Investigation of Antimicrobial Peptides in Lipid Membranes by Solid-State NMR

Yuan Zhang
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Chemistry Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/10291

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Investigation of Antimicrobial Peptides in Lipid Membranes by Solid-state NMR

by

Yuan Zhang

A dissertation submitted to the graduate faculty
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Chemistry

Program of Study Committee:
Mei Hong, Co-Major Professor
Klaus Schmidt-Rohr, Co-Major Professor
Emily Smith
Xueyu Song
Surya Mallapragada

Iowa State University
Ames, Iowa
2011

Copyright © Yuan Zhang, 2011. All rights reserved.
# Table of Contents

Acknowledgement ............................................................................................................. v

Abstract ...................................................................................................................... vii

Chapter 1 Introduction ................................................................................................... 1
  1.1 Antimicrobial Peptides .......................................................................................... 1
  1.2 Defensins .............................................................................................................. 2
  1.3 Protein Structure Determination by Solid State NMR .......................................... 6
  1.4 Thesis Organization ............................................................................................ 8
  1.5 Copyright Permission ....................................................................................... 9

Chapter 2 NMR Methodology ...................................................................................... 17
  2.1 Protein Structure Determination by Solid-state NMR ....................................... 17
  2.2 The Homonuclear Double Quantum Filter Experiment ..................................... 24
  2.3 Dipolar Correlation Experiments for Measuring Backbone (\(\phi, \psi\)) Torsion

Chapter 3 Resonance Assignment and Three-Dimensional Structure
Determination of a Human Alpha-Defensin, HNP-1, by Solid-State NMR ................. 36
  Abstract .................................................................................................................... 36
  Introduction ............................................................................................................. 37
  Results .................................................................................................................... 38
Chapter 4 3D $^{13}$C-$^{13}$C-$^{13}$C Correlation NMR for De Novo Distance Determination of Solid Proteins and Application to a Human Alpha Defensin ................................................................. 80

Abstract ................................................................................................................. 80
Introduction .............................................................................................................. 81
Materials and Methods ......................................................................................... 82
Results and Discussion ......................................................................................... 84
Conclusion .............................................................................................................. 95
Acknowledgment .................................................................................................. 95

Chapter 5 Membrane-Bound Structure and Topology of a Human Alpha Defensin Indicates A Dimer Pore Mechanism for Membrane Disruption................................................................................. 100

Abstract ................................................................................................................. 100
Introduction .............................................................................................................. 101
Materials and Methods ......................................................................................... 103
Results .................................................................................................................... 107
Discussion .............................................................................................................. 121
Supporting Information ......................................................................................... 133

Chapter 6 Bicelle Alignment and Orientation Simulation of HNP-1 in Bicelle Systems ......................................................................................................................... 137
Acknowledgements

I am truly grateful to all people who helped me get through the past five years of study and research. The first and foremost people I want to thank are my thesis advisors Dr. Mei Hong and Dr. Klaus Schmidt-Rohr. They led me into the wonderful world of solid-state NMR. Mei is an excellent mentor who provides the lab a positive atmosphere for research and shapes me as a scientist. Every time when I feel my research is going toward the dead end, she will always be the person who provides me the enlightening advices and encourages me to open my mind. I am also thankful to Klaus for his deep and thorough theoretical lessons and helpful discussion.

I would like to thank my committee members, Dr. Emily Smith, Dr. Xueyu Song and Dr. Surya Mallapragada for offering their opinions and guidance on my scientific course of study. I would also like to thank my past and current group members not only for helping me through my scientific life but also for being really good friends for the past five years: Dr. Rajee Mani, Dr. Ming Tang, Dr. Wenbin Luo, Dr. Timorthy Doherty, Dr. Sarah Cady, Dr. Yongchao Su, Dr. Shenhui Li, Dr. Hongwei Yao, Fanghao Hu, Marilu Perez, Tuo Wang, Aaron Liao, Byungsu Kwon, Jonathan Williams, Keith Fritzsching and all the members of the Schimidt-Rohr group. Thanks to Dr. Wuyuan Lu and Dr. Jing Li at the Institute of Human Virology for supplying us with high quality peptide and kindly teaching me protein expression technique. I am also thankful to Dr. Jacek Lubkowski in the National Cancer Institute for teaching me crystallization skills. I greatly appreciate the assistance from many members of the Iowa State University Department of Chemistry staff – Dr. Shu Xu and Dr. Dave Scott in the NMR facility, Bev, Carlene, Renee, Lynette and Mary in the Chemistry Office.

I would like to express thanks to my friends here. To my chemistry peers, I would not forget the time when we worked hard together to live through the tough courses and exams in the first year. To my friends outside Gilman who are too numerous to name, thanks for sharing the countless wonderful time with me as well as providing me selfless help. You guys make Ames such a great place and I will definitely miss here in the future.
And thanks to my friends far away from Ames: Zilu Wang and Yiqing Wang. Thank you for being good friends with me for such a long time and helping me survive through those tough times.

Last but not least, I thank my beloved parents for their unconditional love and support through my life. Thanks to them, I realize that there is always a place where I can take a rest whenever I feel tired. To them I sincerely dedicate this thesis.
Abstract

Solid-state NMR spectroscopy is an important tool for studying the chemical and three-dimensional structures of organic and inorganic solids because of its intrinsic atomic-level structural information, nonperturbing nature, and the large range of dynamic time scales. It is especially powerful in studying insoluble and noncrystalline membrane proteins, which are difficult to analyze by traditional X-ray crystallography or solution NMR techniques. In this thesis, various NMR techniques are used to study the structure and dynamics of membrane proteins within lipid bilayers.

The main protein we are focusing on is human neutrophil peptide 1 (HNP-1). It is a small cysteine-rich cationic antimicrobial protein found in human neutrophils. It forms the first line of defense by the innate immune system of humans against pathogens. The antimicrobial activity of HNP-1 is believed to be caused by disruption of the microbial cell membrane and various models of HNP-membrane interaction have been proposed. However, none of these mechanistic models are based on structure information from the lipid bilayer. Therefore, understanding the peptide structure in the presence of membrane and its interaction with the lipids will shed light on the antimicrobial mechanism of HNP-1. As a first step, we have calculated the minimum-energy structures of uniformly $^{13}$C, $^{15}$N-labeled microcrystalline HNP-1 based on all NMR torsion angle and distance restraints determined by various 2D and 3D correlation techniques. The solid-state NMR structure has close similarity to the crystal structures of the HNP family. Then we reconstituted HNP-1 into DMPC/DMPG lipid bilayers. We confirmed that the protein is predominantly dimerized at high protein/lipid molar ratios by $^{19}$F spin diffusion experiments. Various methods under magic-angle spinning (MAS) such as $^{13}$C-$^{31}$P REDOR, $^1$H spin diffusion and $^{13}$C DIPSHIFT have been utilized to study the interaction of HNP-1 with lipid bilayers. The experimental results strongly support a “dimer pore” topology of HNP-1 in which the polar top of the dimer lines an aqueous pore while the hydrophobic bottom faces the lipid chains.
The second focus of this thesis is the oriented bicelle alignment, we have studied the alignment of bicelles with different lipids combinations, long- to short-chain lipid ratios, hydration levels and a phase diagram was generated. We also show that the orientation of a protein called HNP-1, which has 3 β-strands and dimerizes, can be simulated by predicting the correlation of the $^{15}$N anisotropic chemical shift and the N-H dipolar coupling.
Chapter 1

Introduction

1.1 Antimicrobial Peptides

Antimicrobial peptides are small polypeptides that constitute part of the innate immune system of many plants and animals to rapidly kill invading microbial pathogens, such as Gram negative and Gram positive bacteria, mycobacteria, enveloped virus, fungi and even transformed or cancerous cells. They are divided into subgroups on the basis of the amino acid composition and structure. For instances, Dermeidins from humans are anionic peptides that are rich in glutamic and aspartic acids [1]; Cecropins from insects [2] and LL37 from humans [3] are linear cationic $\alpha$-helical peptides that are characterized by a lack of cysteine; Tachyplesin from the horseshoe crab [4] and protegrin-1 in porcine leukocytes [5] are cationic peptides that are enriched in arginines and that contain cysteines to form disulfide bonds. Despite this structural diversity, all antimicrobial peptides contain hydrophilic amino acid residues aligned along one side and hydrophobic amino acid residues aligned along the opposite side, with the underlying structural principle that polar and non-polar regions in the molecule are spatially separated (‘amphiphatic’ structure) [6].

The modes of action by which antimicrobial peptides kill bacteria are varied. One way is to allow the AMPs to penetrate into the cell to bind intracellular molecules that are crucial to cell survival [7]. This model includes inhibition of cell wall synthesis, alternation of the cytoplasmic membrane, activation of autolysin, which is an enzyme that hydrolyzes the components of a biological cell, inhibition of DNA, RNA and protein synthesis, and inhibition of certain enzymes. Another way to kill the microbes is to disrupt their cell membranes. Various models have been proposed for this membrane disruption [8]. The first model, ‘barrel-stave’, involves several peptides forming a barrel to span the lipid bilayer. This model has been used to explain pore formation and step-wise conductivity increases in single-channel measurements of alamethicin [9]. The
second model is the ‘carpet’ model, in which peptides first assemble on the bilayer surface with the hydrophobic side embedded into the membrane, causing membrane thinning. Once a critical concentration is reached, the membrane is micellized and starts to leak. This model is proposed based on the observations of the antimicrobial peptide dermaseptin, revealed by a dissipation of diffusion potential and a release of entrapped calcein from SUV (small unilamellar vesicles) [10]. The third model is the toroidal pore model, in which peptide aggregation and lipid disorder lead to the formation of a stable channel. This model has been shown to occur for protegrin-1 [11,12]. The last model is the in-plane diffusion model, in which the peptides are bound to the surface of lipid membrane and diffuse around the bilayer normal, which cause areas of local negative curvature strain and sufficient disruption to form transient pores [13,14].

There are several factors that determine the selectivity of antimicrobial peptides. First, the microbial membranes have more anionic phospholipids, especially for the outer leaflet, while the host cell membranes consist of zwitterionic phospholipids. This makes the microbial membrane more negatively charged, thus it is more susceptible to antimicrobial peptides with cationic residues [15]. Second, cholesterol is normally widely distributed in the mammalian cell membranes but absent in bacterial cell membranes. People have observed the presence of cholesterol will generally reduce the activities of the antimicrobial peptides by either rigidifying the lipid bilayers or interacting with the peptides and therefore protect the cells from attack by the AMPs [16]. Besides, the more negative transmembrane potential of bacterial cells also facilitates the insertion of positively charged peptides into the membrane [17].

1.2 Defensins

Defensins are small (3-5 kDa), cysteine-rich, cationic proteins stabilized by several (usually three) disulfide bonds [18-20]. They are found in both vertebrates and invertebrates and are active against bacteria, fungi and many enveloped and non-enveloped viruses [21,22]. They are classified into three categories based on the disulfide-linkage patterns. In \( \alpha \)-defensins, three disulfide bridges have the topology, Cys\(^{I}\)-Cys\(^{VI}\), Cys\(^{II}\)-Cys\(^{IV}\), Cys\(^{III}\)-Cys\(^{V}\) [23]. They are expressed primarily in neutrophils as well
as in NK cells and certain T-lymphocyte subsets. β-defensins are the most widely distributed, being secreted by leukocytes and epithelial cells of many kinds. The disulfide linkage arrangement of β-defensins is Cys\(^1\)-Cys\(^5\), Cys\(^II\)-Cys\(^IV\), Cys\(^III\)-Cys\(^VI\) [24]. θ-defensins are cyclic peptides and are found in some non-human primates, but not in human, gorilla, bonobo and chimpanzee [25-27]. So far, six α-defensins have been identified in humans. The first four defensins are often called neutrophile peptides (HNP1-4) [18,28,29] and defensins 5 and 6 are frequently referred to as intestinal defensins (HD5-6) [30,31]. HNP1-3 differs only by a single N-terminal residue [22]. The relative potencies of human α-defensins against Gram-negative bacterium E. coli are HNP4 > HNP2 > HNP1 = HNP3 [32].

The overall fold of the α-defensin monomer is attributed to the three β-strands arranged into antiparallel β-sheet, the long loop connecting β1 and β2 and the β-hairpin formed by β2 and β3. The structures of HNP1-3 are similar, which are dimerized in the crystal through intermolecular H-bonds between the two β2 strands, extending the triple-stranded β-sheet to a six-strand β-sheet. The dimer has a basket shape and is amphipathic, with a polar top and an apolar base [19,20].

