid-state NMR spectra of aligned samples, it is necessary to establish an effective analytical framework to transform the data from solid-state NMR spectra into the protein structures.

In Cartesian coordinate system, a point in space can be represented by a row vector  $r$ :

$$r = (x \quad y \quad z) \quad (1.15)$$

Here,  $x$ ,  $y$ , and  $z$  are the Cartesian coordinates of the point. Assuming the Cartesian basis as a standard right-handed coordinate system and the rotation counterclockwise, the basic rotation matrices in three dimension space are:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} R_y(\theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} R_z(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.16)$$

To rotate an old coordinate system into a new one with coordinates of the same vector, we can use the rotation matrix  $R_{tot}$  described by the product of three rotation matrices:

$$\begin{aligned}
R_{tot} &= R_Z(\alpha)R_Y(\beta)R_Z(\gamma) \\
&= \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{pmatrix} \begin{pmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \cos(\alpha)\cos(\beta)\cos(\gamma) - \sin(\alpha)\sin(\gamma) & -\cos(\alpha)\cos(\beta)\sin(\gamma) - \sin(\alpha)\cos(\gamma) & \cos(\alpha)\sin(\beta) \\ \sin(\alpha)\cos(\beta)\cos(\gamma) + \cos(\alpha)\sin(\gamma) & -\sin(\alpha)\cos(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) & \sin(\alpha)\sin(\beta) \\ -\sin(\beta)\cos(\gamma) & \sin(\beta)\sin(\gamma) & \cos(\beta) \end{pmatrix}
\end{aligned} \tag{1.17}$$

So the coordinates of the vector in the new system can be expressed by:

$$(x' \ y' \ z') = (x \ y \ z)R_{tot} \tag{1.18}$$

In this analytical framework, a molecular frame associated with the peptide plane is introduced, the x-axis is chosen along the NH bond, and z-axis is perpendicular to the plane containing NH and NC' bonds. The orientation of applied magnetic field relative to the molecular frame of the  $n$ th peptide plane is described by angles  $\alpha_n$  and  $\beta_n$ . Hence,  $^{15}\text{N}$  chemical shift anisotropy and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling can be written as:

$$\nu(^{15}\text{N}) = \sigma_{11} \sin^2 \beta \sin^2(\alpha - \gamma) + \sigma_{22} \cos^2 \beta + \sigma_{33} \sin^2 \beta \cos^2(\alpha - \gamma) \tag{1.19}$$

$$\nu(^1\text{H} - ^{15}\text{N}) = \pm \chi \frac{3 \sin^2 \beta \cos^2 \alpha - 1}{2} \tag{1.20}$$

Aiming at calculating the protein structure directly from its solid-state NMR spectrum in an ab initio fashion, we should determine the minimum number of experimental angular constraints that are necessary to determine a complete atomic-resolution structure of a protein.

However, in solid-state NMR, both spatial and angular constraints are the observables in the measurement, so a coordinate system, which is called rank-1 irreducible spherical basis, capable of dealing with angular information is more convenient. (Figure 1.16) In the

project, therefore, we choose this basis to relate the protein structure to its multidimensional solid-state NMR spectrum.

Since the bond lengths  $x^2 + y^2 + z^2$  remain constant, the spherical basis is obtained from the Cartesian basis by the transformation matrix T:

$$(x \ y \ z) \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -\frac{x+iy}{\sqrt{2}} & z & \frac{x-iy}{\sqrt{2}} \end{pmatrix} \quad T \equiv \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix} \quad (1.21)$$

$$= r \begin{pmatrix} -\frac{\sin \theta}{\sqrt{2}} e^{i\phi} & \cos \theta & \frac{\sin \theta}{\sqrt{2}} e^{-i\phi} \end{pmatrix}$$

Consequently, the rotation matrix is transformed into the so-called ‘‘Rank-1 Wigner Rotation Matrix’’ (Arfken 1985):

$$T^+ R_{tot} T \equiv \mathbf{D}(\alpha, \beta, \gamma) \quad (1.22)$$

$$= \begin{pmatrix} e^{-i\alpha} \frac{1 + \cos \beta}{2} e^{-i\gamma} & -e^{-i\alpha} \frac{\sin \beta}{\sqrt{2}} & e^{-i\alpha} \frac{1 - \cos \beta}{2} e^{i\gamma} \\ \frac{\sin \beta}{\sqrt{2}} e^{-i\gamma} & \cos \beta & -\frac{\sin \beta}{\sqrt{2}} e^{i\gamma} \\ e^{i\alpha} \frac{1 - \cos \beta}{2} e^{-i\gamma} & e^{i\alpha} \frac{\sin \beta}{\sqrt{2}} & e^{i\alpha} \frac{1 + \cos \beta}{2} e^{i\gamma} \end{pmatrix}$$

In this irreducible spherical basis, all the frequencies obtained in the experiment can be rewritten in a unified manner as a much more compact quadratic form.

$$\nu = \vec{Y}(\beta, \alpha) \left[ \mathbf{D}(\Omega_{MP}) \mathbf{M} \mathbf{D}^{-1}(\Omega_{MP}) \right] \vec{Y}^+(\beta, \alpha) \quad (1.23)$$

Here, Wigner rotation matrix  $\mathbf{D}(\Omega_{\text{MP}})$  describes the transformation from the molecular frame to the principal axis system of each tensor. The superscript “+” denotes the Hermitian conjugate. The row vector of unnormalized spherical harmonics  $\mathbf{Y}(\beta, \alpha)$  is given by the following formulation:

$$\vec{\mathbf{Y}}(\beta, \alpha) = \left( -\frac{\sin \beta}{\sqrt{2}} e^{i\alpha} \quad \cos \beta \quad \frac{\sin \beta}{\sqrt{2}} e^{-i\alpha} \right) \quad (1.24)$$

In Eq. (1.23),  $\mathbf{M}$  is the corresponding interaction matrix. For chemical shift anisotropy,  $\mathbf{M}$  is described by the principal components  $\sigma_{11}$ ,  $\sigma_{22}$ , and  $\sigma_{33}$ , given that  $\sigma_{33} > \sigma_{22} > \sigma_{11}$ :

$$\mathbf{M} = \begin{pmatrix} \frac{\sigma_{11} + \sigma_{22}}{2} & 0 & \frac{\sigma_{22} - \sigma_{11}}{2} \\ 0 & \sigma_{33} & 0 \\ \frac{\sigma_{22} - \sigma_{11}}{2} & 0 & \frac{\sigma_{11} + \sigma_{22}}{2} \end{pmatrix} \quad (1.25)$$

For dipolar couplings,  $\mathbf{M}$  is expressed by the product of the dipolar coupling constant and a single diagonal matrix:

$$\mathbf{M} = \chi \begin{pmatrix} -1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{pmatrix} \quad (1.26)$$

The rotation of the subsequent peptide planes is achieved by the so-called propagator matrix. Mathematically, the propagator matrix  $\mathbf{P}(\Phi_n, \Psi_n)$  given by the product of two rank-1 Wigner rotation matrices shown in Eq. (1.27) brings the molecular frame associated with the  $n$ th peptide plane into coincidence with the molecular frame of the  $(n+1)$ th peptide plane. So the orientation of the succeeding  $(n+1)$ th peptide plane relative to the applied magnetic field is

calculated via operating the propagator matrix on the spherical harmonics of the  $n$ 'th peptide plane.

$$\mathbf{P}(\Phi_n, \Psi_n) = \mathbf{D}(\alpha_{NC_\alpha}, \Phi_n, \gamma_{tetra}) \times \mathbf{D}(0, -\Psi_n - \pi, \gamma_{C_\alpha C}) \quad (1.27)$$

$$\vec{Y}(\beta_{n+1}, \alpha_{n+1}) = \vec{Y}(\beta_n, \alpha_n) \mathbf{P}(\Phi_n, \Psi_n) \quad (1.28)$$

In the propagator matrix, the first Euler angle in the first Wigner matrix is the angle between the y-axis of the molecular frame and the N-C $_{\alpha}$  bond of the  $n$ th peptide plane,  $\alpha_{NC_\alpha} = 151.8^\circ$ ; the third Euler angle in the first Wigner matrix is the tetrahedral angle,  $\gamma_{ideal} = 109.47^\circ$ , but in real protein typically  $\gamma = 110^\circ - 112^\circ$ ; the third Euler angle in the second Wigner matrix is the angle between the C $_{\alpha}$ -C bond and the y-axis of the molecular frame of the  $(n+1)$ th peptide plane,  $\gamma_{C_\alpha C} = 34.9^\circ$ . These numerical values of Euler angles are assumed to be constants for each peptide residue of the structure, so as long as the torsion angles  $\Phi_n$  and  $\Psi_n$  are determined, the backbone conformation of the complete protein can be constructed residue by residue. The peptide plane can be also described by three directional vectors  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  with formulation expressed by Eq. (1.29) in the irreducible spherical basis.

$$\mu_k \equiv \begin{pmatrix} -\frac{x_k + iy_k}{\sqrt{2}} & z_k & \frac{x_k - iy_k}{\sqrt{2}} \end{pmatrix} \quad (1.29)$$

Here,  $\mu_1$  denotes the directional vector ending at carbonyl carbon,  $\mu_2$  ending at nitrogen and  $\mu_3$  ending at the  $\alpha$ -carbon.  $(x_k, y_k, z_k)$  are the vector coordinates measured relative to the molecular frame, related to  $\mu_k$  ( $k=1, 2, 3$ ) via Eq. (1.21). Therefore, the orientation of the first

peptide plane relative to the laboratory frame can be calculated using Eq. (1.30), while Eq. (1.31) works for the succeeding residues.

$$V_k^{(1)T} = \mu_k \mathbf{D}(0, \beta_1, \pi - \alpha_1)^{-1} \quad (1.30)$$

$$V_k^{(n+1)T} = V_k^{(n)T} \mathbf{P}(\Phi_n, \Psi_n, \omega_n)^{-1} \quad (1.31)$$

## 1.4 Prediction of Helical Membrane Proteins Structures using Rosetta

### 1.4.1 De Novo Prediction using Membrane Ab Initio in Rosetta

Many research groups analyzed the available alpha-helical membrane protein structures statistically and summarized the amino acid environmental preferences within hydrophobic, amphiphilic (interface) and polar layers of the membrane. Based on this information, Yarov-Yarovoy et al. have adapted original Rosetta de novo structure prediction method to predict the helical transmembrane protein structures. (Vladimir Yarov-Yarovoy 2006) In the tests on 12 membrane proteins with known structure, such as rhodopsin, V-type Na<sup>+</sup>-ATPase, bacteriorhodopsin and lactose permease transporter, 51-145 residues were predicted with RMSD less than 4Å to the corresponding native structure.

The energy of predicted structure is calculated by Eq.1.32 (Vladimir Yarov-Yarovoy 2006), considering residue-environment interaction, residue-residue interaction, steric overlapping, packing density of membrane proteins and strand pairings. The energy maximizes exposure of hydrophobic residues within the membrane and minimizes hydrophobic exposure outside the membrane. The membrane protein database is constituted by 28 helical transmembrane protein structures in PDB.

$$E_{total} = E_{env} + E_{pair} + E_{clash} + E_{density} + E_{strand} \quad (1.32)$$

First of all, the amino acid sequence is used to generate three- and nine-residue peptide fragments on Robetta server (<http://rosetta.bakerlab.org/>), which is more convenient and accurate than the traditional PSIPRED (Jones 1999), JUFO (Meiler 2003) and SAM-T99 (Kevin Karplus 2001).

Different from de novo prediction for soluble protein structures, besides the FASTA primary sequence file, three-residue and nine-residue peptide fragments files, transmembrane topology prediction file and lipophilicity prediction file are also necessary and significant in membrane ab initio simulation.

Transmembrane region is predicted on OCTOPUS server (<http://octopus.cbr.su.se/>) with primary sequence in FASTA format as input file, and expressed by *i*, *o* and *M* in the output file. (Elofsson 2008) *M*, standing for membrane, is assigned to residues  $< 13 \text{ \AA}$  from the membrane center. *i* and *o* are assigned to residues in loop regions, which is  $13\text{-}23 \text{ \AA}$  from the membrane center. *inside* (*i*) or *outside* (*o*) depends on the side of the membrane they reside on. Figure 1.17 demonstrates the definition of membrane and the allowed region for transmembrane hairpins by OCTOPUS. Then the transmembrane topology prediction file, i.e. SPAN file, is generated using the OCTOPUS file. The number of transmembrane helices and the start and end residue numbers of each single predicted transmembrane helix are shown in the SPAN file. Finally, with sequence file and SPAN file as input, the predicted lipophilicity file, i.e. LIPS4 file, is produced.

The membrane ab initio protocol sets the initial membrane center as the center of the mass of the protein and the membrane normal unit vectors as the average direction of all the helices with predicted transmembrane topology file. Two adjacent transmembrane helices near the

middle of the protein are embedded into the membrane first, then a new helix is added randomly to the N- or C-terminus. All the remaining helices are folded continuously in this manner until the whole structure is embedded into the membrane. The orientation of membrane normal and relative position between membrane center and mass center of protein are changed to search the optimal embedding position, i.e. the position with the lowest energy, of the corresponding helical transmembrane protein. Figure 1.18 illustrates the definition of different layers of membrane environment by the membrane ab initio protocol. The membrane is represented by parallel planes including inner hydrophobic, outer hydrophobic, interface and polar layers.

#### **1.4.2 High-Resolution Prediction using Adapted All-Atom Energy Function**

To achieve high-resolution prediction of helical transmembrane protein structures, Barth and coworkers ameliorated the all-atom energy function and increased the accuracy of predicted structures to RMSD less than 2.5Å for membrane protein domain with less than 150 residues. (P. Barth 2007; P. Barth 2009) The implicit solvation model was modified to illustrate membrane environment and the original solvation energy  $E_{solv}$  was substituted with  $E_{mbenv}$  and  $E_{mbsolv}$ . (P. Barth 2007)  $E_{mbenv}$  represents the change in solvation free energy of each atom when they are transferred along membrane normal from outer polar solvent (water) to inner hydrophobic core of membrane bilayer.  $E_{mbsolv}$  depicts the change in solvation free energy of an atom when it is buried into the protein.  $E_{menv\_smooth}$  item is the statistically derived smooth membrane protein environmental potential based on the statistically derived rotamer pair potential. Adaptation was also made to hydrogen bonding potential (Tanja Kortemme 2003) for membrane proteins. Because water molecules are strong hydrogen bond donors and

acceptors but the acyl chains of membrane lipids is hydrophobic and have no polar groups, the effect of the membrane environment on the strength of the hydrogen bonds has to be considered.

## **1.5 Research Summary**

Over the course of my PhD work, I have been working on structure determination using angular constraints from solid-state NMR experiments and the validation of calculated membrane protein structures using Rosetta. First, we have developed approaches to calculate membrane protein structures solely from heteronuclear dipolar couplings, i.e. from the “shiftless” solid-state NMR spectra of oriented samples (Yuanyuan Yin 2011). Second, to improve the prediction quality and determine the optimal embedding positions of proteins in the biological membrane, we used positional constraints in structural fitting calculation and applied Rosetta all-atom energy function to validate the calculated structures. Finally, oligomeric protein complexes play an important role in many biological processes. Currently, we are developing a method to determine the oligomerization state and relative orientation of each single peptide chain of the oligomer proteins in the biological membrane. The outcomes of these three projects are summarized in the following chapters.

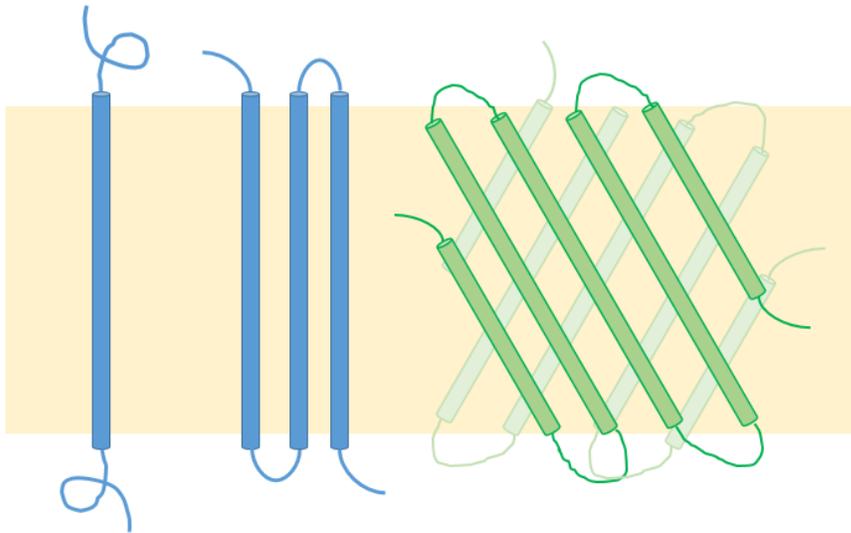


Figure 1.1 Schematic representation of the structure types of transmembrane proteins, (left) single  $\alpha$ -helix, (middle) bundles of  $\alpha$ -helices, and (right)  $\beta$ -barrel. The yellow part represents the membrane. (Adapted from [http://en.wikipedia.org/wiki/Transmembrane\\_protein](http://en.wikipedia.org/wiki/Transmembrane_protein))

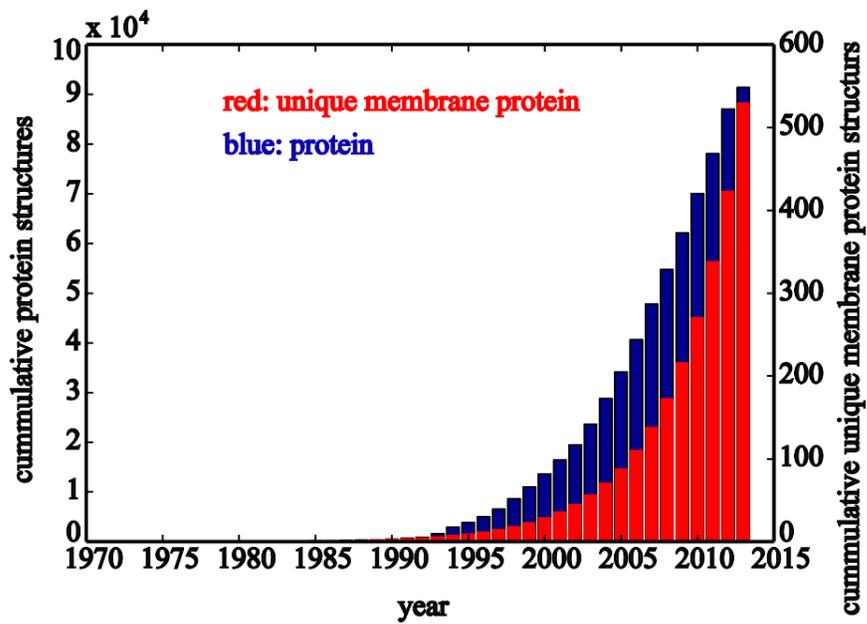


Figure 1.2 Yearly growth of all protein structures and unique membrane protein structures.

(WHITE 2004) The number of all protein structures is in blue, and the number of unique membrane proteins is in red.

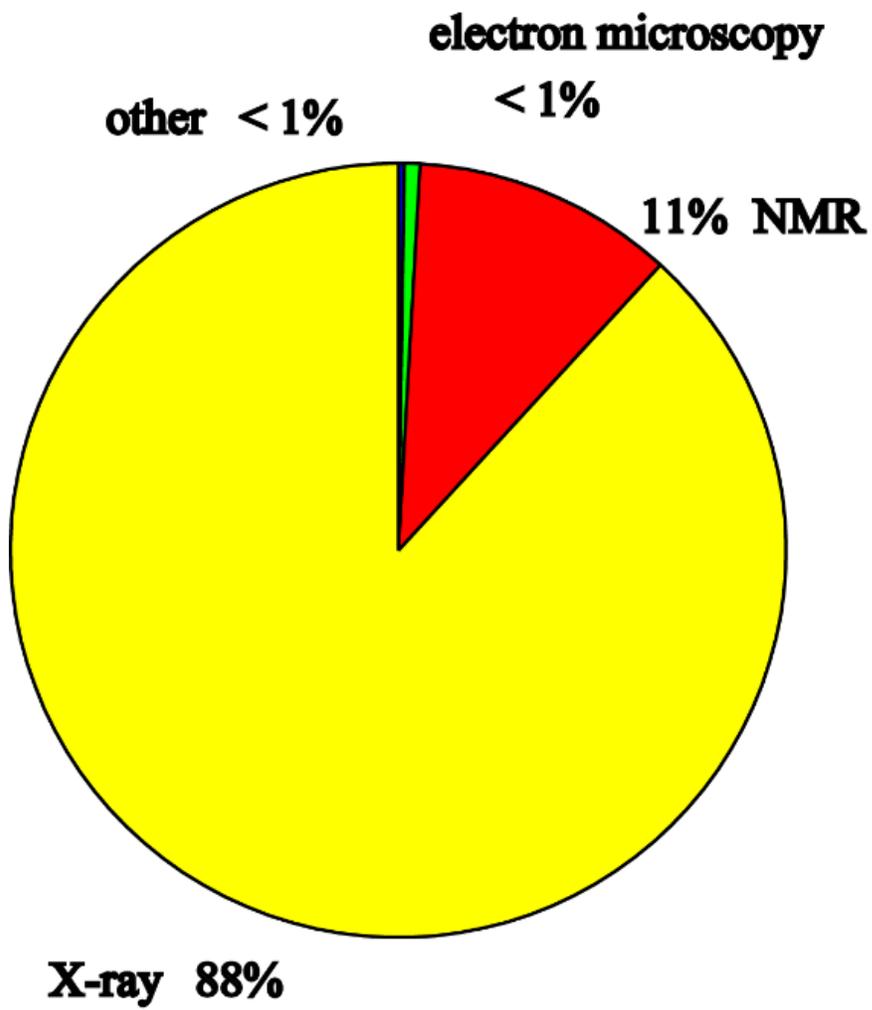


Figure 1.3 Distribution of all experimental methods used in protein structural studies.



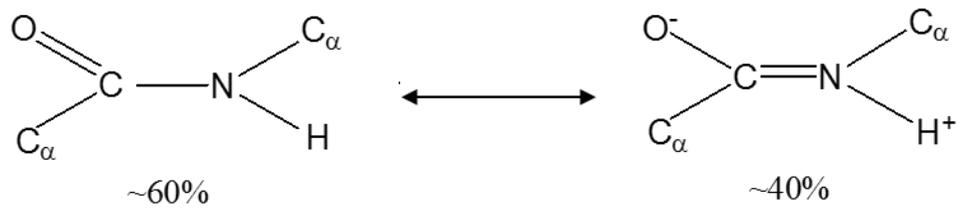


Figure 1.5 Resonance forms of the peptide group, single-bonded form accounts for ~60%, and double-bonded form accounts for ~40%.

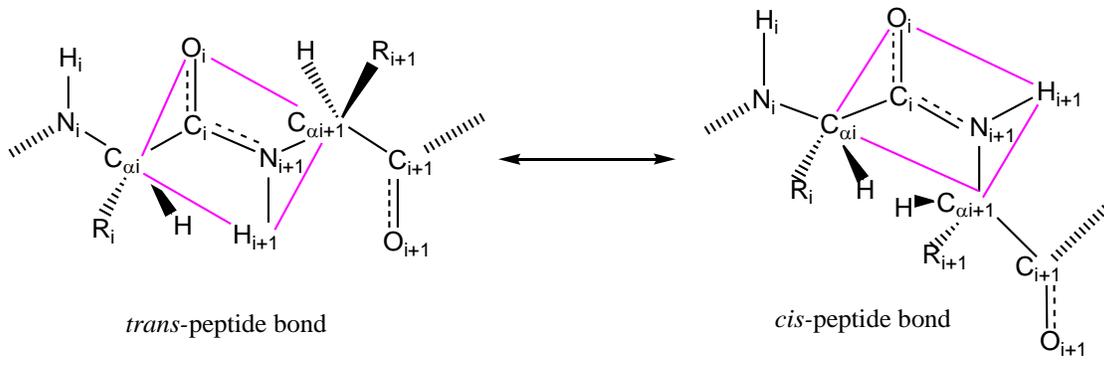


Figure 1.6 *Trans* and *cis* isomers of the peptide plane.

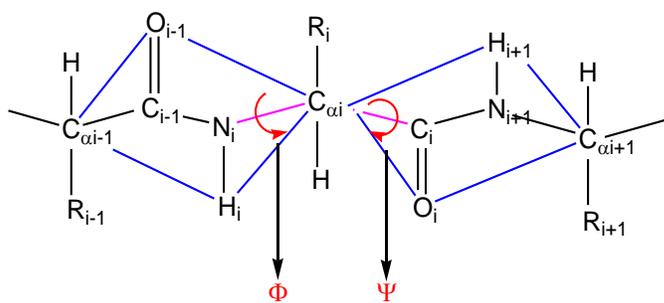


Figure 1.7 Representation of torsion angles  $\Phi$  and  $\Psi$ , respectively.

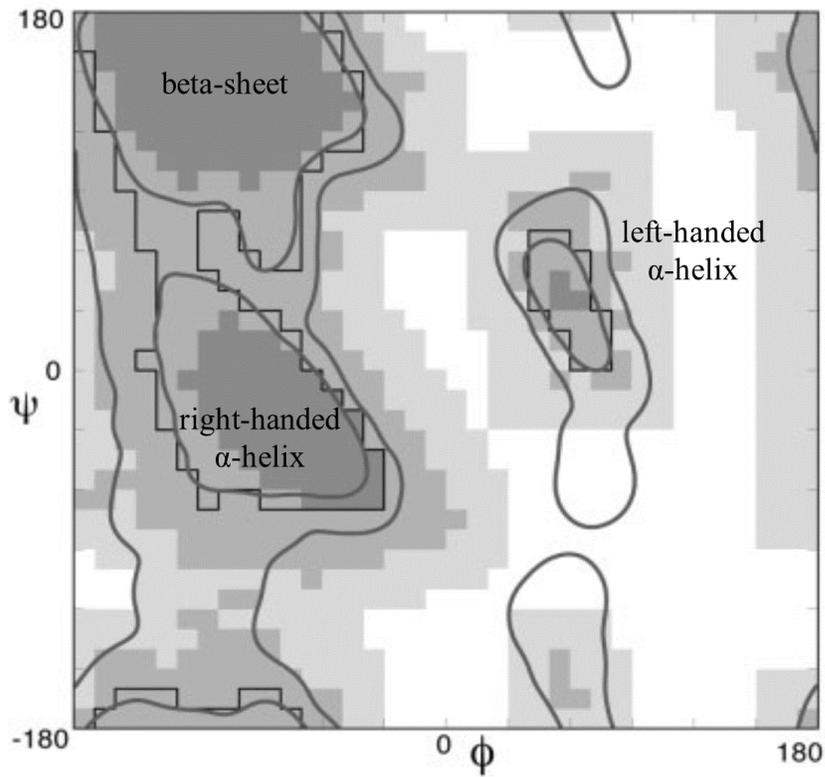


Figure 1.8 Torsion angles  $\phi/\psi$  definition. The areas shaded in dark, medium and light gray are the “core”, “allowed” and “generously allowed” regions, respectively. The stepped black outline encloses the single “strictly allowed” region. (Lovell, Davis et al. 2003)

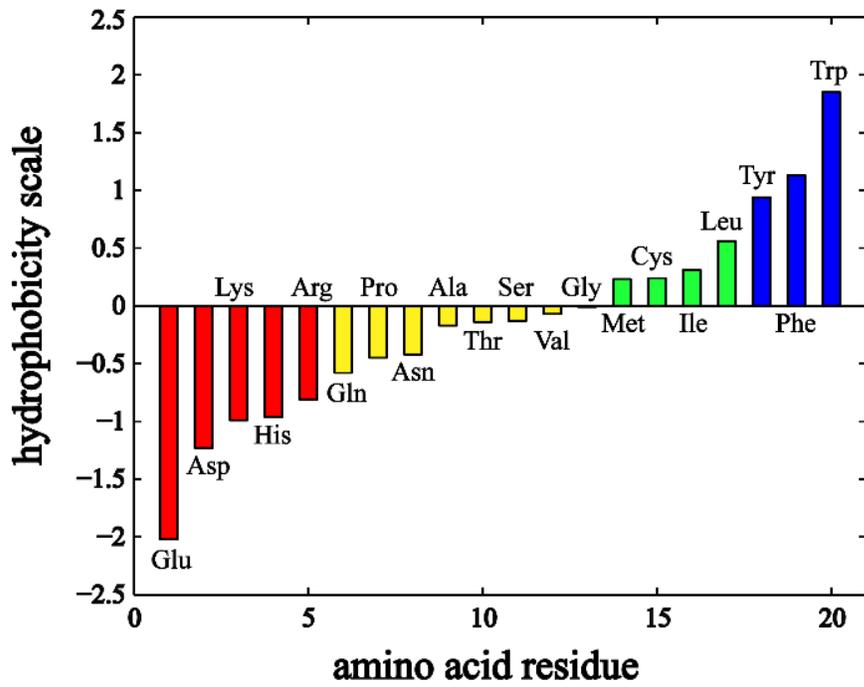


Figure 1.9 Experimentally determined hydrophobicity scale for the 20 standard amino acids.

(White 1996)

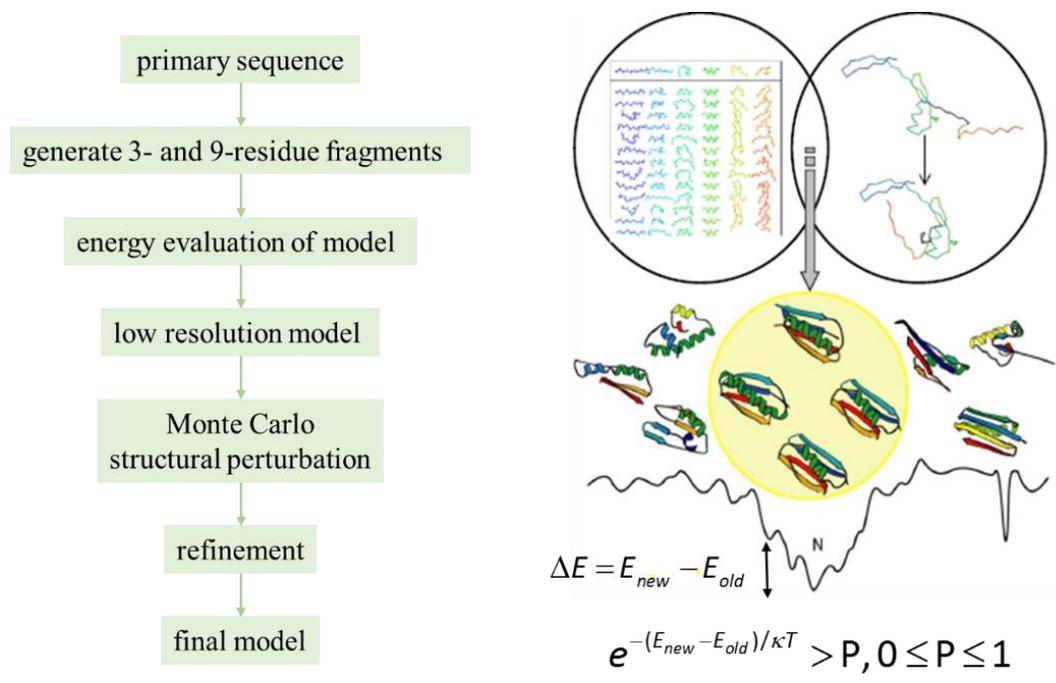
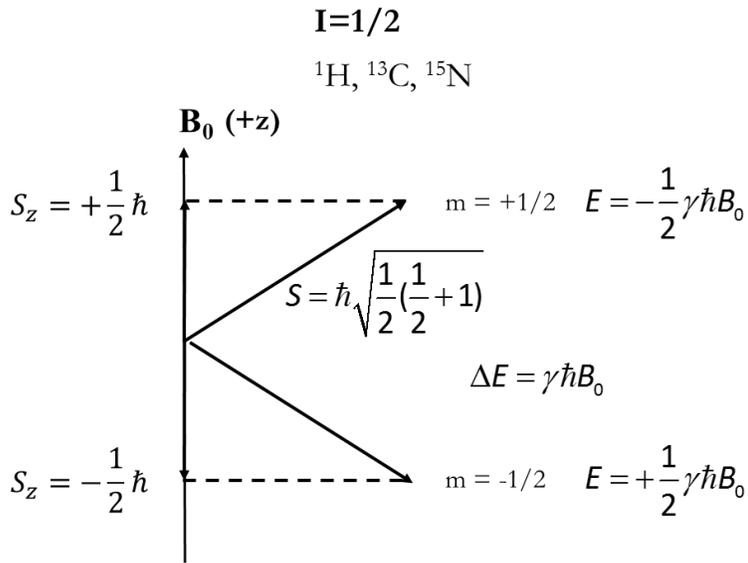


Figure 1.10 Rosetta general sampling algorithm.

A.



B.

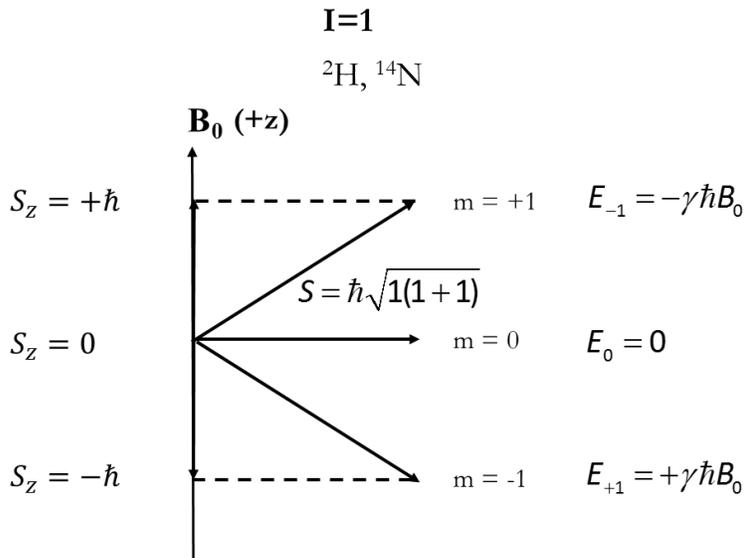


Figure 1.11 Splitting of nuclei spin states in an external magnetic field  $\mathbf{B}_0$  when the angular momentum quantum number  $I$  equals  $\frac{1}{2}$  (A) and  $I$  equals 1 (B), respectively.

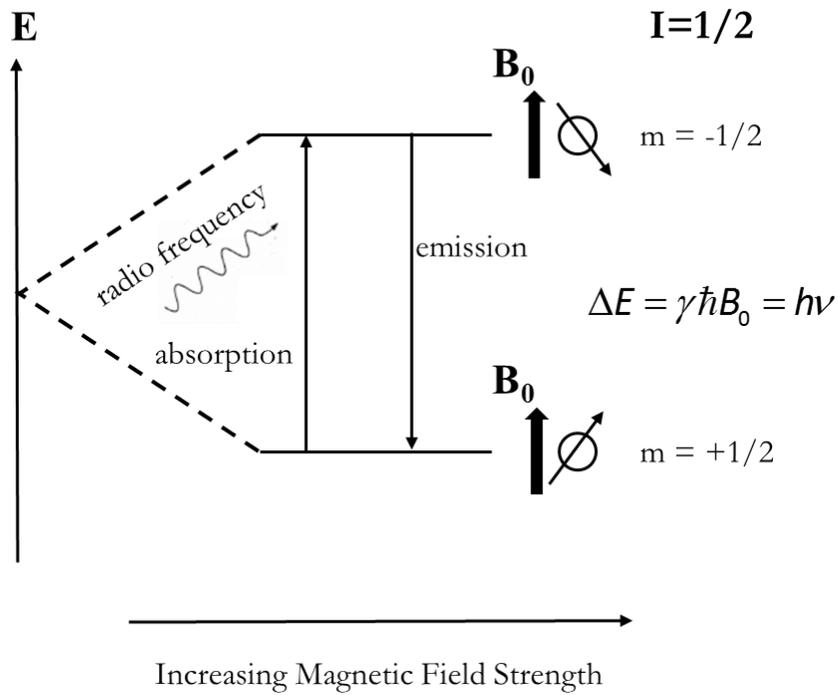


Figure 1.12 Nuclear energy levels in an external magnetic field  $\mathbf{B}_0$  when the angular momentum quantum number  $I$  equals  $1/2$ .

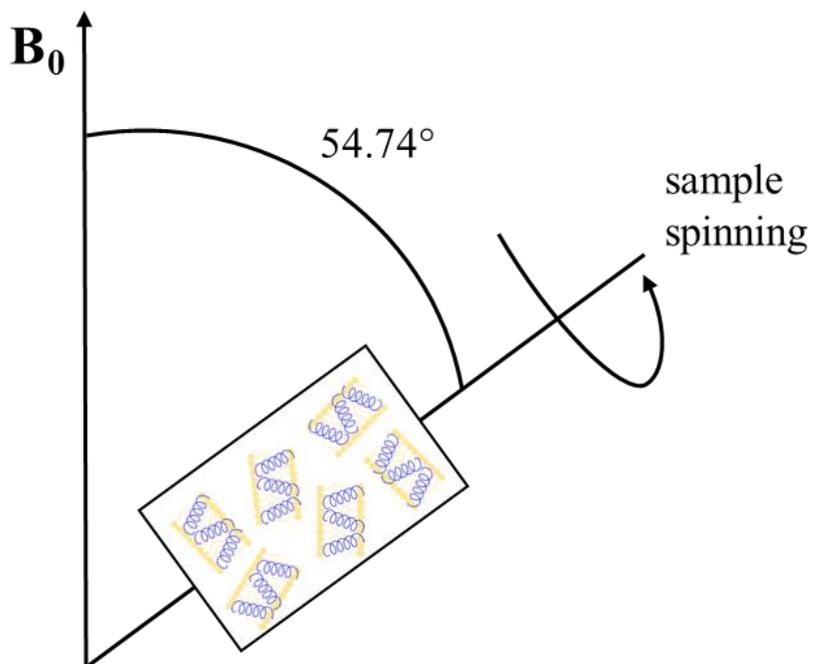


Figure 1.13 Schematic Diagram of magic angle spinning solid state NMR. Samples are rotated along magic angle ( $54.74^\circ$ ) with respect to the external magnetic field  $B_0$ .

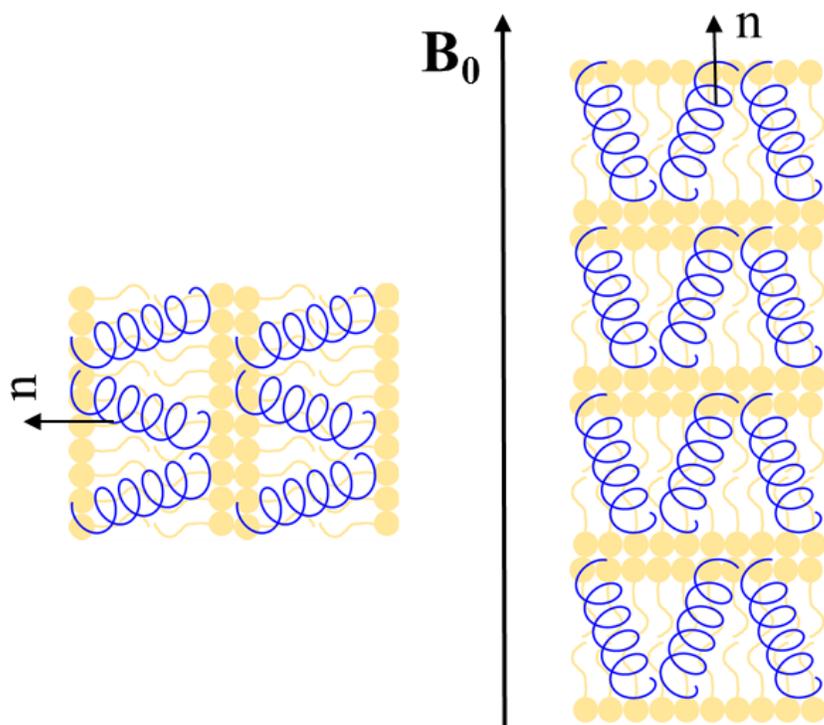


Figure 1.14 Representation of a uniaxially oriented membrane protein sample in lipid bilayer.

$\mathbf{n}$  describes the unit vector of the membrane normal.

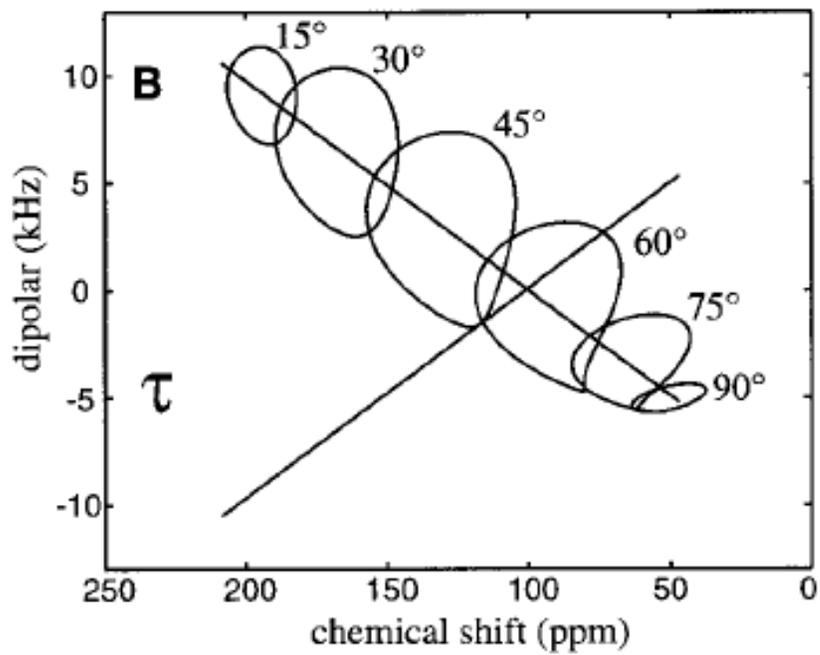


Figure 1.15 PISA wheel patterns as a function of tilt angle. (J. Wang 2000)

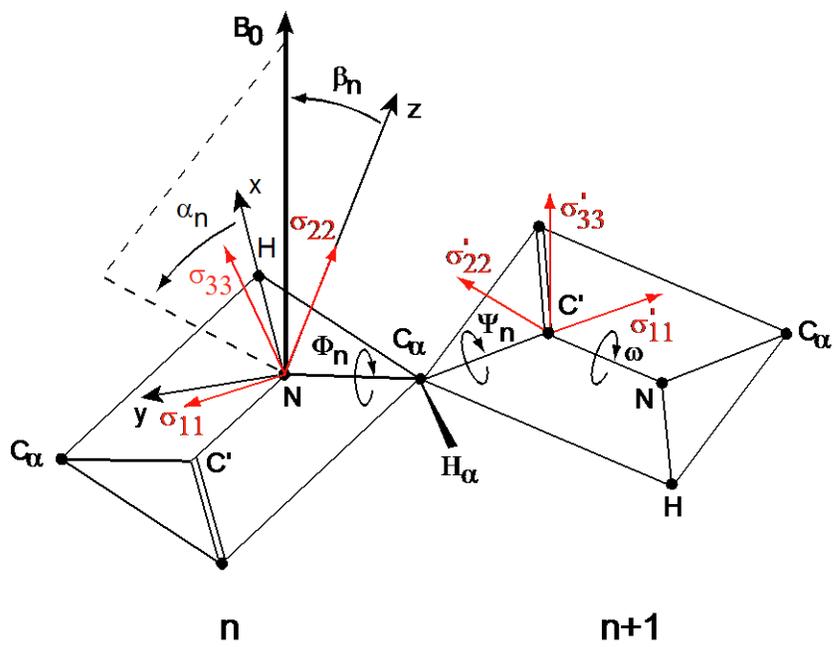


Figure 1.16 Description of the molecular frame (MF) associated with the peptide plane.

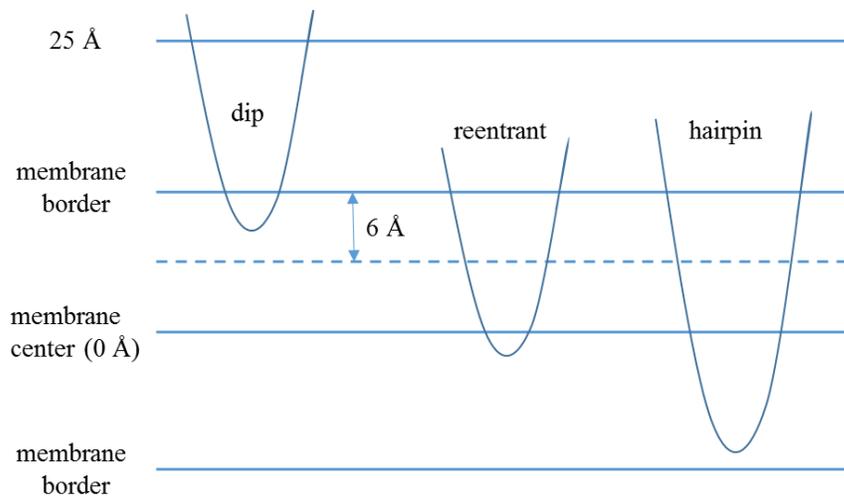


Figure 1.17 Membrane definition in OCTOPUS method.

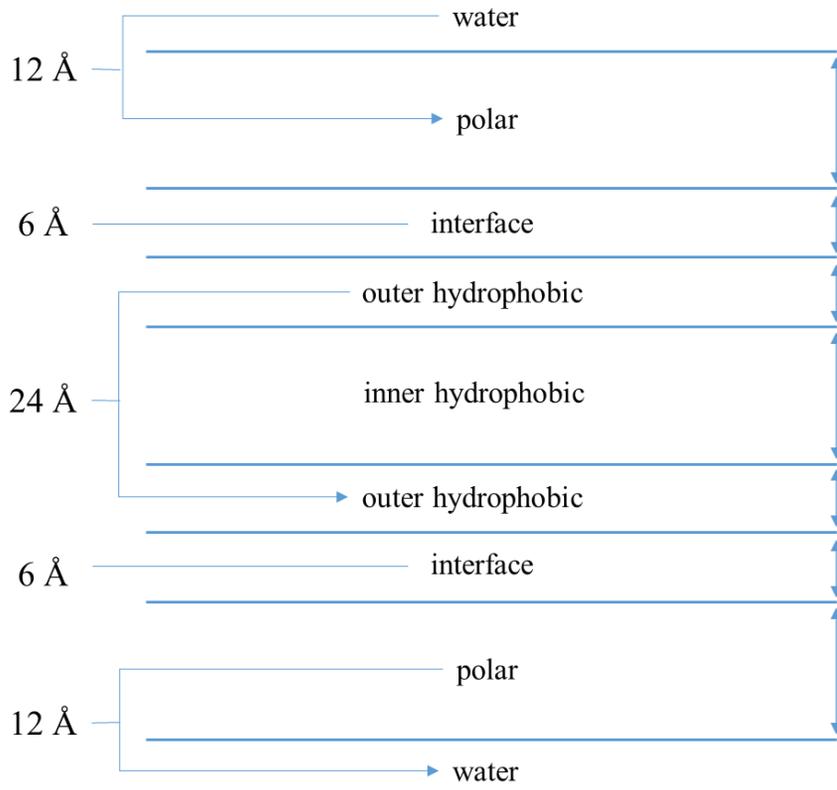


Figure 1.18 Definition of membrane layers in Rosetta membrane ab initio protocol.

## Chapter 2

# Structure Determination in “Shiftless” Solid State NMR of Oriented Protein Samples

Adapted from: *Journal of Magnetic Resonance* 2011, 212(1), 64-73.

### 2.1 Introduction

In the past two decades, solid-state NMR of uniaxially aligned samples has been successfully applied to three-dimensional structure determination of membrane protein. (Opella, Stewart et al. 1987; Cross and Opella 1994; Opella, Kim et al. 1994; De Angelis, Howell et al. 2006; Opella, Zeri et al. 2008; Traaseth, Shi et al. 2009; Mukesh Sharma 2010) Due to substantial improvements in the decoupling pulse sequences (Wu, Ramamoorthy et al. 1994; Ramamoorthy 1998; Fung, Khitrin et al. 2000; Dvinskikh, Yamamoto et al. 2006; Nevzorov and Opella 2007) and methods of alignment (De Angelis and Opella 2007; Park, Loudet et al. 2008; Park and Opella 2010), oriented-sample NMR has demonstrated its effectiveness as a complementary method to traditional X-ray diffraction and solution NMR spectroscopy for studying membrane proteins.

In solid-state NMR of oriented samples, membrane proteins are aligned uniaxially in the phospholipid bilayer environment, so that the experimental observables are orientationally dependent, and the structural information is directly contained in the NMR spectra. The observables typically include heteronuclear dipolar couplings and chemical shift anisotropy (CSA) associated with the isotopically labeled  $^{15}\text{N}$  or  $^{13}\text{C}$  sites. Previously, protein backbone

structures were determined by assembling the final structure from fragments of peptide planes by calculating the possible orientations of the magnetic field relative to the molecular frame associated with each residue (Opella, Stewart et al. 1987; Stewart, Valentine et al. 1987; Bertram, Asbury et al. 2003). The orientational degeneracies were reduced by examining the chemical and steric allowances as suggested by the Ramachandran plot, chemical shift values, the angles calculated from the dipole-dipole and quadrupolar interactions, and bond angles and distances between the specific atoms (Opella, Stewart et al. 1987; Stewart, Valentine et al. 1987; Denny, Wang et al. 2001; Mads Bak 2002; Bertram, Asbury et al. 2003). To improve resolution and reduce the degeneracy of the structural solutions, measurements involving the alpha carbon ( $^{13}\text{C}_\alpha$ ), such as the chiral  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  dipolar couplings, may need to be introduced (Nevzorov and Opella 2006) in addition to the most widely used  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar couplings. New triple-resonance pulse sequences (Sinha, Grant et al. 2007) involving  $^{15}\text{N}$  and  $^{13}\text{C}$  correlations and “shiftless” acquisition techniques (Wu and Opella 2008) are especially encouraging in this regard.

So far, only the planar form of the *trans* peptide unit has been used as the standard geometry for structure calculations in oriented-sample NMR. However, back in 1968 Ramachandran pointed out the significance of out-of-plane distortions in the polypeptide chains, especially in cyclic peptides, which was supported by energy calculations (Ramachandran 1968; G.N.Ramachandran, A.V.Lakshminarayanan et al. 1973). A survey on the torsion angles  $\omega$  describing the rotation around the amide bond in peptides and proteins was implemented by considering the Cambridge Structural Database (CSD) of small molecules and the Brookhaven Protein Databank (PDB), which showed that for all good-quality crystal

structures, the average twist was around  $0.5^\circ$  (MacArthur and Thornton 1996). Scheiner and Kern have indicated that certain environmental effects such as the formation of hydrogen bonds contributed to the non-planar conformation as well (Scheiner and Kern 1977). Consequently, out-of-plane deviations of the peptide linkages should be included in the refinement of the calculated protein structures.

Another potential problem affecting the accuracy of the calculated structures is the possible variations in the magnitude of the principal components and orientation of the shielding tensor along the polypeptide chains (Hartzell, Whitfield et al. 1987; Oas, Hartzell et al. 1987). A more recent work (Cornilescu and Bax 2000) provides a comprehensive review of the variation of CSA in soluble proteins. In a quantum mechanical study on the chemical shift tensors of peptides (Walling, Pargas et al. 1997), the least shielded component of the  $^{15}\text{N}$  chemical shift tensor,  $\sigma_{33}$  was found to deviate substantially from the peptide plane particularly in helical domains, and the principal value of the  $^{13}\text{C}_\alpha$  chemical shift tensor along the  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  bond was found to be very sensitive to the torsion angles  $\Phi$  and  $\Psi$ . In addition to the variations in the tensor values, there is also experimental accuracy (uncertainty) in the determination of the values themselves. Lee et al. (Lee, Santos et al. 1999) reported the  $^{15}\text{N}$  chemical shift tensor parameters for the  $^{15}\text{N}$ -Gly-18 residue to be  $\sigma_{11\text{N}}=42\pm 2$  ppm,  $\sigma_{22\text{N}}=73\pm 2$  ppm,  $\sigma_{33\text{N}}=215\pm 2$  ppm,  $\alpha_{\text{N}}=30\pm 10^\circ$ ,  $\beta_{\text{N}}=22\pm 2^\circ$  (here,  $\alpha_{\text{N}}$  and  $\beta_{\text{N}}$  describe the orientation of the chemical shift tensor relative to the molecular frame); whereas the parameters for  $^{15}\text{N}$ -Phe-16 were  $\sigma_{11\text{N}}=55\pm 2$  ppm,  $\sigma_{22\text{N}}=80\pm 2$  ppm,  $\sigma_{33\text{N}}=220\pm 2$  ppm,  $\alpha_{\text{N}}=45\pm 15^\circ$ ,  $\beta_{\text{N}}=22\pm 3^\circ$ . These findings indicated a significant effect of the local chemical and geometrical environment on the magnitude and orientation of the  $^{15}\text{N}$  CSA (Lee, Wittebort et al. 1998).

While the principal components of the CSA tensor can be directly measured for each specific amino-acid site of the protein using either the Herzfeld-Berger method (Herzfeld and Berger 1980) as in MAS NMR (Franks, Zhou et al. 2005; Wylie, Sperling et al. 2007), or relaxation methods as in solution NMR (Loth, Pelupessy et al. 2005), the orientation of the tensor relative to each peptide plane is somewhat less precisely determined.

Nonetheless, in structure calculations from oriented sample NMR data, the  $^{15}\text{N}$  CSA tensor is usually assumed to have the same orientation and average values of the principal components for all amino-acid residues:  $\sigma_{11\text{N}}=64$  ppm,  $\sigma_{22\text{N}}=77$  ppm,  $\sigma_{33\text{N}}=217$  ppm (Wu, Ramamoorthy et al. 1995; Nevzorov and Opella 2003) (for glycines these values are assumed to be  $\sigma_{11\text{N}}=41$  ppm,  $\sigma_{22\text{N}}=64$  ppm,  $\sigma_{33\text{N}}=210$  ppm). For alpha-helical structures with the highly constrained torsion angles  $\Phi$  and  $\Psi$  and well-defined hydrogen bonding, the above-mentioned tensor variations and uncertainties do not generally represent a problem since the arising orientational solutions can be easily sorted out (Stewart, Valentine et al. 1987; Marassi and Opella 2000; Denny, Wang et al. 2001; Marassi and Opella 2002; Bertram, Asbury et al. 2003; Page, Kim et al. 2008). Structures of macroscopically aligned proteins having up to 50 residues can be obtained even in the presence of short connecting loops by imposing weak packing and hydrophobic restraints (De Angelis, Howell et al. 2006). However, the variations and/or uncertainty in the tensor parameters can become an important issue in calculating backbone conformations with less restrained torsion angles such as the connecting interhelical loops or  $\beta$ -type structures. This becomes evident from the fact that the orientational restraints (dipolar couplings, CSA) provide only one or two angles that define the orientation of the peptide plane relative to the laboratory frame; whereas *three* Euler

angles are necessary to unequivocally define the orientation of a rigid body. Therefore, the calculated structures can branch off at a point where the multiple solutions corresponding to the relative orientations of the adjacent peptide units are distributed uniaxially around the main magnetic field. It has been shown (Nevzorov and Opella 2006) that in the absence of uncertainty in the tensor values and experimental error there is a mathematically unique structural solution to a three-dimensional solid-state NMR spectrum (including  $^{15}\text{N}$  CSA, and  $^1\text{H}$ - $^{15}\text{N}$  and  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  dipolar couplings). However, the above-mentioned uncertainties are still inevitably contained in the experimental data. In the present paper we evaluate the effect of experimental error on the accuracy and convergence of the calculated structures. Resolving these important issues could potentially give way to a complete *de novo* structure determination of membrane proteins of arbitrary topology as experimentally determined dipolar couplings for the doubly-labeled proteins become available.

## 2.2 NMR Observables in the Spherical Basis and the Algorithm

In oriented-sample NMR spectroscopy, the structural information is obtained directly from angular-dependent observables; therefore, the Cartesian basis is not very efficient. To deal with the constraints in the angular space and minimize the number of required calculations we choose the irreducible spherical basis of rank 1 to relate the protein structure to its multidimensional solid-state NMR spectra. The molecular frame (MF) depicted in Figure 1 is associated with an individual peptide plane, in which the x-axis is along the NH bond, and the z-axis is perpendicular to the plane determined by the N-H and C'-N bonds (where C' denotes the carbonyl atom). The orientation of the magnetic field  $B_0$  relative to the MF is described by the angles  $\alpha$  and  $\beta$ .

To avoid explicit calculation of the trigonometric functions in the NMR frequencies, a more compact quadratic form was previously employed (Nevzorov and Opella 2006):

$$v = \vec{Y}(\beta, \alpha) \left[ \mathbf{D}(\Omega_{MP}) \mathbf{M} \mathbf{D}^{-1}(\Omega_{MP}) \right] \vec{Y}^{\dagger}(\beta, \alpha) \quad (2.1)$$

Here  $\vec{Y}(\beta, \alpha)$  denotes the row vector of the unnormalized spherical harmonics,

$$\vec{Y}(\beta, \alpha) = \left( -\frac{\sin \beta}{\sqrt{2}} e^{i\alpha} \quad \cos \beta \quad \frac{\sin \beta}{\sqrt{2}} e^{-i\alpha} \right). \text{ The rank-1 Wigner rotation matrix } \mathbf{D}(\Omega_{MP})$$

describes the transformation from the MF to the principal axis system of each tensor, and the superscript “+” denotes the Hermitian conjugate. The matrix  $\mathbf{M}$  corresponds to the specific type of interactions (e.g. CSA or dipolar) as defined below.

To include deviations from the planarity of peptide units, the following propagator matrix is introduced by employing a product of three Wigner rotation matrices:

$$\mathbf{P}(\Phi_n, \Psi_n, \omega_n) = \mathbf{D}(\alpha_{NC_\alpha}, \Phi_n, \gamma_{tetra}) \mathbf{D}(0, -\Psi_n - \pi, \gamma_{C_\alpha C}) \mathbf{D}(-\alpha_{NC'C_\alpha}, \pi - \omega_n, -\pi/2 - \gamma_{HNC'}) \quad (2.2)$$

The three torsion angles that result in the secondary and tertiary conformations of the polypeptide chains are denoted as  $\Phi_n$ ,  $\Psi_n$ , and  $\omega_n$  (Fig. 2.1). The remaining angles reflect the local peptide bond geometry. The angle  $\alpha_{NC_\alpha}$  denotes the angle between the N-C $_{\alpha}$  bond and the y-axis of the MF corresponding to the  $n$ 'th plane,  $\alpha_{NC_\alpha} = 151.8^\circ$ ;  $\gamma_{tetra}$  is the tetrahedral angle, typically  $110\text{-}112^\circ$  in real proteins instead of the ideal value  $109.47^\circ$ ;  $\gamma_{C_\alpha C}$  is the angle between the C'-C $_{\alpha}$  bond and the y-axis of MF of the  $(n+1)$ 'th plane,  $\gamma_{C_\alpha C} = 34.9^\circ$ . The other angles can be also calculated from the standard peptide plane geometry (Engl and Huber

1991), and are given by:  $\alpha_{NC'C_\alpha}=115.6^\circ$ ;  $\gamma_{HNC'}=119.5^\circ$ . Here, the bond angles are assumed constant for all amino acid residues of the protein.

The orientation of the magnetic field  $B_0$  relative to the MF of the next peptide plane can be calculated in a recursive manner by operating the above propagator matrix  $\mathbf{P}(\Phi, \Psi, \omega)$  from the right on the row vector of the spherical harmonics corresponding to the orientation of  $B_0$  with respect to the preceding frame:

$$\vec{Y}(\beta_{n+1}, \alpha_{n+1}) = \vec{Y}(\beta_n, \alpha_n) \mathbf{P}(\Phi_n, \Psi_n, \omega_n) \quad (2.3)$$

We note that the Wigner rotation matrices in Eq. (2.2) can be expressed in terms of the angular momentum operators  $\mathbf{L}_y$  and  $\mathbf{L}_z$  in the following manner (G. 1985):

$$\mathbf{D}(\phi, \theta, \psi) = e^{-i\phi\mathbf{L}_z} e^{-i\theta\mathbf{L}_y} e^{-i\psi\mathbf{L}_z} \equiv \mathbf{Z}(-\phi)\mathbf{Y}(-\theta)\mathbf{Z}(-\psi) \quad (2.4)$$

Here, the individual angular momentum operators are defined as:

$$\mathbf{L}_x \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{L}_y \equiv \frac{i}{\sqrt{2}} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{L}_z \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad (2.5)$$

and a short-hand notation is used to designate rotation about the corresponding axis, i.e.  $\mathbf{Z}(-\phi) \equiv \exp(-i\phi\mathbf{L}_z)$ , etc. In the propagator matrix, Eq. (2.2), the first and the third Euler angles of each Wigner rotation matrix are constant, thus reflecting the highly restrained covalent bond angles, while the second angle (corresponding to a rotation generated by the operator  $\mathbf{L}_y$ ) is variable. However, the non-diagonal form for  $\mathbf{L}_y$ , Eq. (2.5), indicates that the number of elements that one needs to calculate is much greater than those required for the diagonal

operator,  $\mathbf{L}_z$ . Consequently, by using the following transformation rules for the rotation operators,

$$\mathbf{X}\left(\frac{\pi}{2}\right)\mathbf{Y}(\theta)\mathbf{X}\left(-\frac{\pi}{2}\right)=\mathbf{Z}(\theta), \quad \mathbf{X}\left(\frac{\pi}{2}\right)\mathbf{Z}(\varphi)\mathbf{X}\left(-\frac{\pi}{2}\right)=\mathbf{Y}(-\varphi) \quad (2.6)$$

one can diagonalize the variable part for the rotations corresponding to the torsion angles  $\Phi$ ,  $\Psi$ , and  $\omega$  in the propagator, Eq. (2.2). This is easily accomplished by inserting an extra transformation  $\mathbf{T} = \mathbf{X}(\pi/2)$  in the quadratic form, yielding the transformed  $Q$ -basis:

$$\nu = \vec{Q}(\beta, \alpha) [\mathbf{D}_Q(\Omega_{MP}) \mathbf{M}_Q \mathbf{D}_Q^{-1}(\Omega_{MP})] \vec{Q}^+(\beta, \alpha) \quad (2.7)$$

Here the spherical harmonics  $\vec{Q}(\beta, \alpha)$  replace  $\vec{Y}(\beta, \alpha)$  of Eq. (2.1) and together with the transformed matrices  $\mathbf{D}_Q(\Omega_{MP})$  and  $\mathbf{M}_Q$  are given by:

$$\begin{aligned} \vec{Q}(\beta, \alpha) &= \vec{Y}(\beta, \alpha) \mathbf{T} \\ \mathbf{D}_Q(\Omega_{MP}) &\equiv \mathbf{T}^{-1} \mathbf{D}(\Omega_{MP}) \mathbf{T}, \quad \mathbf{M}_Q \equiv \mathbf{T}^{-1} \mathbf{M} \mathbf{T} \end{aligned} \quad (2.8)$$

For any CSA, the interaction matrix  $\mathbf{M}$  can be expressed in terms of the principal components  $\sigma_{11}$ ,  $\sigma_{22}$  and  $\sigma_{33}$  ( $\sigma_{33} > \sigma_{22} > \sigma_{11}$ ) as:

$$\mathbf{M} = \begin{pmatrix} \frac{\sigma_{11} + \sigma_{22}}{2} & 0 & \frac{\sigma_{22} - \sigma_{11}}{2} \\ 0 & \sigma_{33} & 0 \\ \frac{\sigma_{22} - \sigma_{11}}{2} & 0 & \frac{\sigma_{11} + \sigma_{22}}{2} \end{pmatrix} \quad (2.9)$$

For instance, for  $^{15}\text{N}$  CSA, the corresponding Euler angles are given by:  $\Omega_{MP}^{15N} = (\gamma_N, \pi/2, \pi/2 + \alpha_N)$ , where  $\gamma_N=15-20^\circ$  is the angle between the NH bond and the  $\sigma_{33}$  axis of the  $^{15}\text{N}$  CSA,  $\alpha_N=0-25^\circ$  is the angle between the  $\sigma_{22}$  axis of  $^{15}\text{N}$  CSA and the normal of the peptide plane; for  $^1\text{H}$  CSA,  $\Omega_{MP}^{1H} = (-\pi/2, -\pi/2, \pi/2)$ . It should be noted, however, that in real polypeptides and proteins, the orientation of the  $^1\text{H}$  CSA frame is highly variable (Cornilescu and Bax 2000).

For the heteronuclear dipolar couplings, the interaction matrix is the product of the corresponding dipolar coupling constant  $\chi \equiv \mu_0 \gamma_1 \gamma_2 h / 16 \pi^3 r_{12}^3$  (in Hz) and a diagonal matrix:

$$\mathbf{M} = \chi \begin{pmatrix} -1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{pmatrix} \quad (2.10)$$

where  $\gamma_1$  and  $\gamma_2$  are the gyromagnetic ratios of the two interacting spins, and  $r_{12}$  is the interspin distance. For the  $^1\text{H}$ - $^{15}\text{N}$  dipolar couplings, the related Euler angles are given by  $\Omega_{MP}^{1H-15N} = (0, \pi/2, 0)$ ; for  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings, the Euler angles for the corresponding transformation are given by  $\Omega_{MP}^{13C_\alpha-15N} = (-\alpha_{HNC_\alpha}, \pi/2, 0)$ , where  $\alpha_{HNC_\alpha}=118.2^\circ$ ; for  $^{13}\text{C}'$ - $^{15}\text{N}$  dipolar couplings, the Euler angles are given by  $\Omega_{MP}^{13C'-15N} = (\alpha_{HNC'}, \pi/2, 0)$ , where  $\alpha_{HNC'}=119.5^\circ$ .

The inclusion of the  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  dipolar couplings, which represent an additional chiral restraint, is necessary to further reduce the number of the orientational solutions. The

corresponding transformation associated with the  $^1\text{H}_\alpha\text{-}^{13}\text{C}_\alpha$  dipolar couplings is given by (Nevzorov and Opella 2006):

$$\mathbf{D}\left(\Omega_{MP}^{^1\text{H-}^{13}\text{C}_\alpha}\right) = \mathbf{D}\left(\alpha_{NC_\alpha}, \Phi - \pi/3, \pi/2 - \gamma_{ideal}\right) \mathbf{D}(0, -\pi/2, 0) \quad (2.11)$$

It should be emphasized that the overall matrix in the square bracket given by Eq. (2.7) as well as all the non-diagonal rotation matrices need to be calculated only once, and the protein

chain geometry is contained in the spherical harmonics  $\vec{Q}(\beta, \alpha)$ , which vary throughout the backbone via the new recurrence relation:

$$\vec{Q}(\beta_{n+1}, \alpha_{n+1}) = \vec{Q}(\beta_n, \alpha_n) \mathbf{P}_Q(\Phi_n, \Psi_n, \omega_n) \quad (2.12)$$

where  $\mathbf{P}_Q(\Phi, \Psi, \omega) = \mathbf{T}^{-1} \mathbf{P}(\Phi, \Psi, \omega) \mathbf{T}$ .

Equations (2.7-2.12) can be used to determine the torsion angles  $\Phi$  and  $\Psi$  associated with the adjacent residues along the polypeptide backbone iteratively in a sequential manner. Experimental uncertainty can be treated by randomizing the values for the input frequencies within a pre-defined range (in Hz) in order to obtain different structural solutions for each fit within a pre-defined tolerance. The solutions for the torsion angles  $\Phi$  and  $\Psi$  are further restricted by the differential Ramachandran plots (Lovell, Davis et al. 2003) to automatically keep only the allowed regions. The normalized Ramachandran plots have been chosen by residue type, and include glycines, prolines, pre-prolines, and the general-type plots (De Angelis, Howell et al. 2006) as shown in Figure 2.2 (with the cutoff for the lowest contour chosen as  $10^{-3}$ ).

The algorithm [Eqs. (2.7-2.12)] depicted by the flowchart of Fig. 2.3 has been programmed in MATLAB (Mathworks, Inc.), and the solutions for the torsion-angles are found by minimizing the difference between the back-calculated frequencies and the simulated NMR data from the known PDB structures. In minimizing the root-mean square (rms) deviations relative to the input frequencies, all CSA and dipolar couplings have been scaled so that their maximum values would correspond to that of the  $^1\text{H}$ - $^{15}\text{N}$  coupling (10 kHz) to achieve equal weight in the fitting. If a pair of  $\Phi$  and  $\Psi$  satisfying the experimental restraints and the pre-defined tolerance (i.e. falling within the “experimental uncertainty”, in Hz) is found, the search will proceed to the next pair; otherwise, if no acceptable solution is found, the algorithm will go back several residues (typically 3-5 in the calculations) to search for different solutions. Since MATLAB employs a Simplex-type minimizer (i.e. *fminsearch*), the starting values for the torsion angles are randomized at each residue except those in  $\alpha$ -helical regions from  $-180^\circ$  to  $+180^\circ$  to find possible alternative solutions for every iteration. For residues in the  $\alpha$ -helical regions,  $\Phi_0 = -60 \pm 30^\circ$ ,  $\Psi_0 = -45 \pm 30^\circ$ . To avoid the algorithm getting stuck at a certain unfavorable combination of the input data or at a certain residue an automatic restart has been implemented. Any number of frequency constraints can be introduced in the calculations.

## 2.3 Results and Discussion

### 2.3.1 Structural Fitting of Simulated Three-Dimensional Spectra with $^{15}\text{N}$ CSA, and $^1\text{H}$ - $^{15}\text{N}$ and $^1\text{H}$ - $^{13}\text{C}_\alpha$ Dipolar Couplings

Relative convenience of the uniform  $^{15}\text{N}$  labeling has rendered two-dimensional solid-state NMR spectra involving  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar couplings as a useful tool for the studying of the helical domains of membrane proteins. However, just two spectral dimensions are generally insufficient to calculate tertiary structures of arbitrary topology solely from the oriented-sample NMR data; an important additional angular constraint, the  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  dipolar coupling, needs to be introduced (Nevzorov and Opella 2006). A simulated three-dimensional solid-state NMR spectrum of protein G (PDB ID: 2GB1) is shown in Figure 2.4A. When the experimental uncertainty is assumed to be only 0.01Hz in all dimensions, and no randomization of the input data is performed, back calculation of 100 structures yielded a converged set of solutions as shown in Figure 2.4B. The calculated RMSD's relative to the initial structure are about  $10^{-5}\text{\AA}$ , indicating a mathematically unique solution. Based on the analysis of the structure of protein G (PDB file 2GB1), the torsion angle  $\omega$  is assumed here to be  $178^\circ$  to describe the deviation from the peptide planarity. The torsion angles  $\Phi$  and  $\Psi$  have been restricted by residue type as described in the previous Section, and  $\Phi$  and  $\Psi$  associated with residues 11 and 37 (our numbering), which correspond to the original residues Glu15 and Gly41, have not been restricted as they would be considered outliers.

However, when just the  $^{15}\text{N}$  CSA dimension of the input data is randomly varied within  $\pm 1$  ppm or  $\pm 50$  Hz (at 500 MHz  $^1\text{H}$  NMR frequency), RMSD's of some of the calculated

structures become more than  $10\text{\AA}$  as shown in Fig. 2.5. This variation in the CSA was chosen to reflect the above-mentioned uncertainties in both the principal components and, especially, the orientation of the CSA tensor (see Appendix A for additional simulations). Even a slight randomization of the angle  $\gamma_N$  to within  $\pm 1^\circ$  results in a highly non-uniform variations of the observed  $^{15}\text{N}$  CSA frequencies of up to  $\pm 3$  ppm; To assess the effect of experimental uncertainty in the tensor parameters only, the input data have been simulated by assuming the average tensor values for all amino acids, and the back-calculation has been performed by using the same tensor values. An average uniform  $\pm 1$  ppm uncertainty was assumed for representative simulations. A two-dimensional projection of the simulated three-dimensional spectra shown in Fig. 2.5A demonstrates that such small variations in the CSA frequencies (the dots show the fitted values with 100 Hz tolerance) with respect to the original input data (circles) cause a complete scatter in the calculated structures (100 representative structures are shown in Figure 2.5B). Such a divergence in the structural solutions is most likely due to the multitude of possible orientational solutions for the weakly restrained torsional conformations in the loop regions that are uniaxially distributed around the magnetic field, as was mentioned in the Introduction.

### **2.3.2 Structural Fitting of the Three-Dimensional “Shiftless” Spectra with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ Dipolar Couplings**

On the other hand, heteronuclear dipolar couplings, which only depend on the covalent bond angles and lengths of the relevant atoms, would seem to represent a much more reliable constraint for the structural fitting than the CSA. The average bond lengths and angles are highly conserved and can be accurately determined (Engh and Huber 1991). Therefore, the

$^{15}\text{N}$  CSA has been replaced by the  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings to examine the performance of the “shiftless” (Sinha, Grant et al. 2007; Wu and Opella 2008) solid-state NMR data for the structural fitting of protein G. Figure 2.4C shows the simulated spectrum. Just as in the case of uniform  $^{15}\text{N}$  CSA, a mathematically unique solution as in Fig. 2.4B can be obtained from the back calculation of the simulated spectra without taking into account any experimental uncertainty. However, when an “experimental uncertainty” of 15 Hz is assumed in all three dimensions, the back-calculated structures begin to diverge (Fig. 2.6A). More than one half of the RMSDs are around 7Å. To illustrate the quality of the fit to the synthetic NMR data, we also plot the rms deviations of the fitted frequencies in Hz (over 1000 fits) relative to the “experimental” (i.e. calculated from the PDB coordinates) data for each residue as shown in Fig. 2.6B. The average rms deviation for all residues is around 7 Hz, and residues 27, 31 (at the ends of the  $\alpha$ -helix; our numbering) and 44 (in a  $\beta$ -turn) show somewhat larger deviations.

### 2.3.3 Structural Fitting of Four-Dimensional Data with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ , $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ and $^{13}\text{C}'$ - $^{15}\text{N}$ Dipolar Couplings

An additional dimension, the  $^{13}\text{C}'$ - $^{15}\text{N}$  dipolar interaction, has been included to improve the convergence of the calculated structures in the presence of “experimental” uncertainty in the measured couplings. When the data uncertainty is assumed to be 15 Hz, a converged set of solutions containing 1000 structures can be obtained. The structural RMSDs are less than 2Å as shown by the histogram of Fig. 2.7A. The average deviation of the calculated frequencies from the simulated “experimental” data is around 7Hz, while residues 11, 16, 33 and 44 (our numbering) are consistently fit less well. The location of residues 16, 33 and 44 is in the  $\beta$ -

turn, and residue 11 is at the beginning of a  $\beta$ -sheet, respectively. In OS NMR, the accuracy of the measured dipolar couplings ultimately depends on the stability of the dipolar scaling factor and experimental linewidths (about 200 Hz), so even greater uncertainties may need to be considered. When the tolerance for the fitting of the input data is increased to 25Hz (and up to 50 Hz) in all four dimensions, the back-calculated structures begin to diverge, and about 40% of the RMSDs are around 7Å or greater with the average deviation per residue of 12 Hz and higher (see Appendix A).

Accurate fitting of the residues in the linkage regions or at the beginning/end of an  $\alpha$ -helix or  $\beta$ -sheet is important since small deviations at these locations could lead to an entirely different tertiary conformation. Therefore, more information pertaining to the linkage residues may be required to obtain convergent structural solutions for proteins consisting of several secondary structure elements. Specifically, checking for satisfied hydrogen bonds, hydrophobic restraints (Standley, Eyrich et al. 1999), and electrostatic potential (Dong, Olsen et al. 2008) of the entire protein structure within the membrane may be helpful in eliminating the unfavorable solutions (Im, Feig et al. 2003; Dong, Olsen et al. 2008; Shi, Traaseth et al. 2009) and treating the data for the more dynamic loop regions. Moreover, the ab-initio Rosetta method (Bonneau, Tsai et al. 2001), may also be helpful in selecting the plausible backbone conformations by comparing the calculated segments with the known protein structures of similar sequences.

Additional important restraints include long-range distance constraints (Standley, Eyrich et al. 1999), which are not yet directly accessible in OS NMR but are obtainable from the methyl-methyl NOE data (Gardner, Rosen et al. 1997) in the liquid state, site-directed

labeling EPR (Borbat and Freed 1999; Altenbach, Kusnetzow et al. 2008) and MAS (Jaroniec, Tounge et al. 2001; Luo, Mani et al. 2007). As an illustrative example, we have calculated structures from the simulated 3D data including the  $^1\text{H}$ - $^{15}\text{N}$ ,  $^1\text{H}$ - $^{13}\text{C}_\alpha$ , and  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings and two distance restraints between the  $\text{C}_\alpha$  atoms of residues Ile6 and Thr53 and residues Glu15 and Thr44 (as determined from the original structure 2GB1; see Appendix A for additional information). A structure was accepted if both distance restraints were satisfied within  $\pm 1\text{\AA}$ . Figure 2.8 shows that with a tolerance of 25 Hz, 312 structures (out of 1000 total) satisfying the above distance restraints have the RMSD's of less than  $2\text{\AA}$ . When the tolerance is further increased to 50 Hz, 33 out of 1000 back-calculated structures that satisfy the distance restraints have the RMSDs less than  $2.5\text{\AA}$  relative to the original structure. Therefore, the number of unfavorable structures can be reduced dramatically by including the distances between two or more pairs of  $\text{C}_\alpha$  atoms as an additional restraint even with larger uncertainties in the input data. The optimal positioning of these restraints depends on the expected protein topology, but longer-distance restraints are generally more stringent (Standley, Eyrich et al. 1999).

#### **2.3.4 Structural Fitting of a Helical Hairpin Derived from Bacteriorhodopsin**

To illustrate the general applicability of the algorithm to other topologies, we have applied it to the simulated data for the light-driven proton pump bacteriorhodopsin, which could be considered as a structural paradigm for G-protein coupled receptors. Bacteriorhodopsin from *Halobacterium salinarum* is composed of seven  $\alpha$ -helical transmembrane chains, and was solved by X-ray crystallography with a resolution of  $1.55\text{\AA}$  (Luecke, Schobert et al. 1999) (PDB ID 1C3W). Two  $\alpha$ -helices (containing residues Asp104-Gly155) have been chosen for

a representative calculation. Three-dimensional “shiftless” spectra including  $^1\text{H}$ - $^{15}\text{N}$ ,  $^1\text{H}$ - $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings have been simulated. The torsion angles have been restricted by the Ramachandran plots having the lowest probability cutoff of  $1 \times 10^{-3}$  for the loop region (residues Leu127-Val130), and within  $\pm 30^\circ$  relative to  $\Phi = -60^\circ$ ,  $\Psi = -45^\circ$  for the two helical regions. Interestingly, when the helical hairpin is preferentially aligned about the z-axis (cf. Fig. 2.9A), which would correspond to its native orientation in the membrane, back-calculation from the synthetic spectrum yields a converged set of solutions with the tolerance up to 120 Hz. (Results with higher tolerances for the input data are presented in Appendix A.) By contrast, when the original structure is rotated by 90 degrees (cf. Fig. 2.9C), the back-calculated structures begin to diverge with a tolerance of only 25 Hz. Figure 2.9C shows the top view of the back-calculated structures to illustrate the scatter in the structural solutions. This indicates that error propagates differently into the calculated structures depending on their overall orientation, as expected for inhomogeneously broadened frequency anisotropies. It should be also noted that the analysis using MOLMOL 2.0 (Koradi, Billeter et al. 1996) of the PDB structure 1C3W has revealed that the twist angle  $\omega$  for the peptide linkages is around  $176^\circ$ , especially in the connecting loop region. This may imply that deviations from planarity in membrane proteins may need to be treated more differently than for soluble proteins, which exhibit the average value of around  $178^\circ$ . This merits additional investigation as more high-resolution data on membrane proteins become available in the future.

## 2.4 Conclusions

We have examined the influence of the variations in the magnitude and orientation of angular-dependent solid-state NMR observables on the structural fitting of macroscopically aligned proteins. Even though the principal components of the  $^{15}\text{N}$  CSA can be determined with high precision, a somewhat undefined orientation of the  $^{15}\text{N}$  CSA tensor results in a considerable divergence of the calculated protein structures.

Nevertheless, multidimensional “shiftless” solid-state NMR spectra involving solely the heteronuclear dipolar couplings associated with the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  spin-bearing nuclei can perform as an efficient tool for the determination of three-dimensional protein structures. This was illustrated by obtaining converged sets of solutions from the synthetic datasets for protein G (a soluble protein) and for bacteriorhodopsin (a transmembrane protein) with a simulated experimental uncertainty of up to 15 Hz and 120 Hz, respectively. Non-planarity of the peptide unit as defined by the third torsion angle,  $\omega$  has been included in the present version of the algorithm. Moreover, the formulation of the algorithm entirely in the torsion-angle space simultaneously provides both the Ramachandran (Laskowski 1993) and  $\text{C}_\alpha$ -geometry validation (Lovell, Davis et al. 2003) “on-a-fly”. At present, conformations of transmembrane regions having around 50 residues can be calculated from OS NMR data. With the addition of more constraints, especially those derived from the heteronuclear couplings to  $^{13}\text{C}$  spins and long-range distance restraints, “shiftless” solid-state NMR spectra will have the potential of resolving the structures of more complex polypeptides and proteins solely based on the experimental data even in the presence of larger experimental uncertainties.