It is generally believed that the killing of microorganisms by defensins is a consequence of disruption of the microbial membrane [33]. Many antimicrobial assays and lipid vesicle experiments have been reported to understand the mechanism of action [34,35], and a number of biochemical studies indicate pore formation by human α-defensins in anionic lipid membranes. For example, dye release and fluorescence spectroscopy experiments found that HNP1 caused both fusion and lysis of negatively charged lipid vesicles through electrostatic interactions [36]. Electron micrographs of human parasite Trypanosoma cruzi cells in the presence of micromolar concentration of HNP1 showed 25 nm sized pores in the cellular and flagellar membranes, through which HNP1 appears to enter the trypanosome cells, causing subsequent DNA fragmentation and cell destruction [37]. Vesicle leakage experiments also showed that HNP-induced pores increasing in number with the concentration of the anionic lipid [38].
Based on the crystal structure of HNP3, several models have been proposed to explain the antimicrobial mechanism of human α-defensins. The first model is called the wedge model [19], in which the dimer is inserted into the membrane either partly or completely, with its hydrophobic base embedded into the hydrocarbon region of the membrane. The second model is called the dimer pore model [19]. In this model, two dimers form a membrane-spanning pore with the hydrophobic base facing the lipids and the top view shows a basket shape. The general pore model [19] is similar to the dimer pore model, with the dimers rotated by 90º around the horizontal axis and its side view shows the basket shape of the dimer. A multimeric pore model was also proposed with six to eight HNP dimers forming a large pore with an inner diameter of 20 – 25 Å and the dimers are oriented so that the Arg residues are located in two rings parallel to the membrane surface [39].

To obtain high-resolution information about intermolecular interaction and dynamics of HNP-1 in the lipid membrane, solid-state NMR spectroscopy is an excellent technique because of its advantage in the amorphous and insoluble systems compared to solution NMR and X-ray diffraction. In order to distinguish which one of the models mentioned above explains the antimicrobial activity of HNP-1, the following aspects of the complex system consisting of HNP1 and lipid bilayers have to be investigated.

**Depth of Insertion**

For antimicrobial peptides, their depth of insertion into the lipid membrane often provides important clues to their mechanism of action. For example, in the carpet model and the in-plane diffusion model, the peptides reside on the membrane surface, while for the barrel-stave model and the toroidal pore model they insert all the way to the hydrophobic center of the lipid bilayer. In the case of HNP1, the measurement of depth of insertion can differentiate wedge model from the other models. Several NMR methods are developed to determine the depth of insertion of membrane proteins. A lipid-to-protein 1H spin diffusion experiment is employed for immobilized membrane proteins such as PG-1 and TP-1 [40-43]. In this experiment, the intensity buildup of lipid chain-peptide cross peaks as a function of mixing times gives a semi-quantitative measurement
of the minimum distance separation between the protein and the hydrophobic core of the bilayers. For mobile membrane peptides that have similar mobility as the lipids, the Paramagnetic Relaxation Enhancement (PRE) effect is used to extract quantitative distance information. Since paramagnetic ions binding to the phosphate groups on the membrane surface can broaden the nuclear spin signals by enhancing their $T_2$ relaxation in a distance-dependent manner [44], the electron-nucleus distances can be measured semi-quantitatively by comparing the PRE effect of the peptide with that of the lipids. Our group has determined the penetratin position in the lipid bilayers using this method [45].

**Site-Specific Quantitative Depth of Insertion**

Revealing the contact between the peptide and the lipids helps to explain how the peptide binds the lipid bilayer. A well-established $X^{-31}P$ ($X=^{13}C, ^{15}N$, etc.) REDOR technique can be carried out to measure residue specific and accurate depths of insertion [46]. Our group measured the depth of V6 $^{13}CO$ and G10 $^{13}CO$ in TP-I bound to DMPC/DMPG membranes by $^{13}C^{-31}P$ REDOR on trehalose-protected lipid membranes and found that TP-I lies at the interface between water and the hydrophobic chains, parallel to the membrane surface [14].

**Oligomeric Structure of Membrane Proteins**

Membrane-lytic peptides are thought to oligomerize to inflict maximal damage to the lipid membrane. It is thus important to determine the protein’s quaternary structure in addition to the secondary and tertiary structures to understand the structural basis of membrane protein function. Previously our group developed a $^{19}F$ spin diffusion NMR technique based on the CODEX technique [47] to determine both the oligomeric number and long-range intermolecular distances in membrane protein assemblies. The fraction of magnetization residing on the initial spin without the orientational change is the inverse of the number of spins in the cluster [48,49]. The rate of the magnetization decay reflects the distance between the monomers [50]. For HNP1, it has been proposed that several dimers need to aggregate together to achieve pore formation on the lipid membrane.
[19,39]. Therefore, understanding the oligomerization of HNP1 in the lipid bilayers should shed light on its antimicrobial mechanism.

**Protein Dynamics**

The hydrated lipid membranes are highly mobile systems. Lipids undergo fast lateral diffusion and uniaxial rotational diffusion, thus molecular motion for membrane proteins is very common and is often related to the function and lipid-interaction of proteins. The most robust solid-state NMR strategy for obtaining site-resolved dynamics information employs 2D chemical shift and dipolar correlation spectroscopy to measure heteronuclear X-H dipolar couplings (DIPSHIFT or 2D LG-CP) [51,52]. Since the distance of X-H bond is fixed, the dipolar coupling strength is also fixed in the absence of motion; thus any reduction from the rigid-limit value is indicative of motional averaging. Our group has measured C-H and N-H dipolar couplings in several membrane proteins and peptides to determine the amplitude of motion [41,53].

### 1.3 Protein Structure Determination by Solid State NMR

The main methods for high-resolution protein structure determination are X-ray crystallography and solution NMR. However, there is a wide variety of proteins that are insoluble in water, which makes the X-ray crystallography and solution NMR unsuitable. Over the past few years, there have been remarkable advances in solid-state NMR spectroscopy which make the structure characterization of insoluble proteins possible. Excellent NMR line widths can be achieved for microcrystalline proteins using magic angle spinning (MAS) methods. For instances, the $^{13}$C line width of the microcrystallized ubiquitin can be as narrow as 0.5-0.6 ppm and the $^{15}$N line widths are 0.3-0.5 ppm [54].

In recent years, solid-state NMR has accumulated a considerable track record with respect to *de novo*-structure determination of proteins, which are mostly on nano- or microcrystalline preparation of proteins and on protein fibrils. Oschkinat and coworkers determined the structure of SH3 domain with 286 carbon-carbon and 6 nitrogen-nitrogen inter-residue constraints and the C$\alpha$ coordinates of this structure show a root mean square deviation of 1.6 Å [55]. Loquet et al. published a structure of the domain-swapped dimer
of the Crh domain with 1.33 Å RMSD by using 643 distance constraints, combining carbon-carbon and proton-proton distances [56]. Frank et al. determined the GB1 structure through 7826 distance constraints, resulting in a backbone RMSD of 0.31 Å [57]. Investigations of membrane protein systems are also making progress. The group of Baldus has published a number of interesting structural models of potassium channels derived from solid-state NMR, which enabled a deeper understanding of potassium channel activations and inhibition [58]. The group of Rienstra has published assignments of the membrane protein DsbB [59]. In general, the improvement in solid-state NMR methodology will lead to reasonably well determined structures that are useful to the scientific community.

The structure determination paradigm is overall robust. First, $^{13}$C and $^{15}$N chemical shifts were measured on uniformly $^{13}$C, $^{15}$N labeled proteins by well-defined solid-state NMR correlation experiments, such as NCACX and NCOCX, providing both inter- and intra-residue correlations [60-62]. Correlations between $\alpha$-carbons have also proved to be of high value for sequential assignment. These chemical shift values provide information about the secondary structure of the protein.

With all the chemical shifts assigned, the next step is to obtain inter-residue distance constraints. Various methods have been developed for this purpose. The most frequently used method is DARR or PDSD with different mixing times [63]. Another type of experiment is CHHC, which has very short CP and $^1$H-$^1$H exchange times, to obtain inter-strand distance constraints [64]. Griffin’s group developed the PAR technique, which can obtain long range constraints more easily than DARR [65]. A 3D CCC technique was developed by our group in order to obtain both short range and long range correlations and improve the resolution by adding another dimension [66]. Different labeling schemes are also applied. Instead of uniformly labeled protein, 1,3- and 2-$^{13}$C-glycerol are used as carbon sources during the growth of the bacteria to generate partially labeled protein [55,67-69]. In this way, the spectra are less crowded, which is helpful for removing ambiguities and resolving resonance overlap. Another very promising approach is based on the clever use of paramagnetic centers, either present in the molecules or added to the sample [70]. Since these paramagnetic centers often cause
low intensities or pseudocontact shifts, by analyzing these effects, one can estimate the distance from various residues to the paramagnetic center and therefore obtain more distance constraints.

Torsion angle restraints can also be measured in proteins. Hong et al have developed a method to determine the torsion angle $\phi$ in peptides, based on the measurement of the relative orientation of the N-HN and C$\alpha$-H$\alpha$ bonds [71]. And the $\psi$ torsion angle can be measured by fitting the time-course of C$\alpha$-C’ double-quantum coherence dephasing [72]. The group of Rienstra has demonstrated that relative orientations of molecular fragments can be determined from dipolar line shapes of proteins. By using vector angle restraints, they improved the backbone RMSD of GB1 from 1.01 Å to 0.31 Å [57].

1.4. Thesis Organization

This thesis is composed mainly of published journal papers on the structure and antimicrobial mechanism of the human neutrophil peptide 1. Some papers contains supplementary information that was only published in the online edition of the journal, and this data is located at the end of its respective chapter.

Chapter 2 describes the solid state NMR theory and techniques related to this work. Double quantum filter theory is briefly reviewed. Chapter 3 shows the determination of the three-dimensional structure of HNP-1 in a microcrystalline state outside the lipid membrane. A suite of 2D and 3D magic-angle spinning experiments were applied to complete the $^{13}$C and $^{15}$N chemical shift assignment. Combining the torsion angle constraints derived from chemical shifts and inter-residue distance constraints, a high resolution solid-state NMR structure of HNP-1 was determined.

Chapter 4 demonstrates a 3D $^{13}$C-$^{13}$C-$^{13}$C (CCC) magic-angle spinning correlation technique for obtaining two orders of magnitude more long-range distance restraints. Combining these distance restraints and the chemical-shift-derived torsion angles, we obtained a de novo high-resolution NMR structure of HNP-1.
Chapter 5 determined the conformation, dynamics, oligomeric state and topology of HNP-1 in DMPC/DMPG lipid bilayers by various solid-state NMR techniques. The results strongly support a “dimer pore” topology of HNP-1 and thus provide the first glimpse into the membrane-bound structure and mechanism of action of human-α-defensins.

Chapter 6 shows the investigation of the alignment of bicelles with various conditions. We also show in this chapter that the orientation of HNP-1 can be determined in principle through 2D correlation spectra of $^{15}$N anisotropic chemical shift and N-H dipolar coupling.

Finally, two appendices are located at the end of the thesis. Appendix A summarizes the protein expression, purification and crystallization of HNP-1 in detail. Appendix B gives the input code for protein structure determination by Xplor-NIH.

1.5 Copyright Permission

Chapter 3, 4, 5 are reprints of published papers. The copyright permissions do not require a license for re-use in a thesis or dissertation. Figure 2.3, Figure 5.5a and Figure 5.6a are reprints from the referenced paper. Permissions have been obtained from the following publishing groups.

Chapter 3 and 4: Elsevier, Inc.

Chapter 5: American Chemical Society

Figure 2.3: Springer, license number: 2738430062616

Figure 5.5a: Elsevier, Inc. license number: 2738400453944

Figure 5.6a: Elsevier, Inc. license number: 2738400676772
References


Chapter 2

NMR Methodology

2.1 Protein Structure Determination by Solid-state NMR

2.1.1 Heteronuclear Magnetization Transfer Method -- SPECIFIC CP

3D NCACX and NCOCX are two important experiments for $^{13}$C and $^{15}$N resonance assignment of protein by solid state NMR. In these experiments, the $^{15}$N-$^{13}$C SPECIFIC CP is a crucial step. The cross-polarization experiment [1,2] is plagued by its extreme sensitivity to the exact match condition at one of the sidebands of the Hartman-Hahn condition $|\omega_{1S}| - |\omega_{1I}| = n|\omega_r|$ [3] under fast magic angle spinning (MAS). Here $\omega_{1S}$ and $\omega_{1I}$ denote the $^{13}$C and $^{15}$N RF field amplitudes and $\omega_r$ is the spinning (angular) frequency and $n$ is a small integer. This condition must be met within the magnitude of the dipolar coupling frequency. Therefore, the chemical shift differences make it impossible to achieve exact match for all spin pairs simultaneously. However, this chemical shift differences can be utilized to achieve frequency specific transfer such as N $\rightarrow$ CO or N $\rightarrow$ CA.

Figure 2.1 shows the theory of SPECIFIC CP [4]. If the carrier is close to the CA chemical shift (~55 ppm), then $\omega_{\text{eff,CA}} \approx \omega_{1,\text{CA}}$. But for CO, since it has a large chemical shift difference from CA (18 kHz on a 600 MHz spectrometer and 12 kHz on a 400 MHz NMR), its $\omega_{\text{eff,CO}}$ can be calculated by the equation: $\omega_{\text{eff,CO}}^2 = \omega_{1,\text{CO}}^2 + \omega_r^2$. Therefore, if the $\omega_{\text{eff,CA}}$ matches the $^{15}$N-$^{13}$C CP matching condition $|\omega_{\text{eff,CA}}| - |\omega_{\text{eff,N}}| = n|\omega_r|$, the $\omega_{\text{eff,CO}}$ cannot satisfy this condition. Thus we will obtain only the N $\rightarrow$ CA transfer and vice versa. The MLF spectra show that we can obtain clean NCA or NCO spectra using the correct condition.
It turns out that the SPECIFIC CP condition is quite sensitive that even a difference of 0.2 dB, which translates to ~2% difference in $\omega_1$, difference can affect the intensity significantly. Usually the efficiency of SPECIFIC CP, as indicated by the percentage between the $^{13}$C CP and the $^{13}$C detected SPECIFIC CP, is ~30% for NCA and ~50% for NCO.