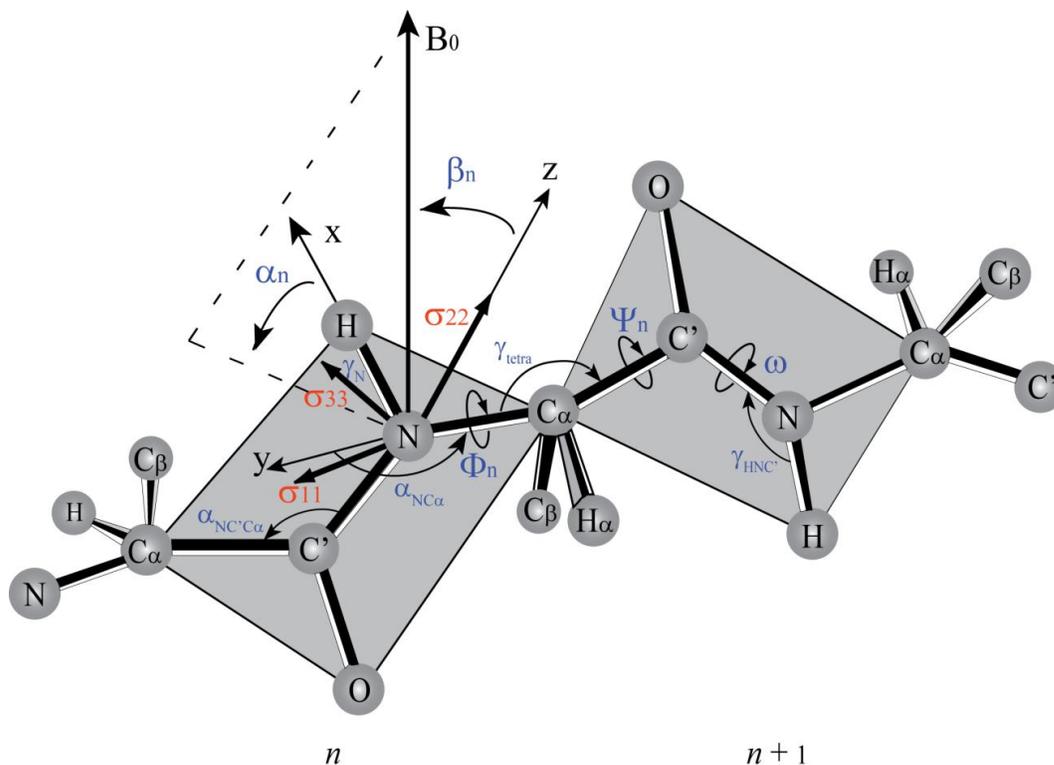


Figure 2.1 Local molecular frame associated with a peptide plane. The orientation of the main magnetic field  $\mathbf{B}_0$  relative to the  $n$ 'th peptide plane is described by two angles  $\alpha_n$  and  $\beta_n$ . An approximate orientation of the  $^{15}\text{N}$  CSA tensor relative to the peptide plane is represented by arrows originating from the N atom. Two adjacent amino acid residues are linked by the torsion angles  $\Phi_n$  and  $\Psi_n$ . Deviation from peptide planarity is given by the angle  $\omega$ . The angle  $\alpha_{NC_\alpha}$  denotes the angle between the N- $C_\alpha$  bond and the y-axis of the MF corresponding to the  $n$ 'th plane,  $\gamma_N$  is the angle between the NH bond and the  $\sigma_{33}$  axis of the  $^{15}\text{N}$  CSA,  $\gamma_{\text{tetra}}$  is the tetrahedral angle,  $\gamma_{C'C}$  is the angle between the  $C'-C_\alpha$  bond and the y-axis of MF of the  $(n+1)$ 'th plane,  $\alpha_{NC_\alpha}$  and  $\gamma_{HNC'}$  are given by the standard peptide plane geometry.

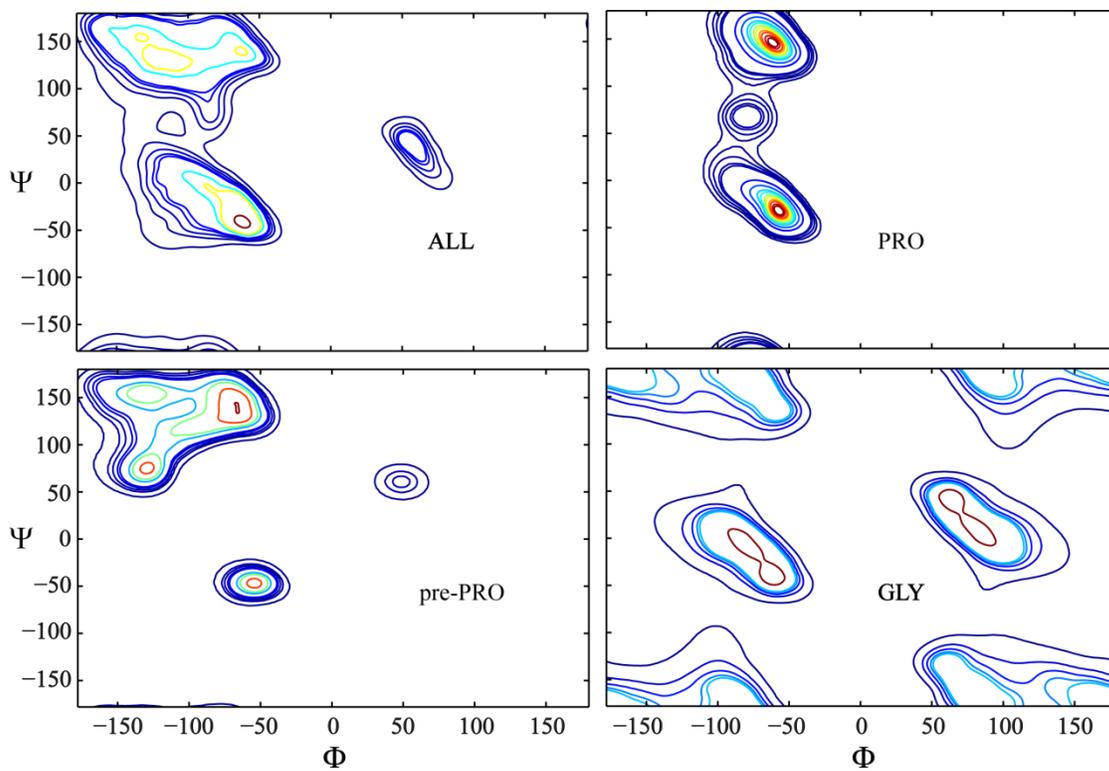


Figure 2.2 Normalized Ramachandran plots for different residue types, including glycines, prolines, pre-prolines and the general-type residues with the lowest probability density cutoff of  $10^{-3}$ .

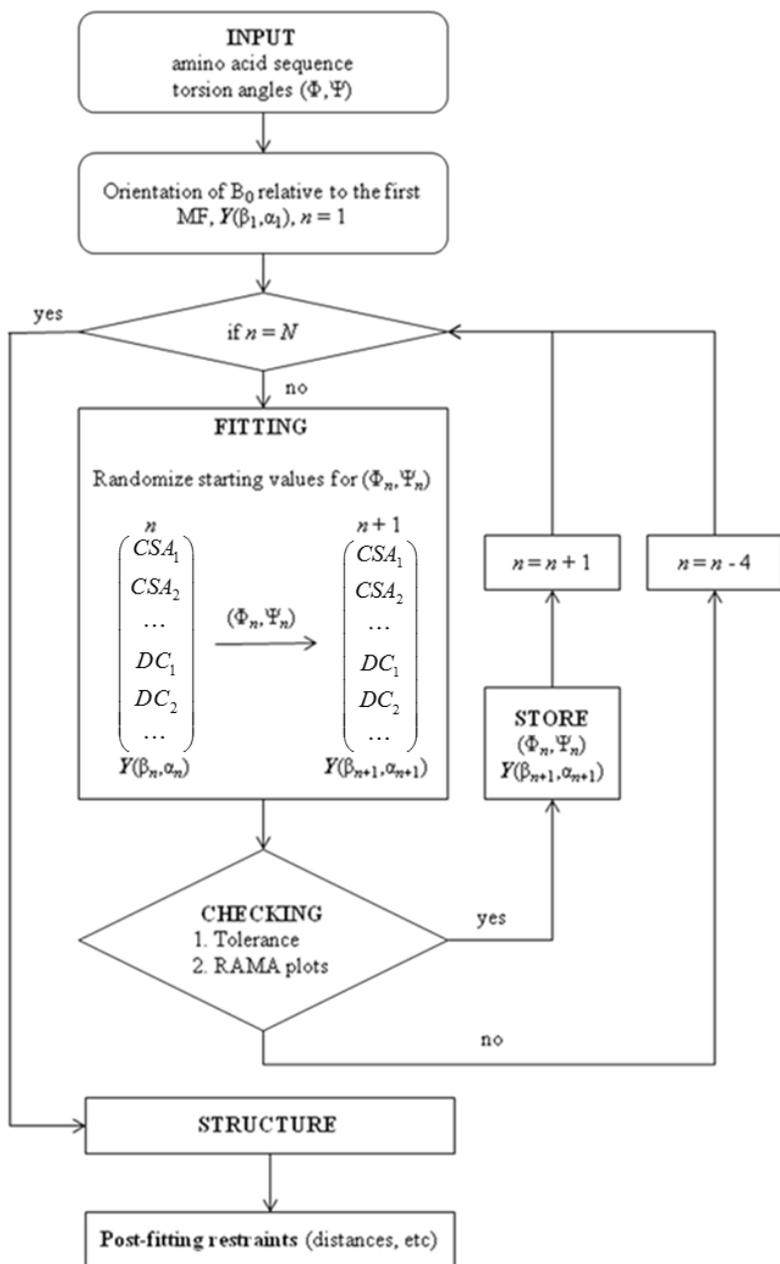


Figure 2.3 A flowchart for the structural fitting algorithm. The figure describes the procedure for structure determination by minimizing the difference between the calculated data and the synthetic data generated from the PDB file including the Ramachandran-plot restrictions for the torsion angles  $\Phi$  and  $\Psi$  and post-fitting filtering. For further details cf. the text.

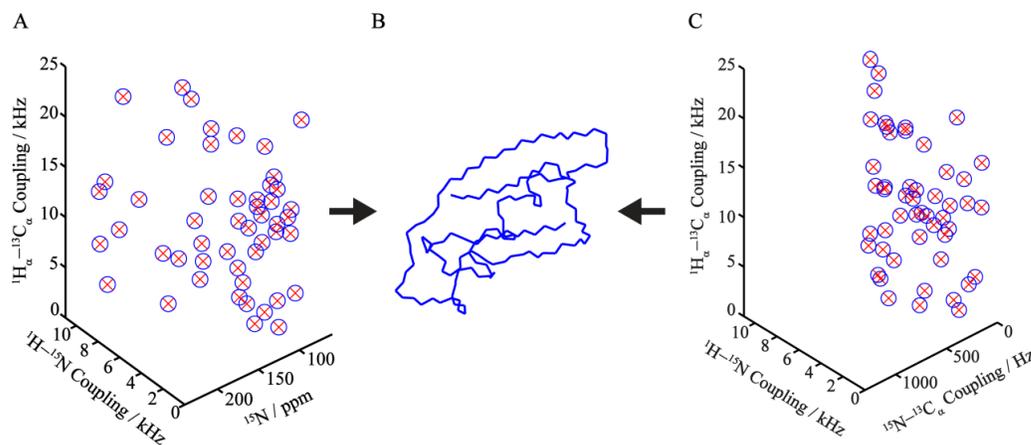


Figure 2.4 (A) Structural fitting of the simulated three-dimensional solid-state NMR spectrum of protein G obtained from PDB coordinates 2GB1 including  $^{15}\text{N}$  CSA,  $^1\text{H}\text{-}^{15}\text{N}$  dipolar couplings and  $^1\text{H}_\alpha\text{-}^{13}\text{C}_\alpha$  dipolar couplings. No uncertainty in the input data (circles) is considered. The fitted frequencies are depicted as crosses. (B) A mathematically unique structural solution is obtained that coincides with the original structure. (C) Fitting of simulated three-dimensional “shiftless” solid-state NMR spectra created by substituting the  $^{15}\text{N}$  CSA dimension by the  $^{13}\text{C}_\alpha\text{-}^{15}\text{N}$  dipolar couplings. No experimental uncertainty is assumed. The same unique solution as from fitting the data in part A is obtained.

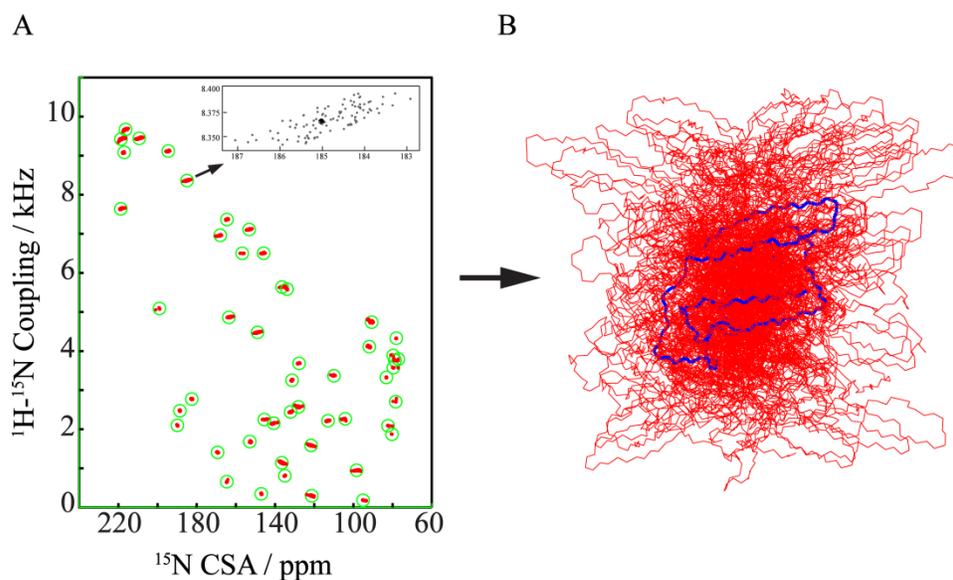


Figure 2.5 (A) Two-dimensional projection of the simulated spectra (circles) with the same dimensions as in Fig. 2.4 with  $^{15}\text{N}$  CSA input data randomly varied within 1 ppm for each simulation. The tolerance for the fitting of the input data is 100Hz. The fitted frequencies are given by dots (no randomization was applied to the other two dimensions). Inset shows a magnified distribution of the fitted frequencies. (B) Back-calculated structures deviate substantially for each simulation. The RMSDs for some of the calculated 100 structures relative to the original structure are more than 10 Å.

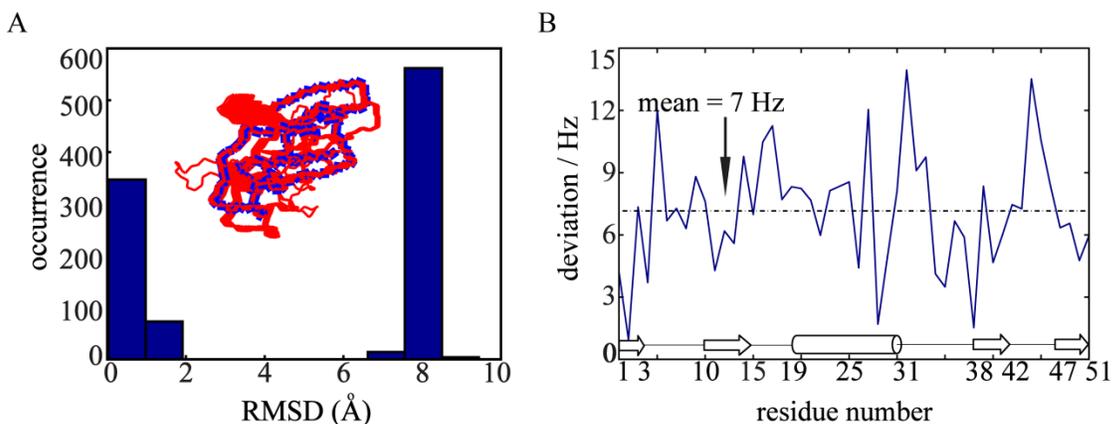


Figure 2.6 (A) Structural fitting of simulated three-dimensional “shiftless” solid-state NMR spectra containing  $^1\text{H}$ - $^{15}\text{N}$ ,  $^1\text{H}_\alpha$ - $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings. Histogram of the RMSDs relative to the starting structure shows that when the tolerance for the fitting of the input data is 15 Hz, more than one-half of the RMSDs for 1000 calculated structures are around 7 Å. (B) The root-mean square deviation (in Hz) at each residue for 1000 fits with respect to the synthetic data simulated from the PDB coordinates. The cylinder and the arrows at the bottom represent the  $\alpha$ -helix and two  $\beta$ -sheets, respectively.

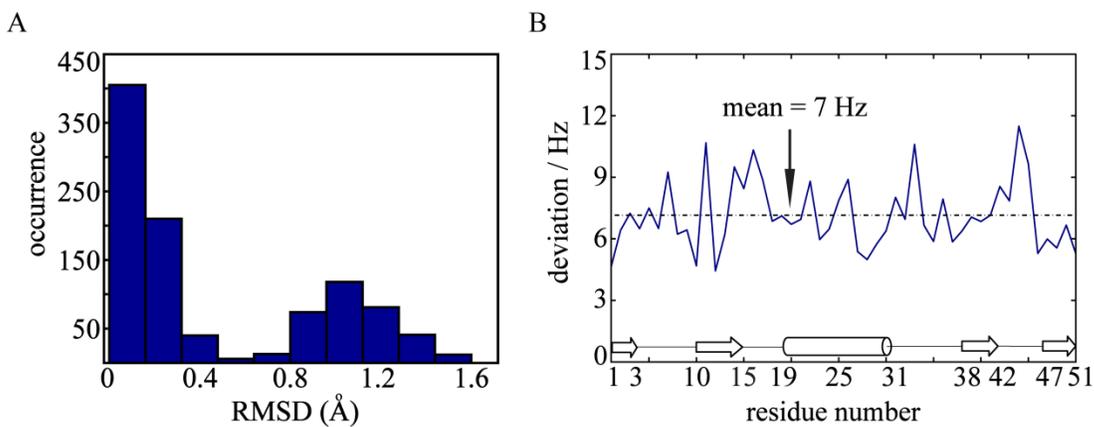


Figure 2.7 (A) Structural fitting of simulated four-dimensional data for protein G including additional  $^{13}\text{C}'$ - $^{15}\text{N}$  dipolar couplings. With the tolerance for the fitting of the input data of 15 Hz, all back-calculated structures are converged with RMSDs less than  $2\text{\AA}$ . (B) The root-mean square deviations (in Hz) at each residue for 1000 fits with respect to the simulated data from the PDB coordinates.

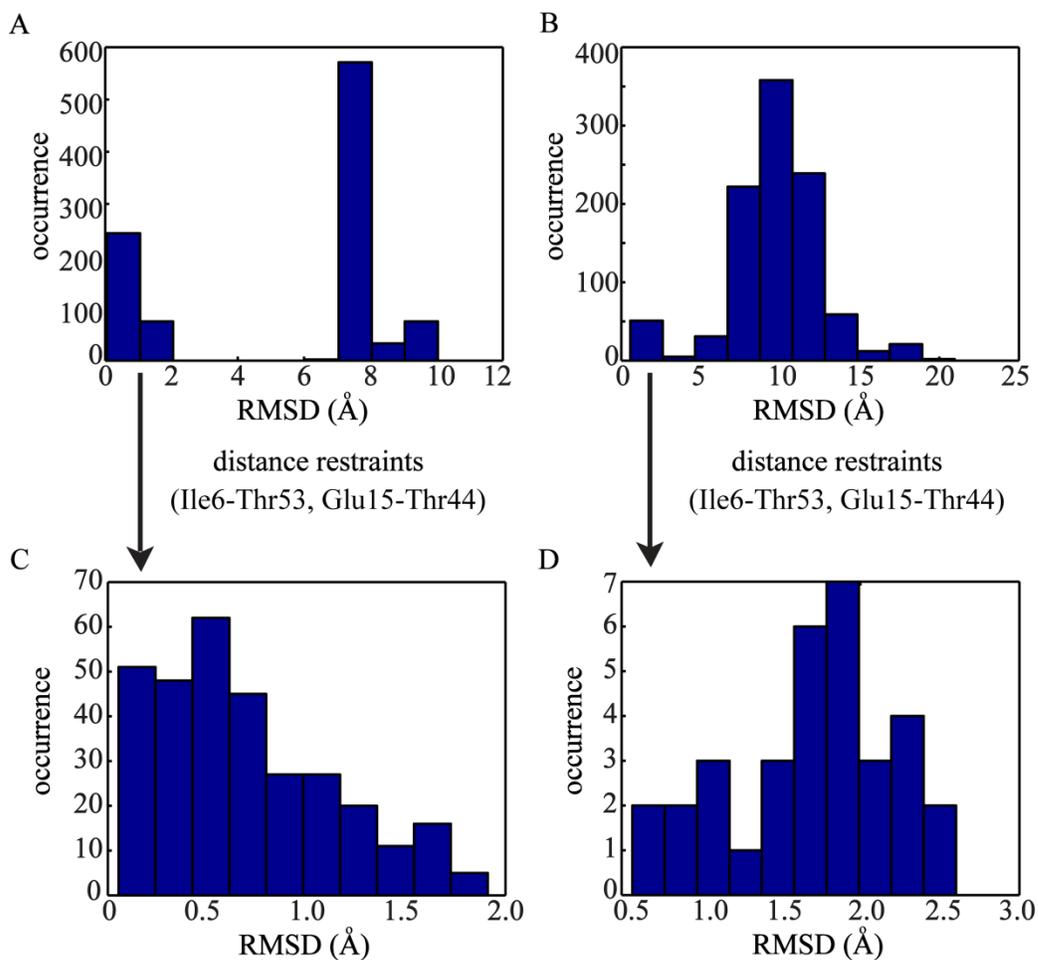


Figure 2.8 (A) and (B): Histograms of the RMSDs of 1000 back-calculated structures obtained from three-dimensional “shiftless” solid state NMR spectra of protein G (2GB1) with a tolerance of 25 Hz (A) and 50 Hz (B). (C) and (D): Histograms of the RMSDs for the structures satisfying the distance restraints between the  $C_{\alpha}$  carbons of residues Ile6 and Thr53, and of residues Glu15 and Thr44 for the 25 Hz tolerance (C) and 50 Hz tolerance (D). The inclusion of distance restraints dramatically improves the convergence of the fits with larger tolerances (greater experimental uncertainties).

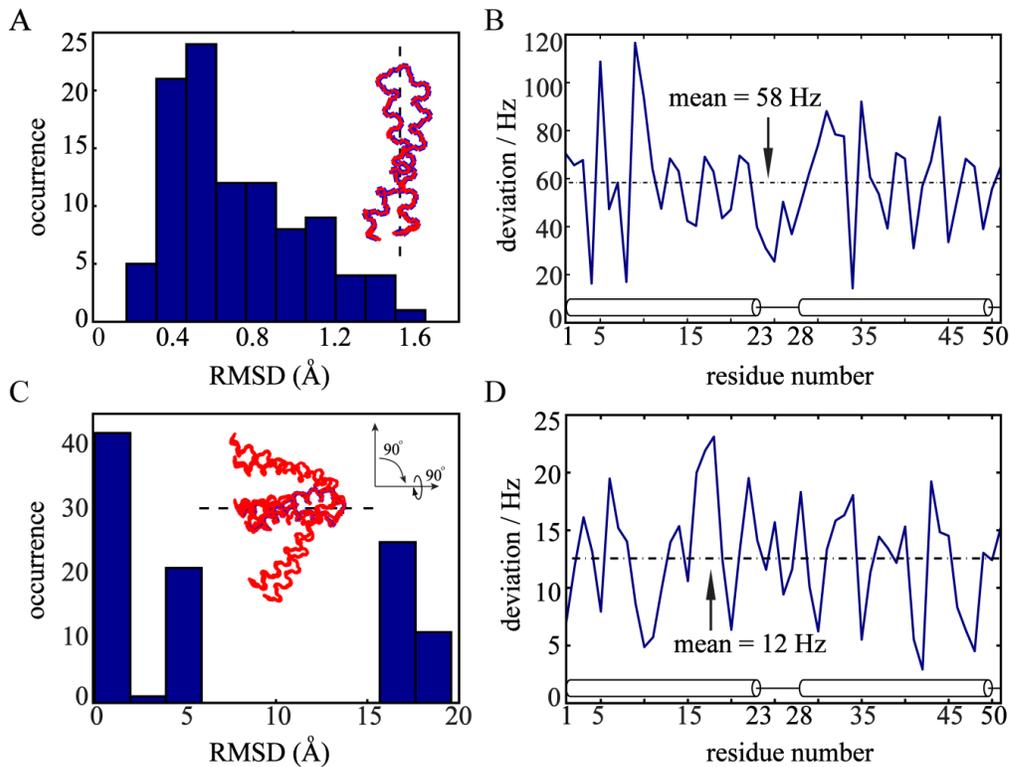


Figure 2.9 (A) Structural fitting of simulated three-dimensional “shiftless” spectra of two transmembrane  $\alpha$ -helices (residues 104 to 155) of bacteriorhodopsin. When the initial structure is predominantly aligned about the z-axis, with the tolerance for the fitting of the input data of 120Hz back calculation yields a converged set of solutions having RMSDs of less than  $2\text{\AA}$ . (B) The mean of the deviation (in Hz) at each residue for 100 calculated structures with respect to the simulated data from the PDB coordinates. The two cylinders at the bottom depict the  $\alpha$ -helices. (C) When the initial structure is rotated by about 90 degrees, with the tolerance of 25Hz, about 60% of the RMSDs for the back-calculated structures are more than  $5\text{\AA}$ . Rotation arrows show the view angles of the calculated structures relative to the original orientation. (D) The rms deviation (in Hz) at each residue for 100 calculated structures shown in part C with respect to the simulated data.

## **Chapter 3**

### **Validation of SSNMR Structures using Rosetta**

#### **3.1 Introduction**

Due to the degeneracy of orientational solutions given by the quadratic form of the mathematical expressions for the observable frequencies and the inevitably present experimental error, there is always a large amount of uncertainty produced by structure calculations using solid state NMR experimental data alone. Therefore, it becomes challenging to choose the most topologically and biologically reliable model as the “native” structure even with torsion angles and distance restraints. Here, we will use knowledge-based Rosetta all-atom energy to screen the calculated structures and determine the lowest-energy set of structures. Since the membrane environment is described by an implicit solvent model and the solvation energy of all atoms depends on their vertical coordinate in this heterogeneous membrane-aqueous model, we can evaluate the optimal immersion depth of the target protein within the biological membrane. (Themis Lazaridis 2003) At the same time, the angular restraints from solid-state NMR will preserve the orientation of the individual peptide planes relative to the membrane.

#### **3.2 Application of Membrane Ab Initio Protocol in Rosetta**

First, the membrane ab initio protocol was applied to bacteriorhodopsin and a mercury transporter protein, which are the classic representatives in membrane protein research, to test the capabilities of Rosetta. The transmembrane region is predicted by the OCTOPUS

server (<http://octopus.cbr.su.se/>) with the primary sequence in FASTA format as input file. Then the transmembrane topology prediction file, i.e. SPAN file, is generated using the OCTOPUS file. The number of transmembrane helices and the start and end residue numbers of each single predicted transmembrane helix are shown in the SPAN file. Finally, with the primary sequence file and the SPAN file as input, the predicted lipophilicity file, i.e. LIPS4 file, is produced. The three- and nine-residue peptide fragments are generated on Robetta server (<http://rosetta.bakerlab.org/>) with the sequence file as input. All the RMSDs mentioned below correspond to  $C_{\alpha}$ -RMSDs, i.e. root mean square deviation between alpha carbon atoms after superimposing the reference structure and the target structure. Here, the unit of energy is not the conventional kcal/mol, but REU, i.e. the Rosetta energy unit. The knowledge-based potentials are statistical potentials and the real energy can be obtained by the reverse Boltzmann equation. (Chrysi Konstantinou Kirtay 2005) The structure with the lowest energy in REU is routinely considered to be the most stable one and, therefore, more similar to the native structure.

### **3.2.1 Membrane Ab Initio Calculation of Bacteriorhodopsin (1C3W, 1PY6)**

As a classic membrane protein representative, bacteriorhodopsin (PDB ID: 1C3W) (Luecke, Schobert et al. 1999) was first tested in membrane ab initio protocol. Because of the missing conformation between residue 156 and residue 162, we split this seven-helix transmembrane protein into two parts and named them as 1C3WD (first 5 helices) and 1C3WE (last 2 helices) separately. Here, we conducted a de novo protein folding using the membrane ab initio protocol in Rosetta3.4.

The amino acid sequence and OCTOPUS predicted topology files for 1C3WD are











residues as shown in Table 3.1. Membrane ab initio simulation was executed with all these modified OCTOPUS files and 5000 structures were generated in each calculation. The plots of energy versus RMSD and the predicted models with the lowest energy for each modified OCTOPUS topology are shown in Figure 3.6.

Negative energies appeared in subplots 8-15 where the modification of “*M*” to “*o*” in OCTOPUS topology file was applied to from residue Leu39 to residue Leu47. In 2H3O, the loop region is from residues Leu43 to Leu48. When the two helices are separated at Leu39, the first helix is short and the number of interacting helical residues are nearly half of the number in native structure, so the lowest energy obtained with this OCTOPUS topology file (around -20) is also about half of the lowest energy of the model close to the native structure (around -40). When the two helices are separated at residue Thr40, there are 12 residues assigned with “*M*” as the first helix. Since there are 3.6 residues in each turn of alpha helix, one new turn could be predicted in the first helix, and thus the interacting helical residues between two helices would increase and the Rosetta energy will decrease. Similarly, when the helix intersection appears at residue Gly41 or Tyr42, a new turn could be predicted in the first helix as well. As observed in subplots 9-11, the shape of the energy versus RMSD plots are similar but more predicted models with negative energies were produced by the membrane ab initio calculation while the predicted loop position went closer to the actual loop region. When the modification brought the corresponding residues into the real loop region, the RMSDs of the lowest-energy predicted models relative to 2H3O reduced to around 2.86 Å. Combining the results shown in the subplots in Figure 3.6, we can estimate the whole loop region and accordingly obtain an outline of the conformation of the target

membrane protein. Consequently, a modification in the OCTOPUS file can correct the occasional deviation in topology prediction and improve the quality of membrane ab initio prediction in Rosetta.

#### **3.2.2.4 Effect of Hydrogen Bonding on Conformation Prediction of Transmembrane Helices**

Hydrogen bonding is a driving force for membrane protein folding. The internal hydrogen bonding of backbone atoms in alpha-helices and beta-barrels lowers the energy cost of peptide bond partitioning in transferring from the polar water environment to the apolar environment of the lipids, and makes them stable conformational motifs in a membrane. In a hydrophobic membrane bilayer hydrogen bonds are much stronger due to no competitive hydrogen bonding potential from the solvent and, therefore, are significant in stabilizing transmembrane protein structures. (Fang Xiao Zhou 2000; White 2006; Bowie 2011)

In Rosetta all-atom energy function, each potential term has a weight factor. Here we modified the weight of hydrogen bonding potential term in the all-atom energy function to examine the effect of hydrogen bonding on local conformation refinement of transmembrane helices. 10000 structure decoys were generated in each calculation.

For structures predicted with original OCTOPUS file, the lowest RMSD to 2H3O decreased gradually from 10.352 Å to 4.882 Å as the weight of hydrogen bonding potential increased from one to one hundred fold of the original value (i.e. 1x1.16 to 100x1.16, where 1.16 is the original weight of hydrogen bonding potential item in the Rosetta all-atom energy function).

But for the structures predicted with modified OCTOPUS file, since it already holds a relatively stable conformation, enhancement in hydrogen bonding potential has little impact

on the improvement of quality of de novo structure prediction. The RMSD values of the refined structures stayed at around 3.5 Å. Accordingly, the hydrogen bonding potential can be used to evaluate the accuracy of predicted topology and improve the quality of protein folding when the initial topology is not correct.

### 3.3 Structure Validation by PISEMA Back-Calculation

Without experimental verification, a modeled structure with the “best” geometry and conformation as predicted by pure computation could not be deemed to be the one closest to the “native” structure. To check the reliability of simulated structures from membrane ab initio protocol in Rosetta, we back-calculated the PISEMA spectra of Rosetta-predicted MerFt to compare the back-calculated  $^{15}\text{N}$  chemical shift anisotropy and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling frequencies with their corresponding solid state NMR experimental data.

In membrane ab initio calculation, the orientation of the membrane normal is not fixed, so we had to rotate the membrane plane to parallelize the membrane normal to  $\mathbf{B}_0$  (Figure 3.8) before back-calculating the PISEMA spectrum.

In the Rosetta output file, we can find the coordinates of points N, M, P and Q. The vector from O, the origin of the coordinate system, to N represents the membrane normal vector. M, P and Q determine a straight line parallel to vector  $\mathbf{ON}$ , where P is the membrane center, M and Q are 15Å away from P, separately. The red lines perpendicular to  $\mathbf{MPQ}$  stand for the membrane boundary. With the coordinates of N, M, P and Q we can turn the membrane horizontally via 1) translation of P to origin O, 2) rotation around z axis by  $\phi$ , and 3) rotation around y axis by  $\theta$ . The definition of  $\phi$  and  $\theta$  follows the definition in the traditional spherical

coordinate system shown in Figure #. The new coordinates of all atoms after membrane horizontalization can be calculated by Eq. 3.1

$$V_{new} = R_y(\theta)R_z(\phi)(V_{old} - P) \quad (3.1)$$

Here,

$$R_y(\theta) = \begin{pmatrix} \cos(\pi - \theta) & 0 & \sin(\pi - \theta) \\ 0 & 1 & 0 \\ -\sin(\pi - \theta) & 0 & \cos(\pi - \theta) \end{pmatrix} \quad R_z(\phi) = \begin{pmatrix} \cos(-\phi) & -\sin(-\phi) & 0 \\ \sin(-\phi) & \cos(-\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\theta = \arccos((M_z - P_z)/15)$$

$$\phi = \arctan((M_y - P_y)/(M_x - P_x))$$

The coordinates are expressed as  $M=(M_x; M_y; M_z)$  and  $P=(P_x; P_y; P_z)$ .

Following this rotational transformation, the PISEMA spectrum of the Rosetta-predicted lowest energy protein structure can be calculated within the new coordinates. The amide NH bond length is assigned 1.01 Å as the value used by Rosetta. The values of the principal components of  $^{15}\text{N}$  CSA for all amino acid residues except glycine and proline are  $\sigma_{11\text{N}}=64$  ppm,  $\sigma_{22\text{N}}=77$  ppm,  $\sigma_{33\text{N}}=217$  ppm. For glycine these values are  $\sigma_{11\text{NG}}=41$  ppm,  $\sigma_{22\text{NG}}=64$  ppm,  $\sigma_{33\text{NG}}=210$  ppm, and for proline  $\sigma_{11\text{NP}}=38$  ppm,  $\sigma_{22\text{NP}}=127$  ppm,  $\sigma_{33\text{NP}}=231$  ppm. (Kevin W. Waddell 2005) The angle between the NH bond and the  $\sigma_{33}$  axis of the  $^{15}\text{N}$  chemical shift tensor is  $18.5^\circ$ . The angle between the  $\sigma_{22}$  axis of  $^{15}\text{N}$  CSA and the normal of the peptide plane is  $25^\circ$  for all residues.  $\alpha_{\text{HNC}_\alpha}=118.2^\circ$ ,  $\alpha_{\text{HNC}'_\alpha}=119.5^\circ$ ,  $\alpha_{\text{NC}'_\alpha}=115.6^\circ$ . The tetrahedral angle is  $110.5^\circ$ . The chemical shift anisotropy and heteronuclear dipolar coupling frequencies

for the samples in the perpendicular bicelles can be converted from frequencies of the sample in the lipid bilayer by Eq. 3.2.

$$\nu_{bicelle} = \left(1 + \frac{S_0}{2}\right) * (\sigma_{11} + \sigma_{22} + \sigma_{33}) / 3 - \frac{S_0}{2} * \nu_{bilayer} \quad (3.2)$$

Here,  $S_0$  is the order parameter (Isabelle Marcotte 2005) of bicelle, which is assigned to 0.77 in the calculation. For chemical shift anisotropy,  $\sigma_{11}$ ,  $\sigma_{22}$  and  $\sigma_{33}$  are the principal CSA tensor values of the components of the corresponding amino acid residues, while for heteronuclear dipolar couplings their values are zero and the absolute value of the second term in Eq. 3.2 is adopted.

The  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling of MerFt (De Angelis, Howell et al. 2006) obtained in solid state NMR experiments are listed in Table 3.2 . Obvious deviation between calculated and experimental data can be observed at loop residues. (Figure 3.9) Though the calculated frequencies of helical residues form a similar PISA wheel to the experimental data, the correlation parameters for both  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling are only around 0.5. In practice, even a little change in the tilt angle of helices in the membrane will cause a substantial shift in  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling frequencies. On the other hand, as Rosetta searches the optimal relative orientation between membrane and embedded proteins, the protein is immobilized while the membrane is locally moved. In order to compare the calculated frequencies and experimental data, the transformation of membrane orientation needs to be operated on each simulated structure.

Hence, it is nearly impossible to find a de novo predicted transmembrane protein model with the tilt angle close to the NMR-derived structure within  $1^\circ$  directly from membrane ab initio

simulation results. Confronted with this difficulty, we have developed an alternative method by fixing the membrane planes and moving each protein structure calculated solely from the NMR data along the membrane normal to determine the optimal embedding position.

### 3.4 Determination of Transmembrane Helices Embedding Positions in Membrane

In Rosetta, the membrane environment is modeled by two parallel planes separated by 60Å, consisting of water layer, polar layer, interface layer, outer hydrophobic layer and inner hydrophobic layer. (Figure 1.16) To determine the most biologically reasonable position of protein in the heterogeneous membrane-aqueous environment, we fixed the membrane normal to parallel to z-axis, with the membrane bilayer border being parallel to xy plane, and the membrane center coincident with the origin of the coordinate system. Then the first atom of the target protein is moved to the origin. To sufficiently consider the effect of membrane environment on the protein, we chose *membrane\_highres\_Menv\_smooth.wts* as the weight file to calculate all-atom Rosetta energy. Instead of  $E_{sol}$ , which is used to calculate the solvation energy in pure aqueous environment,  $E_{mbenv}$ ,  $E_{mbsolv}$  and  $E_{menv\_smooth}$  are introduced to describe the change in potential caused by membrane surroundings.  $E_{mbenv}$  represents the change in solvation free energy of each atom when they are transferred along membrane normal from outer polar solvent (water) to inner hydrophobic core of membrane bilayer (cyclohexane).  $E_{mbsolv}$  depicts the change in solvation free energy of an atom when it is buried into the protein. (P. Barth 2007)  $E_{menv\_smooth}$  is the statistically derived smooth membrane protein environmental potential adapted from the statistically derived rotamer pair potential.  $E_{mbenv}$  decreases gradually to zero as the protein is moved out of membrane, indicating that when all the atoms are outside the hydrophobic layer this energy item does not work

anymore. Therefore, the target protein is moved from  $z = +70 \text{ \AA}$  to  $z = -70 \text{ \AA}$  in order to analyze the change of potential in the procedure of crossing the membrane with an effective  $E_{mbenv}$ . Each time the structure is moved  $1 \text{ \AA}$ , and the all-atom energy is calculated for each position.

First, we tested this method with 1C3WD, the first 5 helices of bacteriorhodopsin 1C3W. The change in all-atom Rosetta energy along the embedding positions of helices in the membrane is described by Figure 3.11. The maximum energy appears when most residues except the longest loop region are outside the hydrophobic core of membrane, and the second maximum energy belongs to the position where only short loop regions are inside the membrane. The structure is stabilized inside the membrane to the biggest degree (with the minimum energy) when most helical residues are in the hydrophobic core of membrane and the loop residues are in the interface layer.

For a helical hairpin 1C3WE, the tendency of change in all-atom Rosetta energy and the corresponding embedding position of helices is described in Figure 3.12.

For bacteriophage Pf1 (PDB ID: 2KSJ), the minimum energy appears when most helical residues are inside the membrane and the charged residues in the loop are in the polar environment outside the membrane. Burial of charged residues in the hydrophobic layer of membrane was more penalized than exposure of hydrophobic and apolar residues to the polar environment. (Figure 3.13)

For MerFt, we calculated the structures using solid state NMR experimental data, i.e.  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling frequencies, and filtered results with distance constraints. The distance between Gly32 and Gly56, distance between Val28 and Ala61, and vertical

distance between Val28 and Ala61 are selected as the constraints to screen simulated structures.

Finally, when the helical distances Gly32-Gly56 and Val28-Ala61 expanded to 5-30 Å and the allowed vertical distance between Val28 and Ala61 increased to 30 Å, negative full-atom Rosetta energies were obtained. Figure 3.14 shows a calculated structure with its lowest Rosetta all-atom energy -11.03 at the indicated embedding position. At this immersion depth, most of the hydrophobic residues in loop region (Tyr42, Leu43, Tyr45, Val46, Leu47 and Leu48) are inside the membrane core, while Asp44, the only electrically charged loop residue, is close to the interface of the membrane-aqueous heterogeneous environment system. Among the amino acid residues outside the bottom membrane boundary, Gln63 and Gln67 are polar residues, Arg64, Lys65 and Arg66 are positively charged, and Asp69 is negatively charged. The strong electrostatic interaction between these residues and the aqueous environment can make the protein stabilized to the largest degree.

The RMSD value is 4.777 Å in comparison to 2H3O, and 4.222 Å to 2LJ2. The practical distance between Gly32 and Gly56 is 12.1 Å, and the distance between Val28 and Ala61 is 12.5 Å. The vertical distance between Val28 and Ala61 is 4.674 Å. (The distances are between amide nitrogen atoms and measured in PyMOL.) The correlation coefficients between back-calculated frequencies and experimental data are very close to 1.0, indicating the back-calculated PISEMA spectrum has an excellent fit to the two dimensional solid state NMR spectrum. The RMSD for  $^{15}\text{N}$  CSA is 2.844 ppm and the RMSD for  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling is 0.207 kHz. The deviation is mainly caused by residues Gly35, Ser 37, Gly56 and Lys65. These are either polar or charged residues. The differences between principal

components of CSA tensor of glycine and those of other amino acid residues also bring discrepancy in back-calculated values. Because of dependence on both angular constraints and distance constraints, more residues are involved in deviation of  $^{15}\text{N}$  CSA values.

For the calculated MerFt structure shown in Figure 3.14, we also examined its quality using MolProbity, which is developed by the Richardson group at Duke University aiming at evaluating the quality of both protein and nucleic acid models at global and local levels. (V. B. Chen 2010)

Not only contacts between backbone atoms but placement of atoms on side chains are diagnosed as well. The quality statistics is represented by MolProbity score consisting of clashscore, percentage Ramachandran not favored and percentage bad side chain rotamers. Percentile scores are given to clashscore and Molprobity score to describe the comparison between the resolution of input structure and the resolution of the comparable structures in PDB. 100<sup>th</sup> percentile is the best among structures of comparable resolution while 0<sup>th</sup> percentile is the worst. The MolProbity score and the percentile score for each model are shown in the corresponding figures.

The MolProbity score and clashscore of 2H3O and 2LJ2 (Figure 3.16) are calculated as control. We can see that the calculated structure shown in Figure 3.14 has a lower MolProbity score than 2H3O and 2LJ2, implying a more stable conformation with less conflict between both backbone atoms and side-chain atoms. Considering the value of Rosetta all-atom energy, RMSDs comparing to 2H3O and 2LJ2, and reasonability of positions of amino acid residues in the overall heterogeneous membrane-aqueous

environment, the model and the corresponding immersion depth shown in Figure 3.14 should be the most biologically and topologically reliable.

### **3.5 Conclusions**

Rosetta membrane ab initio protocol is useful in predicting the conformation of membrane proteins with good accuracy for structures with less than 250 residues. However, considering the complexity of a real protein structure and the impact of surrounding biological environment, the predicted model and the predicted embedding position in the membrane by pure computation sometimes may be in striking disagreement with the experimental NMR data, as shown by back-calculating of the spectra of MerFt protein as an example.

By fixing the membrane protein structures to the tilt angle determined by solid state NMR experiments and translating the structures along the membrane normal, we calculated their Rosetta all-atom energies to find their optimal immersion depth in the heterogeneous membrane-aqueous environment that is simultaneously consistent with the observed NMR data. The MerFt structure calculated with experimental data and selected by distance constraints Gly32-Gly56 as well as Val28-Ala61 has the lowest Rosetta all-atom energy when most helical residues and hydrophobic loop residues are inside the hydrophobic core of the membrane and the charged/polar residues are outside the membrane in the interface and aqueous environment. The evaluation by back-calculated PISEMA spectra showed an excellent agreement between the experimental data and back-calculated frequencies for the  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar couplings. MolProbity scores and clashscores also indicate that the conflicts among both backbone atoms and side-chain atoms are not significant.

Therefore, the combination of structure calculation with angular constraints from solid state NMR experiments and structure validation by Rosetta all-atom energy could be an effective method to select the consensus structures and determine their optimal embedding position within the membrane. With the availability of more solved membrane protein conformations in PDB, we will be able to perform a more accurate prediction and validation of membrane protein structures.

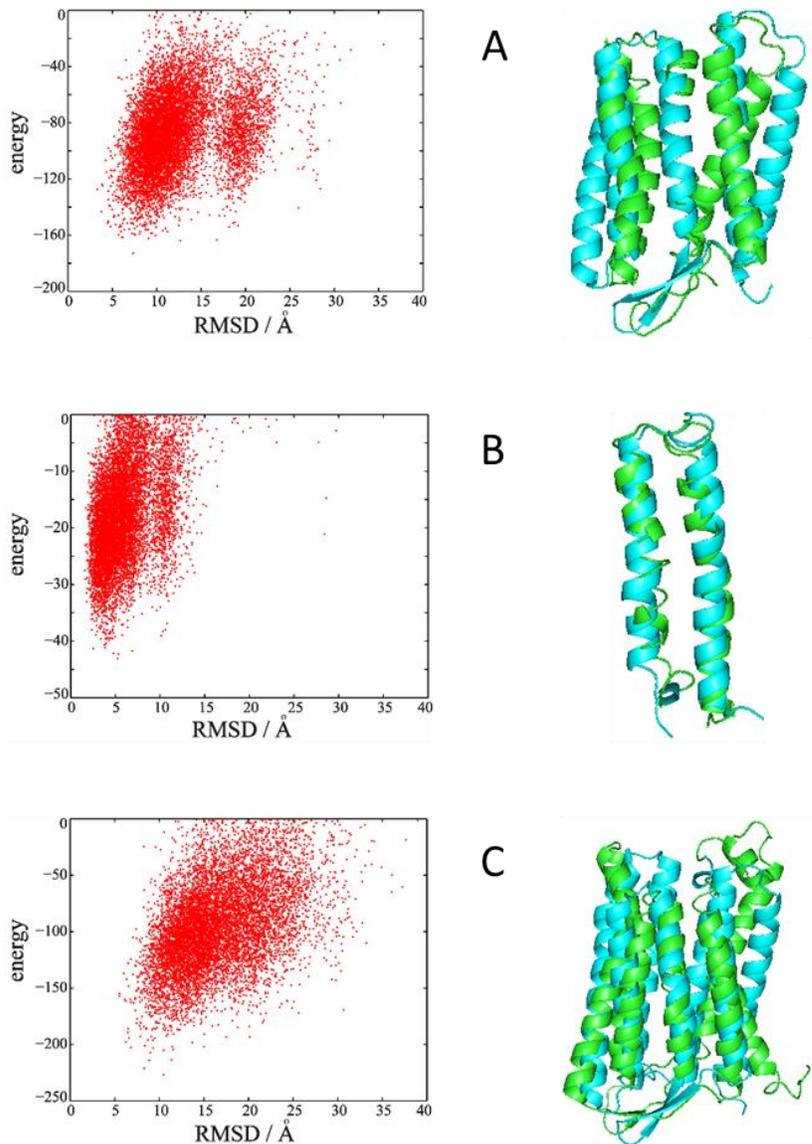


Figure 3.1 Structure prediction. (A) 1C3WD. The superposition between the native (blue) and the lowest-energy predicted structure (green) is represented. The RMSD over 152  $C_{\alpha}$  atoms is 5.725 Å. (B) 1C3WE. The superposition between the native (blue) and the lowest-energy predicted structure (green) is represented. The RMSD over 70  $C_{\alpha}$  atoms is 2.516 Å. (C) 1PY6, chain A. The superposition between the native (blue) and the lowest-energy predicted structure (green) is represented. The RMSD over 227  $C_{\alpha}$  atoms is 6.23 Å.

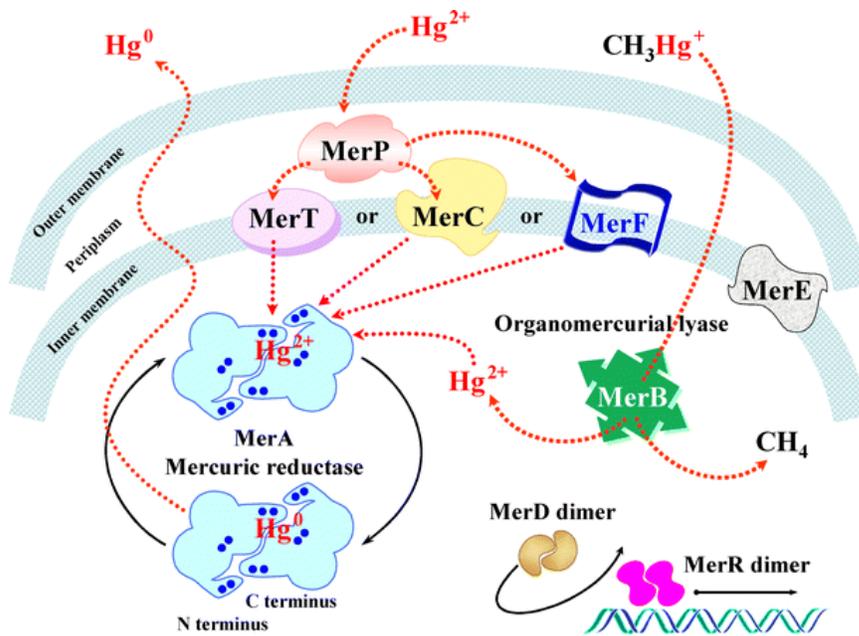


Figure 3.2 The products of the *mer* operon(s). MerR is the transcriptional regulatory protein; MerA is mercuric reductase enzyme; MerB is organomercurial lyase enzyme; MerC is an alternative (to MerT) membrane uptake protein; MerD is a co-repressor or “chaperone”’s involved in transcriptional regulation; MerE is a hypothesized membrane protein that has not been studied; MerF is an additional alternative (to MerT) membrane uptake protein. Transport across the inner and outer membranes without indicated proteins are either by diffusion through the lipid bilayer or via yet unidentified proteins. (Simon Silver 2005)

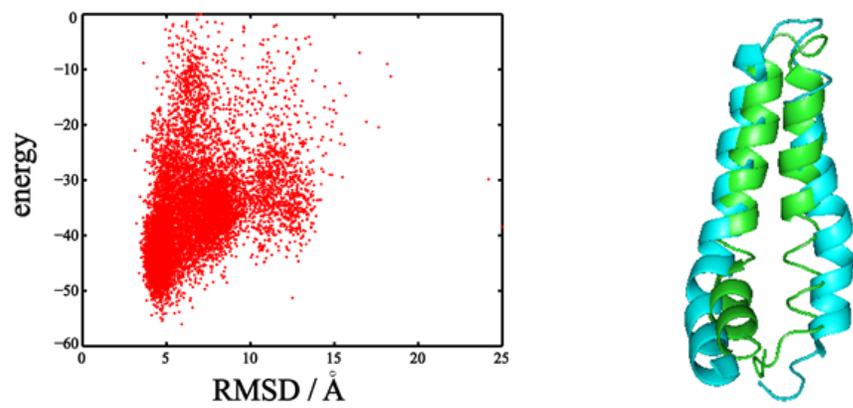


Figure 3.3 The membrane ab initio energy versus RMSD for MerF.

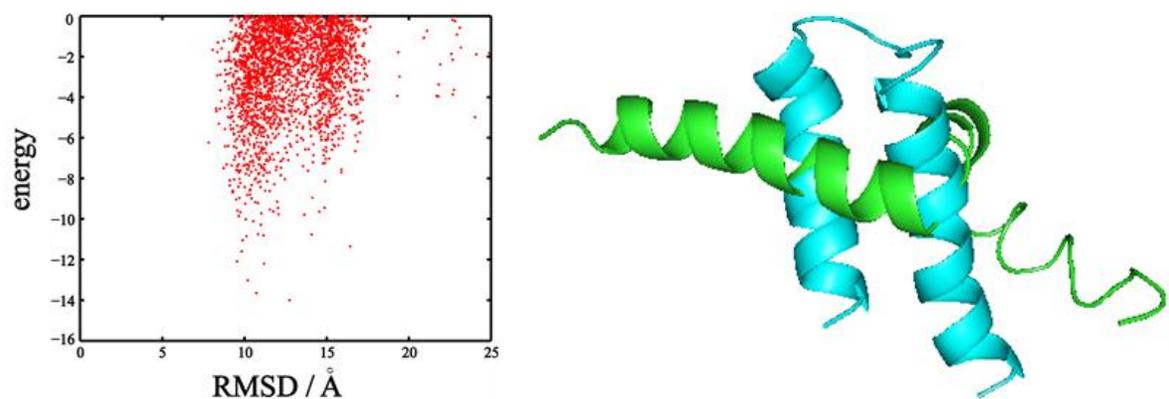


Figure 3.4 The membrane ab initio energies calculated with original OCTOPUS topology versus RMSD for MerFt.

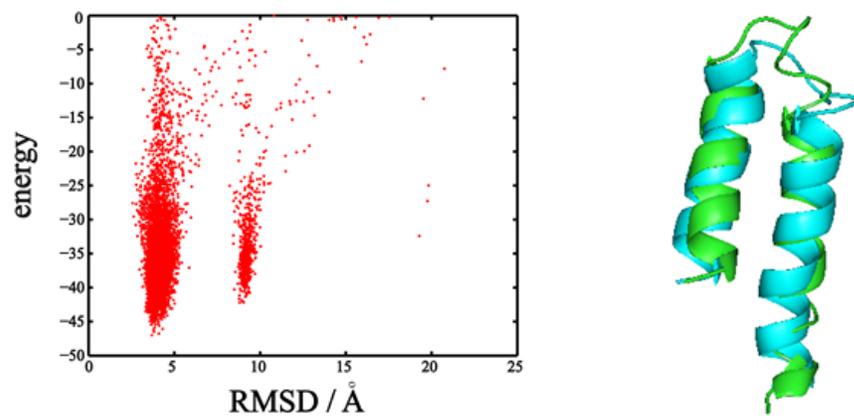
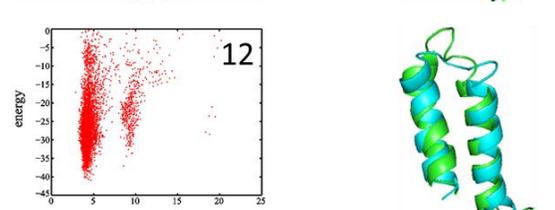
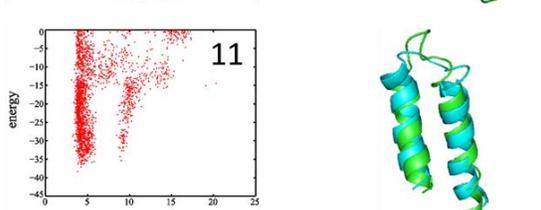
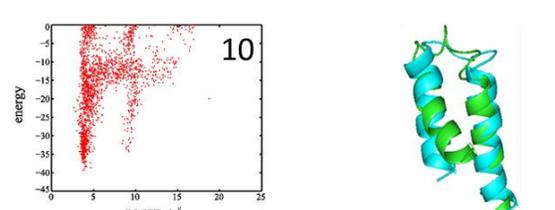
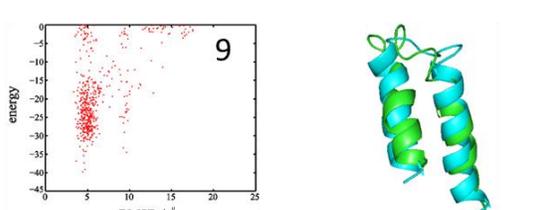
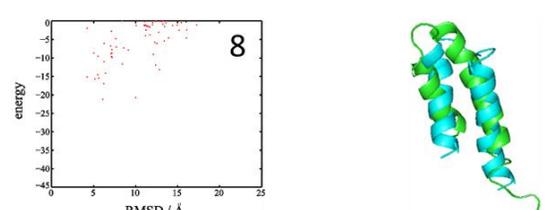
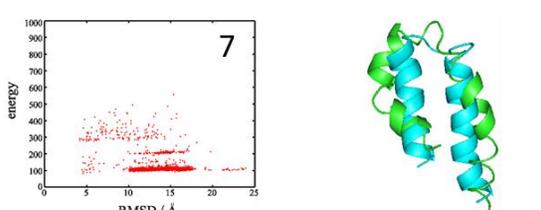
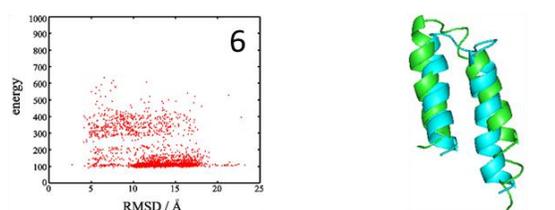
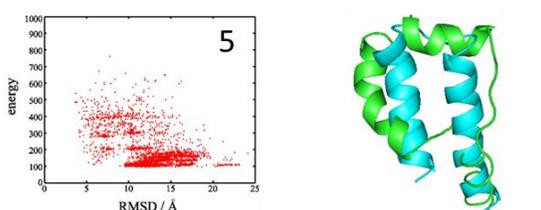
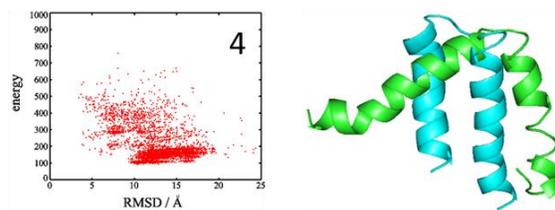
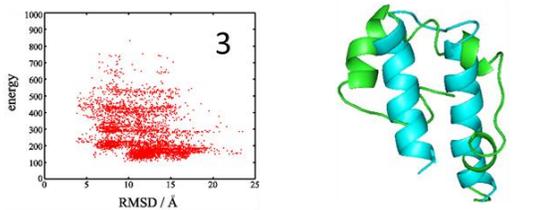
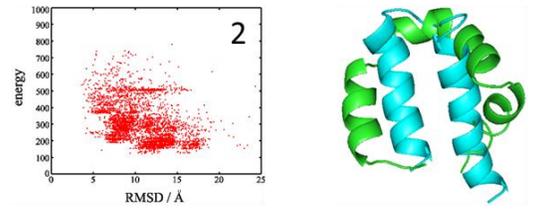
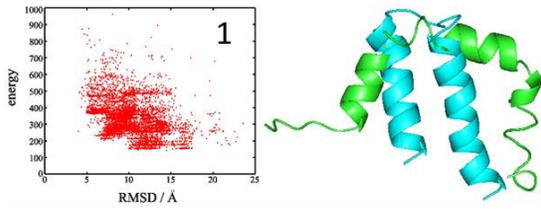
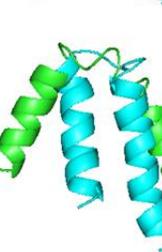
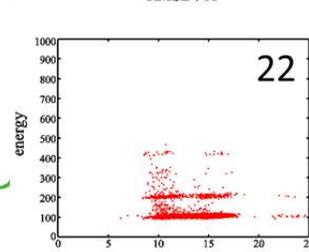
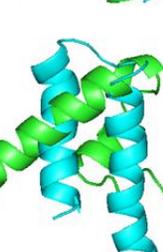
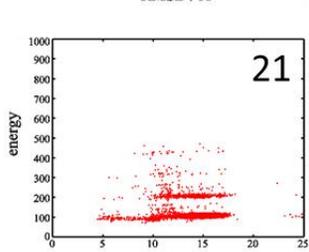
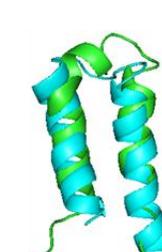
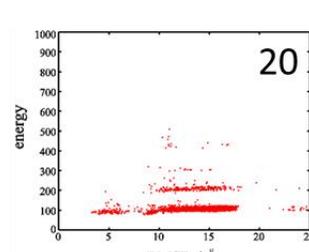
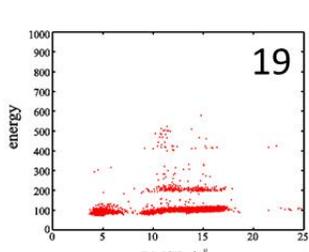
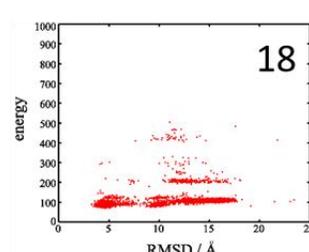
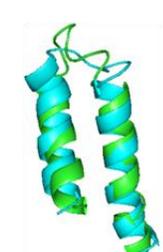
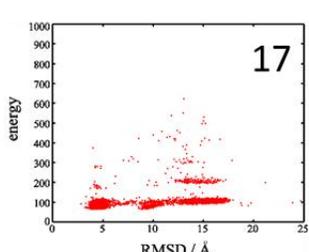
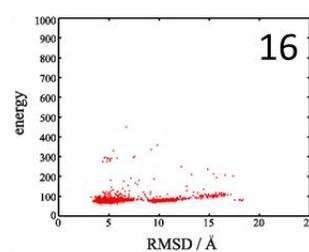
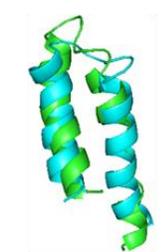
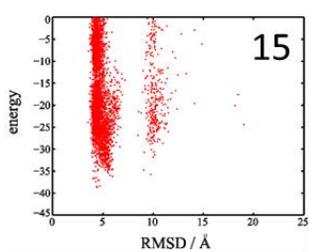
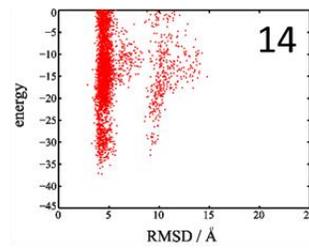
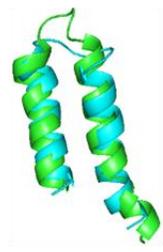
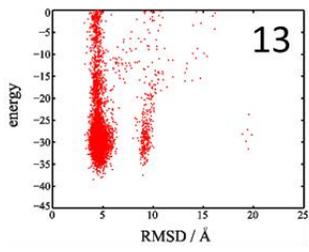


Figure 3.5 The membrane ab initio energies calculated with modified OCTOPUS topology versus RMSD for MerFt.

Figure 3.6 (Left) The membrane ab initio energy calculated with a series of modified OCTOPUS predicted topology shown in Table 3.1 versus RMSD plot for MerFt. (Right) The lowest energy model predicted with each modified OCTOPUS topology. The best structures correspond to the OCTOPUS sequences of 12 of Table 3.1.





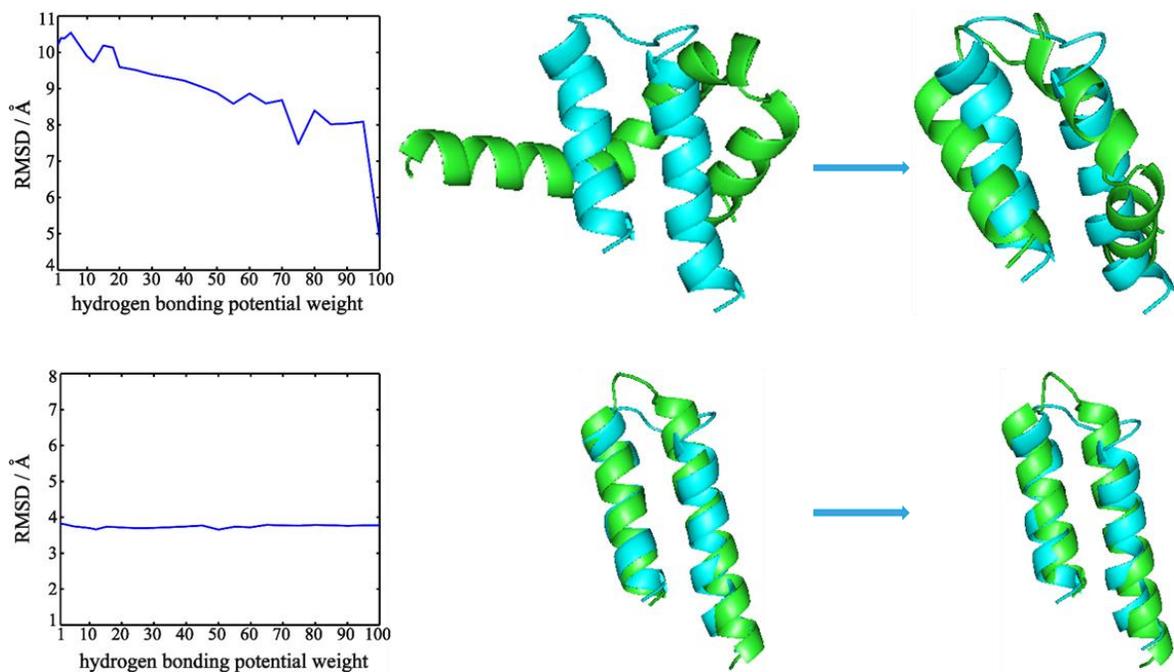


Figure 3.7 Tendency of RMSDs of MerFt structures predicted by membrane ab initio protocol in Rosetta3.4 along the change of the weight for the hydrogen bonding potential. (Top) Prediction with original OCTOPUS topology file. (Bottom) Prediction with modified OCTOPUS topology file. The blue structure is 2H3O, and the green structures are models predicted by Rosetta. The green structures on the left were obtained using original weight for hydrogen bonding potential, while those on the right were obtained using 100 fold of the weight.

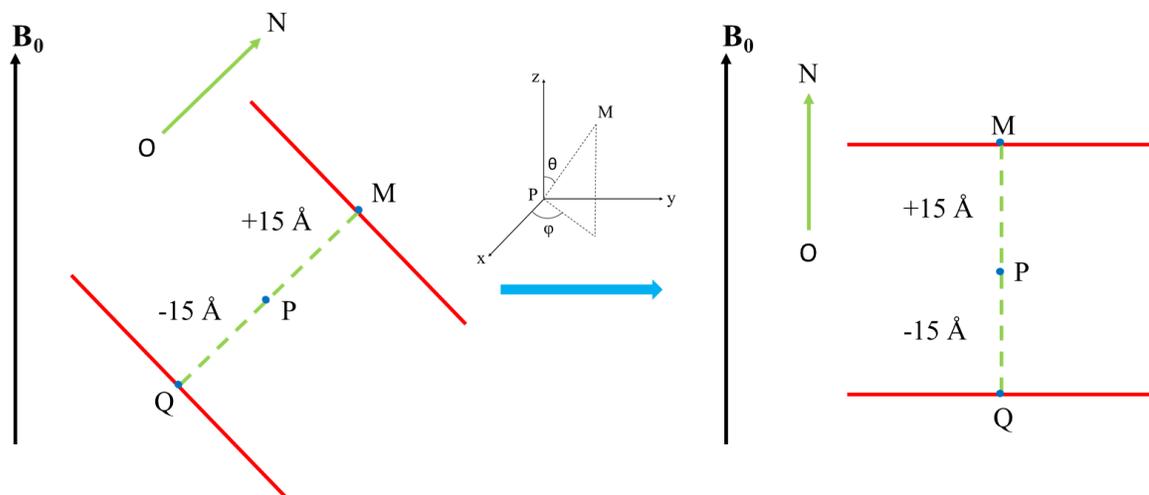


Figure 3.8 Schematic diagram of membrane rotation. The coordination of M, P, Q and N are obtained from Rosetta output file. The red lines represent the membrane boundary. The green arrow line represents the membrane normal.  $\mathbf{B}_0$  shows the orientation of the external magnetic field.

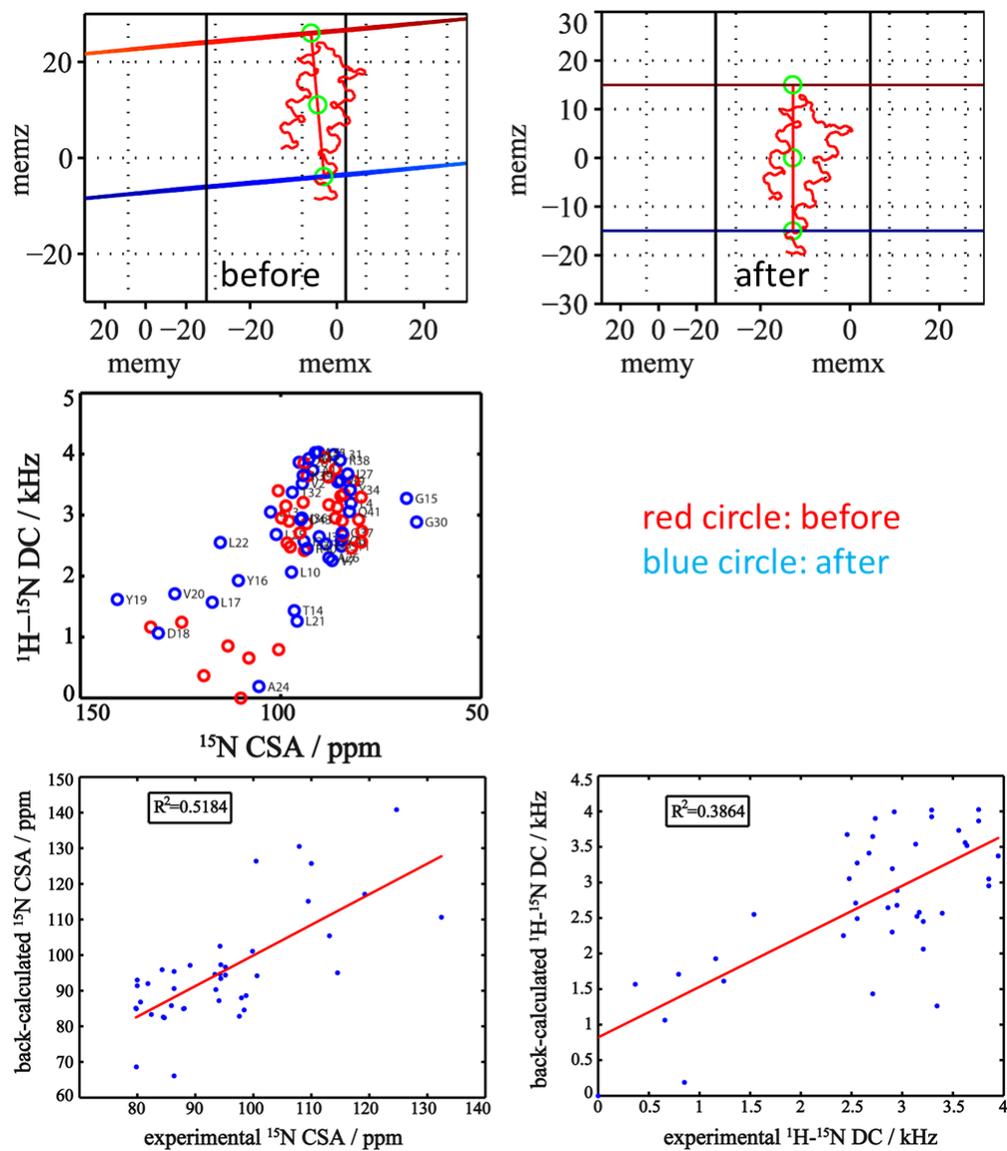


Figure 3.9 Comparison of experimental data from solid state NMR experiments and back-calculated frequencies for structures predicted by membrane ab initio protocol in Rosetta. (A) Before membrane rotation (B) After membrane rotation (C) Correlation between experimental and back-calculated  $^{15}N$  CSA (D) Correlation between experimental and back-calculated  $^1H-^{15}N$  dipolar coupling. The  $R^2$  correlation coefficients are listed for each constraint.

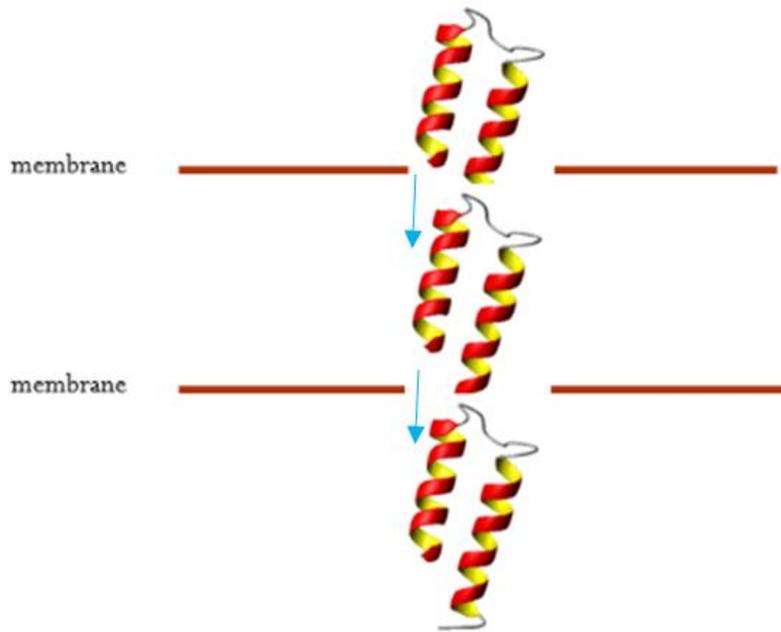


Figure 3.10 Schematic diagram for translation of an NMR-derived structure across lipid bilayer along the orientation of membrane normal.

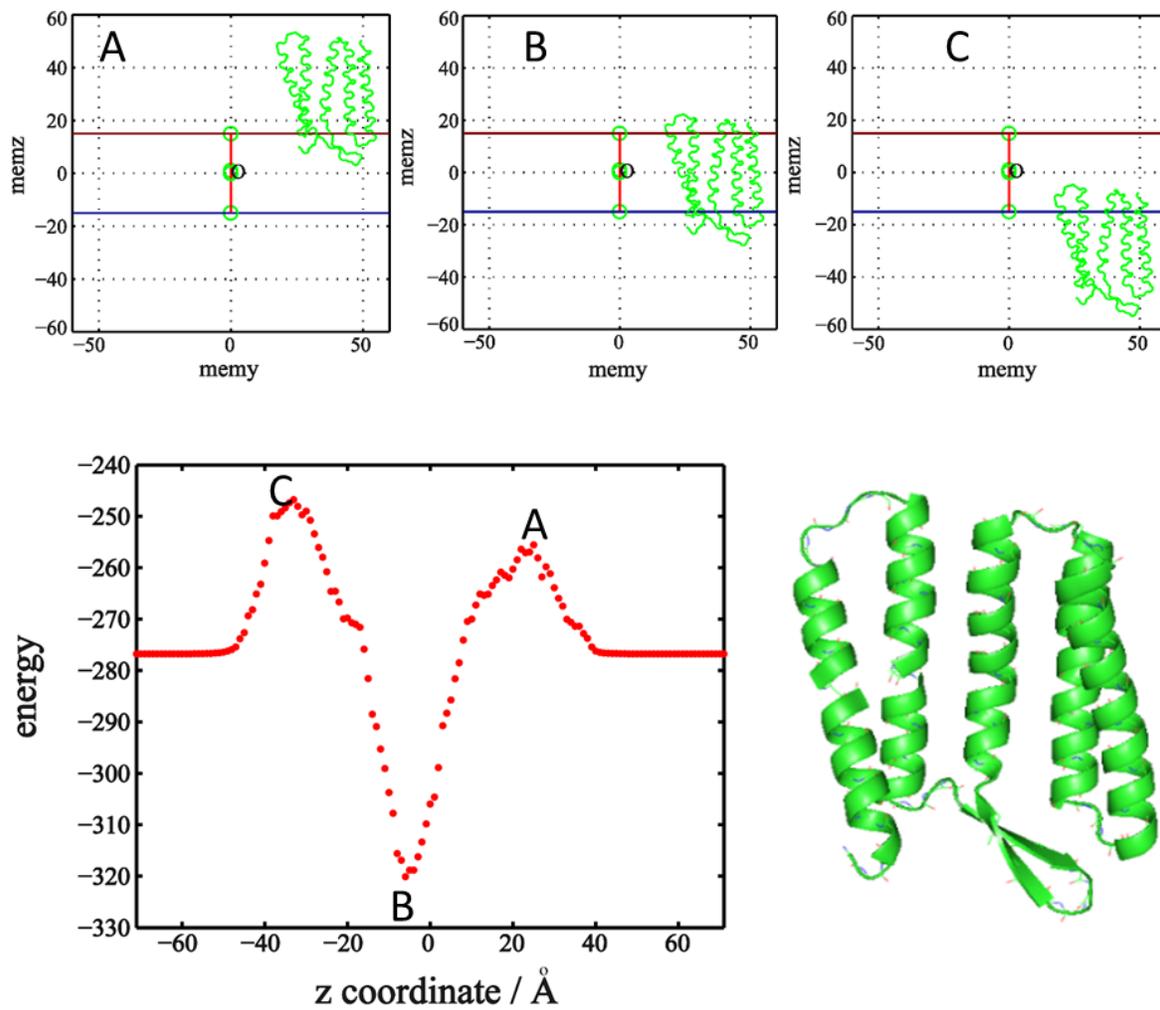


Figure 3.11 Change in all-atom energy of 5-helix 1C3WD along z-axis. The maximum and the second maximum energy corresponds to the embedding position shown in (C) and (A), separately. The minimum energy corresponds to (B).

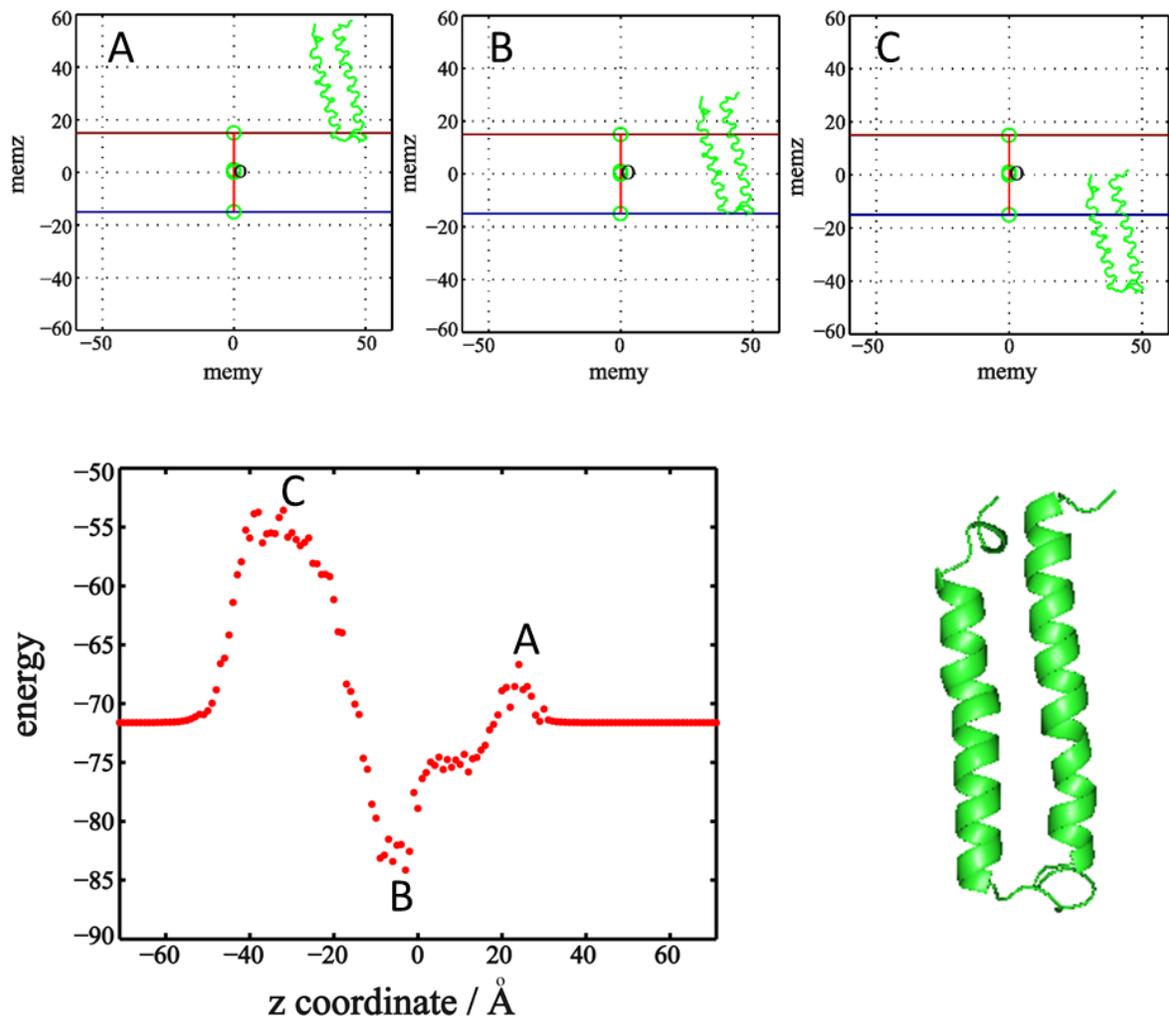


Figure 3.12 Change in all-atom energy of 2-helix (hairpin) 1C3WE along z-axis. The maximum and the second maximum energy corresponds to the embedding position shown in (C) and (A), separately. The minimum energy corresponds to (B).