2.1.2 3D NCACX and NCOCX and Resonance Assignment
A 3D N-C-C correlation technique can be used to obtain the sequential $^{13}$C and $^{15}$N chemical shift assignment of a protein [5,6]. In this experiment, the magnetization was first transferred from $^1$H to $^{15}$N by a normal Hartman-Hahn $^1$H-$^{15}$N CP with a 1 ms CP contact time. A $^{15}$N $t_1$ evolution time was applied after the CP to give the F1 dimension. Then the magnetization was selectively transferred from $^{15}$N to either $^{13}$CA or $^{13}$CO by SPECIFIC CP. The contact time is 3-4 ms, longer than $^1$H-$^{13}$C CP, to give complete magnetization transfer. After the SPECIFIC CP, there is a $^{13}$C $t_2$ evolution time to give the F2 dimension for $^{13}$CA or $^{13}$CO. A DARR (Dipolar Assisted Resonance Recoupling) block was implemented to drive the magnetization transfer from the backbone $^{13}$C to the sidechain by $^{13}$C-$^{13}$C dipolar coupling. The spin diffusion time is variable. Previous experiences have shown that 40-60 ms is enough for a microcrystalline sample. The membrane sample may need longer time (100 ms) because it has more mobility, which weakens the $^{13}$C-$^{13}$C dipolar coupling and makes it more difficult to obtain $^{13}$C-$^{13}$C correlation. The 1D test of NCACX and NCOCX shows it has 5-10% intensity of a normal $^{13}$C CP experiment. The 3D experiment needs 4 days to one week to obtain a sufficient signal-to-noise ratio.

![Figure 2.2. 3D NCACX and NCOCX pulse sequence.](image)

Both 3D NCACX and NCOCX spectra are necessary for sequential chemical shift assignment of a protein. The NCACX gives intra-residue assignment and NCOCX gives inter-residue assignment. One can make a “backbone walk” by combining these two spectra. First, we need a residue which has a unique $^{13}$C or $^{15}$N chemical shift as a starting
point. For example, for the microcrystalline HNP-1 sample, its Ala2 residue has a NH3 group and its 15N chemical shift is ~41 ppm which is “folded” to ~110 ppm in the NCACX spectrum. So we found the Ala2 13Cα, 13Cβ and 13CO in the 110 ppm 15N plane of NCACX. Then we went to the NCOCX spectrum and found the 13C chemical shifts of Ala2 in the correct 13CO plane. Since the NCOCX gives 15Ni+1 – 13COi – 13CXi correlation, the 15N chemical shift of these peaks is assigned to the next residue, Cys3. After this, we returned to the NCACX spectrum, went to the Cys3 15N plane and found the chemical shifts of the Cys3 backbone and sidechain carbons. Same as the Ala2 residue, we then returned to the NCOCX spectrum again to find the 15N chemical shift of Tyr4.

Although sequential assignment is obtained by “backbone walk” because it uses the backbone 15N, 13CO and 13Cα chemical shifts, the sidechains of amino acids are very useful as well. There are a number of amino acids that have unique sidechain 13C chemical shifts. For example, alanine, leucine and isoleucine can be easily identified by their distinct correlation patterns involving their isolated methyl group signals (16-20 ppm, isoleucine even has another methyl which gives a more upfield chemical shift: ~12 ppm). The characteristic Cβ-Cα cross-peaks at the downfield end of the aliphatic region allow the identification of serines (65 ppm – 60 ppm) and threonines (70 ppm – 60 ppm). The glutamates and glutamines can be identified starting from their Cδ chemical shifts, isolated at the downfield end of the spectrum (182 ppm for Glu and 177 ppm for Gln). Glycines can be recognized by their isolated 15N-Cα and 15N-CO cross peaks (15N: ~110 ppm, 13CO: ~170 ppm, 13Cα: ~45 ppm) and cysteines can be identified by their isolated 13Cα-13Cβ (~40 ppm) correlations. Prolines can be recognized on the basis of their typical Cα (62 ppm) and Cδ (47 ppm) chemical shifts. They also have a downshield 15N chemical shift of ~134 ppm. But since this 15N doesn’t have any 1H attached, its signal is difficult to observe, which makes prolines “breaking point” in resonance assignment.

2.1.3 Torsion Angle Prediction by TALOS

The 13C and 15N chemical shifts allow us to determine the backbone (φ, ψ) torsion angles of a protein using the well-known empirical relation between N, Cα, Cβ and CO
chemical shifts and protein secondary structure [7]. The TALOS program can be used for this purpose [8].

The idea behind TALOS is that if one can find some triplet of residues in a protein of known structure with similar secondary shifts and sequence to a triplet in the target protein, then the $\phi$ and $\psi$ angles in the known structure will be useful predictors for the torsion angles in the target protein. This program is a database system for empirical prediction of $\phi$ and $\psi$ backbone torsion angles using a combination of five chemical shifts (H$\alpha$, C$\alpha$, C$\beta$, CO, N) for a given protein sequence. During the prediction, it searches a high-resolution structure database for the 10 best matches to the secondary chemical shifts of a given residue in the target protein along with its two flanking residues (a triplet). If there is consensus of $\phi$ and $\psi$ angles among the 10 best database matches, then it uses these database triplet structures to form a prediction of the backbone angles of the target residue. If the 10 best matches have mutually inconsistent values of $\phi$ and $\psi$, the matches are declared ambiguous and no prediction is made for the central residue. What is more, a prediction is considered “good” if at least 9 out of the 10 best database matches fell in the same well-populated region of the Ramachandran map as the $\phi$ and $\psi$ values from the crystal structure. Conversely, a prediction was considered “bad” or incorrect if it greatly deviated from the observed $\phi$ or $\psi$ angles from the crystal structure.

In 2009, the Bax group developed an enhanced version of the TALOS system named TALOS+ [9]. It adds an artificial neural network (ANN) classification scheme to this database mining approach. The neural network analyzes the chemical shifts and sequence to estimate the likelihood of a given residue being in a sheet, helix, or loop conformation. This ANN classification information is combined with the database mining results to increase the number of residues where useful backbone angle predictions can be made. As a result, the TALOS+ rules for defining consistent (“good”) predictions are slightly stricter. Three conditions need to be satisfied: 1. All 10 best database matches fall in a “consistent” region of the Ramachandran map. 2. The confidence of the ANN 3-state $\phi$/$\psi$ distribution prediction for this residue (defined as the difference between the probabilities of the two most favored predicted states) must be above 0.6. 3. The order
parameter $S^2$ values predicted by the RCI method, which is based on chemical shift data to give a quantitative mapping of protein backbone mobility [10], is larger than 0.5.

Figure 2.3. The flowchart of TALOS+ database search procedure [9].

2.1.4 XPLOR-NIH

Xplor-NIH is a generalized package for biomolecular structure determination from experimental NMR data combined with known geometric data [11,12]. This is achieved by seeking the minimum of a target function comprising terms from the experimental NMR restraints, covalent geometry and non-bonded contacts using a variety of minimization procedures including molecular dynamics in Cartesian and torsion angle space, Monte Carlo methods and conventional gradient-based minimization. The following potential energy terms are frequently used in NMR structure determination:

1. Covalent geometry. These terms ensure that fixed bond lengths, bond angles, planar atom groups and chiral centers are correct.
2. Non-bonded contacts. These terms include a repulsive quartic van der Waals term, which is used solely to prevent atoms from overlapping.

3. NMR observables. A number of potential energy terms are included in Xplor-NIH that correspond to NMR observables. For structure determination by solid state NMR, the most important terms are torsion angle restraints generated from chemical shifts and $^{13}$C-$^{13}$C and $^{15}$N-$^{15}$N distance restraints, which are derived from the long mixing time spin diffusion spectra. Sequential Cα–Cα distances are fixed by the peptide plane geometry to be 3.8 Å and are thus placed in the range of 3.5–4.1 Å. Sequential N–N distances depend on the $\phi$ torsion angle and fall within the range of 2.5–3.6 Å. In one representative restraints sorting method, all inter-residue peaks first observed in the 100-ms $^{13}$C–$^{13}$C DARR spectrum were assigned to the range 2.5–5.4 Å, those first appearing in the 200-ms DARR spectrum were assigned to 2.5–6.3 Å, and those inter-residue peaks appearing in the 2D CHHC spectra were assigned to 2.5–7.5 Å [8]. In addition, the few interresidue correlations observed in the 40-ms $^{13}$C–$^{13}$C DARR spectrum were assigned to the range 2.5–4.8 Å. All non-sequential $^{15}$N–$^{15}$N contacts were placed in the 3.0- to 6.0-Å range.

4. Knowledge-based potentials of mean force. These terms seek to bias conformational space sampling during simulated annealing to regions that are known to be physically realizable.

The structure calculations proceed in two stages. For the annealing process, an ensemble of 200 structures was calculated by performing molecular dynamics at 3500 K for 40 ps, followed by slow cooling from 3500 K to 25 K in 12.5 K increments, with 0.4 ps of dynamics at each temperature using a soft square NOE potential. In the second stage, each structure was refined by repeating the annealing protocol with 10 ps of initial annealing using a hard square NOE potential with the $k_{\text{NOE}}$ force constant held at 30 kcal. After structure refinement, the ensemble of 10 lowest energy structures was selected to represent the final structure of HNP-1.
2.2 The Homonuclear Double Quantum Filter Experiment

![Figure 2.4](image)

**Figure 2.4.** The format of double-quantum filtered experiment. Double-quantum coherence is excited then converted to observable single-quantum coherence.

In any system with strongly dipolar-coupled spin pairs, it is in principle possible to excite double-quantum coherences involving the pair of coupled spins. The experiments rely on the fact that excitation efficiency for the double-quantum coherences depends upon the strength of the dipolar coupling. Since we cannot observe double-quantum coherences directly, we convert any excited double-quantum coherence into observable, single-quantum coherence. This experiment is expressed as double-quantum filtered experiment. This kind of experiment can be applied to give directly-bonded carbon correlation and therefore suppress the overlapping lipid signals.

Most solid-state NMR experiments are run under magic-angle-spinning (MAS) for the purpose of high sensitivity and resolution. However, MAS also suppresses all dipolar coupling, which is essential for double-quantum coherence excitation. Therefore, ‘dipolar recovery’ methods are applied to reintroduce the dipolar coupling between the spins of interest. Here I will outline the theory behind the operation of these methods. The following treatment is referred to Solid-State NMR Spectroscopy by Duer [13] and Multidimensional Solid-State NMR and Polymers by Schmidt-Rohr and Spiess [14].

### 2.2.1 Dipolar Hamiltonian in Terms of Spherical Tensor Operators

The first-order average homonuclear dipolar Hamiltonian in terms of spherical tensor operators in the rotating frame of reference for dipolar coupling between two spins, I and J, is
Here the spin part of this Hamiltonian is given by
\[ \hat{T}_{20} = \frac{1}{\sqrt{6}} (3I_z J_z - \vec{I} \cdot \vec{J}) \].
The spatial part \( \Lambda_{20} \) is related with the dipolar coupling tensor in the laboratory frame via
\[
\Lambda_{20} = \sum_{m'=-2}^{2} D^{2}_{m'm}(\alpha_{PL}, \beta_{PL}, \gamma_{PL}) \lambda_{2m'} = D^{2}_{00}(\alpha_{PL}, \beta_{PL}, \gamma_{PL}) \lambda_{20} = \frac{1}{2} (3 \cos^2 \beta_{PL} - 1) \lambda_{20}
\]
Here \( D^{2}_{m'm}(\alpha_{PL}, \beta_{PL}, \gamma_{PL}) \) is the Wigner rotation matrix with the Euler angles \((\alpha_{PL}, \beta_{PL}, \gamma_{PL})\) specifying the coordinate transformation from the principal axis frame (PAF) to the laboratory frame and has been already well defined. Generally, \( D^{L}_{m'm}(\alpha, \beta, \gamma) \) is defined as
\[
D^{L}_{m'm}(\alpha, \beta, \gamma) = d^{L}_{m'm}(\beta) e^{-im'\alpha} e^{-im\beta}.
\]
\( d^{L}_{m'm}(\beta) \) is the ‘reduced rotational matrix’ element. Here \( d^{L}_{00}(\beta) \) is defined as
\[
d^{L}_{00}(\beta) = \frac{1}{2} (3 \cos^2 \beta - 1). \lambda_{20} \) is the dipolar coupling tensor in the principal axis frame (PAF), defined as \( \lambda_{20} = -\sqrt{6}d \), where \( d \) is the dipolar coupling constant for the IS spin pair [15,16].