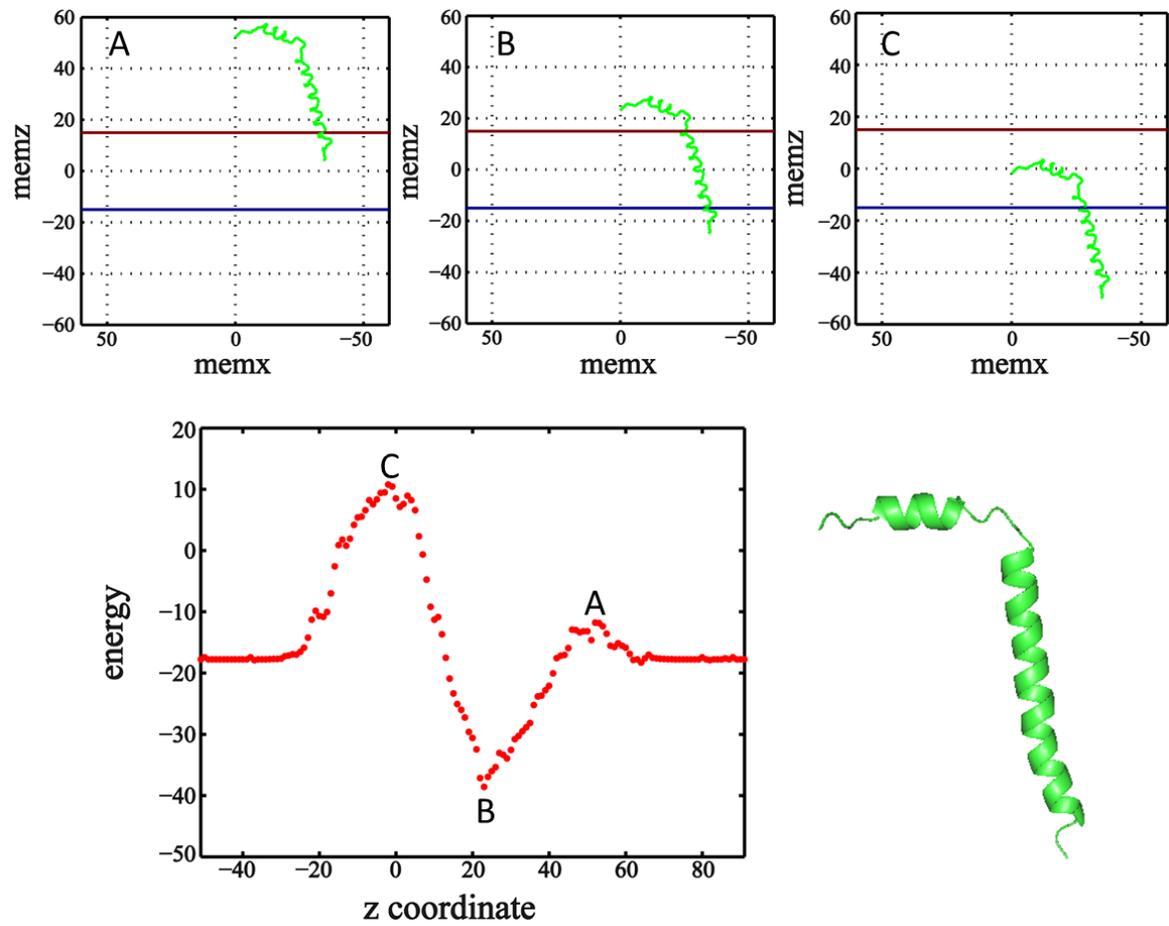
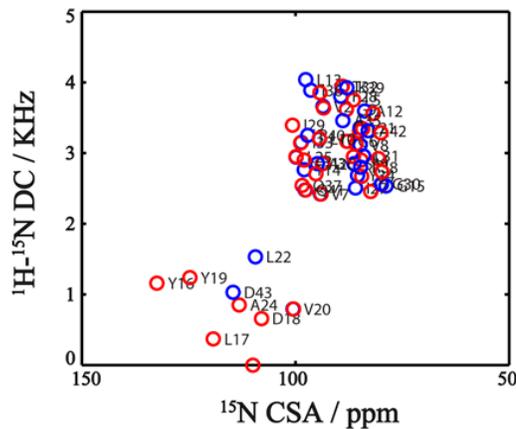
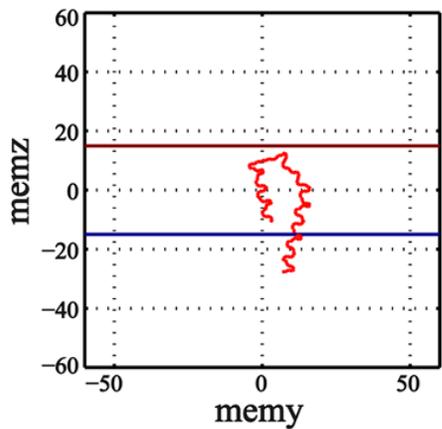


Figure 3.13 Change in all-atom energy of bacteriorhodopsin Pf1 (2KSJ) along z-axis.



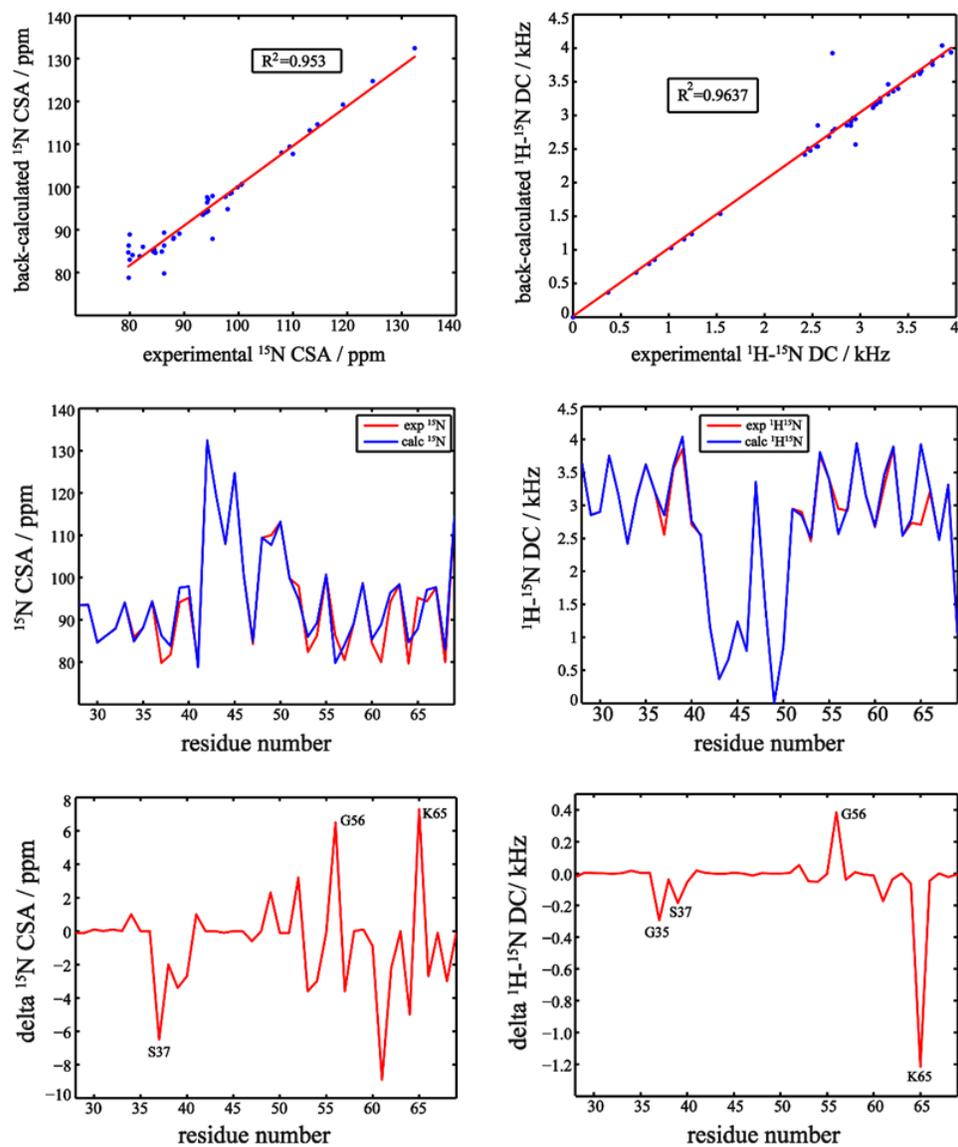


Figure 3.15 Comparison between experimental data and back-calculated frequencies for  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling. The correlation coefficient  $R^2$  are very close to 1.0 indicating that there is an excellent fit of back-calculated frequencies to the experimental data. The two plots in the middle describe the coincidence at each residue. The two plots at the bottom illustrate the calculated difference in  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling at each residue.

## 2H3O

### Summary statistics

All-Atom Contacts	Clashscore, all atoms:	165.22	0 <sup>th</sup> percentile* (N=1784, all resolutions)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	0	0.00%	Goal: <1%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	40	97.56%	Goal: >98%
	MolProbity score <sup>^</sup>	2.77	33 <sup>rd</sup> percentile* (N=27675, 0Å - 99Å)	
	C $\beta$ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	0 / 171	0.00%	Goal: 0%
	Bad backbone angles:	0 / 212	0.00%	Goal: <0.1%

## 2LJ2

### Summary statistics

All-Atom Contacts	Clashscore, all atoms:	14.02	54 <sup>th</sup> percentile* (N=1784, all resolutions)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	4.17%	Goal: <1%
	Ramachandran outliers	3	5.17%	Goal: <0.05%
	Ramachandran favored	51	87.93%	Goal: >98%
	MolProbity score <sup>^</sup>	2.72	35 <sup>th</sup> percentile* (N=27675, 0Å - 99Å)	
	C $\beta$ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	0 / 239	0.00%	Goal: 0%
	Bad backbone angles:	0 / 297	0.00%	Goal: <0.1%

the part in 2LJ2 with the same sequence as 2H3O

### Summary statistics

All-Atom Contacts	Clashscore, all atoms:	13.02	58 <sup>th</sup> percentile* (N=1784, all resolutions)	
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	2	6.06%	Goal: <1%
	Ramachandran outliers	2	4.88%	Goal: <0.05%
	Ramachandran favored	36	87.80%	Goal: >98%
	MolProbity score <sup>^</sup>	2.82	30 <sup>th</sup> percentile* (N=27675, 0Å - 99Å)	
	C $\beta$ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	0 / 174	0.00%	Goal: 0%
	Bad backbone angles:	0 / 214	0.00%	Goal: <0.1%

Figure 3.16 Summary of MolProbity results for MerFt (2H3O) (top), MerF (2LJ2) (middle), and the partial structure in 2LJ2 with the same sequence as 2H3O (bottom).



Table 3.2  $^{15}\text{N}$  CSA and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling frequencies of MerFt (residue 27-68).

residue name	residue number	$^{15}\text{N}$ CSA (ppm)	$^1\text{H}$ - $^{15}\text{N}$ DC (kHz)
L	27	84.945	3.312
V	28	93.373	3.638
I	29	93.523	2.858
L	30	84.653	2.903
L	31	86.344	3.754
G	32	87.96	3.168
V	33	94.08	2.420
V	34	85.871	3.132
G	35	88.08	3.621
L	36	94.366	3.206
S	37	79.837	2.556
A	38	81.8	3.557
L	39	94.237	3.853
T	40	95.17	2.709
G	41	79.837	2.556
Y	42	132.424	1.161
L	43	119.203	0.367
D	44	107.927	0.659
Y	45	124.705	1.239
V	46	100.478	0.793
L	47	84.282	3.344
L	48	109.448	1.537
P	49	110	0
A	50	113.125	0.852
L	51	99.844	2.948
A	52	97.958	2.901
I	53	82.403	2.457
F	54	86.344	3.754
I	55	100.629	3.395
G	56	86.341	2.949
L	57	80.53	2.920
T	58	89.079	3.946
I	59	98.742	3.146
Y	60	84.472	2.672
A	61	79.973	3.290
I	62	94.237	3.853
Q	63	98.396	2.540
R	64	79.735	2.735
K	65	95.17	2.709
R	66	94.366	3.206
Q	67	97.579	2.477
A	68	79.973	3.290
D	69	114.526	1.025

Table 3.3 Distance constraints used in calculated MerFt structures screening.

distance constraints	Test 1	Test 2	Test 3
Gly32-Gly56	8-15 Å	5-30 Å	5-30 Å
Val28-Ala61	8-15 Å	5-30 Å	5-30 Å
Val28-Ala61 (z)	< 1.5 Å	< 15 Å	< 30 Å

## Chapter 4

### Future Directions

#### 4.1 Homomeric Ion Channels

Homomeric protein assemblies behave as channels, containers and molecular rulers in many organelles. The study of their conformation can lead to a better understanding of their function in many biological processes. Nevertheless, the packing of chains and relative orientation in oligomer proteins have been less studied than those of monomeric proteins. (Hong 2006) Importantly, symmetry-related homomeric structures usually cannot be distinguished by NMR spectroscopy alone, thus requiring additional methods to determine the oligomerization state of a membrane protein. Symmetric modeling in Rosetta can be used to predict the structures with  $C_n$ ,  $D_n$  and icosahedral symmetry types. (Frank DiMaio 2011) RosettaOligomers method was developed to determine solution structures of oligomer using chemical shift, sparse NOE and RDC frequencies. (Nikolaos G. Sgourakis 2011) Symmetry docking protocol for membrane proteins is still under development.

The M2 protein of the influenza A virus is a tetrameric transmembrane protein. Many efforts have been made on revealing the binding manner among the chains in this oligomer protein and its functional mechanism as a proton channel. (Mukesh Sharma 2010; Sarah D. Cady 2010) Currently, I am working on the determination of optimal oligomerization state for M2 (PDB ID: 2L0J) in the biological membrane by combining protein symmetric docking

protocol (Rhiju Das 2009; Chaudhury S 2011; P. Benjamin Stranges 2011) and membrane ab initio protocol in Rosetta3.4.

## 4.2 Application of Symmetry Docking to M2 (2L0J)

Because M2 is in  $C_4$  symmetric mode, the symmetry definition file C4.symm is generated first by

```
make_symmdef_file_denovo.py -symm_mode cn -nsub 4 > C4.symm
```

with the output file as

```
symmetry_name c4
subunits 4
recenter
number_of_interfaces 2
E = 4*VRT0001 + 4*(VRT0001:VRT0002) + 2*(VRT0001:VRT0003)
anchor_residue COM
virtual_transforms_start
start -1,0,0 0,1,0 0,0,0
rot Rz 4
virtual_transforms_stop
connect_virtual JUMP1 VRT0001 VRT0002
connect_virtual JUMP2 VRT0002 VRT0003
connect_virtual JUMP3 VRT0003 VRT0004
set_dof BASEJUMP x(50) angle_x(0:360) angle_y(0:360) angle_z(0:360).
```

The symmetry docking is performed with the whole molecule of 2L0J and the helical part of 2L0J (named as 2L0JT), respectively.

For the whole molecule 2L0J, the amino acid sequence is

```
SNASSDPLVVAASIIGILHLILWILDRLFFKSIYRFFEHLKRG
```

, while for helical part 2L0JT the amino acid sequence is

```
SSDPLVVAASIIGILHLILWILDRL .
```

The results obtained by performing the scripts in Appendix B are shown in Figure 4.1 and Figure 4.2. Three representative points are selected to analyze. The predicted models are

demonstrated in side view and top view, separately, so we can observe both the packing state of peptide chains and the hole formed by the four chains. In the plot of energy versus RMSD, the data did not form a traditional funnel but two points with similar lowest energy and one point with lowest RMSD but higher energy. Further calculations and more information from known homomeric protein assemblies are needed before an accurate prediction can be obtained.

Currently, in the solid state NMR experiment one can isotopically label Ala30 and Ile35 on helix 1 and helix 2, and then calculate their distance based on the experimental data. So we can, in principle, use the distance between A30 on helix 1 and I35 on helix 2, and the distance between I35 on helix 1 and A30 on helix 2 to evaluate the quality of predicted structures. Since both  $^{15}\text{N}$  and  $^{13}\text{C}'$  can be used to label the residue, we measured both the  $^{15}\text{N}$ - $^{15}\text{N}$  and  $^{13}\text{C}'$ - $^{13}\text{C}'$  distances for structure evaluation.

The predicted structures with GDTMM of more than 0.7 are selected for distance comparison. The histograms of these distances are shown in Figure 4.4. The standard deviation of h1A30\_h2I35 distances is more than 0.9 Å, while the standard deviation of h1I35\_h2A30 distances is only 0.31 Å. This may suggest that the residue Ala30 on helix 1 and the residue Ile35 on helix 2 are more dynamic and, therefore, are more involved in biological activities. However, additional results from both computation and experiments would be needed to explain the interaction among the helices in a homomeric protein assembly.

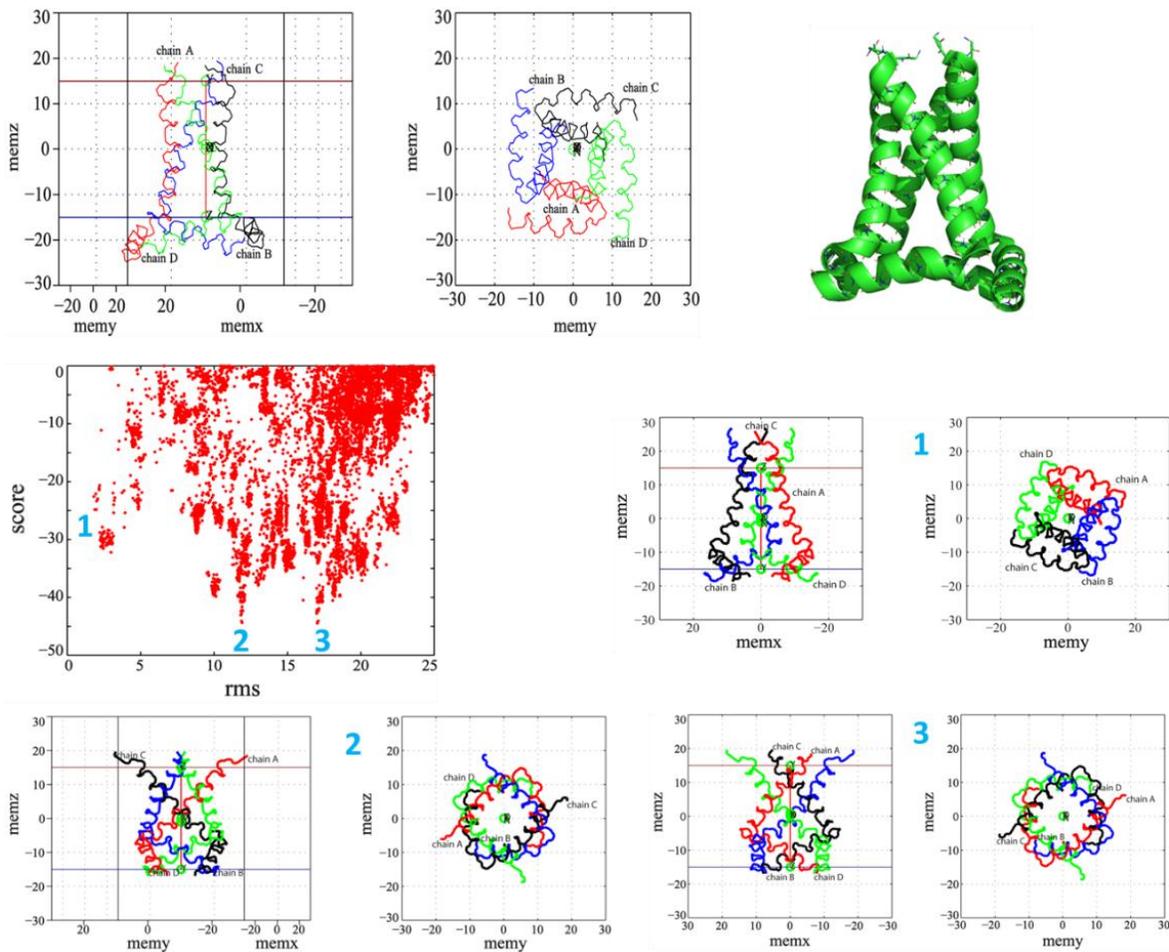


Figure 4.1 Symmetry docking results for 2L0J. The side view and top view for structures represented by 1, 2 and 3 are shown on the bottom right.

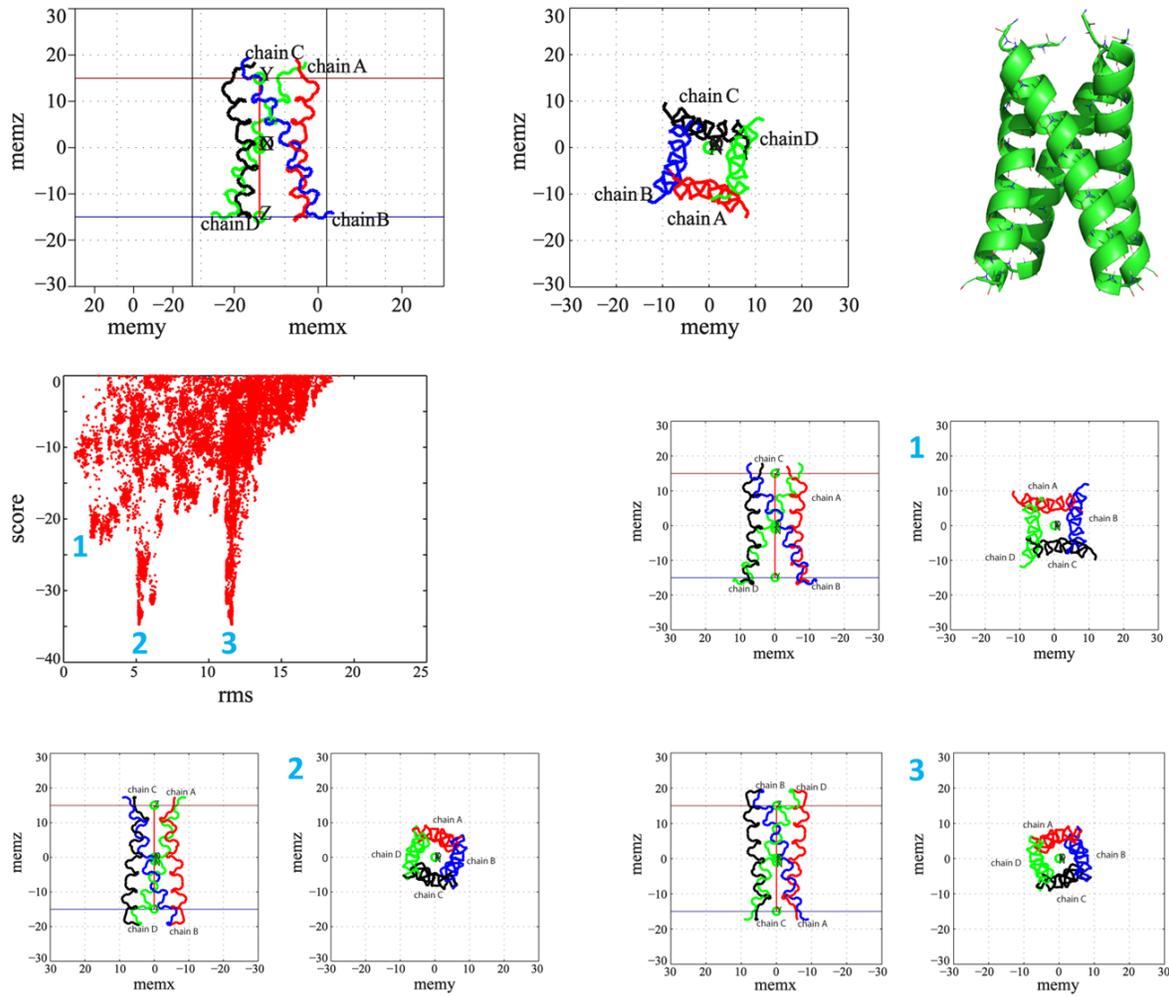


Figure 4.2 Symmetry docking results for helical part of 2L0J. The side view and top view for structures represented by 1, 2 and 3 are shown on the bottom right.

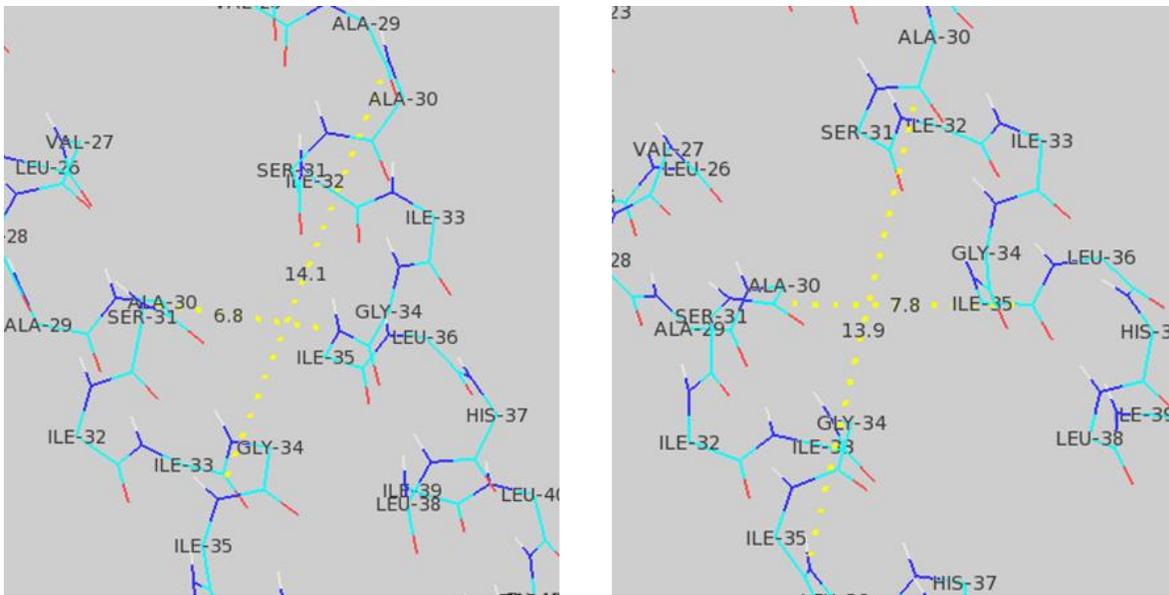


Figure 4.3 Illustration of h1A30\_h2I35 and h1I35\_h2A30 distances measured in PyMOL.

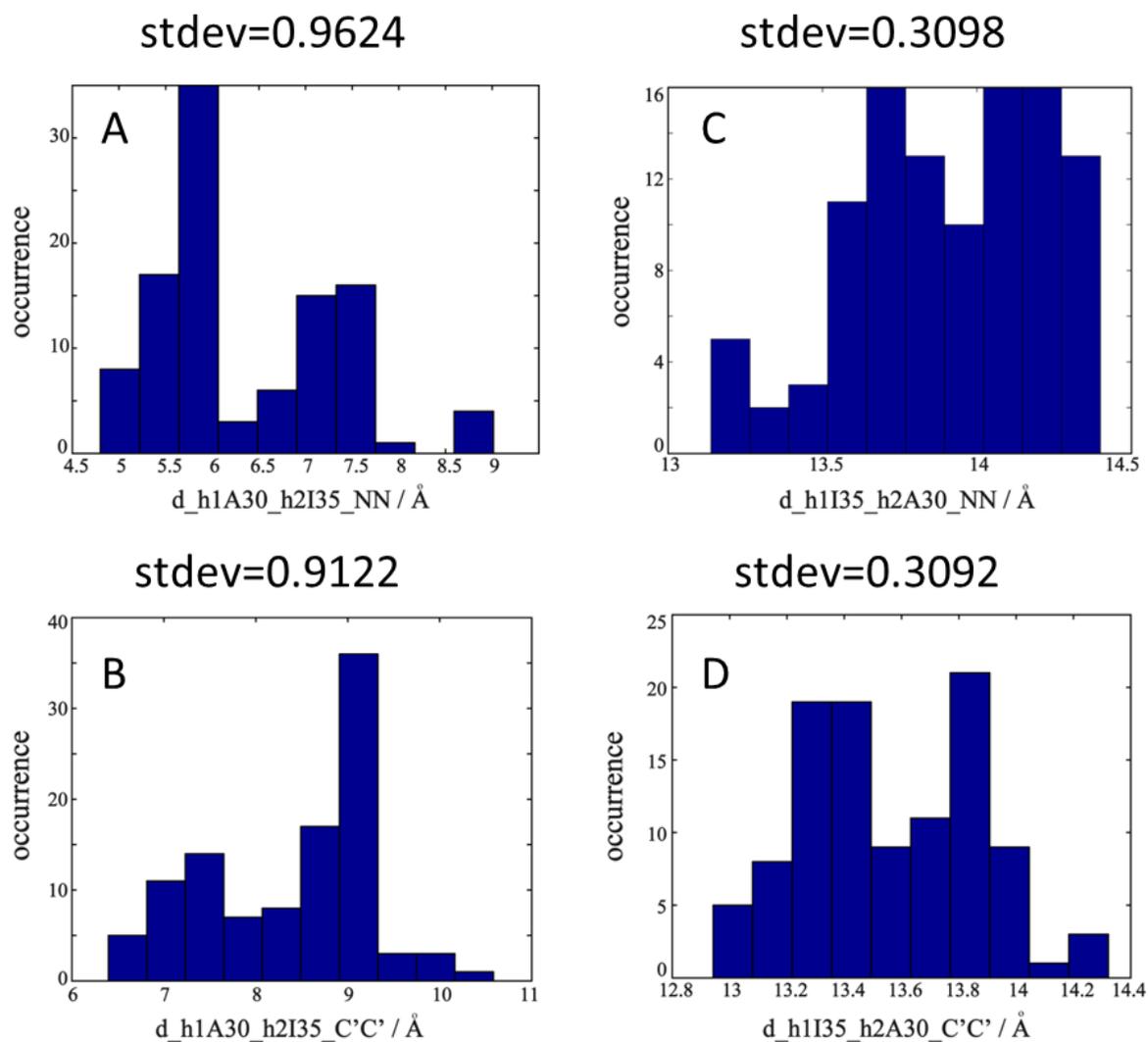


Figure 4.4 (A) and (B): Histograms of distances between Ala30 on helix 1 and Ile35 on helix 2. (C) and (D): Histograms of distances between Ile35 on helix 1 and Ala30 on helix 2. (A) and (C) describe the distribution of N-N distances. (B) and (D) describe the distribution of C'-C' distances.

Table 4.1 Summary of symmetry docking results for 2LOJ.

ID	Energy	RMSD	GDTMM	MaxSub
1	-23.01	2.098	0.789	164
2	-43.931	11.898	0.25	41
3	-43.882	17.04	0.26	49

Table 4.2 Summary of symmetry docking results for helical part of 2L0J.

ID	Energy	RMSD	GDTMM	MaxSub
1	-11.321	0.798	0.982	100
2	-34.732	5.17	0.25	25
3	-34.732	11.578	0.28	40

Table 4.3 Summary of distances measured in native structure, simulated whole structure and simulated helical part of M2 (2LOJ). (unit: Å)

ID	$d_{h1A30\_h2I35\_NN}$	$d_{h1A30\_h2I35\_CC}$	$d_{h1I35\_h2A30\_NN}$	$d_{h1I35\_h2A30\_CC}$
native 2LOJ	6.8	7.8	14.1	13.9
calculated 2LOJ 1	5.6842	6.7034	13.225	13.0174
calculated 2LOJ 2	17.9876	17.8243	7.9652	7.2805
calculated 2LOJ 3	7.9087	7.2242	17.9331	17.7574
calculated 2LOJT 1	5.9749	7.5165	13.6352	13.373
calculated 2LOJT 2	12.8572	12.4498	11.8818	12.4872
calculated 2LOJT 3	11.8451	12.4411	12.8303	12.4147

## Chapter 5

### Conclusions

Solid-state NMR spectroscopy has proved to be a useful tool for studying membrane proteins in their native-like environment without the necessity of obtaining high-quality crystalline samples or the requirement of using cryogenic temperatures during the experiments.

Chemical shift anisotropy and heteronuclear dipolar couplings derived from solid-state NMR frequencies are orientationally dependent and, therefore, both distances and bond orientations can be calculated directly from these angular-dependent NMR observables.

Due to the significant presence of experimental uncertainty in the chemical shift tensor parameters, we have developed an efficient approach to calculate membrane protein structures solely from the heteronuclear dipolar couplings associated with the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  spin-bearing nuclei from multidimensional “shiftless” solid-state NMR spectra of oriented samples. Converged solutions were obtained for protein G (a soluble protein) and for bacteriorhodopsin (a transmembrane protein) with a simulated experimental uncertainty of up to 15 Hz and 120 Hz, respectively.

The predicted membrane protein structures and their embedding positions in the membrane via pure computation can be sometimes in striking disagreement with the experimental NMR data. Therefore, we have developed a method to improve the prediction quality. Positional constraints in the structural fitting, Rosetta all-atom energies and MolProbity scores are combined to evaluate the reliability of the NMR-derived structures and to predict their

optimal immersion depth in the heterogeneous membrane-aqueous environment consistent with the observed NMR data.

Biological importance of homomeric protein assemblies and the limitations in determining the symmetry-related homomeric structures by NMR spectroscopic method alone have prompted us to study the oligomerization states of membrane proteins using Rosetta docking protocols. With the constraints obtained from isotopically labeled amino acid residues, a combination of solid-state NMR spectroscopy and Rosetta simulation can potentially develop into a useful tool for the studies of the conformations of large protein complexes.

With further development and advancement in solid-state NMR experimental instruments, pulse sequences, and computational algorithms, atomic-resolution structures of membrane proteins will be obtainable using a single, uniformly ( $^{15}\text{N}$ ,  $^{13}\text{C}$ )-labeled sample as performed routinely for soluble proteins in liquid NMR.

## REFERENCES

A. Ramamoorthy, Y. W., Dong-Kuk Lee (2004). "PISEMA Solid-State NMR Spectroscopy." Annual Reports on NMR Spectroscopy **52**: 1-52.

A. T. Brünger, P. D. A., G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination." Acta Cryst. **D54**(5): 905-921.

Alexander A Nevzorov, S. J. O. (2003). "A "Magic Sandwich" pulse sequence with reduced offset dependence for high-resolution separated local field spectroscopy." Journal of Magnetic Resonance **164**(1): 182-186.

Alexander D. MacKerell Jr., N. B., Nicolas Foloppe (2000). "Development and current status of the CHARMM force field for nucleic acids." Biopolymers **56**(4): 257-265.

Almeida FC, O. S. (1997). "fd coat protein structure in membrane environments: structural dynamics of the loop between the hydrophobic trans-membrane helix and the amphipathic in-plane helix." Journal of Molecular Biology **270**(3): 481-495.

Altenbach, C., A. K. Kusnetzow, et al. (2008). "High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation." Proc. Natl. Acad. Sci. U. S. A. **105**: 7439-7444.

Ana Carolina Zeri, M. F. M., Alexander A. Nevzorov, and Stanley J. Opella (2003). "Structure of the coat protein in fd filamentous bacteriophage particles determined by solid-state NMR spectroscopy." Proceedings of the National Academy of Sciences of the United States of America **100**(11): 6458-6463.

Andréa M.A. Nascimento, E. C.-S. (2003). "Operon mer: Bacterial resistance to mercury and potential for bioremediation of contaminated environments." Genetics and Molecular Research **2**(1): 92-101.

Andrew, E. R. (1981). "Magic Angle Spinning in Solid State n.m.r. Spectroscopy." Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **299**(1452): 505-520.

Anna A. De Angelis, A. A. N., Sang Ho Park, Stanley C. Howell, Anthony A. Mrse, and Stanley J. Opella (2004). "High-Resolution NMR Spectroscopy of Membrane Proteins in Aligned Bicelles." J. Am. Chem. Soc. **126**(47): 15340-15341.

Arfken, G. B. (1985). Mathematical methods for physicists, 3rd ed. Orlando, Academic Press.

B. R. Brooks, C. L. B. I., A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus (2009). "CHARMM: The biomolecular simulation program." Journal of Computational Chemistry **30**(10): 1545–1614.

Baker, B. K. a. D. (2000). "Native protein sequences are close to optimal for their structures." Proceedings of the National Academy of Sciences of the United States of America **97**(19): 10383-10388.

Bax, A. (1994). "Multidimensional nuclear magnetic resonance methods for protein studies." Current Opinion in Structural Biology **4**: 738-744.

Bernard R. Brooks, R. E. B., Barry D. Olafson, David J. States, S. Swaminathan, Martin Karplus (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." Journal of Computational Chemistry **4**(2): 187-217.

Bertram, R., T. Asbury, et al. (2003). "Atomic refinement with correlated solid-state NMR restraints." J. Magn. Reson. **163**: 300-309.

Bibhuti B. Das, H. J. N., George J. Lu, Woo Sung Son, Ye Tian, Francesca M. Marassi, and Stanley J. Opella (2012). "Structure Determination of a Membrane Protein in Proteoliposomes." J. Am. Chem. Soc. **134**(4): 2047–2056.

Bin Qian, S. R., Rhiju Das, Philip Bradley, Airlie J. McCoy, Randy J. Read and David Baker (2007). "High-resolution structure prediction and the crystallographic phase problem." Nature **450**(7167): 259-264.

Bonneau, R., J. Tsai, et al. (2001). "Rosetta in CASP4: progress in ab initio protein structure prediction." Proteins Suppl **5**: 119-26.

Borbat, P. P. and J. H. Freed (1999). "Multiple-quantum ESR and distance measurements." Chem. Phys. Lett. **313**: 145-154.

Bowie, J. U. (2011). "Membrane Protein Folding: How important are hydrogen bonds?" Curr. Opin. Struct. Biol. **21**(1): 42-49.

Brian Kuhlman, G. D., Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, David Baker (2003). "Design of a Novel Globular Protein Fold with Atomic-Level Accuracy." Science **302**(5649): 1364-1368.

C. B. Anfinsen, E. H., M. Sela, and F. H. White, Jr. (1961). "THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN." Proceedings of the National Academy of Sciences of the United States of America **47**(9): 1309-1314.

Caffrey, M. (2003). "Membrane protein crystallization." Journal of Structural Biology **142**(1): 108-132.

Carol A. Rohl, C. E. M. S., Dylan Chivian, David Baker (2004). "Modeling structurally variable regions in homologous proteins with rosetta." Proteins: Structure, Function, and Bioinformatics **55**(3): 656-677.

Carol A. Rohl, C. E. M. S., Kira M.S. Misura, David Baker (2004). "Protein Structure Prediction Using Rosetta." Methods in Enzymology **383**: 66-93.

Chaudhury S, B. M., Weitzner BD, Muthu P, Bergman H, Gray JJ (2011). "Benchmarking and analysis of protein docking performance in Rosetta v3.2." PLoS ONE **6**(8): e22477.

Chrysi Konstantinou Kirtay, J. B. O. M. a. J. A. L. (2005). "Knowledge Based Potentials: the Reverse Boltzmann Methodology, Virtual Screening and Molecular Weight Dependence." QSAR & Combinatorial Science **24**(4): 527-536.

Cornilescu, G. and A. Bax (2000). "Measurement of Proton, Nitrogen, and Carbonyl Chemical Shielding Anisotropies in a Protein Dissolved in a Dilute Liquid Crystalline Phase." J. Am. Chem. Soc. **122**: 10143-10154.

Creighton, T. E. (1984). Proteins: Structures and Molecular Principles. New York, W.H.Freeman and Company.

Cross, T. A. and S. J. Opella (1994). "Solid-state NMR structural studies of peptides and proteins in membranes." Curr. Opin. Struct. Biol. **4**: 574-581.

D Benjamin Gordon, S. A. M., Stephen L Mayot (1999). "Energy functions for protein design." Current Opinion in Structural Biology **9**(4): 509-513.

Daniel J Mandell, E. A. C. a. T. K. (2009). "Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling." Nature Methods **6**(8): 551 - 552.

David A. Case, T. E. C. I., Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz Jr., Alexey Onufriev, Carlos Simmerling, Bing Wang, Robert J. Woods (2005). "The Amber biomolecular simulation programs." Journal of Computational Chemistry **26**(16): 1668–1688.

David S. Thiriot, A. A. N., Lena Zagayanskiy, Chin H. Wu, Stanley J. Opella (2004). "Structure of the Coat Protein in Pf1 Bacteriophage Determined by Solid-state NMR Spectroscopy." Journal of Molecular Biology **341**(3): 869–879.

David Van Der Spoel, E. L., Berk Hess, Gerrit Groenhof, Alan E. Mark, Herman J. C. Berendsen (2005). "GROMACS: Fast, flexible, and free." Journal of Computational Chemistry **26**(16): 1701–1718.

De Angelis, A. A., S. C. Howell, et al. (2006). "Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy." Journal of the American Chemical Society **128**(37): 12256-12267.

De Angelis, A. A. and S. J. Opella (2007). "Bicelle samples for solid-state NMR of membrane proteins." Nature Protocols **2**(10): 2332-2338.

Denny, J. K., J. Wang, et al. (2001). "PISEMA Powder Patterns and PISA Wheels." J. Magn. Reson. **152**: 217-226.

Dong, F., B. Olsen, et al. (2008). "Computational Methods for Biomolecular Electrostatics." Methods Cell Biol. **84**: 843-870.

Dvinskikh, S. V., K. Yamamoto, et al. (2006). "Heteronuclear isotropic mixing separated local field NMR spectroscopy." J. Chem. Phys. **125**: 034507.

E. R. ANDREW, A. B. a. R. G. E. (1958). "Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed." Nature **182**: 1659.

E. Munck, J. T. P. D. a. (1969). Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois., University of Illinois Press.

Elofsson, H. V. a. A. (2008). "OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar." Bioinformatics **24**(15): 1662-1668.

Engh, R. A. and R. Huber (1991). "Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement." Acta. Cryst. **A47**: 392-400.

Eyal Neria, S. F., and Martin Karplus (1996). "Simulation of activation free energies in molecular systems." Journal of Chemical Physics **105**(5): 1902.

Fang Xiao Zhou, M. J. C., William P. Russ, Axel T. Brunger and Donald M. Engelman (2000). "Interhelical hydrogen bonding drives strong interactions in membrane proteins." Nature Structural Biology **7**(2): 154-160.

Filmore, D. (2004). "It's a GPCR world -- Cell-based screening assays and structural studies are fueling G-protein coupled receptors as one of the most popular classes of investigational drug targets." MODERN DRUG DISCOVERY **7**(11): 24-28.

Frances C. Bernstein, T. F. K., , Grahame J.B. Williams, Edgar F. Meyer Jr., Michael D. Brice, John R. Rodgers, Olga Kennard, Takehiko Shimanouchi, Mitsuo Tasumi (1977). "The protein data bank: A computer-based archival file for macromolecular structures." Journal of Molecular Biology **112**(3): 535–542.

Frank DiMaio, A. L.-F., Phil Bradley, David Baker, Ingemar André (2011). "Modeling Symmetric Macromolecular Structures in Rosetta3." PLoS ONE **6**(6): e20450.

Franks, W. T., D. H. Zhou, et al. (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 the American Chemical Society **127**(35): 12291-12305.

Fung, B. M., A. K. Khitrin, et al. (2000). "An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids." J. Magn. Reson. **142**: 97-101.

G., A. (1985). Mathematical methods for physicists. Orlando, Academic Press.

G.N. Ramachandran, C. R., V. Sasisekharan (1963). "Stereochemistry of polypeptide chain configurations." Journal of Molecular Biology **7**(1): 95–99.

G.N.Ramachandran, A.V.Lakshminarayanan, et al. (1973). "Theory of The Non-planar Peptide Unit." Biochim. Biophys. Acta **303**: 8-13.

Gardner, K. H., M. K. Rosen, et al. (1997). "Global Folds of Highly Deuterated, Methyl-Protonated Proteins by Multidimensional NMR." Biochemistry **36**: 1389-1401.

Gautam Dantas, B. K., David Callender, Michelle Wong, David Baker (2003). "A Large Scale Test of Computational Protein Design: Folding and Stability of Nine Completely Redesigned Globular Proteins." Journal of Molecular Biology **332**(2): 449–460.

- Grisshammer, C. G. T. a. R. (1996). "Heterologous expression of G-protein-coupled receptors." Trends in Biotechnology **14**(11): 426-430.
- Grisshammer R., T. C. G. (1995). "Overexpression of integral membrane proteins for structural studies." Q. Rev. Biophys. **28**(3): 315-422.
- Guidotti, S.-H. L. a. G. (2009). "Purification of Membrane Proteins." Methods in Enzymology **463**: 619-629.
- H. M. Berman, T. B., T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook and C. Zardecki (2002). "The Protein Data Bank." Acta Cryst. **D58**(6): 899-907.
- H.J.C. Berendsen, D. v. d. S., R. van Drunen (1995). "GROMACS: A message-passing parallel molecular dynamics implementation." Computer Physics Communications **91**(1-3): 43-56.
- Hartzell, C. J., M. Whitfield, et al. (1987). "Determination of the <sup>15</sup>N and <sup>13</sup>C Chemical Shift Tensors of L-[<sup>13</sup>C]Alanyl-L-[<sup>15</sup>N]alanine from the Dipole-Coupled Powder Patterns." J. Am. Chem. Soc. **109**: 5966-5969.
- HEIJNE, E. W. a. G. V. (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms " Protein Science **7**: 1029-1038.
- Herzfeld, J. and A. E. Berger (1980). "Sideband intensities in NMR spectra of samples spinning at the magic angle." J. Chem. Phys. **73**: 6021-6030.
- Hong, M. (2006). "Oligomeric Structure, Dynamics, and Orientation of Membrane Proteins from Solid-State NMR." Structure **14**(12): 1731-1740.
- Im, W., M. Feig, et al. (2003). "An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins." Biophysical Journal **85**(5): 2900-2918.
- Isabelle Marcotte, M. A. (2005). "Bicelles as model membranes for solid- and solution-state NMR studies of membrane peptides and proteins." Concepts in Magnetic Resonance Part A **24A**(1): 17-37.
- J. R. Quine, T. A. C., Michael S. Chapman, Richard Bertram (2004). "Mathematical aspects of protein structure determination with NMR orientational restraints." Bulletin of Mathematical Biology **66**(6): 1705-1730.

J. Wang, J. D., C. Tian, S. Kim, Y. Mo, F. Kovacs, Z. Song, K. Nishimura, Z. Gan, R. Fu, J.R. Quine, T.A. Cross (2000). "Imaging Membrane Protein Helical Wheels." Journal of Magnetic Resonance **144**(1): 162–167.

Jack Kyte, R. F. D. (1982). "A simple method for displaying the hydrophobic character of a protein." Journal of Molecular Biology **157**(1): 105–132.

James Thompson, D. B. (2011). "Incorporation of evolutionary information into Rosetta comparative modeling." Proteins: Structure, Function, and Bioinformatics **79**(8): 2380–2388.

Jaroniec, C. P., B. A. Tounge, et al. (2001). "Frequency Selective Heteronuclear Dipolar Recoupling in Rotating Solids: Accurate  $^{13}\text{C}$ - $^{15}\text{N}$  Distance Measurements in Uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled Peptides." J. Am. Chem. Soc. **123**: 3507-3519.

Jaume Torres, T. J. S. a. M. S. (2003). "Membrane proteins: the 'Wild West' of structural biology." TRENDS in Biochemical Sciences **28**(3): 137-144.

Jay W. Ponder, D. A. C. (2003). "Force Fields for Protein Simulations." Advances in Protein Chemistry **66**: 27-85.

Jeffrey J. Gray, S. M., Chu Wang, Ora Schueler-Furman, Brian Kuhlman, Carol A. Rohl, David Baker (2003). "Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations." Journal of Molecular Biology **331**(1): 281–299.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." Journal of Molecular Biology **292**(2): 195–202.

Kaiser, R. (1963). "Use of the Nuclear Overhauser Effect in the Analysis of High-Resolution Nuclear Magnetic Resonance Spectra." Journal of Chemical Physics **39**(10): 2435.

Karplus, R. L. D. J. a. M. (1994). "Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains." Nature Structural Biology **1**(5): 334 - 340.

Kazumi Shimono, M. G., Takashi Kikukawa, Seiji Miyauchi, Mikako Shirouzu, Naoki Kamo, and Shigeyuki Yokoyama (2009). "Production of functional bacteriorhodopsin by an Escherichia coli cell-free protein synthesis system supplemented with steroid detergent and lipid." Protein Science **18**(10): 2160—2171.

Kevin Karplus, R. K., Christian Barrett, Spencer Tu, Melissa Cline, Mark Diekhans, Leslie Grate, Jonathan Casper, Richard Hughey (2001). "What is the value added by human intervention in protein structure prediction?" Proteins: Structure, Function, and Bioinformatics **45**(Supplement 5): 86–91.

Kevin R. MacKenzie, J. H. P., and Donald M. Engelman (1997). "A Transmembrane Helix Dimer: Structure and Implications." Science **276**(5309): 131-133.

Kevin W. Waddell, E. Y. C., and Richard J. Wittebort (2005). "Single-Crystal Studies of Peptide Prolyl and Glycyl  $^{15}\text{N}$  Shielding Tensors." J. Am. Chem. Soc. **127**(25): 9030–9035.

Kim T. Simons, C. K., Enoch Huang, David Baker (1997). "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions." Journal of Molecular Biology **268**(1): 209–225.

Kim T. Simons, I. R., Charles Kooperberg, Brian A. Fox, Chris Bystroff, David Baker (1999). "Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins." Proteins: Structure, Function, and Bioinformatics **34**(1): 82-95.

Kira M. S. Misura, D. C., Carol A. Rohl, David E. Kim, and David Baker (2006). "Physically realistic homology models built with ROSETTA can be more accurate than their templates." Proceedings of the National Academy of Sciences of the United States of America **103**(14): 5361–5366.

Koradi, R., M. Billeter, et al. (1996). "MOLMOL: A program for display and analysis of macromolecular structures." J. Mol. Graphics **14**: 51-55.

Laskowski, R. A. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Crystallogr. **26**: 283-291.

Lee, D. K., J. S. Santos, et al. (1999). "Nitrogen-15 chemical shift anisotropy and  $^1\text{H}$ - $^{15}\text{N}$  dipolar coupling tensors associated with the phenylalanine residue in the solid state." Chem. Phys. Lett. **309**: 209-214.

Lee, D. K., R. J. Wittebort, et al. (1998). "Characterization of  $^{15}\text{N}$  Chemical Shift and  $^1\text{H}$ - $^{15}\text{N}$  Dipolar Coupling Interactions in a Peptide Bond of Uniaxially Oriented and Polycrystalline Samples by One-Dimensional Dipolar Chemical Shift Solid-State NMR Spectroscopy." J. Am. Chem. Soc. **120**: 8868-8874.

Loth, K., P. Pelupessy, et al. (2005). "Chemical Shift Anisotropy Tensors of Carbonyl, Nitrogen, and Amide Proton Nuclei in Proteins through Cross-Correlated Relaxation in NMR Spectroscopy." J. Am. Chem. Soc. **127**: 6062-6068.

Lovell, S. C., I. W. Davis, et al. (2003). "Structure validation by C alpha geometry: phi,psi and C beta deviation." Proteins-Structure Function and Genetics **50**(3): 437-450.

- Lowe, I. J. (1959). "Free Induction Decays of Rotating Solids." Physical Review Letter **2**(7): 285-287.
- Lucio Frydman, J. S. H. (1995). "Isotropic Spectra of Half-Integer Quadrupolar Spins from Bidimensional Magic-Angle Spinning NMR." J. Am. Chem. Soc. **117**(19): 5367-5368.
- Luecke, H., B. Schobert, et al. (1999). "Structure of Bacteriorhodopsin at 1.55 Angstrom Resolution." J. Mol. Biol. **291**: 899-911.
- Luo, W., R. Mani, et al. (2007). "Side-Chain Conformation of the M2 Transmembrane Peptide Proton Channel of Influenza A Virus from  $^{19}\text{F}$  Solid-State NMR." J. Phys. Chem. B **111**: 10825-10832.
- M. Hohwy, C. P. J., B. Reif, C. M. Rienstra, and R. G. Griffin (2000). "Local Structure and Relaxation in Solid-State NMR: Accurate Measurement of Amide N–H Bond Lengths and H–N–H Bond Angles." J. Am. Chem. Soc. **122**(13): 3218-3219.
- MacArthur, M. W. and J. M. Thornton (1996). "Deviations from Planarity of the Peptide Bond in Peptides and Proteins." J. Mol. Biol. **264**: 1180-1195.
- Mads Bak, R. S., Thomas Vosegaard, Niels Chr. Nielsen (2002). "Specification and Visualization of Anisotropic Interaction Tensors in Polypeptides and Numerical Simulations in Biological Solid-State NMR." Journal of Magnetic Resonance **154**(1): 28-45.
- Marassi, F. M. and S. J. Opella (2000). "A Solid-State NMR Index of Helical Membrane Protein Structure and Topology." J. Magn. Reson. **144**: 150-155.
- Marassi, F. M. and S. J. Opella (2002). "Using Pisa pies to resolve ambiguities in angular constraints from PISEMA spectra of aligned proteins." J. Biomol. NMR **23**: 239-242.
- Marassi, S. J. O. a. F. M. (2004). "Structure Determination of Membrane Proteins by NMR Spectroscopy." Chem. Rev. **104**(8): 3587-3606.
- Markus Sällman Almén, K. J. N., Robert Fredriksson and Helgi B Schiöth (2009). "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin." BMC Biology **7**(50): 1-14.
- Maxwell, E. L. H. a. D. E. (1952). "Spin Echo Measurements of Nuclear Spin Coupling in Molecules." Physical Review **88**(5): 1070-1084.
- Meiler, J. (2003). "JUFO3D: Secondary structure prediction for proteins from low resolution tertiary structure. ."

Mukesh Sharma, M. Y., Hao Dong, Huajun Qin, Emily Peterson, David D. Busath, Huan-Xiang Zhou, Timothy A. Cross (2010). "Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer." Science **330**(6003): 509-512

Naomi Siew, A. E., Leszek Rychlewski and Daniel Fischer (2000). "MaxSub: an automated measure for the assessment of protein structure prediction quality." Bioinformatics **16**(9): 776-785.

Nevzorov, A. A. and S. J. Opella (2003). "Structural fitting of PISEMA spectra of aligned proteins." J. Mag. Reson. **160**: 33-39.

Nevzorov, A. A. and S. J. Opella (2006). Analytical Framework for Protein Structure Determination by Solid-State NMR of Aligned Samples. Handbook of Magnetic Resonance. G. A. Webb. The Netherlands, Kluwer.

Nevzorov, A. A. and S. J. Opella (2007). "Selective averaging for high-resolution solid-state NMR spectroscopy of aligned samples." J. Magn. Reson. **185**: 59-70.

Nikolaos G. Sgourakis, O. F. L., Frank DiMaio, Ingemar André, Nicholas C. Fitzkee, Paolo Rossi, Gaetano T. Montelione, Ad Bax, and David Baker (2011). "Determination of the Structures of Symmetric Protein Oligomers from NMR Chemical Shifts and Residual Dipolar Couplings." J. Am. Chem. Soc. **133**(16): 6288–6298.

Oas, T. G., C. J. Hartzell, et al. (1987). "The Amide <sup>15</sup>N Chemical Shift Tensors of Four Peptides Determined from <sup>13</sup>C Dipole-Coupled Chemical Shift Powder Patterns." J. Am. Chem. Soc. **109**: 5962-5966.

Opella, S. J., Y. Kim, et al. (1994). "Experimental nuclear magnetic resonance studies of membrane proteins." Methods Enzymol. **239**: 536-560.

Opella, S. J., P. L. Stewart, et al. (1987). "Protein structure by solid-state NMR spectroscopy." Q. Rev. Biophys. **19**: 7-49.

Opella, S. J., A. C. Zeri, et al. (2008). "Structure, Dynamics, and Assembly of Filamentous Bacteriophages by Nuclear Magnetic Resonance Spectroscopy." Annu. Rev. Phys. Chem. **59**: 635-657.

Overhauser, A. W. (1953). "Polarization of Nuclei in Metals." Physical Review **92**(2): 411-415.

P. Barth, B. W. a. D. B. (2009). "Prediction of membrane protein structures with complex topologies using limited constraints." Proceedings of the National Academy of Sciences of the United States of America **106**(5): 1409–1414.

P. Barth, J. S., and D. Baker (2007). "Toward high-resolution prediction and design of transmembrane helical protein structures." Proceedings of the National Academy of Sciences of the United States of America **104**(40): 15682–15687.

P. Benjamin Stranges, M. M., Michael J. Miley, Ashutosh Tripathy, and Brian Kuhlman (2011). "Computational design of a symmetric homodimer using beta-strand assembly." Proceedings of the National Academy of Sciences **108**(51): 20562-20567.

Page, R. C., S. Kim, et al. (2008). "Transmembrane Helix Uniformity Examined by Spectral Mapping of Torsion Angles." Structure **16**: 787-797.

Palczewski, K. (2006). "G Protein–Coupled Receptor Rhodopsin." Annu. Rev. Biochem. **75**: 743-767.

Park, S. H., C. Loudet, et al. (2008). "Solid-state NMR spectroscopy of a membrane protein in biphenyl phospholipid bicelles with the bilayer normal parallel to the magnetic field." J. Magn. Reson. **193**: 133-138.

Park, S. H. and S. J. Opella (2010). "Triton X-100 as the "Short-Chain Lipid" Improves the Magnetic Alignment and Stability of Membrane Proteins in Phosphatidylcholine Bilayers for Oriented-Sample Solid-State NMR Spectroscopy." J. Am. Chem. Soc. **132**: 12552-12553.

Philip Bradley, K. M. S. M., David Baker (2005). "Toward High-Resolution de Novo Structure Prediction for Small Proteins." Science **309**(5742): 1868-1871

Ramachandran, G. N. (1968). "Need for Nonplanar Peptide Units in Polypeptide Chains." Biopolymers **6**(10): 1494-1496.

Ramamoorthy, A. (1998). "Phase-alternated composite pulses for zero-field NMR spectroscopy of spin 1 systems." Mol. Phys. **93**: 757-766.

Rhiju Das, I. A., Yang Shen, Yibing Wu, Alexander Lemak, Sonal Bansal, Cheryl H. Arrowsmith, Thomas Szyperski, and David Baker (2009). "Simultaneous prediction of protein folding and docking at high resolution." Proceedings of the National Academy of Sciences **106**(45): 18978-18983.

Richard Bonneau, C. E. M. S., Carol A Rohl, Dylan Chivian, Phillip Bradley, Lars Malmström, Tim Robertson, David Baker (2002). "De Novo Prediction of Three-dimensional Structures for Major Protein Families." Journal of Molecular Biology **322**(1): 65–78.

Roland L. Dunbrack Jr, M. K. (1993). "Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction." Journal of Molecular Biology **230**(2): 543–574.

- Roland L. Dunbrack Jr., F. E. C. (1997). "Bayesian statistical analysis of protein side-chain rotamer preferences." Protein Science **6**(8): 1661–1681.
- Salem Faham, D. Y., Emiko Bare, Sarah Yohannan, Julian P. Whitelegge, James U. Bowie (2004). "Side-chain Contributions to Membrane Protein Structure and Stability." Journal of Molecular Biology **335**(1): 297–305.
- Sang Ho Park, A. A. M., Alexander A. Nevzorov, Anna A. De Angelis, Stanley J. Opella (2006). "Rotational diffusion of membrane proteins in aligned phospholipid bilayers by solid-state NMR spectroscopy." Journal of Magnetic Resonance **178**(1): 162–165.
- Sang Ho Park, F. M. M., David Black, Stanley J. Opella (2010). "Structure and Dynamics of the Membrane-Bound Form of Pf1 Coat Protein: Implications of Structural Rearrangement for Virus Assembly." Biophysical Journal **99**(5): 1465 - 1474.
- Sanguk Kim, T. A. C. (2004). "2D solid state NMR spectral simulation of 310,  $\alpha$ , and  $\pi$ -helices." Journal of Magnetic Resonance **168**(2): 187–193.
- Sarah D. Cady, K. S.-R., Jun Wang, Cinque S. Soto, William F. DeGrado and Mei Hong (2010). "Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers." Nature **463**(4): 689-692.
- Scheiner, S. and C. W. Kern (1977). "Theoretical Studies of Environmental Effects on Protein Conformation. 1. Flexibility of the Peptide Bond." J. Am. Chem. Soc. **99**: 7042-7050.
- Schnell JR, C. J. (2008). "Structure and mechanism of the M2 proton channel of influenza A virus." Nature **451**(7178): 591-595.
- Shi, L., N. J. Traaseth, et al. (2009). "A refinement protocol to determine structure, topology and depth of insertion of membrane proteins using hybrid solution and solid-state NMR restraints." J. Biomol. NMR **44**: 195-205.
- Simon Silver, L. T. P. (2005). "A bacterial view of the periodic table: genes and proteins for toxic inorganic ions." Journal of Industrial Microbiology and Biotechnology **32**(11-12): 587-605.
- Sinha, N., C. V. Grant, et al. (2007). "Triple resonance experiments for aligned sample solid-state NMR of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled proteins." J. Magn. Reson. **186**: 51-64.
- Standley, D. M., V. A. Eyrich, et al. (1999). "A Branch and Bound Algorithm for Protein Structure Refinement from Sparse NMR Data Sets." J. Mol. Biol. **285**: 1691-1710.

Stanley C. Howell, M. F. M., and Stanley J. Opella (2005). "NMR Structure Determination of a Membrane Protein with Two Transmembrane Helices in Micelles: MerF of the Bacterial Mercury Detoxification System." Biochemistry **44**(13): 5196–5206.

Stephanie J. Hirst, N. A., Hassane S. Mchaourab, Jens Meiler (2011). "RosettaEPR: An integrated tool for protein structure determination from sparse EPR data." Journal of Structural Biology **173**(3): 506–514.

Stewart, P. L., K. G. Valentine, et al. (1987). "Structural-Analysis of Solid-State Nmr Measurements of Peptides and Proteins." Journal of Magnetic Resonance **71**(1): 45-61.

Stewart, P. L., K. G. Valentine, et al. (1987). "Structural Analysis of Solid-State NMR Measurements of Peptides and Proteins." J. Magn. Reson. **71**: 45-61.

Tanja Kortemme, A. V. M., David Baker (2003). "An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes." Journal of Molecular Biology **326**(4): 1239–1259.

Tara Hessa, H. K., Karl Bihlmaier, Carolina Lundin, Jorrit Boekel, Helena Andersson, IngMarie Nilsson, Stephen H. White and Gunnar von Heijne (2005). "Recognition of transmembrane helices by the endoplasmic reticulum translocon." Nature **433**(7024): 377-381.

Tate, C. G. (2001). "Overexpression of mammalian integral membrane proteins for structural studies." FEBS Letters **504**: 94-98.

Themis Lazaridis, M. K. (1999). "Effective energy function for proteins in solution." Proteins: Structure, Function, and Bioinformatics **35**(2): 133-152.

Themis Lazaridis, M. K. (2003). "Effective energy function for proteins in lipid membranes." Proteins: Structure, Function, and Bioinformatics **52**(2): 176–192.

Thomas Vosegaard, N. C. N. (2002). "Towards high-resolution solid-state NMR on large uniformly <sup>15</sup>N- and [<sup>13</sup>C,<sup>15</sup>N]-labeled membrane proteins in oriented lipid bilayers." Journal of Biomolecular NMR **22**(3): 225-247.

Toyoshima C, N. M., Nomura H, Ogawa H. (2000). "Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution." Nature **405**(6787): 647-655.

Traaseth, N. J., L. Shi, et al. (2009). "Structure and topology of monomeric phospholamban in lipid membranes determined by a hybrid solution and solid-state NMR approach." Proc. Natl. Acad. Sci. U. S. A. **106**: 10165-10170.

V. B. Chen, W. B. A., J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Cryst. **D66**(1): 12-21.

Vladimir Yarov-Yarovoy, J. S., David Baker (2006). "Multipass membrane protein structure prediction using Rosetta." Proteins: Structure, Function, and Bioinformatics **62**(4): 1010–1025.

W-K. Rhim, D. D. E., and R. W. Vaughan (1973). "Analysis of multiple pulse NMR in solids." Journal of Chemical Physics **59**(7): 3740-3749.

Walling, A. E., R. E. Pargas, et al. (1997). "Chemical Shift Tensors in Peptides: A Quantum Mechanical Study." J. Phys. Chem. A **101**: 7299-7303.

WHITE, S. H. (2004). "The progress of membrane protein structure determination." Protein Science **13**: 1948-1949.

White, S. H. (2006). "How Hydrogen Bonds Shape Membrane Protein Structure." Advances in Protein Chemistry **72**: 157–172.

White, W. C. W. a. S. H. (1996). "Experimentally determined hydrophobicity scale for proteins at membrane interfaces." Nature Structural Biology **3**(10): 842 - 848.

William J. Wedemeyer, D. B. (2003). "Efficient minimization of angle-dependent potentials for polypeptides in internal coordinates." Proteins: Structure, Function, and Bioinformatics **53**(2): 262-272.

Wimley, W. C. (2003). "The versatile b-barrel membrane protein." Current Opinion in Structural Biology **13**: 404-411.

Wu, C. H. and S. J. Opella (2008). "Shiftless nuclear magnetic resonance spectroscopy." J. Chem. Phys. **128**(5): 052312.

Wu, C. H., A. Ramamoorthy, et al. (1995). "Simultaneous Characterization of the Amide <sup>1</sup>H Chemical Shift, <sup>1</sup>H-<sup>15</sup>N Dipolar, and <sup>15</sup>N Chemical Shift Interaction Tensors in a Peptide Bond by Three-Dimensional Solid-State NMR Spectroscopy." J. Am. Chem. Soc. **117**: 6148-6149.

Wu, C. H., A. Ramamoorthy, et al. (1994). "High-Resolution Heteronuclear Dipolar Solid-State NMR-Spectroscopy." J. Magn. Reson., Ser. A **109**: 270-272.

Wylie, B. J., L. J. Sperling, et al. (2007). "Chemical-Shift Anisotropy Measurements of Amide and Carbonyl Resonances in a Microcrystalline Protein with Slow Magic-Angle Spinning NMR Spectroscopy." J. Am. Chem. Soc. **129**: 5318-5319.

Yalini Arinaminpathy, E. K., Donald M. Engelman and Mark B. Gerstein (2009). "Computational analysis of membrane proteins: the largest class of drug targets." Drug Discovery Today **14**: 1130-1135.

Yang Shen, O. L., Frank Delaglio, Paolo Rossi, James M. Aramini, Gaohua Liu, Alexander Eletsky, Yibing Wu, Kiran K. Singarapu, Alexander Lemak, Alexandr Ignatchenko, Cheryl H. Arrowsmith, Thomas Szyperski, Gaetano T. Montelione, David Baker and Ad Bax (2008). "Consistent blind protein structure generation from NMR chemical shift data." Proceedings of the National Academy of Sciences of the United States of America **105**(12): 4685-4690.

Yang Shen, R. V., David Baker, Ad Bax (2009). "De novo protein structure generation from incomplete chemical shift assignments." Journal of Biomolecular NMR **43**(2): 63-78.

Yu L, S. C., Song D, Shen J, Xu N, Gunasekera A, Hajduk PJ, Olejniczak ET (2005). "Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex." Biochemistry **44**(48): 15834-15841.

Yuanyuan Yin, A. A. N. (2011). "Structure Determination in Shiftless Solid State NMR of Oriented Protein Samples." Journal of Magnetic Resonance **212**(1): 64-73.

## APPENDICES

## Appendix A

### A.1 Variations in the observable $^{15}\text{N}$ CSA values due to uncertainties in the principal values of the CSA tensor and its orientation

Typically, for oriented-sample (OS) NMR structure calculations, the principal components of the  $^{15}\text{N}$  CSA values are approximated by the same “average” values:  $\sigma_{11\text{N}}=64$  ppm,  $\sigma_{22\text{N}}=77$  ppm,  $\sigma_{33\text{N}}=217$  ppm (Wu, Ramamoorthy et al. 1995; Nevzorov and Opella 2003) for all amino-acid residues (for glycines these values are assumed to be  $\sigma_{11\text{N}}=41$  ppm,  $\sigma_{22\text{N}}=64$  ppm,  $\sigma_{33\text{N}}=210$  ppm). However, several findings (Marassi and Opella 2000; Marassi and Opella 2002; De Angelis, Howell et al. 2006; Page, Kim et al. 2008) have demonstrated that the variation of either the principal components or the orientation of the CSA tensor is influenced by the local chemical and geometrical environment of the polypeptide chains. The Rienstra group measured residue-specific experimental values for the principal values of  $^{15}\text{N}$  CSA for  $\beta 1$  immunoglobulin binding protein G by magic-angle spinning (MAS) NMR, and demonstrated that for different secondary structures, such as  $\beta$ -strand,  $\alpha$ -helical, and turn residues, the isotropic  $^{15}\text{N}$  chemical shifts were very different ( $\beta$ -strand:  $121.0\pm 6.5$  ppm,  $\alpha$ -helical:  $119.8\pm 2.5$  ppm, turn:  $117.2\pm 8.0$ ) (Franks, Zhou et al. 2005; Wylie, Sperling et al. 2007). In addition, Loth et.al reported the uncertainty in the orientation of the  $^{15}\text{N}$  CSA tensors along the protein backbone defined by the angle  $\beta_N$  ( $\gamma_N$  in our notation) ranging from  $12^\circ$  to  $25^\circ$  (Loth, Pelupessy et al. 2005). Nevertheless, to assess the convergence of the back-calculated structures by the method of orientational constraints, it is permissible to simulate the “experimental” data with the same average tensor values as long as the same values are

consistently used for the spectral fitting. Here we estimate the effect of tensor variations on the observable range of the  $^{15}\text{N}$  CSA in protein G by randomizing  $\sigma_{11N}$ ,  $\sigma_{22N}$ ,  $\sigma_{33N}$  within  $\pm 1$  ppm relative to their average values, and also the angle  $\gamma_N=18.5^\circ$  within  $\pm 1^\circ$ , which defines the orientation of the tensor relative to the peptide plane (the local molecular frame).

Figure A.1 shows the effect of the  $\pm 1$  ppm random variation in the magnitudes of the three principal components of the  $^{15}\text{N}$  CSA. A two-dimensional projection of the simulated three-dimensional spectra is displayed in the top portion of the figure. The calculated observable CSA values are shown by red dots and the fixed-tensor values correspond to the centers of the blue circles. The histograms of the deviation between the simulated CSA (calculated 1000 times for the randomized principal values) and the fixed-tensor data at each residue show that the deviations have a Gaussian-like distribution with an average half-width of 1 ppm. When the orientation of the CSA tensor (as given by the angle  $\gamma_N=18.5^\circ$ ) is randomly varied within  $\pm 1^\circ$ , these deviations are translated into the observed CSA values non-homogeneously. The histograms in Figure A.2 indicate that the half-widths of the distributions range from 0.5 ppm to 3 ppm. Some distributions have a Gaussian type, and some are more uniform. Figure A.3 shows the superimposed effect, in which  $\sigma_{11N}$ ,  $\sigma_{22N}$ ,  $\sigma_{33N}$  and  $\gamma_N$  are randomly varied within  $\pm 1$  ppm and  $\pm 1^\circ$ , respectively. Therefore, even if the principal components of the CSA tensor could be determined for each amino acid to within  $\pm 1$  ppm accuracy, much greater uncertainty can arise from the variability of the angle  $\gamma_N$ , which ultimately affects the accuracy in the determination of protein structures when fixed CSA values are employed.

## **A.2 Structural Fitting of Protein G of Four-Dimensional Data with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ , $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ and $^{13}\text{C}'$ - $^{15}\text{N}$ Dipolar Couplings with Larger Experimental Uncertainty**

The introduction of a fourth dimension,  $^{13}\text{C}'$ - $^{15}\text{N}$  dipolar couplings, does improve the convergence of the calculated structures when the “experimental” uncertainty is taken into account. (See Figure 2.7) However, when the tolerance for the input data is increased to 25Hz and even up to 50Hz in all the four dimensions, the back-calculated structures begin to diverge. The histograms in Supplementary Figure S4 show that at least 50% of the RMSDs are above 4Å.

### A.3 Structural Fitting of Three-Dimensional Data for Protein G with $^1\text{H}$ - $^{15}\text{N}$ , $^1\text{H}$ - $^{13}\text{C}_\alpha$ , and $^{13}\text{C}_\alpha$ - $^{15}\text{N}$ Dipolar Couplings Including Distance Restraints

When the distances between the  $\text{C}_\alpha$  atoms of residues Ile6 and Thr53 and residues Glu15 and Thr44 are included as additional restraints, the back-calculated structures with the RMSDs less than  $2\text{\AA}$  (for a tolerance of 25 Hz) and  $2.5\text{\AA}$  (for a tolerance of 50 Hz) can be successfully filtered out. As shown in Figure 2.8, for a tolerance of 25 Hz, 319 out of the total 1000 structural fits have the RMSDs less than  $2\text{\AA}$ , and the number of the back-calculated structures that satisfy the distance restraints within  $\pm 1\text{\AA}$  is 312; for a tolerance of 50 Hz, 49 out of the total 1000 structural fits have the RMSDs less than  $2.5\text{\AA}$ , and 33 back-calculated structures satisfy the distance restraints.

Figure A.5A and A.5B show that for the 25Hz tolerance fits, the distance restraint for the  $\text{C}_\alpha$  atoms of residues Glu15 and Thr44 has a more significant effect on the structural convergence than that for residues Ile6 and Thr53. From the histograms it can be seen that about 72% of the Ile6-Thr53 distances are already within  $\pm 1\text{\AA}$  relative to the target value, while only 38% of the Glu15-Thr44 distances fall within the same range, which ultimately affects the total percentage of the back-calculated structures having the RMSDs less than  $2\text{\AA}$ . However, when the tolerance is increased to 50Hz (cf. Fig. A.5C and A.5D), about 14% of Ile6-Thr53 distances and about 21% of Glu15-Thr44 distances are within  $\pm 1\text{\AA}$ , which are both larger than the final percentage (3%) of the accepted structures with the RMSDs less than  $2.5\text{\AA}$ . As the fitting tolerance (and experimental uncertainty) is increased, more distance restraints will likely be necessary to obtain convergent sets of solutions.

#### **A.4 Structural Fitting of a Helical Hairpin Derived from Bacteriorhodopsin (1C3W) using Three-Dimensional Data with Larger Experimental Uncertainty**

Three-dimensional “shiftless” spectra including  $^1\text{H}$ - $^{15}\text{N}$ ,  $^1\text{H}$ - $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\alpha$ - $^{15}\text{N}$  dipolar couplings simulated from a helical hairpin from bacteriorhodopsin (1C3W) yielded a converged set of solutions with the tolerance up to 120 Hz. To investigate the influence of larger “experimental” uncertainty on the helical domains, simulations were performed with the tolerance of 200 Hz, 250 Hz, 300 Hz and 350 Hz respectively. Figure A.6 shows that when the tolerance is increased to 350 Hz, the back-calculated structures begin to exhibit evident deviations, with about 30% of structures having the RMSDs above 5Å.

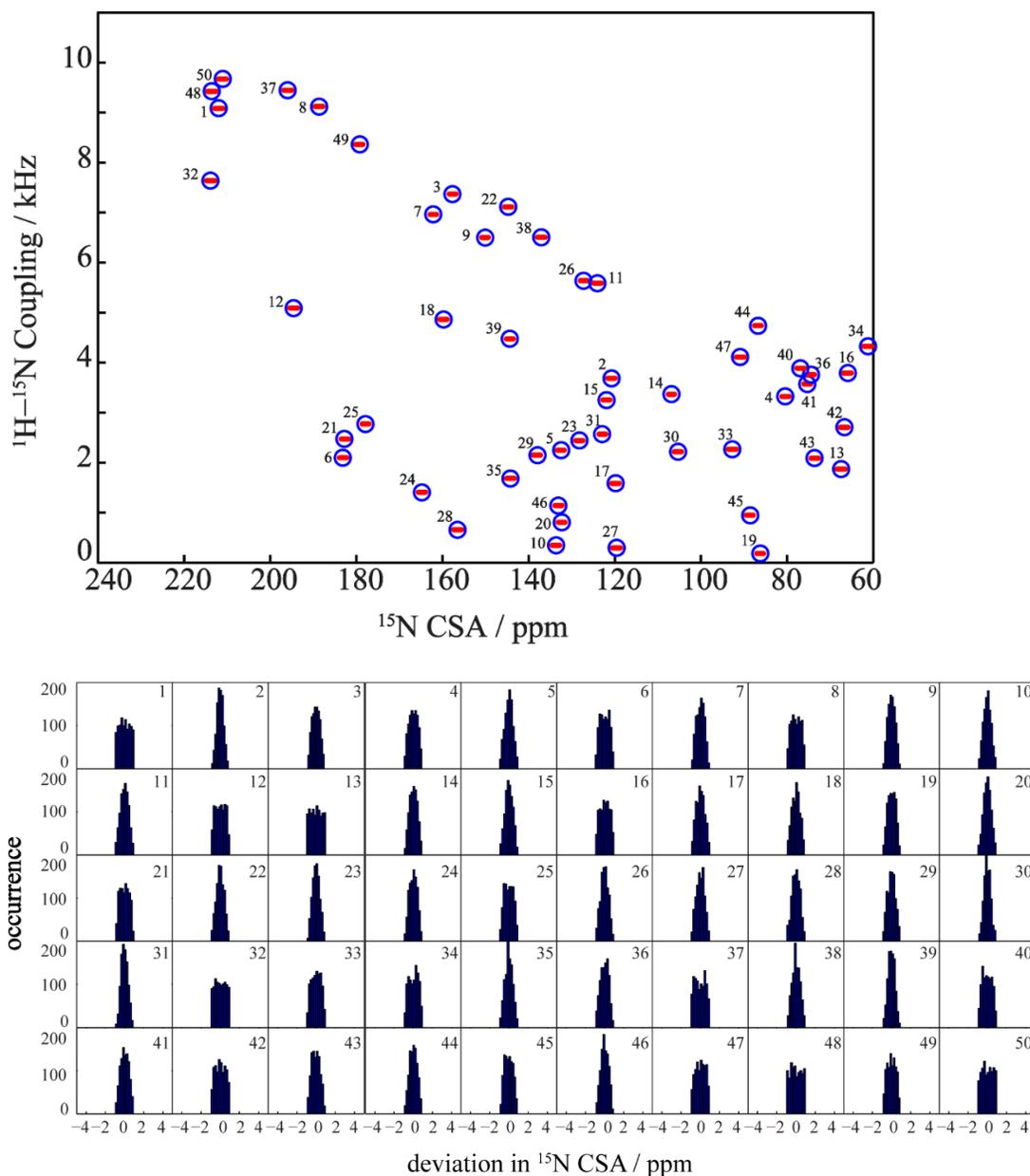


Figure A.1 Effect of the variations in the principal components of the  $^{15}\text{N}$  CSA tensor on the observed CSA values for individual  $^{15}\text{N}$  sites of protein G ( $\sigma_{11N}$ ,  $\sigma_{22N}$  and  $\sigma_{33N}$  were all randomly varied to within  $\pm 1$  ppm and 1000 simulations were performed). Red dots show variations of the observed CSA in the 2D spectrum relative to the fixed-tensor data corresponding to the centers of blue circles. The numbering of the residues is indicated in the spectrum. Histograms show distributions of the deviations from the fixed-tensor data for individual residues.

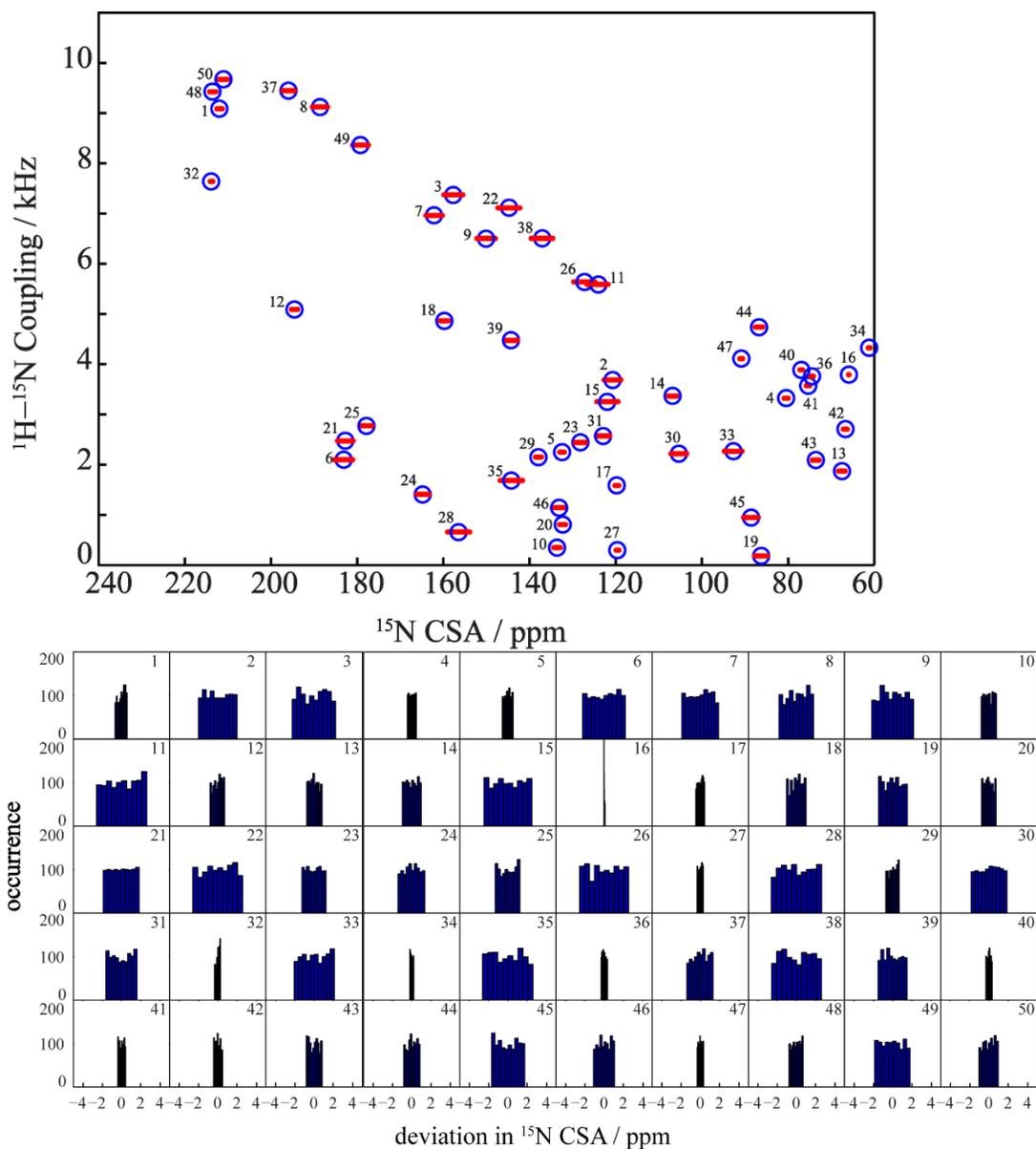


Figure A.2 Effect of the orientation of the  $^{15}\text{N}$  CSA tensor on the observed CSA values for protein G. Red dots show variations of the observed CSA in the 2D spectrum relative to the fixed-tensor data corresponding to the centers of blue circles. The numbering of the residues is indicated in the spectrum. The angle  $\gamma_N$  was randomly varied within  $\pm 1^\circ$  for each  $^{15}\text{N}$  site in protein G for 1000 simulations. Histograms show deviations from the fixed-tensor data.