**Figure 2.5.** The relationship between the laboratory frame (x, y, z) and the magic-angle rotor frame \((x_R, y_R, z_R)\). The \( z_R \) axis in the rotator frame is tilted by \( \theta_R (54.7^\circ) \) from the z axis of the laboratory frame. The \( x_R \) and \( y_R \) axis are precessing around the \( z_R \) axis with a frequency of \( \omega_r \).

We need to consider the effects of magic-angle spinning (MAS). To do this, we need to describe an extra coordinate frame, the rotor frame, which we designate \((x_R, y_R, z_R)\). To express the dipolar-coupling tensor in the rotor frame, we first find the dipolar-
coupling tensor with respect to the rotor frame, which is given by
\[ \Lambda_{2m}^R(R) = D_{0m}^2(\alpha_{PR}, \beta_{PR}, \gamma_{PR})\lambda_{20}. \]
Then we need to relate the tensor to the laboratory frame \((\alpha_{RL} = \omega_t, \beta_{RL} = \theta, \gamma_{RL} = 0)\), which can be found in a similar manner:

\[
\Lambda_{2m}^R(lab) = \sum_{m=-2}^{+2} D_{m,m}^2(\omega_t, \theta, 0)\Lambda_{2m}^R
= \sum_{m=-2}^{+2} D_{m,m}^2(\omega_t, \theta, 0)D_{0m}^2(\alpha_{PR}, \beta_{PR}, \gamma_{PR})\lambda_{20}
= \sum_{m=-2}^{+2} \exp(im\omega_t)d_{m0}^2(\theta)D_{0m}^2(\alpha_{PR}, \beta_{PR}, \gamma_{PR})\lambda_{20}
= \sum_{m=-2}^{+2} \exp(im\omega_t)\omega_{IJ}^{(m)}
\]

Here \(\omega_{IJ}^{(m)} = \sum_{m=-2}^{+2} d_{m0}^2(\theta)d_{m,m}^2(\alpha_{PR}, \beta_{PR}, \gamma_{PR})\lambda_{20}\cdot d_{m0}^2(\theta)\) is a \(\theta\)-dependent factor whose polynomial depends on the \(m\) value.

Thus the first-order average homonuclear dipolar Hamiltonian under magic-angle spinning in spherical tensor form (for a two-spin system) is

\[
\hat{H}_{dd}(t) = \hat{T}_{20} \sum_{m=-2}^{+2} \exp(im\omega_t)\omega_{IJ}^{(m)}
\]

### 2.2.2 Analysis of the C7 Pulse Sequence

![Figure 2.6](image)

**Figure 2.6.** The C7 pulse sequence includes \(n\) rf pulse cycles and is timed to fit into \(N\) rotor periods [17].
The double-quantum C7 method is an efficient way for homonuclear dipolar recoupling. It has been used successfully for double-quantum excitation. The C7 pulse sequence [17] consists of a sequence of \( n \) rf pulse cycles \( C \), which is timed to fit into \( N \) periods of the sample rotation exactly. Each rf pulse cycle consists of \((2\pi \rho, 2\pi \rho)\) where the phase \( \Phi_\rho \) is given by \( \Phi_\rho = 2\pi \rho / n \) (\( \rho = 0, 1, 2, \ldots, n-1 \)) and the length of each cycle is \( \tau_c = N\tau_\rho / n \) [18].

If the pulse sequence starts at time 0, and the pulse cycle \( C_\rho \) starts at time \( t_\rho^0 \), then a time \( \tau \) into the cycle, the dipolar Hamiltonian is given by

\[
\hat{H}_{dd}(t_\rho^0 + \tau) = \hat{T}_{20} \sum_{m=-2}^{2} \exp(i\omega_r(t_\rho^0 + \tau))\omega^{(m)}_{IS}
\]

\[
= \hat{T}_{20} \sum_{m=-2}^{2} \exp(i\omega_r N\tau_\rho / n) \exp(i\omega_r \tau)\omega^{(m)}_{IS}
\]

\[
= \hat{T}_{20} \sum_{m=-2}^{2} \exp(i2\pi m\rho / n) \exp(i\omega_r \tau)\omega^{(m)}_{IS}
\]

Where we have \( t_\rho^0 = N\tau_\rho / n \), since each cycle is of length \( N\tau_\rho / n \).

Figure 2.7. The relationship between the rotating frame (\( x, y, z \)) and the toggling frame (\( x_{\log}, y_{\log}, z_{\log} \)) employed in the analysis of the C7 pulse sequence. A rotation of \( -(\pi/2 - \Phi_\rho) \) about \( z \) brings \( y \) into coincidence with the pulse axis in the rotating frame. A rotation of \( -\beta_{rf} \) about the pulse axis.
generates the $z_{\text{tog}}$ of the toggling frame. Finally, a rotation of $(\pi/2 + \Phi_p)$ about $z_{\text{tog}}$ generates the $x_{\text{tog}}$ and $y_{\text{tog}}$ axes of the toggling frame.

Now we transform the spin coordinates in the Hamiltonian describing the spin system to a toggling frame in which the effect of the rf pulse vanishes during the pulse cycle. This frame is one which is rotated from the normal rotating frame by $-\beta_{\text{rf}}(\tau)$ about the axis of the pulse, where $\beta_{\text{rf}}$ is the nutation angle of the pulse, $\omega_1 \tau$. The axis of the pulse is in the x-y plane of the rotating frame at an angle $\Phi_p$ from the x-axis. The rotating frame is therefore rotated into the toggling frame via the Euler angles $(-((\pi/2) - \Phi_p), -\beta_{\text{rf}}(\tau), (\pi/2) - \Phi_p))$. So the dipolar Hamiltonian at time $\tau$ into the pulse cycle expressed in the new toggling frame is $\hat{H}^*(t^0_{\text{p}} + \tau)$, given by

$$\hat{H}^*(t^0_{\text{p}} + \tau) = \hat{R}^{-1}(-((\pi/2) - \Phi_p), -\beta_{\text{rf}}(\tau), (\pi/2) - \Phi_p)) \hat{H}_{dd}(t^0_{\text{p}} + \tau)$$

$$\times \hat{R}(-((\pi/2) - \Phi_p), -\beta_{\text{rf}}(\tau), (\pi/2) - \Phi_p))$$

The rotation operator $R$ is the operator which rotates the spin coordinate rotating frame into the toggling frame [13]. It acts on the spin operators of the dipolar Hamiltonian. For this case

$$\hat{R}^{-1} \hat{T}_{2q} \hat{R} = \sum_{q=-2}^{+2} D_{q0}^2(-((\pi/2) - \Phi_p), -\beta_{\text{rf}}(\tau), (\pi/2) - \Phi_p)) \hat{T}_{2q}$$

$$= \sum_{q=-2}^{+2} \exp(i q (\pi/2 - \Phi_p)) d_{q0}^2(-\beta_{\text{rf}}(\tau)) \hat{T}_{2q}$$

$$= \sum_{q=-2}^{+2} i^q \exp(i q 2\pi p / n) d_{q0}^2(-\beta_{\text{rf}}(\tau)) \hat{T}_{2q}$$

Thus, the Hamiltonian in the toggling frame can be rewritten as

$$\hat{H}^*(t^0_{\text{p}} + \tau) = \sum_{q=-2}^{+2} \sum_{m=-2}^{+2} \exp(i 2\pi (Nm - q) p / n) \omega_{qm}(\tau) \hat{T}_{2q}$$

Where $\omega_{qm}(\tau)$ is given by
\[ \omega_{qm}(\tau) = i^q \exp(i m \omega \tau) d_q^2(-\beta_{rf}(\tau)) \omega_{IS}^m \]

The average Hamiltonian to first order over the \( p \)th pulse cycle can be formed:

\[
\hat{H}_p^{(0)} = \sum_{q=-2}^{2} \sum_{m=-2}^{2} \exp(i 2\pi (Nm - q) p / n) \bar{\omega}_{qm} \hat{T}_{2q}^{IS}
\]

Where

\[
\bar{\omega}_{qm} = \frac{1}{\tau_c} \int_0^{\tau_c} \omega_{qm}(\tau) d\tau
\]

The average Hamiltonian over the whole pulse sequence is then found by averaging the above equation over all the \( n \) cycles in the pulse sequence:

\[
\hat{H}^{(0)} = \frac{1}{n} \sum_p \hat{H}_p^{(0)}
\]

\[
= \frac{1}{n} \sum_p \sum_{q=-2}^{2} \sum_{m=-2}^{2} \exp(i 2\pi (Nm - q) p / n) \bar{\omega}_{qm} \hat{T}_{2q}^{IS}^n \]

\[
= \sum_{q=-2}^{2} \sum_{m=-2}^{2} \bar{\omega}_{qm} \hat{T}_{2q}^{IS} \delta(Nm - q = n\mu)
\]

Here \( \mu \) is an integer. Therefore, to reintroduce the dipolar coupling interaction, \( \delta(Nm - q = n\mu) \) needs to be 1, which means \( Nm - q = n\mu \). And we also wish to 1) Excite double-quantum coherence with the pulse sequence. For that we need to retain \( \hat{T}_{2\pm 2} \) terms in the average Hamiltonian, which makes \( q = \pm 2 \). 2) Suppress all remaining chemical shift terms, then \( q \neq \pm 1 \). 3) Each double-quantum term \( q \) is associated with one spatial component \( m \). If we have \((N, n) = (2, 4)\), then \((m, q, \mu)\) can be \((1, 0, 2), (-1, -1, 2)\) etc. Since \( q = 2 \) couples to both \( m = 1 \) and \( m = -1 \), which does not satisfy condition 3), \((N, n) = (2, 4)\) would not work. If \((N, n) = (2, 5)\), then \((m, q, \mu)\) can be \((1, 2, 0), (-1, -2, 0), (2, -1, 1)\) and \((-2, 1, -1)\). Note here we have \( q = \pm 1 \), which means we still have the chemical shift terms. So this condition does not work as well. To satisfy all three conditions, it turns out the simplest sequence is \( n = 7 \) and \( N = 2 \), while with \((m, q, \mu)\) to be \((1, 2, 0)\) and \((-1, -2, 0)\). The \( C_7^2 \) sequence therefore generates double-quantum coherence in a dipolar coupled
spin pair. The pulse length of each $2\pi$ rf pulse in the cycle is $\frac{N\tau_r}{2n} = \frac{\tau_r}{7}$. Thus the field strength $\omega_1 = \frac{7 \cdot 2\pi}{\tau_r} = 7\omega_r$.

2.2.3 Analysis of the SPC-5 pulse sequence

Figure 2.8. The SPC-5 pulse sequence.

The SPC-5 sequence consists of two C7 type cycles with an overall phase shift of $\Psi[19]$. The Euler angle between the rotating frame and the toggling frame of the second pulse cycle is $(-\frac{\pi}{2} - \Phi_p + \Psi), \beta_{rf}(\tau), (\pi/2 - \Phi_p + \Psi))$. Thus the total Hamiltonian for each interaction (chemical shift, dipolar coupling, etc.) is

$$
\vec{H}^{(0)} = \frac{1}{2n} \sum_{p=0}^{n-1} \sum_{q=-2}^{2} \sum_{m=-2}^{2} (1 + \exp(-iq\Psi)) \exp(i2\pi(Nm - q)p/n)\vec{a}_{qm}^{IS}\vec{T}_{2q}^{IS}
$$

When $\Psi = \pi$ and $q = \pm 1$, the $(1 + \exp(-iq\Psi))$ term is zero, which means when $\Psi = \pi$, all the $q = \pm 1$ terms are eliminated. This provides a wider choice of the $n$ and $N$ number. For the C7 type cycles, when $N = 2$ and $n = 5$, there are 4 ($q$, $m$) combinations: $(2, 1), (-2, -1), (-1, 2)$ and $(1, -2)$. Since we want to suppress the $q = \pm 1$ terms, this combination does not work. But for the SPC-5 cycles, the $q = \pm 1$ terms are eliminated by the overall phase shift of $\pi$, which makes the $N = 2$ and $n = 5$ a good combination for double quantum excitation.

As a result, the length of the $2\pi$ pulses in the SPC-5 sequence is $\frac{N\tau_r}{2n} = \frac{\tau_r}{5}$ and the field strength is $\omega_1 = \frac{5 \cdot 2\pi}{\tau_r} = 5\omega_r$. Therefore, if one is using 5 kHz MAS frequency, the radio frequency power needs to be 35 kHz for C7 sequence while for SPC-5 sequence, it only needs to
be 25 kHz. In the real condition, the optimized radio frequency power is \(~2\) dB softer than the calculated value.

### 2.3 Dipolar Correlation Experiments for Measuring Backbone (φ, ψ) Torsion Angles

In recent years solid-state nuclear magnetic resonance has proven crucial for investigating the structure of macroscopically disordered biological solids, especially for those inaccessible to the more traditional methods such as X-ray crystallography and solution NMR. While people are still trying to make progress to improve both the accuracy and range of the internuclear distance measuring techniques, some alternative experimental approaches can also be applied to increase structural resolution. For example, besides distance restraints, the (φ, ψ) torsion angles, which give secondary structure information, can also be used as a powerful approach for structure determination. People commonly use TALOS program to generate empirical (φ, ψ) angle prediction from chemical shifts. The direct measurement of the (φ, ψ) angles can increase the accuracy of this kind of restraints.

#### 2.3.1 HNCH

The φ angles can be measured using the HNCH technique, which correlates the \(^{1}H\text{N}^{15}N\) and \(^{13}C\alpha^{1}H\alpha\) dipolar couplings. This experiment yields \(^{1}H\text{N}-N-C\alpha-H\alpha\) angle (φ\(_{H}\)), which is related to the φ-angle according to φ = φ\(_{H}\) + 60º [20].

![HNCH pulse sequence](image)

**Figure 2.9.** (a) HNCH pulse sequence. (b) Simulated HNCH dephasing curves for φ\(_{H}\) from 180º to 120º with \(\omega_r/2\pi = 8\) kHz.
Figure 2.8 shows the pulse sequence for HNCH. After the transverse $^{13}$C magnetization is created by cross polarization, it evolves under the influence of the recoupled $^{15}$N-$^{13}$C dipolar interaction for a period $\tau$. A series of rotor-synchronized $^{15}$N and $^{13}$C 180° pulses was implemented after CP to recouple the N-C dipolar coupling. These pulses perturb MAS averaging of the anisotropic dipolar interaction so that antiphase magnetization $C_yN_z$ is created. Then a $^{15}$N 90° pulse converts $C_yN_z$ into a combination of double- and zero-quantum coherences, $C_yN_x$. Next, the $^1$H homonuclear dipolar coupling is removed by a MREV-8 pulse train, which defines the evolution time $t_1$, which has a maximum length of one rotor period. During this time, the magnetization evolves under the N-H and C-H dipolar coupling interaction. A pair of 180° pulses is placed at the center of this constant time period ($\tau_{\text{const}}$) to refocus the $^{15}$N and $^{13}$C chemical shifts. At the end of the $\tau_{\text{const}}$, a second $^{15}$N 90° pulse and N-C recoupling period reconvert the N-C multiple-quantum coherences into observable $^{13}$C magnetization, which is then detected. The transformations of the spin density operators under the pulse sequence are outlined below.

$$
C_x \rightarrow C_yN_z f(\tau) \xrightarrow{90^\circ - 15N} C_yN_x f(\tau) \xrightarrow{\tau} C_yN_x f(\tau) \cos(\Psi_{CH}(t_1)) \cos(\Psi_{NH}(t_1))
$$

$$
\xrightarrow{90^\circ - 15N} C_yN_z f(\tau) \cos(\Psi_{CH}(t_1)) \cos(\Psi_{NH}(t_1)) \xrightarrow{\tau} -C_x f^2(\tau) \cos(\Psi_{CH}(t_1)) \cos(\Psi_{NH}(t_1))
$$

Here $\Psi_{CH}(t_1)$ and $\Psi_{NH}(t_1)$ are the MAS phase angles accumulated during $t_1$ due to the C-H and N-H dipolar interactions. They depend on the relative orientations of the two dipolar coupling tensors, whose unique axes are along the Cα-Hα and N-HN bonds and therefore reflect the torsion angle $\Phi_H$.

### 2.3.2 NCCN
Figure 2.10. (a) NCCN pulse sequence. (b) Simulated NCCN dephasing curves for $\psi$ from 180° to 120°.

NCCN is a robust method for measuring the backbone $\psi$ torsion angles of proteins, which correlates the $^{15}N_i-^{13}C\alpha_i$ and $^{13}CO_i-^{15}N_{i+1}$ dipolar couplings to obtain the relative bond orientation [21,22]. $^{13}C\alpha$-$^{13}CO$ double quantum coherence was excited using the SPC-5 sequence [23]. After that, the $^{15}N$-$^{13}C$ heteronuclear dipolar couplings were recoupled by a REDOR sequence [24], which yield a double quantum coherence dephasing curve whose shape provides information about the relative orientation of the $^{15}N$-$^{13}C$ internuclear vectors, and hence the $^{13}C$-$^{13}C$ torsion angle.

References


Chapter 3

Resonance Assignment and Three-Dimensional Structure
Determination of a Human Alpha-Defensin, HNP-1, by Solid-State NMR


2010, vol. 397, pp. 408-422

Yuan Zhang, Tim Doherty, Jing Li, Wuyuan Lu, Cyril Barinka, Jacek Lubkowski, and Mei Hong

Abstract

Human α-defensins (HNPs) are immune defense mini-proteins that act by disrupting microbial cell membranes. Elucidating the three-dimensional structures of HNPs in lipid membranes is important for understanding their mechanisms of action. Using solid-state NMR, we have determined the three-dimensional structure of HNP-1 in a microcrystalline state outside the lipid membrane, which provides benchmarks for structure determination and comparison with the membrane-bound state. From a suite of 2D and 3D magic-angle spinning experiments, $^{13}$C and $^{15}$N chemical shifts were obtained that yielded torsion angle constraints while inter-residue distances were obtained to restrain the three-dimensional fold. Together, these constraints led to the first high-resolution SSNMR structure of a human defensin. The SSNMR structure has close similarity to the crystal structures of the HNP family, with the exception of the loop region between the first and second β-strands. The difference, which is partially validated by direct torsion angle measurements of selected loop residues, suggests possible
Introduction

Antimicrobial peptides (AMPs) are small cationic peptides that constitute part of the innate immune system of many plants and animals to rapidly kill invading microbial pathogens. In mammals the two main classes of antimicrobial peptides are cathelicidins and defensins [1]. While most cathelicidins are α-helical in structure, defensins contain β-strands stabilized by three disulfide bonds [2-4]. Based on the disulfide-linkage patterns, three classes of defensins, α-, β-, and θ-defensins, are identified in vertebrates. Humans have six α-defensins that contain 29-32 residues. Four of these proteins are present in the azurophilic granules of neutrophils and are called human neutrophil peptides 1-4 (HNP 1-4) [2,5], and two are expressed in intestinal Paneth cells and are called HD-5 and HD-6 [6,7]. Except for HD-6, all five human α-defensins have wide-spectrum antimicrobial activities with LD<sub>50</sub> in the µg/ml range [8,9]. Similar to most other AMPs, the main mechanism of action of HNPs is believed to be permeabilization of the microbial cell membrane [10,11].

The presence of three antiparallel disulfide-stabilized β-strands makes defensins larger and more complex proteins than most other AMPs. In particular, they are natural extensions of the two-stranded β-hairpin AMPs such as protegrins from porcine leukocytes and tachyplesins from horseshoe crabs, whose membrane-bound structure and lipid interaction have been characterized in detail by solid-state NMR recently [12-21]. Compared to these smaller β-sheet AMPs, HNPs have weaker antimicrobial activities. Moreover, although HNP 1-3 have nearly identical sequences, differing only in the N-terminus residue, their activities differ by as much as 4-fold [8]. Understanding the activity differences among human defensins requires structural investigations in the lipid membrane. However, so far most high-resolution structures of human defensins come from X-ray crystallography in the absence of any membrane-mimetic solvents [3,4]. A few biophysical studies that investigated the interactions of defensins with lipid conformational variation and flexibility of this segment of the protein, which may regulate HNP interaction with the phospholipid membrane of microbial cells.
membranes have also been reported, but did not contain high-resolution structural information [22-24].

Solid-state NMR magic-angle spinning (MAS) NMR has been recently shown to be able to elucidate the atomic-resolution structure of $^{13}$C, $^{15}$N-labeled proteins in microcrystalline states [25-28], in fibrillar forms [29-31], and in lipid membranes [32,33]. A wide variety of 2D and 3D correlation techniques have been developed to obtain resonance assignment [34,35], torsion angles [36-38] and inter-atomic distances [39]. Therefore, the technology is now in place to determine the full structures of human $\alpha$-defensins in lipid membranes to elucidate the mechanism of action of this important class of immune defense molecules.

![Amino acid sequence and disulfide bond connectivities of HNP-1.](image)

**Figure 3.1.** Amino acid sequence and disulfide bond connectivities of HNP-1.

In this work, we report the three-dimensional structure determination of recombinant HNP-1, whose sequence is shown in **Figure 3.1**, by SSNMR. We have carried out resonance assignment and internuclear distance measurements of HNP-1 in an ordered microcrystalline state, in preparation for structure determination of the protein bound to the lipid membrane, since membrane-bound proteins usually give lower-resolution spectra due to the disorder created by the lipid bilayer. The conformation of microcrystalline HNP-1 also represents the protein structure before binding to the lipid membrane, and is thus necessary for understanding potential conformational changes of HNPs induced by the lipid bilayer. We compare the SSNMR structure with various published crystal structures of HNPs [3,4,40].

**Results**

*Preparation of microcrystalline HNP-1 samples for solid-state NMR experiments*
Figure 3.2. An image of microcrystalline HNP-1. The average size of the crystals is \(\sim 20\ \mu\text{m}\).

The solid-state NMR data presented here were obtained from microcrystalline samples of HNP-1 precipitated from a polyethylene glycol (PEG) solution. Two samples were prepared using different concentrations of the protein stock solution. The high-concentration stock solution (36 mg/ml) yielded microcrystals one hour after mixing with PEG, while the low-concentration stock solution (30 mg/ml) produced microcrystals gradually in a 4-day period, with \(\sim 80\%\) of the protein eventually precipitating from the supernatant. The slower precipitation produced larger crystals of \(\sim 20\ \mu\text{m}\) diameter compared to 2 \(\mu\text{m}\) for the rapidly precipitated sample. Figure 3.2 shows an image of HNP-1 microcrystals obtained from the slower precipitation procedure.
Both microcrystalline HNP-1 samples yielded well-resolved $^{13}$C and $^{15}$N MAS spectra, shown in Figure 3.3. For the rapidly precipitated sample in the 4 mm rotor, the $^{13}$C full widths at half maxima were 0.6 – 1.0 ppm and the $^{15}$N linewidths were 1.3 – 1.7 ppm. For the slowly precipitated sample in the 2.5 mm rotor, the linewidths improved to 0.4 – 0.8 ppm for $^{13}$C and 1.0 – 1.5 ppm for $^{15}$N. Faster spinning frequencies and stronger $^1$H decoupling fields further contributed to the narrower linewidths of the 2.5 mm rotor sample.

Resonance assignment of HNP-1 by 2D and 3D SSNMR
Figure 3.4. Representative regions of 2D $^{13}$C-$^{13}$C correlation spectra of HNP-1. (a) Ala Cα-Cβ region from a 40 ms DARR spectrum. (b) Cys Cα-Cβ region from a 40 ms DARR spectrum. (c) Cys Cα-Cβ region from a 0.8 ms CM$_5$RR spectrum. Blue: positive intensities. Red: negative intensities. (d-f) CO-Cα/Cβ region. (d) 40 ms DARR spectrum, (e) 0.8 ms CM$_5$RR spectrum. (f) 1.5 ms CM$_5$RR. The sign of the cross peak intensities distinguishes the one-bond Gly CO-Cα peaks (red, indicating negative) from the two-bond Cys CO-Cβ peaks (blue, indicating positive).

We first assigned the amino acid spin systems of HNP-1 using 2D $^{13}$C-$^{13}$C correlation experiments, including a DARR experiment with 40 ms mixing, and double-quantum (DQ)-filtered experiments using SPC5 [41] and CM$_5$RR sequences [42]. The 40 ms DARR spectrum allowed the assignment of most amino acid types (Figure 3.4a, b, d), including Cys, Ala, Ile, Arg, Leu, Pro, and Thr, which have distinct cross peak patterns.
The aromatic residues were assigned based on the Cα/Cβ cross peaks to the aromatic carbons. For example, the Cγ of the unique tryptophan W27 resonates at 107.8 ppm (Figure S1), which allowed the rest of the W27 spin system to be identified. For the unique glutamate E14 and glutamine Q23, the sidechain Cδ chemical shifts (181 ppm for E14 and 176 ppm for Q23) served as useful guides for identifying the chemical shifts of the other sidechain carbons.

The DQ-filtered 2D spectra confirmed the DARR assignment by exhibiting opposite intensities for one-bond and two-bond cross peaks. For example, Gly Cα and Cys Cβ, which both resonate at 40-45 ppm, are distinguished based on the negative intensities of the one-bond Gly Cα-CO peaks and the positive intensities of the two-bond Cys Cβ-CO peaks in the CM3RR spectra (Figure 3.4c, e, f). Similarly, while Q23 Cβ has similar chemical shifts to Arg Cγ in the DARR spectrum, in the DQ filtered spectra Q23 Cβ exhibits a positive one-bond Cα-Cβ cross peak, which is distinct from the negative two-bond Arg Cα-Cγ cross peaks. By comparing various 2D 13C-13C spectra, we assigned all amino acid spin systems except for one Arg.