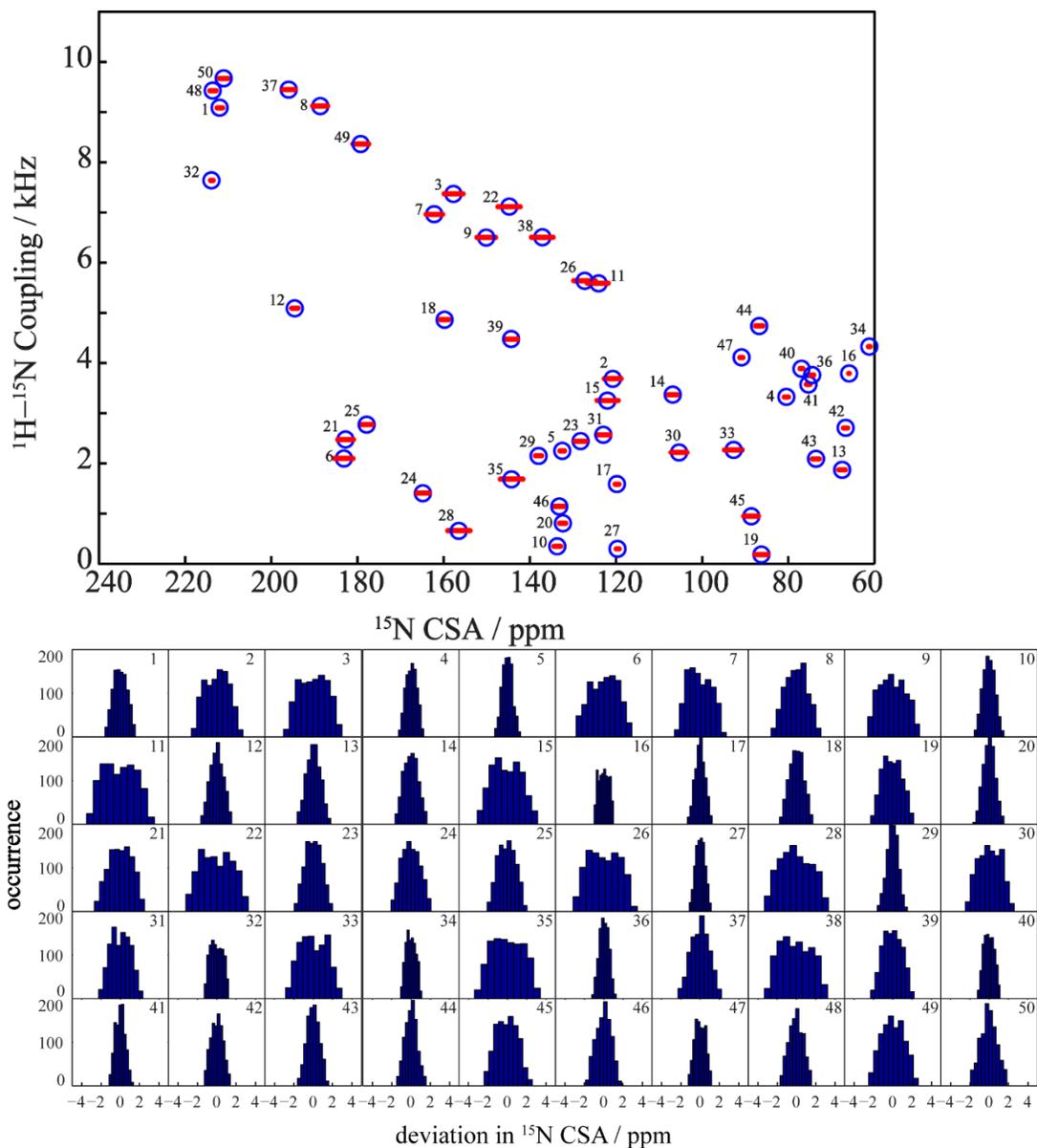


Figure A.3 Combined effect of the variations in both the principal components of the  $^{15}\text{N}$  CSA tensor and its orientation on the individual observed CSA values for protein G. Red dots show variations of the observed CSA in the 2D spectrum relative to the fixed-tensor data corresponding to the centers of blue circles. The numbering of the residues is indicated in the spectrum. The values for  $\sigma_{11N}$ ,  $\sigma_{22N}$  and  $\sigma_{33N}$  were all randomly varied within  $\pm 1$  ppm, and the angle  $\gamma_N$  was randomly varied within  $\pm 1^\circ$ .

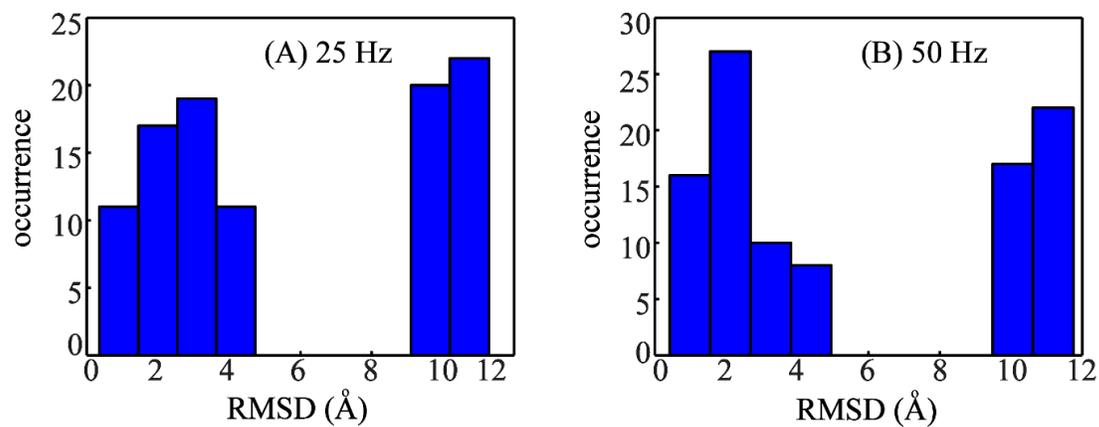


Figure A.4 Histograms of the RMSDs of 1000 back-calculated structures by simulating four-dimensional “shiftless” solid state NMR spectra of protein G (2GB1) with the tolerance of (A) 25Hz (B) 50 Hz.

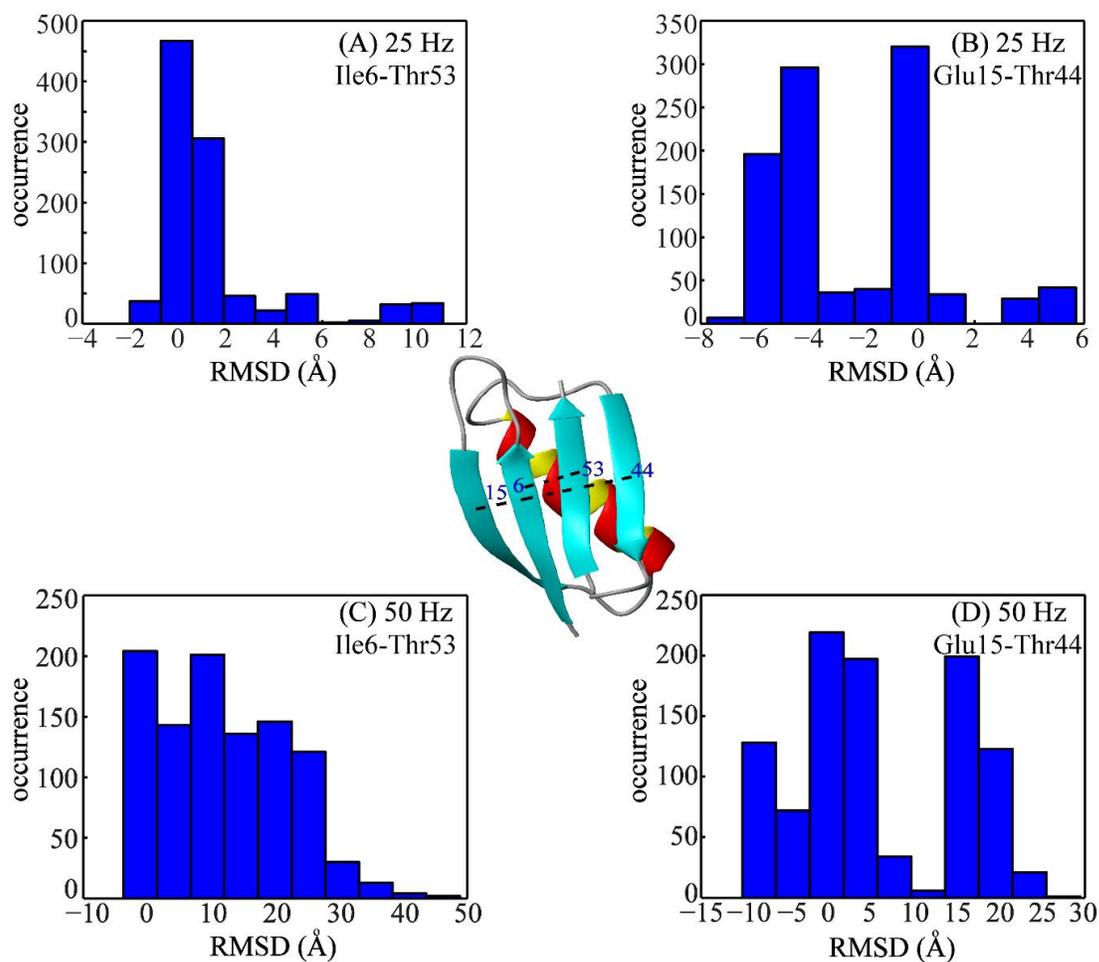


Figure A.5 Histograms of the deviations of the back-calculated distances from the “true” distances (from 2GB1) corresponding to the C $\alpha$  atoms of residues Ile6 and Thr53, and residues Glu15 and Thr44. (A) 25 Hz tolerance, Ile6-Thr53 distances; (B) 25 Hz tolerance, Glu15-Thr44 distances; (C) 50 Hz tolerance, Ile6-Thr53 distances; (D) 50 Hz tolerance, Glu15-Thr44 distances.

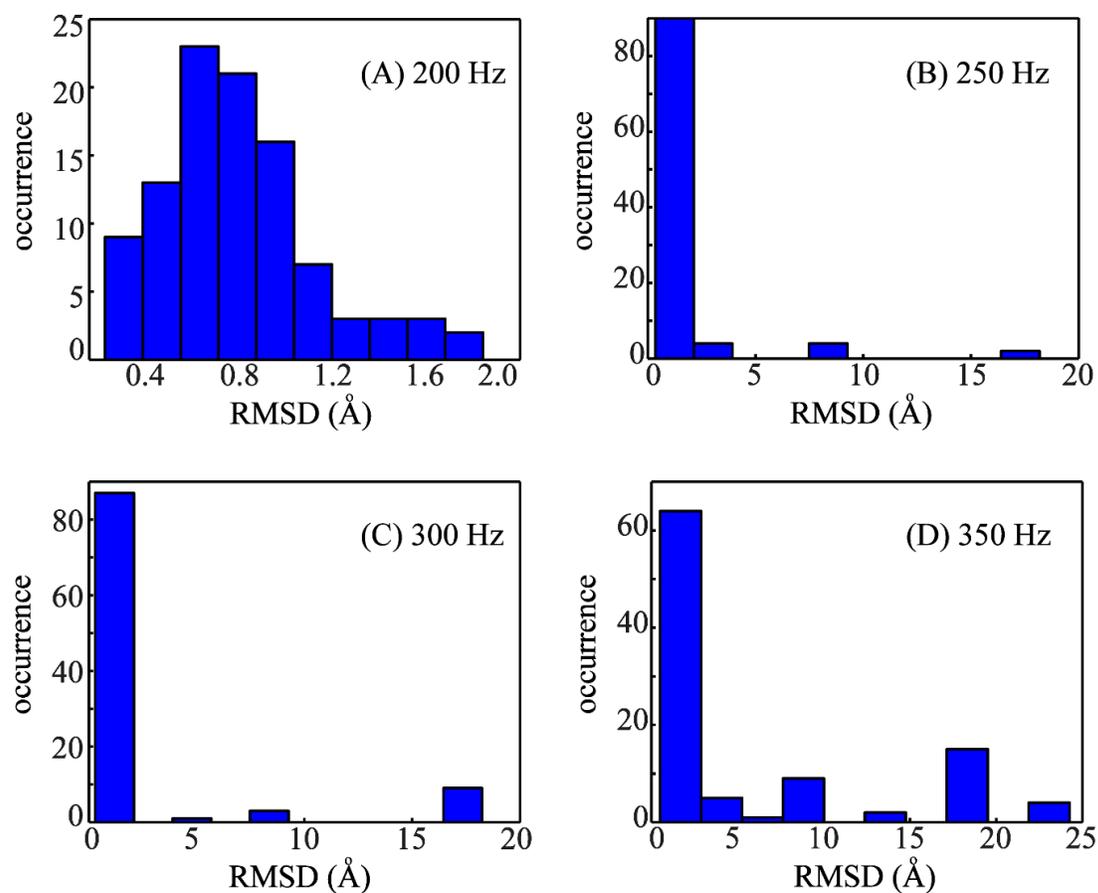


Figure A.6 Histograms of the RMSDs for the back-calculated structures obtained from fitting three-dimensional “shiftless” spectra for a helical hairpin of bacteriorhodopsin (1C3W) with the tolerances of (A) 200 Hz (B) 250 Hz (C) 300 Hz (D) 350 Hz.

## Appendix B

### B.1 Structural Calculation from Solid-State NMR Experimental Data

Calculate the membrane protein structure using chemical shift anisotropy and heteronuclear dipolar couplings obtained in solid-state NMR spectrum using propagation method.

```
clear all
tic
clf
global freq freq_fit freq_frst T_dip chi0
global N n kb
global Q_0
global surf_RAM RAM_crit N_grid
global S0 T_cs T_dipNH Y_tetra Y_a1_g2 Q_00
global B_av B_curr cos_theta_flip
global wt wt_dip
global files

%CSA tensor values for ALL, Gly, and Pro - see below for specific
%tensor orientations

s11N=64;
s22N=77;
s33N=222;

s11NG=41;
s22NG=64;
s33NG=215;

s11NP=38;
s22NP=127;
s33NP=231;

%kb_max -- the total number of structures to calculate
kb_max=1000;

%number of dry runs to optimize tolerance in Hz for each data point
dry_runs=5;

[A1, A2]=xlsread('MerFt_27_68.xlsx');

A1(25,2:3)=[86.3, (2949-100)/1000]; %S37
A1(44,2:3)=[79.8, 2556/1000]; %G56
A1(53,2:3)=[86.3, 3893/1000]; %K65
```

```

A1=abs(A1(15:58,:));
A2=A2(15:58,:);

%ISOTROPIC CHEMICAL SHIFT DATA FROM HSQC - OPTIONAL
sigma_iso=A1(:,6);

[N_max,M_max]=size(A1);

A1(:,3)=1e3*A1(:,3);
%randomization of the experimental data - optional
A1(2:N_max,2)=A1(2:N_max,2)+(2*rand(N_max-1,1)-1)*0;
%residues to be fit except the first one
res_num=A1(2:N_max,1);

freq_frst=A1(1,2:3);
freq=A1(2:N_max,2:3);

RAM_crit=A1(2:N_max,4);
tolerance_all=1*A1(2:N_max,5);

%HERE ONE CAN OVERRIDE USER'S INPUT FOR TOLERANCE - set all to 2000 Hz
tolerance_all=ones(N_max-1, 1)*3.0e3;

%primary sequence
S=char(A2);
S0=S;
S_full=['V';S0];

[N, M]=size(S0);

for n=1:N

    if S(n)=='G'
        sigma_iso(n)=(s11NG+s22NG+s33NG)/3;
    else
        if S(n)=='P'
            sigma_iso(n)=(s11NP+s22NP+s33NP)/3;
        else
            sigma_iso(n)=(s11N+s22N+s33N)/3;
        end
    end
end

end

%Enter the order parameter here for bicelle data; For plates enter -2;
%For flipped bicelles enter -2*So (phage, parallel orientation, z axis)
%So=-2;
%For bicelle, So=0.8
So=0.77;

%experimental frequency to scale 15N CSA in Hz
W0=550;

```

```

%filter final structures at the end of calculation? - 1 - yes, 0 - no
filter_structures=1;

%Defining endpoints for structure filtering
N_Nt=find(res_num==32)+1;
N_Ct=find(res_num==56)+1;

N_Nt_end=find(res_num==28);
N_Ct_end=find(res_num==61);

R_i_low=5;
R_i_high=30;
R_e_low=5;
R_e_high=30;

delta_helices=30.0;

freq_frst(1,1)=abs(2/So*((1+So/2)*sigma_iso(1)-freq_frst(1,1)));
freq_frst(1,2)=abs(2/So*(freq_frst(1,2)));

%for other planes:
for n=1:N-1

    freq(n,1)=abs(2/So*((1+So/2)*sigma_iso(n+1)-freq(n,1)));
    freq(n,2)=abs(2/So*(freq(n,2)));

end

%Loading and defining the grid for the RAMA restraints
N_grid=180;
%ALL aminoacids (TYPE RAM_crit=1)
data_current=load('RAMA/rama500general.data');
data_currentN=data_current(:,3)/sum(data_current(:,3))*100;
surf_RAM(:, :, 1)=reshape(data_currentN,N_grid,N_grid);

%PRO (residues after the proline; TYPE RAM_crit=2)
data_current=load('RAMA/rama500pro.data');
data_currentN=data_current(:,3)/sum(data_current(:,3))*100;
surf_RAM(:, :, 2)=reshape(data_currentN,N_grid,N_grid);

%prePRO (actually used to fit PRO residues in this nomenclature; TYPE
RAM_crit=3)
data_current=load('RAMA/rama500prepro.data');
data_currentN=data_current(:,3)/sum(data_current(:,3))*100;
surf_RAM(:, :, 3)=reshape(data_currentN,N_grid,N_grid);

%GLY (residues after GLY; TYPE RAM_crit=4)
data_current=load('RAMA/rama500gly_sym.data');
data_currentN=data_current(:,3)/sum(data_current(:,3))*100;
surf_RAM(:, :, 4)=reshape(data_currentN,N_grid,N_grid);

%HELIX - generated here; TYPE RAM_crit=5
PHI=linspace(-179, 179, N_grid);

```

```

PSI=linspace(-179, 179, N_grid);
surf_RAM(:,:,5)=zeros(N_grid);
PHI0=-65; PSI0=-40;
dPHI2=20; dPSI2=20;
PHI_max=PHI0+dPHI2;
PHI_min=PHI0-dPHI2;
PSI_max=PSI0+dPSI2;
PSI_min=PSI0-dPSI2;

n1=round((PHI_min/179+1)*(N_grid-1)/2+1);
n2=round((PHI_max/179+1)*(N_grid-1)/2+1);
m1=round((PSI_min/179+1)*(N_grid-1)/2+1);
m2=round((PSI_max/179+1)*(N_grid-1)/2+1);

surf_RAM(m1:m2,n1:n2,5)=ones(m2-m1+1, n2-n1+1);

%3-10 HELIX (or can be adapted to a disordered alpha-helix) - generated
here; TYPE RAM_crit=6

PHI=linspace(-179, 179, N_grid);
PSI=linspace(-179, 179, N_grid);
surf_RAM(:,:,6)=zeros(N_grid);
PHI0=-49; PSI0=-26;
dPHI3=40; dPSI3=40;
PHI_max=PHI0+dPHI3;
PHI_min=PHI0-dPHI3;
PSI_max=PSI0+dPSI3;
PSI_min=PSI0-dPSI3;

n1=round((PHI_min/179+1)*(N_grid-1)/2+1);
n2=round((PHI_max/179+1)*(N_grid-1)/2+1);
m1=round((PSI_min/179+1)*(N_grid-1)/2+1);
m2=round((PSI_max/179+1)*(N_grid-1)/2+1);

surf_RAM(m1:m2,n1:n2,6)=ones(m2-m1+1, n2-n1+1);

%all allowed, RAM_crit=7
PHI=linspace(-179, 179, N_grid);
PSI=linspace(-179, 179, N_grid);
surf_RAM(:,:,7)=0.5*ones(N_grid);

beating=30;

%defining the constants
r_NH=1.01;
r_CH=1.09;
r_CaCo=1.52;
r_CO=1.23;

ga=2.718*10^7; gb=2.675*10^8; gc=6.728e7; hbar=1.055*10^(-34);
chi0=1e-7*ga*gb*hbar/(r_NH*1e-10)^3/2/pi;
chi1=1e-7*gb*gc*hbar/(r_CH*1e-10)^3/2/pi;
chi2=1e-7*gc*gc*hbar/(r_CaCo*1e-10)^3/2/pi;

```

```

gamma1=18.5*pi/180;
gamma2=(90+25)*pi/180;

wt=W0*ga/gb;

HNCA=118.2;
NCCA=115.6;
HNCO=119.5;
CaCoO=121.1;
NCoO=123.2;
tetra=110.5;
tetra_rad=tetra*pi/180;
angle1=270-HNCA;
angle2=270-NCCA-HNCO;
T_dip=[1/4, 0, -3/4; 0, -1/2, 0; -3/4, 0, 1/4];

Iz=diag([1,0,-1]);
Ix=[0, 1, 0; 1, 0, 1; 0, 1, 0]/sqrt(2);
Iy=[0, -i, 0; i, 0, -i; 0, i, 0]/sqrt(2);

Y_a1_g2=expm(-i*pi/180*(angle1+angle2)*Iy);
Y_tetra=expm(-i*tetra*pi/180*Iy);

T_cs_all0=wt*wigner1([gamma1, pi/2, gamma2])*[(s11N+s22N)/2, 0, (s22N-
s11N)/2; 0, s33N, 0; (s22N-s11N)/2, 0, (s11N+s22N)/2]*wigner1([gamma1,
pi/2, gamma2]');
T_cs_G0=wt*wigner1([gamma1, pi/2, gamma2])*[(s11NG+s22NG)/2, 0, (s22NG-
s11NG)/2; 0, s33NG, 0; (s22NG-s11NG)/2, 0,
(s11NG+s22NG)/2]*wigner1([gamma1, pi/2, gamma2]');
T_cs_P0=wt*wigner1([20*pi/180, pi/2, 1*pi/2])*[(s11NP+s22NP)/2, 0, (s22NP-
s11NP)/2; 0, s33NP, 0; (s22NP-s11NP)/2, 0,
(s11NP+s22NP)/2]*wigner1([20*pi/180, pi/2, 1*pi/2]');

T_cs_G=expm(-i*pi/180*angle2*Iy)*expm(i*pi/2*Ix)*T_cs_G0*expm(-
i*pi/2*Ix)*expm(i*pi/180*angle2*Iy);
T_cs_all=expm(-i*pi/180*angle2*Iy)*expm(i*pi/2*Ix)*T_cs_all0*expm(-
i*pi/2*Ix)*expm(i*pi/180*angle2*Iy);
T_cs_P=expm(-i*pi/180*angle2*Iy)*expm(i*pi/2*Ix)*T_cs_P0*expm(-
i*pi/2*Ix)*expm(i*pi/180*angle2*Iy);

T_dipNH=chi0*expm(-i*pi/180*angle2*Iy)*expm(i*pi/2*Ix)*T_dip*expm(-
i*pi/2*Ix)*expm(i*pi/180*angle2*Iy);
T_dipCH=chi1*wigner1(pi/180*[60, 90-tetra,0])*expm(i*pi/2*Ix)*T_dip*expm(-
i*pi/2*Ix)*wigner1(pi/180*[60, 90-tetra,0]');

OPTIONS=optimset('Display', 'off', 'TolX', 1e-4, 'TolFun', 1e-4,
'MaxFunEvals', 1000, 'MaxIter', 1000);

%reconstruction matrix to sum up the backbone
R_matrix=tril(ones(3*N));

for structures=1:(kb_max+dry_runs-1)

```

```

if structures<=dry_runs
    kb=1;
else
    kb=kb+1;
end

    rng default;
    rng('shuffle');

cnt=0;
n=1;

while n<=N-1

    if n==1 & structures==1 & kb==1
        z_prev=1e7;
        if S(1)=='G'
            T_cs=T_cs_G0;
        else
            T_cs=T_cs_all0;
        end

        for trials=1:100

            [plane_orientation, z_curr]=fminsearch('plane_orientation_find2',
[2*pi*rand, pi*rand]);
            if z_curr<z_prev
                z_prev=z_curr;
                plane_orientation_f=plane_orientation;
            end
        end
        'Trying to initiate propagation...'
        z_prev
        theta00=plane_orientation(2);
        phi00=plane_orientation(1);

        theta0=theta00;
        phi0=phi00;

        Y_00=[-sin(theta0)*exp(i*phi0)/sqrt(2), cos(theta0), sin(theta0)*exp(-
i*phi0)/sqrt(2)];
        Q_00=Y_00*expm(-i*pi/180*angle1*Iz)*expm(-i*pi/2*Ix);
        Q_0=Q_00;
    end

    wt_dip=1;

    freq_fit(1)=freq(n,1)+1*(2*rand-1)*0.2;
    freq_fit(2)=freq(n,2)+1*(2*rand-1)*10/1000*1000;

    if S(n+1)=='G'
        T_cs=T_cs_G;
    else

```

```

        if S(n+1)=='P'
            T_cs=T_cs_P;
        else
            T_cs=T_cs_all;
        end
    end
end

z0=1e6;
tolerance=tolerance_all(n);

while z0>tolerance & cnt<beating

    if RAM_crit(n)==1 | RAM_crit(n)==3 | RAM_crit(n)==4 | RAM_crit(n)==7
        start(1)=(2*rand-1)*180;
        start(2)=(2*rand-1)*180;
    end

    if RAM_crit(n)==5 %for alpha-helical regions
        start(1)=-65+(2*rand-1)*dPHI2;
        start(2)=-40+(2*rand-1)*dPSI2;
    end

    if RAM_crit(n)==6 %for 3-10-helical regions
        start(1)=-49+(2*rand-1)*dPHI3;
        start(2)=-26+(2*rand-1)*dPSI3;
    end

    if S(n+1)=='P', wt_dip=0; end
    if RAM_crit(n)==2
        start(1)=-52+(2*rand-1)*30;
        start(2)=(2*rand-1)*180;
    end

    [results, z0]=fminsearch('sigma_redQ2d_RAM2', pi/180*start, OPTIONS);

    cnt=cnt+1;
end

if z0<=tolerance

    if structures<=dry_runs
        tolerance_all_trial(n)=z0;
    end

    rmsdf(n,kb)=z0;
    params(n,:)=results;
    RAM(n, :)=180/pi*params(n, 1:2);

    Q_0(1)=exp(i*params(n,1))*Q_0(1);   Q_0(3)=-conj(Q_0(1));

    Q_0=Q_0*Y_tetra;

    Q_0(1)=exp(-i*(pi+params(n,2)))*Q_0(1);   Q_0(3)=-conj(Q_0(1));

```

```

Q_0=Q_0*Y_a1_g2;
Q_all(n, :)=Q_0;

cnt=0;

n=n+1;

end
if cnt>=beating

    n=n-2;
    if n>1
        Q_0=Q_all(n-1, :);
        cnt=0;
    else
        n=1;
        cnt=0;
        clear RAM;
        clear results;
        clear params;
        Q_0=Q_00;
    end
end

end

if structures<=dry_runs
tolerance_all=tolerance_all_trial+10;
['dry run ', num2str(structures)]
end
%back calculation of the spectrum to check
theta0=theta00;
phi0=phi00;

Y_0=[-sin(theta0)*exp(i*phi0)/sqrt(2), cos(theta0), sin(theta0)*exp(-
i*phi0)/sqrt(2)];

for n=1:N-1

OMEGA1=pi/180*[angle1, RAM(n, 1), tetra];
OMEGA2=pi/180*[0, -RAM(n, 2)-180, angle2];

Y_CH=Y_0*wigner1(pi/180*[angle1, RAM(n, 1)-60, 90-tetra])*wigner1([0, -
pi/2, 0]);
Y_cs_C=Y_0*wigner1(pi/180*[angle1, RAM(n, 1), 0]);

Y_0=Y_0*wigner1(OMEGA1)*wigner1(OMEGA2);

    if S(n+1)=='G'
        T_cs=T_cs_G0;
    else
        if S(n+1)=='P'

```

```

        T_cs=T_cs_P0;
    else
        T_cs=T_cs_all0;
    end
end

thetal=acos(Y_0(2));

phil=log(-Y_0(1)*sqrt(2)/sin(thetal))/i;

calc1=abs(real(Y_0*T_cs*Y_0')/wt);
calc2=abs(real(chi0*(3*sin(thetal)^2*cos(phil)^2-1)/2));
calc3=abs(real(chi1*(3*Y_CH(2)^2-1)/2));

freq_calc(n, :)= [calc1, calc2, calc3];

end

%CALCULATED BACKBONE

wigner_i=wigner1_half([0, theta00, (pi-phi00)]);
th1=pi/180*90;
th2=pi/180*90;
th3=pi/180*90;

f1=pi/180*(-(360-HNCo-NCCa));
f2=pi/180*(-(180-HNCo));
f3=pi/180*(-HNCA);

r1=1.52;
r2=1.33;
r3=1.46;
k=1;

wigner_tot=wigner_i*wigner1_half([f1, th1, 0]);

beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_tot(2,1)));
alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
u_tot(k+0, :)=r1*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t), cos(beta_t)];

wigner_tot=wigner_i*wigner1_half([f2, th2, 0]);

beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_tot(2,1)));
alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
u_tot(k+1, :)=r2*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t), cos(beta_t)];

wigner_tot=wigner_i*wigner1_half([f3, th3, 0]);
beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_tot(2,1)));

```

```

alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
u_tot(k+2, :)=r3*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t),
cos(beta_t)];

k=k+3;

for n=1:N-1

    OMEGA1=pi/180*[angle1, RAM(n,1), tetra];
    OMEGA2=pi/180*[0, -RAM(n,2)-180, angle2];

    wigner_i=wigner_i*wigner1_half(OMEGA1)*wigner1_half(OMEGA2);
    wigner_tot=wigner_i*wigner1_half([f1,th1,0]);

beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_to
t(2,1)));
alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
gamma_t=-real(log(wigner_tot(1,1)*wigner_tot(2,1)*2/sin(beta_t))/i);

u_tot(k+0, :)=r1*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t),
cos(beta_t)];

wigner_tot=wigner_i*wigner1_half([f2,th2,0]);

beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_to
t(2,1)));
alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
u_tot(k+1, :)=r2*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t),
cos(beta_t)];

wigner_tot=wigner_i*wigner1_half([f3,th3,0]);

beta_t=real(acos(wigner_tot(1,1)*wigner_tot(2,2)+wigner_tot(1,2)*wigner_to
t(2,1)));
alpha_t=real(log(wigner_tot(2,1)*wigner_tot(2,2)*2/sin(beta_t))/i);
u_tot(k+2, :)=r3*[cos(alpha_t)*sin(beta_t), sin(alpha_t)*sin(beta_t),
cos(beta_t)];

k=k+3;

end

B_calc(1,:)= [0,0,0];
B_calc(2:3*N+1, :)=R_matrix*u_tot;
RAM_all(:, :, kb)=RAM;

B_all(:, :, kb)=B_calc;
'one more structure done...'
kb
subplot(1,2,1)

freq_calc_bic(:,1)=(1+So/2)*sigma_iso(2:N)-So/2*freq_calc(:,1);

```

```

freq_calc_bic(:,2)=abs(So/2*freq_calc(:,2));

if structures>=dry_runs
plot(freq_calc_bic(:, 1), freq_calc_bic(:, 2), '. r')
hold on
['Freq RMSD for structure ', num2str(kb)]
sum(rmsdf(:,kb))
end

end

freq_bic(:,1)=(1+So/2)*sigma_iso(2:N)-So/2*freq(:,1);
freq_bic(:,2)=abs(So/2*freq(:,2));
plot(freq_bic(:, 1), freq_bic(:, 2), 'o')
hold on
plot(A1(1,2),A1(1,3), 'o')
hold on
text(A1(1,2), A1(1, 3), [' ', S(1), num2str(res_num(1)-1)], 'FontSize', 8,
'HorizontalAlignment','left', 'Clipping', 'on',
'FontName','Times','Color', 'r');
hold on
for n=1:N-1

    if freq_bic(n, 1)<85
        text(freq_bic(n, 1), freq_bic(n, 2)+0*(2*rand-1), [' ', S(n+1),
num2str(res_num(n))], 'FontSize', 8, 'HorizontalAlignment','left',
'Clipping', 'on', 'FontName','Times');
        hold on
    end
    if freq_bic(n, 1)>=85 & freq_bic(n, 1)<=130
        text(freq_bic(n, 1), freq_bic(n, 2)+0*(2*rand-1), [S(n+1),
num2str(res_num(n)), ' '], 'FontSize', 8, 'HorizontalAlignment','right',
'Clipping', 'on', 'FontName','Times');
        hold on
    end

    if freq_bic(n, 1)<200 & freq_bic(n, 1)>130
        text(freq_bic(n, 1), freq_bic(n, 2)+0*(2*rand-1), [' ', S(n+1),
num2str(res_num(n))], 'FontSize', 8, 'HorizontalAlignment','left',
'Clipping', 'on', 'FontName','Times');
        hold on
    end
    if freq_bic(n, 1)>=200
        text(freq_bic(n, 1), freq_bic(n, 2)+0*(2*rand-1), [S(n+1),
num2str(res_num(n)), ' '], 'FontSize', 8, 'HorizontalAlignment','right',
'Clipping', 'on', 'FontName','Times');
        hold on
    end

end
axis([65 146 0e3 5.1e3])
xlabel('15N / ppm');
ylabel('1H-15N Coupling / Hz');

```

```

set(gca, 'Xdir', 'reverse');
toc

k_accpt=0;
k_accpt_add=0;

for kb=1:kb_max

    if filter_structures==1
        r_HH=B_all(3*N_Nt, :, kb)-B_all(3*N_Ct, :, kb);
        inter_helix=sqrt(r_HH*r_HH');
        r_end2end=B_all(3*N_Nt_end, :, kb)-B_all(3*N_Ct_end, :, kb);
        helix_end2end=sqrt(r_end2end*r_end2end');
        r_head2tail=abs(B_all(3*N_Nt_end, 3, kb)-B_all(3*N_Ct_end, 3, kb));
        inter_helix_matrix(kb, :)=inter_helix;
        helix_end2end_matrix(kb, :)=helix_end2end;
        r_head2tail_matrix(kb, :)=r_head2tail;

        if inter_helix>R_i_low & inter_helix<R_i_high & helix_end2end<R_e_high
        & helix_end2end>R_e_low & r_head2tail<delta_helices
            inter_helix;
            helix_end2end;
            r_head2tail;

            k_accpt=k_accpt+1;
            kb_accpt_matrix(k_accpt, :)=kb;
            inter_helix_accpt_matrix(k_accpt, :)=inter_helix;
            helix_end2end_accpt_matrix(k_accpt, :)=helix_end2end;
            r_head2tail_accpt_matrix(k_accpt, :)=r_head2tail;

            B_total(:, :, k_accpt)=B_all(:, :, kb);
            RAM_accpt(:, :, k_accpt)=RAM_all(:, :, kb);
            rmsdf_accpt(:, k_accpt)=rmsdf(:, kb);

        end

    end

end

if k_accpt==0
    'No solutions found'

else

[M_accpt, K_accpt, N_accpt]=size(B_total);
B_av=B_total(:, :, 1);
N_accpt

subplot(1,2,2)

for kb=1:N_accpt

```

```

E_RAM(kb)=0;
for n=1:N-1

n0=round((RAM_accpt(n,1,kb)/179+1)*(N_grid-1)/2+1);
m0=round((RAM_accpt(n,2,kb)/179+1)*(N_grid-1)/2+1);

    E_RAM(kb)=E_RAM(kb)+surf_RAM(m0,n0,RAM_crit(n));

end

B_curr=B_total(:,:,kb);
z_min=1e6;

for iter=1:100
    theta_flip=round(rand)*pi;
    cos_theta_flip=cos(theta_flip);
    [trans, z0]=fminsearch('backbone_fit_flip', 2*rand(1,4)-1, OPTIONS);
    if z0<z_min
phi=trans(1);
theta=theta_flip;
psi=0;
X0=trans(2);
Y0=trans(3);
Z0=trans(4);
z_min=z0;

phi_matrix(kb,:)=phi;
X0_matrix(kb,:)=X0;
Y0_matrix(kb,:)=Y0;
Z0_matrix(kb,:)=Z0;
theta_matrix(kb,:)=theta;
Z_MIN(kb,:)=z_min;

    end
end

z_min

ROT=[-sin(psi)*sin(phi)+cos(theta)*cos(psi)*cos(phi),
sin(psi)*cos(phi)+cos(theta)*cos(psi)*sin(phi), -sin(theta)*cos(psi);
-cos(psi)*sin(phi)-cos(theta)*cos(phi)*sin(psi), cos(psi)*cos(phi)-
cos(theta)*sin(phi)*sin(psi), sin(theta)*sin(psi);
sin(theta)*cos(phi), sin(theta)*sin(phi), cos(theta)];

B_rot(:,:,kb)=(ROT*B_total(:,:,kb).'+[X0*ones(1, 3*N+1); Y0*ones(1,3*N+1);
Z0*ones(1, 3*N+1)]).';

    xlsfilename=sprintf('2H30_B_rot%d',kb);
    xlswrite(xlsfilename,B_rot(:,:,kb));
plot3(B_rot(:,1,kb), B_rot(:,2,kb), B_rot(:,3,kb), '- r', 'MarkerSize',10)
hold on

end

```

```

axis([-30 30 -30 30 -60 50])
daspect([1, 1, 1])
hold on

%calculation of true RMSD
for kb=1:N_accpt
    rmsd(kb)=sqrt(sum((B_av(:,1)-B_rot(:,1,kb)).^2+(B_av(:,2)-
B_rot(:,2,kb)).^2+(B_av(:,3)-B_rot(:,3,kb)).^2)/(3*N+1));
end

sqrt(sum(rmsd.^2/N_accpt))
end

files=dir('2H30_B_rot*.xls');

for i=1:length(files)
    filename=files(i).name;
    B=xlsread(filename);

    outfile=sprintf('output/2H30_%d.pdb',i);
    filnam=fopen(outfile,'w');

    plane=['N '; 'CA'; 'C '];
    plane2=['NCC'];

    atom_count=1;
    frst=res_num(1)-2;

    k=1;
    for n=1:N+1

        if S_full(n)=='A', res='ALA'; end
        if S_full(n)=='G', res='GLY'; end
        if S_full(n)=='V', res='VAL'; end
        if S_full(n)=='L', res='LEU'; end
        if S_full(n)=='I', res='ILE'; end
        if S_full(n)=='S', res='SER'; end
        if S_full(n)=='T', res='THR'; end
        if S_full(n)=='C', res='CYS'; end
        if S_full(n)=='M', res='MET'; end
        if S_full(n)=='P', res='PRO'; end
        if S_full(n)=='D', res='ASP'; end
        if S_full(n)=='N', res='ASN'; end
        if S_full(n)=='E', res='GLU'; end
        if S_full(n)=='Q', res='GLN'; end
        if S_full(n)=='K', res='LYS'; end
        if S_full(n)=='R', res='ARG'; end
        if S_full(n)=='H', res='HIS'; end
        if S_full(n)=='F', res='PHE'; end
        if S_full(n)=='Y', res='TYR'; end
        if S_full(n)=='W', res='TRP'; end
    end
end