**Figure 3.5.** F2-F3 strips of 3D NCACX (blue) and NCOCX (red) spectra for residues G18 to Q23, illustrating sequential resonance assignment. The strip width for the F2 dimension is 2.3 ppm. The F1 15N chemical shifts for the strips are indicated on the right end of each strip.
Next, we obtained sequence-specific assignment from inter-residue NCOCX and intra-residue NCACX 3D experiments [34,35]. Figure 3.5 shows a representative strip of F2-F3 planes of the two 3D experiments for residues G18-Q23. We first identified the $^{15}$N chemical shift of residue $i$ in the NCACX spectrum by comparing the $^{13}$C chemical shifts with those obtained from the 2D CC spectra. The $^{15}$N chemical shift of residue $i+1$ was then obtained from the NCOCX spectrum through its cross peaks with the $^{13}$C chemical shifts of residue $i$. The $^{15}$N plane of residue $i+1$ was then searched in the NCACX spectrum to identify the $^{13}$C chemical shifts of residue $i+1$.

A2, A9, T19 and C30 are clear starting points for assigning a consecutive segment of residues. The N-terminal residue A2 has a distinct $^{15}$N chemical shift of 41.7 ppm, which is folded to 106.9 ppm. Residue A9 has cross peaks with P8, which has characteristic $^{13}$C chemical shifts. Residue T19 has characteristic $\text{C}_\alpha$ and $\text{C}_\beta$ chemical shifts.

The four Ala residues, which are well resolved in the 2D CC spectra, were sequentially assigned in the 3D spectra. For example, A9 was assigned based on its correlation with the unique P8 and with C10, and A12 was assigned based on its correlation with the well defined I11 and G13. For the three Gly residues, G24 was identified by inter-residue correlations Q23N-G24$\text{C}_\alpha$ and R25N-G24$\text{C}_\alpha$ in the 3D spectra and in the 100 ms DARR spectrum, while G18 was assigned based on the T19N-G18$\text{C}_\alpha$ correlations in the 3D spectrum.

The six Cys residues were resolved and assigned based on their cross peaks with their neighboring residues in the NCOCX spectrum. For example, C3 has characteristic cross peaks with A2 $\text{C}_\beta$ and $\text{C}_\alpha$, and C5 has cross peaks with the Y4 spin system. Interestingly, C5 exhibits two $\text{C}_\beta$ chemical shifts differing by 0.9 ppm, suggesting the presence of two conformations at this site. C10 has a $\text{C}_\beta$ chemical shift of 33.4 ppm (on the TMS scale), which is slightly low for an oxidized cysteine. However, its assignment was corroborated by multiple sequential C10 – A9 cross peaks and I11 – C10 cross peaks. For example, C10N-A9C’-A9$\text{C}_\alpha$/C$\beta$ cross peaks were identified in the 3D NCOCX
spectrum, and sequential C10Cα-A9Cβ, C10Cβ-A9Cβ, and A9Cα-C10Cβ cross peaks were also detected in the 100 ms DARR spectrum (Figure S2).

Table 3.1. 13C and 15N chemical shifts of microcrystalline HNP-1 without lipids.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Cα</th>
<th>Cβ</th>
<th>Cγ</th>
<th>Cδ</th>
<th>Cε</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>41.7</td>
<td>49.5</td>
<td>17.5</td>
<td></td>
<td></td>
<td></td>
<td>171.0</td>
</tr>
<tr>
<td>C3</td>
<td>117.8</td>
<td>53.0</td>
<td>46.6</td>
<td></td>
<td></td>
<td></td>
<td>171.4</td>
</tr>
<tr>
<td>Y4</td>
<td>117.6</td>
<td>54.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>172.6</td>
</tr>
<tr>
<td>C5</td>
<td>130.6</td>
<td>52.7</td>
<td>44.5</td>
<td>43.4</td>
<td></td>
<td></td>
<td>171.4</td>
</tr>
<tr>
<td>R6</td>
<td>123.1</td>
<td>53.9</td>
<td>33.7</td>
<td>26.7</td>
<td>41.5</td>
<td></td>
<td>173.6</td>
</tr>
<tr>
<td>I7</td>
<td>122.3</td>
<td>56.5</td>
<td>41.2</td>
<td>28.0</td>
<td>15.3</td>
<td>10.1</td>
<td>172.6</td>
</tr>
<tr>
<td>P8</td>
<td>134.4</td>
<td>62.2</td>
<td>32.9</td>
<td>23.1</td>
<td>22.1</td>
<td>47.5</td>
<td>172.7</td>
</tr>
<tr>
<td>A9</td>
<td>122.6</td>
<td>48.8</td>
<td>20.6</td>
<td></td>
<td></td>
<td></td>
<td>172.9</td>
</tr>
<tr>
<td>C10</td>
<td>115.6</td>
<td>50.8</td>
<td>33.4</td>
<td></td>
<td></td>
<td></td>
<td>173.0</td>
</tr>
<tr>
<td>I11</td>
<td>115.4</td>
<td>57.9</td>
<td>37.5</td>
<td>23.8</td>
<td>15.9</td>
<td>12.2</td>
<td>173.0</td>
</tr>
<tr>
<td>A12</td>
<td>121.9</td>
<td>51.8</td>
<td>16.6</td>
<td></td>
<td></td>
<td></td>
<td>175.2</td>
</tr>
<tr>
<td>G13</td>
<td>109.6</td>
<td>42.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E14</td>
<td>121.3</td>
<td>52.2</td>
<td>31.4</td>
<td>34.4</td>
<td>181.8</td>
<td></td>
<td>173.6</td>
</tr>
<tr>
<td>R15</td>
<td>130.6</td>
<td>52.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>171.4</td>
</tr>
<tr>
<td>R16</td>
<td>122.8</td>
<td>54.5</td>
<td>28.1</td>
<td>27.0</td>
<td>42.4</td>
<td>157.1</td>
<td>175.3</td>
</tr>
<tr>
<td>Y17</td>
<td>128.1</td>
<td>56.3</td>
<td>38.5</td>
<td>131.5</td>
<td></td>
<td></td>
<td>173.5</td>
</tr>
<tr>
<td>G18</td>
<td>107.3</td>
<td>45.0</td>
<td></td>
<td>69.6</td>
<td>21.7</td>
<td></td>
<td>171.9</td>
</tr>
<tr>
<td>T19</td>
<td>124.2</td>
<td>59.4</td>
<td>69.6</td>
<td></td>
<td>21.7</td>
<td></td>
<td>171.7</td>
</tr>
<tr>
<td>C20</td>
<td>122.3</td>
<td>51.6</td>
<td>43.0</td>
<td></td>
<td></td>
<td></td>
<td>171.8</td>
</tr>
<tr>
<td>I21</td>
<td>123.6</td>
<td>59.5</td>
<td>35.8</td>
<td>24.3</td>
<td>16.2</td>
<td>11.4</td>
<td>173.5</td>
</tr>
<tr>
<td>Y22</td>
<td>125.5</td>
<td>54.8</td>
<td>37.7</td>
<td></td>
<td></td>
<td></td>
<td>172.6</td>
</tr>
<tr>
<td>Q23</td>
<td>129.0</td>
<td>54.2</td>
<td>23.8</td>
<td>31.9</td>
<td>176.5</td>
<td></td>
<td>173.9</td>
</tr>
<tr>
<td>G24</td>
<td>103.4</td>
<td>44.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>170.3</td>
</tr>
<tr>
<td>R25</td>
<td>118.4</td>
<td>51.6</td>
<td>32.2</td>
<td>25.5</td>
<td>41.2</td>
<td></td>
<td>172.0</td>
</tr>
<tr>
<td>L26</td>
<td>120.2</td>
<td>51.6</td>
<td>41.2</td>
<td>25.4</td>
<td>23.2</td>
<td>22.4</td>
<td>174.4</td>
</tr>
<tr>
<td>W27</td>
<td>120.7</td>
<td>52.0</td>
<td>29.0</td>
<td></td>
<td></td>
<td></td>
<td>175.3</td>
</tr>
<tr>
<td>A28</td>
<td>127.1</td>
<td>51.0</td>
<td>16.9</td>
<td></td>
<td></td>
<td></td>
<td>172.8</td>
</tr>
<tr>
<td>F29</td>
<td>123.8</td>
<td>56.6</td>
<td>39.8</td>
<td>129.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C30</td>
<td>121.9</td>
<td>53.0</td>
<td>38.6</td>
<td></td>
<td></td>
<td></td>
<td>172.7</td>
</tr>
<tr>
<td>C31</td>
<td>121.1</td>
<td>55.3</td>
<td>47.0</td>
<td></td>
<td></td>
<td></td>
<td>177.5</td>
</tr>
</tbody>
</table>

* The 13C chemical shifts are referenced to TMS. To convert to DSS-referenced chemical shifts, the 13C frequencies in this table should be increased by 1.7 ppm. The 15N chemical shifts are referenced to liquid ammonia.

* The partner cysteine involved in the disulfide bond is indicated in the bracket.
The complete $^{13}$C and $^{15}$N isotropic chemical shifts of HNP-1 are summarized in Table 3.1. Among the six cysteines, the C10 – C30 pair has relatively upfield C$\beta$ chemical shifts. To further verify that C10 and C30 are indeed disulfide-bonded, we compared the C$\alpha$ and C$\beta$ chemical shifts of all HNP-1 cysteines with literature statistical analyses of the dependence of cysteine $^{13}$C chemical shifts on the oxidation state and conformation [43,44]. On the DSS scale, the oxidized cysteines have C$\beta$ chemical shifts of 33 - 51 ppm and C$\alpha$ chemical shifts of 50 - 61 ppm. When the HNP-1 cysteine chemical shifts are also referenced to DSS (thus increasing the $^{13}$C chemical shifts in Table 3.1 by 1.7 ppm), we find that all six cysteines, including C10 and C30, lie in the oxidized range (Figure S3), thus confirming the disulfide-bonded nature of all three pairs.

*Backbone dihedral angles and verification by dipolar correlation experiments*

![Figure 3.6](image)

**Figure 3.6.** HNP-1 ($\phi$, $\psi$) torsion angles determined from solid-state NMR $^{13}$C and $^{15}$N chemical shifts (filled squares). The torsion angles of HNP-3 are shown for comparison (open squares and ribbons) (PDB accession code: 1DFN). Residues with significantly different torsion angles from the crystal structure values of HNP-3 are shaded.
The $^{13}$C and $^{15}$N chemical shifts allowed us to determine the backbone ($\phi$, $\psi$) torsion angles of HNP-1 using the well-known empirical relation between N, C$\alpha$, C$\beta$, CO chemical shifts and protein secondary structure [45]. We used the TALOS program [46] to predict the ($\phi$, $\psi$) angles and compared them with the values found from the crystal structure of HNP-3 [3] (PDB code: 1DFN). The chemical-shift constrained torsion angles indicate three well-defined $\beta$-strands, $\beta_1$, $\beta_2$ and $\beta_3$, separated by a type-II $\beta$-turn at A12-G13 and a type I’ $\beta$-turn at Q23-G24 (Figure 3.6). These secondary structure motifs are consistent with those of the HNP-3 structure, although the boundaries of the loop or turn segments between $\beta$-strands differ slightly between HNP-3 and HNP-1.

Among the 28 ($\phi$, $\psi$) pairs predicted by TALOS, excluding the terminal A2 and C31, 24 $\phi$ angles and 24 $\psi$ angles agree well with the crystal structure values within the experimental uncertainty. Residues with large dihedral angle discrepancies are P8, C10, Y17, G18 and Y22 (Figure 3.6 and Table S1). P8 and C10 both lie in the loop between $\beta_1$ and $\beta_2$ strands and might be expected to be dynamically disordered. However, the HNP-3 crystal structure shows well-defined positions for these residues, with B factors not higher than the average [3]. Direct measurement of C-H order parameters of the HNP-1 sample by NMR (see Figure 3.11 below) also indicates that the loop region in HNP-1 is not dynamic.

To verify the torsion angle differences between TALOS and the crystal structure, we conducted the HNCH experiment to directly measure the $\phi$ angles of C10 and Y22. The experiment measures the $\phi_H$ angle (defined as $H^N$-N-C$\alpha$-H$\alpha$, which is related to $\phi$ by $\phi_H+60^\circ=\phi$) by correlating the N-H and C$\alpha$-H$\alpha$ dipolar couplings [36,47], and is particularly sensitive to the $\beta$-sheet conformation, which has $\phi_H$ angles around 180$^\circ$. To maximize the angular resolution, we implemented the N-H doubled version of the experiment [47]. In principle, a 3D experiment where two chemical shift dimensions provide site resolution while one dipolar dimension gives the angular information is the most desirable. However, the low sensitivity of the 3D experiment and the limited protein amount precluded its implementation. Thus, we used one $^{13}$C chemical shift dimension for site resolution and measured the dipolar evolution at time 0 and half a rotor period.
Simulations have shown that the HNCH dipolar dephasing at the middle of the rotor period is the most sensitive to $\phi_1$ angle differences [48].
Figure 3.7. Direct measurement of φ torsion angles by HNCH. (a) Aliphatic region of the 1D $^{13}$C spectrum. (b) Cα region (44-64 ppm) of the $^{13}$C spectrum, indicating the assignment of the resolved peaks and the $\phi_H$ torsion angles ($\phi_H+60^\circ=\phi$) of each residue. For most residues the $\phi_H$ angles agree between TALOS and the crystal structure within ±10°. But for C10 and Y22, significant deviations exist. (c) Measured HNCH spectra with dipolar evolution time of 0 ($S_0$) and half a rotor period (S, shown in blue). The measured $S/S_0$ values are indicated along with calculated values. The 50.8-ppm peak has an $S/S_0$ value that is consistent with the chemical shift derived φ angles and not the HNP-3 value. (d) Calculated HNCH curves for the 50.8 ppm C10/A28 Cα peak. The A28 $\phi_H$ angle is fixed while the C10 $\phi_H$ angle is either 173° or 132°. The measured HNCH intensity at the middle of the rotor period is consistent with the TALOS $\phi_H$ angle but not the crystal structure value.

Figure 3.7c shows the Cα region of the $^{13}$C spectra with 0 and half a rotor period of HNCH evolution. The peaks were assigned based on the 2D and 3D spectra above. Although single-site resolution is not available, most sites have consistent φ angles between the TALOS prediction and the crystal structure within ±10° (Table S1). The only uncertain residues are C10 and Y22, thus their $\phi_H$ angles can be extracted from the intensities. Specifically, the C10 Cα signal at 50.8 ppm overlaps with only one residue, A28, whose $\phi_H$ angle is -145°±5°, as indicated by both TALOS and the crystal structure (Figure 3.7b). Thus, the calculated HNCH curve for 50.8 ppm is very sensitive to the C10 $\phi_H$ angle (Figure 3.7d). The experimental intensity (0.53) is much more consistent with the TALOS prediction of a β-sheet like $\phi_H$ angle of about 180° rather than a low value of about -130° (Figure S4). Similarly, the measured HNCH intensity for the 54.5-ppm peak, which is a composite of Y22 Cα with five other Cα sites, is more consistent with the TALOS prediction than with the crystal structure, although the distinction between the two possibilities is smaller for this peak due to more resonances overlapping at this position.

For P8, both φ and ψ angles differ between the chemical shift prediction and the crystal structure (Table S1). The lack of $H^N$ at Pro precludes direct measurement of its φ angle. In principle, the ψ angle can be measured using the NCCN technique [37]. At
present the protein amount is insufficient for this experiment, which requires $^{13}$C double-quantum filtration in addition to $^{13}$C-$^{15}$N REDOR. Thus, future experiments are necessary to clarify the conformational discrepancy at P8. Nevertheless, based on distance constraints involving this loop, we hypothesize that the difference may be real instead of due to inaccuracy of the NMR structure determination (see below).

The torsion angle difference for Y17-G18 is also noteworthy. TALOS predicted a relatively ideal $\beta$-strand $\psi$ angle for Y17 whereas the crystal structure gives an unusual $\psi$ angle near 0°. On the other hand, the TALOS Y17 torsion angles have relatively large uncertainties ($\Delta \phi=23^\circ$, $\Delta \psi=20^\circ$) (Table S1). Among the 10 pairs of torsion angles predicted, seven pairs lie in the $\beta$-sheet region while three pairs lie near the $\alpha$-helical region, with $\psi$ angles of about -20°, which are similar to the crystal structure $\psi$ value. Thus, the ensemble of NMR torsion angles for Y17 encompasses the crystal structure value.

For G18, chemical shifts predicted a $\phi$ angle of -120° whereas the crystal structure shows a $\phi$ angle near 180°. Since glycine residues have two H$\alpha$ protons, the HNCH experiment is not ideal for measuring their $\phi$ angles. The TALOS uncertainties for G18 are ($\Delta \phi=35^\circ$, $\Delta \psi=23^\circ$) (Table S1), and three of the predicted pairs of angles lie in the right-handed $\alpha$-helix region of the Ramachandran diagram. Thus, the conformation of this residue is not well predicted from the chemical shifts. On the other hand, G18 is strictly conserved in all human defensins. Based on comparisons of HNP-4, HD5 and HD6 crystal structures, it has been suggested that G18 may act as a hinge for the G18-L29 $\beta$-hairpin [4], which has significantly different orientations in various HNPs. Thus, it remains a possibility that the G18 torsion angle difference between the NMR and crystal structures may reflect real conformational variations at this residue.

Inter-residue distances
To constrain the three-dimensional fold of HNP-1, we measured distances between sequential and non-sequential residues using DARR experiments with long mixing times. Two spin diffusion mixing times, 100 ms and 200 ms, were chosen, which resulted in larger numbers of cross peaks that deteriorated the spectral resolution (Figure 3.8). To bypass this problem, we used a comparative strategy to identify the long-distance correlation peaks. For each inter-residue cross peak, we first found all possible assignments that agree with the peak position to ±0.2 ppm. We then measured the distances of these inter-residue pairs in the HNP-3 crystal structure [3], and selected the assignment with the shortest distance. For these inter-residue cross peaks, since distances between sequential residues are usually the shortest, where ambiguity arises, we gave preferences to sequential distances over medium or long-range distances.

Table 3.2. Non-sequential inter-residue distance restraints.

<table>
<thead>
<tr>
<th>residues</th>
<th>spectrum</th>
<th>type of correlation</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C10Ca-E14Cb</td>
<td>40 ms DARR</td>
<td>medium range: i – (i+4)</td>
<td>2.5 – 4.8</td>
</tr>
<tr>
<td>E14Cγ-C30Ca</td>
<td>40 ms DARR</td>
<td>long range</td>
<td>2.5 – 4.8</td>
</tr>
<tr>
<td>A2Ca-C30Ca</td>
<td>100 ms DARR</td>
<td>long range: β1 – β3</td>
<td>2.5 – 5.4</td>
</tr>
</tbody>
</table>
In this way, we identified 40 inter-residue CC correlations, among which 26 were sequential contacts, 1 was medium range (|i-j| < 4), and 13 were long-range contacts (|i-j| > 4). Non-sequential correlations were observed for residues A2, Y4, R6, I7, C10, E14, R16, I21, L26, W27, F29 and C30 (**Table 3.2**). Six long-range correlation peaks were detected between β1 and β3 strands, and six contacts between β2 and β3 strands. These inter-strand correlations provided important restraints to the three-dimensional fold of the protein.
Figure 3.9. $^{15}\text{N}$$^{15}\text{N}$ 2D PDSD spectrum with a mixing time of 3 s. Black and red assignments indicate sequential and non-sequential correlations, respectively.

We also measured a 2D $^{15}\text{N}$$^{15}\text{N}$ correlation spectrum to obtain additional inter-residue distance restraints (Figure 3.9). The Gly region was the best resolved and provided many useful sequential and long-range cross peaks. P8 also gave well-resolved cross peaks due to its unique imine $^{15}\text{N}$ chemical shift. The 110 – 130 ppm region is more congested, thus we only assigned the intensity maxima and gave them tentative assignments in the same fashion as the CC constraints. The non-sequential NN contacts are included Table 3.2.

Since the above distance extraction strategy used the HNP-3 structure as partial input, potential differences of HNP-1 from HNP-3 can only be discerned based on missing correlations for HNP-1, and mis-assignment cannot be ruled out. To extract distance restraints in a de novo fashion, without input from the HNP-3 structure, we carried out further 3D $^{13}\text{C}$$^{13}\text{C}$$^{13}\text{C}$ correlation experiments, which removed the problem of resonance overlap. As we report in a separate publication, the 3D CCC experiment yielded 270 unique inter-residue distances, among which 129 were sequential, 45 were
medium range, and 96 were long-range constraints. These distances verified and refined the current structure, as we show below.

Three-dimensional structure of HNP-1

Table 3.3. Structure calculation summary of HNP-1.

<table>
<thead>
<tr>
<th>Restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total inter-residue CC and NN restraints60</td>
<td>60</td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium range (1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long range (</td>
<td>i-j</td>
</tr>
<tr>
<td></td>
<td>β1 – β3</td>
</tr>
<tr>
<td></td>
<td>β2 – β3</td>
</tr>
<tr>
<td></td>
<td>β1 – β2</td>
</tr>
<tr>
<td>CC Restraints in the class 2.5 – 4.8 Å</td>
<td>4</td>
</tr>
<tr>
<td>CC Restraints in the class 2.5 – 5.4 Å</td>
<td>18</td>
</tr>
<tr>
<td>CC Restraints in the class 2.5 – 6.3 Å</td>
<td>5</td>
</tr>
<tr>
<td>CC Restraints in the class 2.5 – 8.0 Å</td>
<td>7</td>
</tr>
<tr>
<td>NN restraints in the class 2.5 – 4.0 Å</td>
<td>10</td>
</tr>
<tr>
<td>NN restraints in the class 3.0 – 6.0 Å</td>
<td>10</td>
</tr>
<tr>
<td>Total (φ, ψ) torsional angle restraints</td>
<td>56</td>
</tr>
</tbody>
</table>

| rms deviation of 11 lowest-energy structures from the crystal structure |
|--------------------------------------------------|-------|
| Backbone atoms:                                   | 2.7 Å |
| Heavy atoms:                                      | 3.8 Å |
| $E_{\text{global}}$ (kcal mol$^{-1}$)            | 264   |

Using the distance and angular restraints obtained from the above SSNMR experiments, we calculated the HNP-1 structure using the XPLOR-NIH program. We assigned the distance restraints to specific ranges based on the mixing times for the first appearance of cross peaks that correspond to well-defined secondary structures (Table 3.3). Sequential Cα-Cα distances are fixed by the peptide plane geometry to be 3.8 Å and
were thus placed in the range of 3.5-4.1 Å. Sequential N-N distances depend on the \( \psi \) torsion angle and fall within the range of 2.5 – 3.6 Å. All inter-residue peaks first observed in the 100 ms \(^{13}C\text{-}^{13}C\) DARR spectrum were assigned to the range 2.5 – 5.4 Å, those first appearing in the 200 ms DARR spectrum assigned to 2.5 – 6.3 Å, and those inter-residue peaks appearing in the 2D CHHC spectra assigned to 2.5 – 7.5 Å. In addition, the few inter-residue correlations observed in the 40 ms \(^{13}C\text{-}^{13}C\) DARR spectrum were assigned to the range 2.5 – 4.8 Å. All non-sequential \(^{15}N\text{-}^{15}N\) contacts were placed in the 3.0 – 6.0 Å range.

**Figure 3.10.** SSNMR structures of HNP-1 compared with the X-ray crystal structure of HNP-3. (a) Eleven minimum energy SSNMR structures of HNP-1. (b) Average structure of (a). (c) Crystal structure of HNP-3. The disulfide bonds are shown in orange. (d) Comparison of the \( \beta_1\)-\( \beta_2 \) loop conformation between the crystal structure (red), the current solid-state NMR structure (blue), and the NMR structure with the additional distance restraints from the 3D CCC experiment (green).

In total, we obtained 60 inter-residue distance restraints, 56 (\( \phi, \psi \)) torsion angles, and three disulfide bond restraints, and subjected them to structure calculation by XPLOR-NIH. The 10 lowest-energy structures out of 200 calculated structures (**Figure 3.10a**) cluster closely. Compared to the crystal structure, the average NMR structure has a backbone atomic root-mean-square deviation (RMSD) of 2.7 Å and a heavy-atom RMSD of 3.8 Å. **Figure 3.10** compares the NMR structure ensemble with the X-ray structure of HNP-3. The NMR structure shows three well-defined \( \beta \)-strands connected by
a long loop between the β1 and β2 strands and a tight turn between the β2 and β3 strands. Both the HNP-1 NMR structure and the HNP-3 crystal structure have the same topology: the β3 strand is inserted between the β1 and β2 strands. Beyond these similarities, we found two main differences: the β1-β2 loop has a very different orientation to the rest of the protein in the two structures, and the β2 strand has a different twist and orientation. The origins of these differences will be discussed below.

**Figure 3.11.** 2D $^{13}$C-$^{1}$H LG-CP spectrum to determine the mobility of microcrystalline HNP-1. (a) 2D spectrum, measured under 11 kHz MAS and 293 K. (b-k) 1D dipolar cross sections at selected chemical shifts. Most Cα and Cβ cross sections show rigid-limit values (after taking into account the LG scaling factor of 0.577), except for the 56.2-ppm cross section (c), which has a second coupling that is smaller than the rest.
To ascertain whether large-amplitude motion exists in the HNP-1 backbone, we measured the $^1$H-$^{13}$C one-bond dipolar couplings using a 2D LG-CP experiment (Figure 3.11). The spectrum shows that most backbone C$\alpha$ sites have rigid-limit couplings of $\sim$12.5 kHz after scaling by the homonuclear decoupling scaling factor. The only exception is the signal at 56 ppm, which shows a splitting of 9.8 kHz. This value corresponds to a smaller C-H order parameter of 0.78, consistent with moderate-amplitude dynamics of some of the C$\alpha$ sites, including I7, Y17, and F29.

**Discussion**

*HNP-1 conformational homogeneity and SSNMR structural quality*

The microcrystalline HNP-1 samples are well ordered on the nanometer and micrometer scale. The $^{13}$C linewidths of 0.4 – 0.8 ppm and $^{15}$N linewidths 1.0-1.5 ppm are overall comparable to those observed in other microcrystalline proteins [49] and are better than some amyloid proteins [31]. The $^{15}$N linewidths are particularly sensitive to the conformational heterogeneity. Compared to one of the well-studied globular proteins, ubiquitin [49], whose $^{15}$N linewidths were reported as 0.3 – 0.5 ppm, the microcrystalline HNP-1 still has residual conformational heterogeneity. However, since the eventual goal is to study HNP-1 in the lipid membrane, whose thermal disorder generally promotes conformational heterogeneity in small proteins, it is not essential to produce HNP-1 with the highest conformational homogeneity.