```

```

if n==1, p_init=2; else, p_init=1; end
for p=p_init:3
    if k<=3*N+1
        fprintf(filnam,'ATOM');
        fprintf(filnam,'%7.0f',atom_count);
        fprintf(filnam,' ');
        fprintf(filnam,plane(p, :));
        fprintf(filnam,' ');
        fprintf(filnam,res);
        fprintf(filnam,'%6.0f', n+frst-1);
        fprintf(filnam,'%12.3f', B(k, 1));
        fprintf(filnam,'%8.3f', B(k, 2));
        fprintf(filnam,'%8.3f', B(k, 3));
        fprintf(filnam,'%6.2f', 1.00);
        fprintf(filnam,'%6.2f', 0.00);
        fprintf(filnam,' ');
        fprintf(filnam,plane2(p));
        fprintf(filnam,'%c\n', ' ');

        atom_count=atom_count+1;
        if plane(p, :)=='N ' | plane(p, :)=='C '
            if plane(p, :)=='N ', add_atom='HN'; r_bond=r_NH;
ang1=HNCo; ang2=HNCA; end
            if plane(p, :)=='C ', add_atom='O '; r_bond=r_CO;
ang1=CaCoO; ang2=NCoO; end
            r1=B(k, :)-B(k-1, :);
            r2=B(k+1, :)-B(k, :);

            r_n=cross(r1,r2);
            r_n=r_n/sqrt(r_n(1)^2+r_n(2)^2+r_n(3)^2);

            r0(1,1)=sqrt(r1(1)^2+r1(2)^2+r1(3)^2)*cos(pi/180*(180-ang1));
            r0(2,1)=sqrt(r2(1)^2+r2(2)^2+r2(3)^2)*cos(pi/180*ang2);
            r0(3,1)=0;

A_plane=[r1(1), r1(2), r1(3); r2(1), r2(2), r2(3); r_n(1), r_n(2),r_n(3)];
r=A_plane\r0;
r=r_bond*r/sqrt(r(1)^2+r(2)^2+r(3)^2);

vector_bond(1, :)=B(k, :);
vector_bond(2, :)=vector_bond(1, :)+r.';
        if res=='PRO'
            if add_atom=='O '
                fprintf(filnam,'ATOM');
                fprintf(filnam,'%7.0f',atom_count); fprintf(filnam,' ');
                fprintf(filnam,add_atom); fprintf(filnam,' '); fprintf(filnam,res);
                fprintf(filnam,'%6.0f', n+frst-1); fprintf(filnam,'%12.3f', vector_bond(2,
1)); fprintf(filnam,'%8.3f', vector_bond(2, 2));
                fprintf(filnam,'%8.3f',vector_bond(2, 3)); fprintf(filnam,'%6.2f', 1.00);
                fprintf(filnam,'%6.2f', 0.00); fprintf(filnam,' ');
                fprintf(filnam,add_atom(1)); fprintf(filnam,'%c\n', ' ');
                atom_count=atom_count+1;
            end
        end
    end
end

```

```

        else
            fprintf(filnam,'ATOM');
fprintf(filnam,'%7.0f',atom_count); fprintf(filnam,' ');
fprintf(filnam,add_atom); fprintf(filnam,' '); fprintf(filnam,res);
fprintf(filnam,'%6.0f', n+frst-1); fprintf(filnam,'%12.3f', vector_bond(2,
1)); fprintf(filnam,'%8.3f', vector_bond(2, 2)); fprintf(filnam,'%8.3f',
vector_bond(2, 3)); fprintf(filnam,'%6.2f', 1.00); fprintf(filnam,'%6.2f',
0.00); fprintf(filnam,' '); fprintf(filnam,add_atom(1));
fprintf(filnam,'%c\n', ' ');
            atom_count=atom_count+1;
        end
    end
    end
    k=k+1;
end
end
    fprintf(filnam,'TER '), fprintf(filnam,'%7.0f',atom_count),
fprintf(filnam,' '), fprintf(filnam,' '), fprintf(filnam,res),
fprintf(filnam,'%6.0f\n', n+frst-1);
    atom_count=atom_count+1;
    fclose(filnam);

end

```

## B.2 Determination of the Orientation of the First Peptide Plane

The orientation of the first peptide plane in the target membrane protein is calculated. This

function is called by the main program shown in Appendix B.1.

```
function z=backbone_fit_flip(x)

global B_av B_curr cos_theta_flip
global kb
global N

phi=x(1);
X0=x(2);
Y0=x(3);
Z0=x(4);
ROT=[cos_theta_flip*cos(phi), cos_theta_flip*sin(phi), 0;
     -sin(phi), cos(phi), 0;
     0, 0, cos_theta_flip];

B_rot=(ROT*B_curr.'+[X0*ones(1, 3*N+1); Y0*ones(1, 3*N+1); Z0*ones(1,
3*N+1)]).';
z=sqrt(sum((B_av(:,1)-B_rot(:,1)).^2+(B_av(:,2)-B_rot(:,2)).^2+(B_av(:,3)-
B_rot(:,3)).^2)/(3*N+1)));
```

### B.3 Calculation of Rank-1 Wigner Rotation Matrix

Each entry in Rank-1 Wigner Rotation Matrix is calculated. This function is called by the main program shown in Appendix B.1.

```
function Z=wigner1(OMEGA);  
  
cos_b=cos(OMEGA(2));  
sin_b_sq_2=sin(OMEGA(2))/sqrt(2);  
ex_ia=exp(i*OMEGA(1));  
ex_ig=exp(i*OMEGA(3));  
D33=ex_ia*(1+cos_b)/2*ex_ig;  
D31=ex_ia*(1-cos_b)/2*ex_ig';  
D23=sin_b_sq_2*ex_ig;  
D32=ex_ia*sin_b_sq_2;  
  
Z=[D33', -D32', D31';  
    D23', cos_b, -D23;  
    D31, D32, D33];
```

#### B.4 Calculation of Rank-1/2 Wigner Rotation Matrix

Each entry in Rank-1/2 Wigner Rotation Matrix is calculated. This function is called by the main program shown in Appendix B.1.

```
function Z=wigner1_half(OMEGA);  
  
a=exp(-i*OMEGA(1)/2)*cos(OMEGA(2)/2)*exp(-i*OMEGA(3)/2);  
b=exp(i*OMEGA(1)/2)*sin(OMEGA(2)/2)*exp(-i*OMEGA(3)/2);  
  
Z=[a, -b'; b, a'];
```

## B.5 Comparison of Experimental Data and Simulated Frequencies

The RMSD between experimental data and simulated frequencies is calculated and returned to the main program shown in Appendix B.1.

```
function z=sigma_redQ6d_RAM(x)

global freq
global n
global Q_0
global surf_RAM RAM_crit N_grid
global Y_tetra
global PHI_min PHI_max PSI_min PSI_max
global wt wt_C wt_dip T_cs T_dipCH T_dipNH T_cs_C chi0 T_dipNCa T_dipNC0
T_cs_Co

Q_i(1)=exp(i*x(1))*Q_0(1); Q_i(2)=Q_0(2); Q_i(3)=-conj(Q_i(1));

calc3=real(T_dipCH(1,1)+Q_i(2)*((T_dipCH(2,2)-T_dipCH(1,1))*Q_i(2)+
4*(T_dipCH(1,2)*Q_i(1)))-2*(Q_i(1)^2*T_dipCH(1,3)));
Q_i=Q_i*Y_tetra;
Q_i(1)=exp(-i*(pi+x(2)))*Q_i(1); Q_i(3)=-conj(Q_i(1));

calc1=real(T_cs(1,1)+Q_i(2)*((T_cs(2,2)-
T_cs(1,1))*Q_i(2)+4*(T_cs(1,2)*Q_i(1)))-2*(Q_i(1)^2*T_cs(1,3)));
calc2=real(T_dipNH(1,1)+Q_i(2)*((T_dipNH(2,2)-
T_dipNH(1,1))*Q_i(2)+4*(T_dipNH(1,2)*Q_i(1)))-2*(Q_i(1)^2*T_dipNH(1,3)));
calc4=real(T_cs_Co(1,1)+Q_i(2)*((T_cs_Co(2,2)-
T_cs_Co(1,1))*Q_i(2)+4*(T_cs_Co(1,2)*Q_i(1)))-2*(Q_i(1)^2*T_cs_Co(1,3)));
calc5=real(T_dipNCa(1,1)+Q_i(2)*((T_dipNCa(2,2)-
T_dipNCa(1,1))*Q_i(2)+4*(T_dipNCa(1,2)*Q_i(1)))-
2*(Q_i(1)^2*T_dipNCa(1,3)));
calc6=real(T_dipNC0(1,1)+Q_i(2)*((T_dipNC0(2,2)-
T_dipNC0(1,1))*Q_i(2)+4*(T_dipNC0(1,2)*Q_i(1)))-
2*(Q_i(1)^2*T_dipNC0(1,3)));

dsigma=abs([calc2])-[freq(n,1)];
z=sqrt(dsigma*dsigma');
n0=round((x(1)*180/pi/179+1)*(N_grid-1)/2+1);
m0=round((x(2)*180/pi/179+1)*(N_grid-1)/2+1);

if n0>=1 & m0>=1 & n0<=N_grid & m0<=N_grid
    if surf_RAM(m0,n0,RAM_crit(n))<5e-3
        z=1e5;
    end
else
    z=1e5;
end
```

## B.6 PISEMA Back-Calculation

The coordinates of a calculated protein structure is used as input to back-calculate the corresponding PISEMA spectrum.

```
clear all
clf
global freq_fit N_fit Y_NH freq_2H30
global n N
global angle1 angle2 tetra tetra_rad
global chi0 chi1 chi2

[A1,A2]=xlsread('2H30_2013_4_12_cold_assign_77_new_1176+21.xlsx');
col_i=5; col_f=7;
[N_max,M_max]=size(A1);

fitwave=0;
res_first=2; res_last=45;

So=0.77;

data_avail=0;
if data_avail==1
load Pfl_cold.txt;
res_exp(:,1)=Pfl_cold(:,1);
freq_exp(:,1)=Pfl_cold(:,2);
freq_exp(:,2)=1e3*Pfl_cold(:,3);
end

kb=1;

n_res=1;
for n=1:N_max
    symb=char(A2(n,3));
    res=char(A2(n,4));
    [M_symb, N_symb]=size(symb);

    if N_symb==2
        if symb=='CA'
            B(kb, :)=A1(n, col_i:col_f); kb=kb+1;
            res_num(n_res)=A1(n, col_i-1);

            if res=='ALA', S(n_res)='A'; end
            if res=='GLY', S(n_res)='G'; end
            if res=='VAL', S(n_res)='V'; end
            if res=='LEU', S(n_res)='L'; end
            if res=='ILE', S(n_res)='I'; end
            if res=='SER', S(n_res)='S'; end
        end
    end
end
```

```

    if res=='THR', S(n_res)='T'; end
    if res=='CYS', S(n_res)='C'; end
    if res=='MET', S(n_res)='M'; end
    if res=='PRO', S(n_res)='P'; end
    if res=='ASP', S(n_res)='D'; end
    if res=='ASN', S(n_res)='N'; end
    if res=='GLU', S(n_res)='E'; end
    if res=='GLN', S(n_res)='Q'; end
    if res=='LYS', S(n_res)='K'; end
    if res=='ARG', S(n_res)='R'; end
    if res=='HIS', S(n_res)='H'; end
    if res=='PHE', S(n_res)='F'; end
    if res=='TYR', S(n_res)='Y'; end
    if res=='TRP', S(n_res)='W'; end

    n_res=n_res+1;
end
end

    if N_symb==1 & kb>1
    if symb=='C', B(kb, :)=A1(n, col_i:col_f); kb=kb+1; 'C'; end
    if symb=='N', B(kb, :)=A1(n, col_i:col_f); kb=kb+1; 'N'; end
end
end

r_NH=1.01e-10;
r_CH=1.09e-10;
r_NCa=1.46e-10;
r_CoN=1.33e-10;
r_CaCo=1.52e-10;

ga=2.718*10^7; gb=2.675*10^8; gc=6.728e7; hbar=1.055*10^(-34);
chi0=1e-7*ga*gb*hbar/r_NH^3/2/pi;
chi1=1e-7*gb*gc*hbar/(r_CH)^3/2/pi;
chi2=1e-7*gc*gc*hbar/(r_CaCo)^3/2/pi;
chi3=1e-7*ga*gc*hbar/(r_NCa)^3/2/pi;
chi4=1e-7*ga*gc*hbar/(r_CoN)^3/2/pi;

s11N=64;
s22N=77;
s33N=222;

s11NG=41+0;
s22NG=64+0;
s33NG=215;

s11NP=38;
s22NP=127;
s33NP=231;

s11H=3;
s22H=8;
s33H=17;

```

```

gamma1=18.5*pi/180
gamma2=(90+25)*pi/180;

HNCA=118.2;
NCCA=115.6;
HNCO=119.5;

tetra=110.5;
tetra_ideal=acos(-1/3)*180/pi;
tetra_rad=tetra*pi/180;
angle1=270-HNCA;
angle2=270-NCCA-HNCO;

[N_atoms,M]=size(B);
n=1;

sum_tetra=0;
for kb=1:3:N_atoms-3

    r2=B(kb+2, :)-B(kb+1, :);
    r3=B(kb+3, :)-B(kb+2, :);

    r_n=cross(r3,r2);
    r_n=r_n/sqrt(r_n(1)^2+r_n(2)^2+r_n(3)^2);

    r0(1,1)=sqrt(r2(1)^2+r2(2)^2+r2(3)^2)*cos(pi/180*(180-HNCO));
    r0(2,1)=sqrt(r3(1)^2+r3(2)^2+r3(3)^2)*cos(pi/180*HNCA);
    r0(3,1)=0;

    A_plane=[r2(1),r2(2),r2(3); r3(1),r3(2),r3(3); r_n(1),r_n(2),r_n(3)];
    r=A_plane\r0;

    n_NH=r/sqrt(r(1)^2+r(2)^2+r(3)^2);
    r=r_NH*r/sqrt(r(1)^2+r(2)^2+r(3)^2);

    vector_NH(1, :)=B(kb+2, :);
    vector_NH(2, :)=vector_NH(1, :)+r.';

    freq_pdb(n,2)=abs(-So/2*chi0*(3*n_NH(3)^2-1)/2);

    if kb+4<=N_atoms
        r4=B(kb+4, :)-B(kb+3, :);
        r_c=cross(r3,r4);

        tetra_real=180-180/pi*acos((r3*r4')/sqrt((r3*r3')*(r4*r4')));
        tetrahedrals(n)=tetra_real;
        sum_tetra=sum_tetra+tetra_real;
        r_c=r_c/sqrt(r_c(1)^2+r_c(2)^2+r_c(3)^2);

        r0(1,1)=sqrt(r3(1)^2+r3(2)^2+r3(3)^2)*cos(pi-pi/180*tetra_ideal);
        r0(2,1)=sqrt(r4(1)^2+r4(2)^2+r4(3)^2)*(-cos(pi/180*(tetra_ideal-
tetra_ideal/2)))/sqrt(3);
        r0(3,1)=sqrt(1-cos(tetra_ideal*pi/180)^2/cos(tetra_real/2*pi/180)^2);

```

```

    A_plane=[r3(1), r3(2), r3(3); r4(1), r4(2), r4(3); r_c(1), r_c(2),
r_c(3)];
    r=A_plane\r0;

    n_CH=r/sqrt(r(1)^2+r(2)^2+r(3)^2);
    r=r_CH*r/sqrt(r(1)^2+r(2)^2+r(3)^2);

    vector_CH(1, :)=B(kb+3, :);
    vector_CH(2, :)=vector_CH(1, :)+r.';

    freq_pdb(n+1,3)=abs(-So/2*chi1*(3*n_CH(3)^2-1)/2);

    r0(1,1)=sqrt(r3(1)^2+r3(2)^2+r3(3)^2)*cos(pi-pi/180*tetra);
    r0(2,1)=sqrt(r4(1)^2+r4(2)^2+r4(3)^2)*cos(pi/180*tetra);
    r0(3,1)=-sqrt(1-cos(tetra*pi/180)^2/cos(tetra/2*pi/180)^2);

    A_plane=[r3(1),r3(2),r3(3); r4(1),r4(2),r4(3); r_c(1),r_c(2),r_c(3)];
    r=A_plane\r0;

    n_CC=r/sqrt(r(1)^2+r(2)^2+r(3)^2);
    r=r_CH*r/sqrt(r(1)^2+r(2)^2+r(3)^2);

    vector_CC(1, :)=B(kb+3, :);
    vector_CC(2, :)=vector_CC(1, :)+r.';

end

    r0(1,1)=sqrt(r2(1)^2+r2(2)^2+r2(3)^2)*cos(pi/180*(180-
HNCo+gamma1*180/pi));
    r0(2,1)=sqrt(r3(1)^2+r3(2)^2+r3(3)^2)*cos(pi/180*(HNCo+gamma1
*180/pi));
    r0(3,1)=0;

    A_plane=[r2(1), r2(2), r2(3); r3(1), r3(2), r3(3); r_n(1), r_n(2),
r_n(3)];
    r=A_plane\r0;

    sigma_zz=r/sqrt(r(1)^2+r(2)^2+r(3)^2);

    r0(1,1)=sqrt(r2(1)^2+r2(2)^2+r2(3)^2)*cos(pi/180*(90-
HNCo+gamma1*180/pi))*cos(pi-gamma2);
    r0(2,1)=cos(pi/180*(gamma2*180/pi-90));
    r0(3,1)=0;

    A_plane=[r2(1), r2(2), r2(3); r_n(1), r_n(2), r_n(3); sigma_zz(1),
sigma_zz(2), sigma_zz(3)];
    r=A_plane\r0;

    sigma_yy=r/sqrt(r(1)^2+r(2)^2+r(3)^2);

    if S(n+1)=='G'

```

```

        s11=s11NG;
        s22=s22NG;
        s33=s33NG;
    else if S(n+1)=='P'

        s11=s11NP;
        s22=s22NP;
        s33=s33NP;
    else

        s11=s11N;
        s22=s22N;
        s33=s33N;
    end
end

freq_pdb(n,1)=s11*(1-sigma_zz(3)^2-
sigma_yy(3)^2)+s22*sigma_yy(3)^2+s33*sigma_zz(3)^2;
freq_pdb(n,1)=(1+So/2)*(s11+s22+s33)/3-So/2*freq_pdb(n,1);
freq_pdb(n,4)=s11H*n_NH(3)^2+s22H*r_n(3)^2+s33H*(1-n_NH(3)^2-r_n(3)^2);
freq_pdb(n,4)=(1+So/2)*(s11H+s22H+s33H)/3-So/2*freq_pdb(n,4);
freq_pdb(n,5)=abs(So/2*chi3*(3*r3(3)^2/(r3*r3')-1)/2);
freq_pdb(n,6)=abs(So/2*chi4*(3*r2(3)^2/(r2*r2')-1)/2);

n=n+1;
end

freq_2KSJ=[
100.252 1.128645
80.091 2.624857
89.784 3.881893
93.396 2.978368
94.458 1.634957
79.377 3.094715
97.967 3.82115
99.191 2.990646
82.944 2.109572
83.811 3.788857
99.462 3.369772
94.818 2.161768
64.639 2.722601
90.982 4.097554
100.139 2.19791
85.265 2.364952
73.001 3.307684
93.976 3.719332
];

[M,N]=size(S);
R=tril(ones(3*(N-1)));

if data_avail==0
    freq_exp=freq_pdb;

```

```

end

subplot(2,2,1)

[N_exp, M_exp]=size(freq_exp);
n_first=res_num(1);
for n=25:42

    if S(n+1)~='P'
        plot(freq_exp(n, 1), freq_exp(n, 2), 'o b')
        hold on
        plot(freq_2KSJ(n-24, 1),freq_2KSJ(n-24, 2)*1e3,'o r')
        hold on

    if S(n+1)~='X'
    if freq_exp(n, 1)<200
        text(freq_exp(n, 1), freq_exp(n, 2), [' ', S(n+1),
num2str(res_num(n+1))], 'FontSize', 6, 'HorizontalAlignment','left',
'Clipping', 'on', 'FontName','Times');
        hold on
    end
    if freq_exp(n, 1)>=200
        text(freq_exp(n, 1), freq_exp(n, 2), [S(n+1), num2str(res_num(n+1))],
'], 'FontSize', 6, 'HorizontalAlignment','right', 'Clipping', 'on',
'FontName','Times');
        hold on
    end
    end
end
end
hold on
freq=freq_pdb;
delta_SP=0;
n_exp=1;
if data_avail==1
for n=1:N-1

    if S(n+1)~='P'
        if res_num(n+1)==res_exp(n_exp)
            delta_SP=delta_SP+sqrt(((freq_pdb(n,1)-
freq_exp(n,1))/5)^2+((freq_pdb(n,2)-freq_exp(n,2))/400)^2)/(N-1);

            plot(freq_pdb(n, 1), freq_pdb(n, 2), 'o c', 'MarkerSize', 3,
'MarkerFace', 'c')
            text(freq_pdb(n, 1), freq_pdb(n, 2), [' ', num2str(res_num(n+1))],
'FontSize', 6, 'HorizontalAlignment','left', 'Clipping', 'on',
'FontName','Times', 'Color', 'c');
            n_exp=n_exp+1;
        end
        hold on
    end
end
end
end

```

```

axis([50 150 0e3 5e3])
xlabel('15N / ppm');
ylabel('1H-15N Coupling / Hz');
set(gca, 'Xdir', 'reverse');
box on

subplot(2,2,3)
width=1;
ppm=linspace(50, 250, 1024);
n15_spectrum=zeros(1,1024);
for n=1:N-1

    if S(n+1)~='P'
        n15_spectrum=n15_spectrum+exp(-(freq(n,1)-ppm).^2/2/width^2);
    end
end
plot(ppm, n15_spectrum)
axis([50 250 0 max(n15_spectrum+2)])
set(gca, 'Xdir', 'reverse');
daspect([1 .05 1])

subplot(2,2,2)
plot3(B(:, 1), B(:, 2), B(:, 3), '- ', 'MarkerSize', 10)
hold on
axis([-60 60 -60 60 -60 60])
daspect([1, 1, 1])

memx=[0 0 0];
memy=[0 0 15.0];
memz=[0 0 -15.0];

plot3(memx(1),memx(2),memx(3), 'o g', 'MarkerSize', 6)
hold on
plot3(memy(1),memy(2),memy(3), 'o g', 'MarkerSize', 6)
hold on
plot3(memz(1),memz(2),memz(3), 'o g', 'MarkerSize', 6)
plot3([memx(1);memy(1);memz(1)], [memx(2);memy(2);memz(2)], [memx(3);memy(3);memz(3)], '- r', 'MarkerSize', 6, 'LineWidth', 2)

xlabel('memx')
ylabel('memy')
zlabel('memz')

x=-40:4:40;
y=-40:4:40;
[X,Y]=meshgrid(x,y);
Z1=( (memz(1)-memy(1))*X+(memz(2)-memy(2))*Y-(memz(1)-memy(1))*memy(1)-
(memz(2)-memy(2))*memy(2)-(memz(3)-memy(3))*memy(3))/(-(memz(3)-memy(3)));
    mesh(X,Y,Z1)

hold on

```

```
Z2= ( memz (1) -memy (1) ) *X+ ( memz (2) -memy (2) ) *Y- ( memz (1) -memy (1) ) *memz (1) -  
( memz (2) -memy (2) ) *memz (2) - ( memz (3) -memy (3) ) *memz (3) / ( - ( memz (3) -memy (3) ) ) ;  
mesh (X,Y,Z2)  
  
axis([-30 30 -30 30 -30 30])  
daspect([1, 1, 1])  
axis tight  
grid  
view([150,30])  
box on
```

## B.7 Membrane Horizontalization

The coordinates related to membrane center and membrane normal obtained in Rosetta output file are used as input. After the horizontalization, the membrane center will be moved to the origin of the coordinate system, and the membrane normal will be parallel to the orientation of the external magnetic field.

```
clear all

clf
global B3 B2 B1 B0

files=dir('input_structure.xlsx');

for i=1:length(files)
    filename=files(i).name;
    [A1,A2]=xlsread(filename);

    [A1,A2]=xlsread('2H30_simulated_bb_NCC.xlsx');
    col_i=5; col_f=7;
    [N_max,M_max]=size(A1);

    memx=[X(1); X(2); X(3)];
    memy=[Y(1); Y(2); Y(3)];
    memz=[Z(1); Z(2); Z(3)];
    memn=[n(1); n(2); n(3)];

    B0(:,1:3)=A1(1:N_max,col_i:col_f);
    B0=B0';

    subplot(1,2,1)
    cstring='rgbcmyk';
    plot3(B0(1,:), B0(2,:), B0(3,:), cstring(mod(i,7)+1), 'MarkerSize',
    10, 'LineWidth', 2);
    hold on
    B0=B0';
end

xlabel('memx')
ylabel('memy')
zlabel('memz')

hold on

plot3(memx(1),memx(2),memx(3),'o g', 'LineWidth', 2, 'MarkerSize', 6)
```

```

hold on
plot3(memy(1),memy(2),memy(3),'o g', 'LineWidth',2,'MarkerSize', 6)
hold on
plot3(memz(1),memz(2),memz(3),'o g', 'LineWidth',2,'MarkerSize', 6)
hold on
plot3(memn(1),memn(2),memn(3),'o g', 'LineWidth',2,'MarkerSize', 6)
hold on
plot3([memx(1);memy(1);memz(1)], [memx(2);memy(2);memz(2)], [memx(3);memy(3);memz(3)], '- r', 'MarkerSize', 6, 'LineWidth', 1)
hold on
plot3(0,0,0,'o g', 'LineWidth',2,'MarkerSize', 6)
text(memz(1),memz(2),memz(3)+0.5,'Z', 'FontSize', 6)
text(memy(1),memy(2),memy(3)+0.5,'Y', 'FontSize', 6)
text(memx(1),memx(2),memx(3)+0.5,'X', 'FontSize', 6)
text(memn(1),memn(2),memn(3)-0.5,'N', 'FontSize', 6)
text(0,0,0.5,'O', 'FontSize', 6)
x=-60:4:60;
y=-60:4:60;

[X,Y]=meshgrid(x,y);

Z1=( (memz(1)-memy(1))*X+(memz(2)-memy(2))*Y-(memz(1)-memy(1))*memy(1)-
(memz(2)-memy(2))*memy(2)-(memz(3)-memy(3))*memy(3))/(-(memz(3)-memy(3)));
mesh(X,Y,Z1)
hold on
Z2=( (memz(1)-memy(1))*X+(memz(2)-memy(2))*Y-(memz(1)-memy(1))*memz(1)-
(memz(2)-memy(2))*memz(2)-(memz(3)-memy(3))*memz(3))/(-(memz(3)-memy(3)));
mesh(X,Y,Z2)

axis([-60 60 -60 60 -60 60])
daspect([1, 1, 1])
grid
view([150,30])
box on

memxn=memx-memx;
memyn=memy-memx;
memzn=memz-memx;
memnn=memn-memx;
memon=[0;0;0]-memx;
r=sqrt(memyn(1,1)^2+memyn(2,1)^2+memyn(3,1)^2);
theta=acos(memyn(3,1)/r);
phi=atan(memyn(2,1)/memyn(1,1));

if memyn(1,1)>=0

    R_y_theta=[cos(pi-theta),0,sin(pi-theta);0,1,0;-sin(pi-
theta),0,cos(pi-theta)];
    R_z_phi=[cos(-phi),-sin(-phi),0;sin(-phi),cos(-phi),0;0,0,1];

else

```

```

    R_y_theta=[cos(-theta),0,sin(-theta);0,1,0;-sin(-theta),0,cos(-
theta)];
    R_z_phi=[cos(pi-phi),-sin(pi-phi),0;sin(pi-phi),cos(pi-phi),0;0,0,1];

end

memxn_I=R_y_theta*R_z_phi*memxn;
memyn_I=R_y_theta*R_z_phi*memyn;
memzn_I=R_y_theta*R_z_phi*memzn;
memnn_I=R_y_theta*R_z_phi*memnn;
memon_I=R_y_theta*R_z_phi*memon;

R_y_theta_I=[cos(pi/2-theta),0,sin(pi/2-theta);0,1,0;-sin(pi/2-
theta),0,cos(pi/2-theta)];

subplot(1,2,2)

for kb=1:N_max

    B1(:,kb)=R_y_theta*R_z_phi*[A1(kb,col_i)-memx(1,1);A1(kb,col_i+1)-
memx(2,1);A1(kb,col_i+2)-memx(3,1)];

end

plot3(B1(1,:), B1(2,:), B1(3,:), '- r', 'MarkerSize', 10, 'LineWidth',2)
hold on
plot3(memxn_I(1),memxn_I(2),memxn_I(3), 'o g', 'LineWidth',2, 'MarkerSize',
6)
hold on
plot3(memyn_I(1),memyn_I(2),memyn_I(3), 'o g', 'LineWidth',2, 'MarkerSize',
6)
hold on
plot3(memzn_I(1),memzn_I(2),memzn_I(3), 'o g', 'LineWidth',2, 'MarkerSize',
6)
hold on
plot3([memxn_I(1);memyn_I(1);memzn_I(1)], [memxn_I(2);memyn_I(2);memzn_I(2)
], [memxn_I(3);memyn_I(3);memzn_I(3)], '- r', 'MarkerSize', 6, 'LineWidth',1)
hold on
plot3(memnn_I(1),memnn_I(2),memnn_I(3), 'o g', 'LineWidth',2, 'MarkerSize',
6)
hold on
plot3(memon_I(1),memon_I(2),memon_I(3), 'o g', 'LineWidth',2, 'MarkerSize',
6)
text(memzn_I(1),memzn_I(2),memzn_I(3)+0.5, 'Z', 'FontSize', 6)
text(memyn_I(1),memyn_I(2),memyn_I(3)+0.5, 'Y', 'FontSize', 6)
text(memxn_I(1),memxn_I(2),memxn_I(3)+0.5, 'X', 'FontSize', 6)
text(memnn_I(1),memnn_I(2),memnn_I(3)-0.5, 'N', 'FontSize', 6)
text(memon_I(1),memon_I(2),memon_I(3)+0.5, 'O', 'FontSize', 6)

xlabel('memx')
ylabel('memy')
zlabel('memz')

```

```

x=-60:4:60;
y=-60:4:60;
[X,Y]=meshgrid(x,y);

Z1=( (memzn_I(1)-memyn_I(1))*X+(memzn_I(2)-memyn_I(2))*Y-(memzn_I(1)-
memyn_I(1))*memyn_I(1)-(memzn_I(2)-memyn_I(2))*memyn_I(2)-(memzn_I(3)-
memyn_I(3))*memyn_I(3)) / (-(memzn_I(3)-memyn_I(3)));
mesh(X,Y,Z1)
hold on
Z2=( (memzn_I(1)-memyn_I(1))*X+(memzn_I(2)-memyn_I(2))*Y-(memzn_I(1)-
memyn_I(1))*memzn_I(1)-(memzn_I(2)-memyn_I(2))*memzn_I(2)-(memzn_I(3)-
memyn_I(3))*memzn_I(3)) / (-(memzn_I(3)-memyn_I(3)));
mesh(X,Y,Z2)

axis([-60 60 -60 60 -60 60])
daspect([1, 1, 1])

grid
view([150,30])
box on

```

## B.8 Generation of PDB File with only Backbone $\text{NC}_\alpha\text{C}'$ Atoms

The full-atom PDB file output by Rosetta is processed and only backbone atoms N,  $\text{C}_\alpha$ , and  $\text{C}'$  will be left in the new PDB file, which can be used in PISEMA back-calculation.

```
clear all
clf
global B S add_atom

[A1,A2]=xlsread('2H3O_2013_4_10_cold_assign_77_new_355-12_0001.xlsx');
[N_max,M_max]=size(A1);
col_i=6;
col_f=8;
kb=1;
n_res=1;

for n=1:N_max

    symb=char(A2(n,3));
    res=char(A2(n,4));
    [M_symb, N_symb]=size(symb);

    if N_symb==2
        if symb == 'CA'

            B(kb, :)=A1(n, col_i:col_f); kb=kb+1;
            res_num(n_res)=A1(n, col_i-1);

            if res=='ALA', S(n_res)='A'; end
            if res=='GLY', S(n_res)='G'; end
            if res=='VAL', S(n_res)='V'; end
            if res=='LEU', S(n_res)='L'; end
            if res=='ILE', S(n_res)='I'; end
            if res=='SER', S(n_res)='S'; end
            if res=='THR', S(n_res)='T'; end
            if res=='CYS', S(n_res)='C'; end
            if res=='MET', S(n_res)='M'; end
            if res=='PRO', S(n_res)='P'; end
            if res=='ASP', S(n_res)='D'; end
            if res=='ASN', S(n_res)='N'; end
            if res=='GLU', S(n_res)='E'; end
            if res=='GLN', S(n_res)='Q'; end
            if res=='LYS', S(n_res)='K'; end
            if res=='ARG', S(n_res)='R'; end
            if res=='HIS', S(n_res)='H'; end
            if res=='PHE', S(n_res)='F'; end
            if res=='TYR', S(n_res)='Y'; end
            if res=='TRP', S(n_res)='W'; end
```

```

n_res=n_res+1;
end
end

if N_symb==1 & kb>1
if symb=='C', B(kb, :)=A1(n, col_i:col_f); kb=kb+1; 'C'; end
if symb=='N', B(kb, :)=A1(n, col_i:col_f); kb=kb+1; 'N'; end

end
end

S_bb=char(S);
[N,M]=size(S_bb);
[N_new_max,M_new_max,N_accpt]=size(B);

for i=1:N_accpt
outfilename=sprintf('2H3O_2013_4_10_cold_assign_77_new_355-
12_0001_bb_NCC.pdb',i);
filnam=fopen(outfilename,'w');

plane=['N ','CA','C '];
plane2=['NCC'];

atom_count=1;
frst=res_num(1);

k=1;
B=B(:, :, i);
for n=1:M

if S_bb(n)=='A', res='ALA'; end
if S_bb(n)=='G', res='GLY'; end
if S_bb(n)=='V', res='VAL'; end
if S_bb(n)=='L', res='LEU'; end
if S_bb(n)=='I', res='ILE'; end
if S_bb(n)=='S', res='SER'; end
if S_bb(n)=='T', res='THR'; end
if S_bb(n)=='C', res='CYS'; end
if S_bb(n)=='M', res='MET'; end
if S_bb(n)=='P', res='PRO'; end
if S_bb(n)=='D', res='ASP'; end
if S_bb(n)=='N', res='ASN'; end
if S_bb(n)=='E', res='GLU'; end
if S_bb(n)=='Q', res='GLN'; end
if S_bb(n)=='K', res='LYS'; end
if S_bb(n)=='R', res='ARG'; end
if S_bb(n)=='H', res='HIS'; end
if S_bb(n)=='F', res='PHE'; end
if S_bb(n)=='Y', res='TYR'; end
if S_bb(n)=='W', res='TRP'; end

if n==1, p_init=2; else, p_init=1; end
for p=p_init:3

```

```

if k<=N_new_max
fprintf(filnam,'ATOM');
fprintf(filnam,'%7.0f',atom_count);
fprintf(filnam,' ');
fprintf(filnam,plane(p, :));
fprintf(filnam,' ');
fprintf(filnam,res);
fprintf(filnam,'%6.0f', n+frst-1);
fprintf(filnam,'%12.3f', B(k, 1));
fprintf(filnam,'%8.3f', B(k, 2));
fprintf(filnam,'%8.3f', B(k, 3));
fprintf(filnam,'%6.2f', 1.00);
fprintf(filnam,'%6.2f', 0.00);
fprintf(filnam,' ');
fprintf(filnam,plane2(p));
fprintf(filnam,'%c\n', ' ');
atom_count=atom_count+1;

end

k=k+1;
end
end
end
fprintf(filnam,'TER '), fprintf(filnam,'%7.0f',atom_count),
fprintf(filnam,' '), fprintf(filnam,' '), fprintf(filnam,res),
fprintf(filnam,'%6.0f\n', n+frst-1);
atom_count=atom_count+1;
fclose(filnam);

end

```

## Appendix C

### C.1 Membrane Ab Initio Calculation

#### SPAN File Generation:

```
/work/Rosetta/rosetta_source/src/apps/public/membrane_abinitio/octopus2span.pl
```

```
2H3O.octopus
```

#### LIPS4 File Generation:

```
/work/Rosetta/rosetta_source/src/apps/public/membrane_abinitio/run_lips.pl 2H3O.fasta
```

```
2H3O.span /work/legacy_blast/blast-2.2.26/bin/blastpgp /work/blast/ncbi-blast-2.2.26+/db/nr
```

```
/work/Rosetta/rosetta_source/src/apps/public/membrane_abinitio/alignblast.pl
```

#### Membrane Ab Initio Simulation:

```
#BSUB -o memab_output
```

```
#BSUB -e memab_error
```

```
/work/Rosetta/rosetta_source/bin/membrane_abinitio2.linuxgccrelease
```

```
-in:file:fasta 2H3O.fasta
```

```
-in:file:spanfile 2H3O.span
```

```
-in:file:lipofile 2H3O.lips4
```

```
-in:file:frag3 aa2H3O_03_05.200_v1_3
```

```
-in:file:frag9 aa2H3O_09_05.200_v1_3
```

```
-in:path:database /work/Rosetta/rosetta_database/
```

```
-abinitio:membrane
```

```
-membrane:no_interpolate_Mpair
```

-membrane:Menv\_penalties  
-membrane:normal\_cycles 40  
-membrane:normal\_mag 15  
-membrane:center\_mag 2  
-score:find\_neighbors\_3dgrid  
-nstruct 1000

**Extract Output File:**

/work/NMRPipe/com/extract\_pdb3.com default.out

**RMSD Calculation:**

/work/NMRPipe/com/pdbrms ref.pdb calculated\_structures.pdb

## C.2 Rosetta All-Atom Energy Calculation

### Input PDB Files Generation (move along +z direction):

```
#BSUB -o genpdbzplus_output
#BSUB -e genpdbzplus_error

k=1

while [ $k -le N ]

do

    i=0

    cp 2H3O_-$k.pdb 2H3O_-$k+$i.pdb

    while [ $i -le 70 ]

    do

        j=`expr $i + 1`

        perl modZcoordsplus.pl -pdbfile 2H3O_-$k+$i.pdb > 2H3O_-$k+$j.pdb

        i=`expr $i + 1`

    done

    k=`expr $k + 1`

done
```

### Input PDB Files Generation (move along -z direction):

```
#BSUB -o genpdbzminus_output
#BSUB -e genpdbzminus_error

k=1

while [ $k -le N ]
```

```
do
  i=0
  cp 2H3O_${k}.pdb 2H3O_${k}-${i}.pdb
  while [ $i -le 70 ]
  do
    j=`expr $i + 1`
    perl modZcoordsminus.pl -pdbfile 2H3O_${k}-${i}.pdb > 2H3O_${k}-${j}.pdb
    i=`expr $i + 1`
  done
  k=`expr $k + 1`
done
```

### **All-Atom Energy Calculation:**

```
#BSUB -o score_jd2_2_output
#BSUB -e score_jd2_2_error

/work/Rosetta/rosetta_source/bin/score_jd2.linuxgccrelease

-database /work/Rosetta/rosetta_database/

-in:file:l pdblast

-in:file:spanfile 2H3O.span

-in:file:lipofile 2H3O.lips4

-layer -abinitio:membrane

-membrane:fixed_membrane

-score:weights membrane_highres_Menv_smooth.wts
```

-out:file:silent score\_jd2\_2.out

-out:file:scorefile score\_jd2\_2.sc

### C.3 Symmetry Docking

#### Symmetry Definition File Generation:

```
/work/Rosetta/rosetta_source/src/apps/public/symmetry/make_symmdef_file_denovo.py
```

```
-symm_mode cn
```

```
-nsub 4
```

```
> C4.symm
```

#### Symmetry Docking Simulation:

```
#BSUB -o symdock_mem_nofix_output
```

```
#BSUB -e symdock_mem_nofix_error
```

```
/work/Rosetta/rosetta_source/bin/SymDock.linuxgccrelease
```

```
-database /work/Rosetta/rosetta_database
```

```
-nstruct 10000
```

```
-out:file:silent symdock_mem_nofix.out
```

```
-out:file:scorefile symdock_mem_nofix.sc
```

```
-in:file:s 2L0JT.pdb
```

```
-in:file:native 2L0Jt.pdb
```

```
-in:file:frag3 aa2L0JT03_05.200_v1_3
```

```
-in:file:frag9 aa2L0JT09_05.200_v1_3
```

```
-rg_reweight 0.01
```

```
-score:find_neighbors_3dgrid
```

```
-symmetry:symmetry_definition C4.symm
```

```
-symmetry:initialize_rigid_body_dofs
```

-fold\_and\_dock::rotate\_anchor\_to\_x  
-relax:fast  
-relax:jump\_move  
-abinitio:membrane  
-membrane:no\_interpolate\_Mpair  
-membrane:Menv\_penalties  
-in:file:spanfile 2L0JT.span  
-in:file:lipofile 2L0JT.lips4  
-stage2\_patch score\_membrane\_s2.wts\_patch  
-stage3a\_patch score\_membrane\_s3a.wts\_patch  
-stage3b\_patch score\_membrane\_s3b.wts\_patch  
-stage4\_patch score\_membrane\_s4.wts\_patch  
-membrane:Membed\_init  
-relax:membrane  
-score:weights membrane\_highres\_Menv\_smooth.wts  
-increase\_cycles 4.3  
-evaluation::gdtmm