Variable-temperature 1D $^{13}$C and $^{15}$N spectra showed quantitatively similar intensities and linewidths in the range of 293 – 253 K (Figure S5), indicating that the HNP-1 backbone is largely immobilized at room temperature and only small-amplitude segmental motions exist. This observation is supported by the 2D LG-CP spectrum, which exhibits near rigid-limit C$\alpha$ - H$\alpha$ dipolar couplings for most sites. The persistence of narrow linewidths at mild low temperatures is consistent with the behavior of other microcrystalline proteins but contrasts with that of membrane peptides, which manifest gel-phase-induced line broadening in the same temperature range.
The quality of the HNP-1 structure can be compared with the SSNMR structures of other recently studied globular proteins, such as the α-spectrin SH3 domain [25], ubiquitin [26], and GB1 [50]. In terms of agreement with the crystal structure, the SH3 domain and HNP-1 are comparable: the SH3 domain NMR structure has a backbone RMSD of 2.6 Å from the crystal structure for the β-strand segments, while the HNP-1 NMR structure has an all-segment backbone RMSD of 2.7 Å from the crystal structure of HNP-3. However, this similarity belies the difference that a larger number of restraints were obtained (292) for SH3, which corresponded to ~5 restraints per residue, and all constraints were independently extracted from NMR. In comparison, 116 restraints were obtained for HNP-1, thus about 4 restraints were available per residue. The HNP-1 structure constraints also contain a smaller number of medium and long-range distances (24) compared to SH3 (170 non-sequential restraints).

For ubiquitin, structure determination using a combination of uniformly $^{13}$C-labeled protein and $^{2-13}$C selectively labeled protein resulted in 336 inter-residue contacts for the 76-residue protein [49], among which 149 were sequential, 74 were medium range and 113 were long range. Combined with 122 torsion angle restraints, the high-resolution solid-state NMR structure had a ~1 Å RMSD for secondary structure elements compared to the crystal structure. For the 56-residue GB1, a very high resolution NMR structure was obtained from 888 unique distance restraints and dihedral angles, resulting in a backbone RMSD of 1.32 Å from one of the crystal structures [50].

*Conformation of the inter-strand β1-β2 loop and comparison with other human α-defensins*

Compared to the structures of other human α-defensins [4], the HNP-1 NMR structure has the same overall fold, with three antiparallel β-strands arranged into a β-sheet. A tight β-hairpin connects β2 and β3 strands while a loop connects the β1 and β2
strands, so that the N- and C-termini lie close together. This structural fold separates the charged and hydrophobic regions spatially, thus may facilitate the insertion of the protein into the phospholipid bilayer with the hydrophobic region buried in the membrane interior while the charged region interacting with the polar surface of the bilayer and with water [3].

Within the context of this general similarity, the SSNMR structure of HNP-1 shows two main differences from the crystal structures of various HNPs. First, the β2 strand in the NMR structure has less orientation change between the N- and C-terminal half compared to the crystal structure (Figure 3.10b, c), and the strand twist is also less significant. We attribute these differences largely to insufficient distance restraints between the β2 strand and the rest of the protein (9 distances for 7 residues). When the much larger number of inter-residue distances restraints from the 3D CCC experiment were used for structure calculation, the β2 strand was found to have a more similar twist and orientation to those of the HNP-3 structure. Thus, there appears to be no real conformational difference in the β2 strand of HNP-1 and HNP-3.

The second, more important, difference is the conformation of the loop between residues 17 and R15, which controls the relative orientation of the β1 and β2 strands. In the HNP-3 structure, the loop is bent at P8-A9 and protrudes significantly from the rest of the protein, while in the HNP-1 structure, the P8-A9 segment is extended and the loop is tugged closer to the β2 strand (Figure 3.10d). Several lines of evidence suggest that these are real conformational differences between HNP-1 and HNP-3, rather than artifacts due to insufficient constraints of the NMR structure. The structural inputs for the loop region include the chemical-shift based torsion angles (Table S1) and eleven inter-residue distances (Table 3.2). The P8 and C10 torsion angles were the main factor that determined the presence or absence of a turn at this position. Although we have not directly measured the P8 ψ torsion angle due to sensitivity limitations, the fact that the C10 φ angle is confirmed by HNCH experiments to have a regular β-strand value suggests that there is a significant possibility that the P8-C10 segment in HNP-1 does not adopt a turn conformation as in HNP-3. Second, eleven inter-residue distances
established close contacts between the N-terminal region of the loop (residues I7-P8) and the β3 strand (residues L26-W27), between residue C10 and residues G13-E14 within the loop, and between the end of the loop (residues E14-R15) and the β3 strand (residues F29-C30) (Table 3.2). Thus, this segment has a reasonable number of distance restraints. When the larger number of independently assigned restraints from 3D CCC experiments were inputted for structure calculation, the loop conformational difference largely remained (Figure 3.10d), supporting the hypothesis that the conformational difference is real.

It is noteworthy that a previous 1H solution NMR structure of HNP-1 also concluded a different loop orientation [51]. Similar to the current result, the solution structure showed the HNP-1 loop to be well defined by itself but that its relation to the rest of the protein was different from that of HNP-3. Moreover, when the HNP-1 solution structure was compared with the structures of the related rabbit defensins NP-2 and NP-5, it was revealed that the loop conformation also differs among these three proteins. Thus, the solution NMR study suggested that the β1-β2 loop conformation may be highly sequence-specific and may play a systematic role in regulating the activities of different α-defensins.

What might be the origin of the loop conformation difference between HNP-1 and HNP-3? It cannot result from fast motion, since the 2D LG-CP spectrum (Figure 3.11) shows nearly rigid-limit C-H dipolar couplings and 13C variable-temperature spectra (Figure S5) also indicate little temperature dependence. Thus, this conformational difference must be attributed to intrinsic structural differences caused either by the single-residue change at the N-terminus or by external environmental factors such as the ionic content, hydration level, and intermolecular packing. Since the loop conformation difference was also seen in the solution NMR structure [51] where there was no crystal packing effects or hydration problems, the single-residue difference at the N-terminus is likely the main reason for the loop conformational variability. The N-terminus residue is Ala in HNP-1, Asp in HNP-3, and absent in HNP-2, which thus starts its sequence with Cys. Thus, HNP-3 is more polar than HNP-1 or HNP-2 at the N-terminus.
Correspondingly, the antimicrobial activities of HNP-1 and HNP-2 are similar but the HNP-3 activity is distinct [2]. A recent study of HNP 1-3 against a panel of six bacteria of both Gram-positive and Gram-negative origins showed that the average LD$_{50}$ of HNP-2 is $1.10 \pm 0.25$ times that of HNP-1 [8], indicating very similar potencies, while the average LD$_{50}$ of HNP-3 is $2.20 \pm 1.10$ that of HNP-1, indicating that HNP-3 is about 2-fold less active than HNP-1 [8]. Molecular dynamics simulations comparing the HNP 1-3 structures suggested that Asp$_2$ in HNP-3 may facilitate electrostatic interactions between the two monomers of the dimer, thus making the basket-shaped dimer more compact for HNP-3 than for HNP-1 and 2 [52]. This difference may in turn cause distinct interactions of HNPs with the lipid membrane.

Whatever the exact molecular reason for the loop conformation difference, given the tight structural constraints imposed by the disulfide bonds and the dominant $\beta$-strand motifs for these proteins, one would expect that it is precisely the non-strand residues, without backbone hydrogen bonds, that should modulate the interactions of these proteins with the lipid membrane and tune their antimicrobial activities. Indeed, preliminary spectra of membrane-bound HNP-1 containing site-specific isotopic labels showed that A12 in the loop undergoes an interesting chemical shift change between a more helical conformation and the conformation seen in the microcrystalline state (Figure S6). This data suggests that the $\beta1$-$\beta2$ loop conformation is sensitive to the environment, in addition to the sequence. Further comparative structural studies are necessary to fully understand the structure-activity relationship of the HNPs and elucidate the functionally relevant conformation of the $\beta1$-$\beta2$ loop. Solid-state NMR experiments that determine the conformation, orientation and depth of insertion of HNP-1 in lipid bilayers will be crucial for this purpose.

Materials and Methods

HNP-1 expression and microcrystalline sample preparation

Recombinant HNP-1, whose amino acid sequence is shown in Figure 3.1, was obtained as a cleavage product from its precursor protein, proHNP1, which was
expressed as a GST-fusion protein in *E. coli* and folded. Uniformly $^{13}$C, $^{15}$N-enriched media Spectra 9 (Cambridge Isotope Laboratories) was used to label the protein. Briefly, the fusion protein GST-proHNP1 was expressed in *E. coli* BL21 with IPTG induction. The insoluble inclusion bodies were denatured by 8 M urea and then folded in 2 M urea, 3 mM cysteine and 0.3 mM cystine. The folded fusion protein was dialyzed in a pH 7 buffer containing 20 mM Tris-HCl, 150 mM NaCl and 1.5 mM CaCl$_2$. The fusion protein was then cleaved by thrombin, producing proHNP1, which was purified by reversed-phase HPLC. ProHNP1 was further cleaved by cyanogen bromide to yield the correctly folded HNP-1. The crude HNP-1 was purified by reversed-phase HPLC and analyzed by electrospray ionization mass spectrometry to confirm the mass (3634 Da). The yield of HNP-1 was ~3 mg per liter culture. Antimicrobial assays confirmed the activity of the protein. For example, 100% killing of *S. aureus* is reached at 64 µg/ml HNP-1.

Microcrystalline HNP-1 was precipitated from a polyethylene glycol 400 (PEG-400) solution containing 30 mM cacodylate and 60 mM Li$_2$SO$_4$ at pH 6.5. The PEG-400 solution was added slowly to a protein stock solution to reach a final PEG-400 concentration of 30 wt%. Two protein stock solutions, at concentrations of 30 mg/ml and 36 mg/ml, were used. The protein precipitants were centrifuged in a sealed pipette tip, the supernatant removed, the pipette tip cut open, and the precipitant was centrifuged into a 4 mm MAS rotor. Two HNP-1 samples, one in a 4 mm MAS rotor containing ~4 mg protein and the other in a 2.5 mm MAS rotor (~2.7 mg protein) were used in this study.

*Solid-state NMR spectroscopy*

Most SSNMR experiments were carried out on a Bruker AVANCE-600 (14.1 Tesla) spectrometer (Karlsruhe, Germany) using triple-resonance MAS probes. Most spectra were measured at 253 – 268 K under MAS frequencies of 8 – 15 kHz. 2D CM$_3$RR spectra were acquired on a 900 MHz NMR spectrometer at the Harvard/MIT Center for Magnetic Resonance. The 900 MHz spectra were measured on a 2.5 mm MAS probe under 20 kHz MAS.
Typical radio-frequency (rf) pulse lengths were 3.5 µs for $^{13}$C, 6.0 µs for $^{15}$N, and 2.5-4.0 µs for $^1$H. $^1$H decoupling fields were typically 70 kHz on the 4 mm rotor sample and 100 kHz on the 2.5 mm rotor sample. $^{13}$C chemical shifts were referenced externally to the $\alpha$-Gly $^{13}$CO signal at 176.49 ppm on the TMS scale, and $^{15}$N chemical shifts were referenced to the $^{15}$N signal of N-acetylvaline at 122.0 ppm on the liquid ammonia scale. The TMS scale differs from the DSS scale by 1.7 ppm, thus all $^{13}$C chemical shift values reported here should be increased by 1.7 ppm before comparing with DSS-referenced solution NMR chemical shifts. For torsion angle prediction by the TALOS software, we converted the TMS-based chemical shifts to DSS-referenced values.

2D $^{13}$C-$^{13}$C dipolar-assisted rotational resonance (DARR) experiments [53] were acquired under 8 kHz MAS on the 4 mm rotor sample and 10-15 kHz on the 2.5 mm rotor sample. Four spectra were acquired with mixing times of 20 ms, 40 ms, 100 ms, and 200 ms. For the 4 mm rotor sample, the maximum evolution time for the indirect dimension was 5.6 ms, corresponding to 300 $t_1$ points. The number of scans per $t_1$ slice was 96 and the recycle delay was 2 s, giving an experimental time of ~16 hour per 2D spectrum. For the 2.5 mm rotor sample, the maximum $^{13}$C evolution time was increased to 7.2 – 7.4 ms due to the better homogeneity and spectral resolution of the sample. 380 $t_1$ points were measured for the 20 ms and 40 ms experiments and 580 $t_1$ points were measured for the 100 ms and 200 ms mixing times. The experimental time was 18 – 26 hours per 2D spectrum.

Sequence-specific $^{13}$C and $^{15}$N resonance assignment was carried out by 3D NCOCX and NCACX experiments [34,35,54] under 8 kHz MAS at 268 K. Magnetization transfer between $^{15}$N and $^{13}$CO or $^{15}$N and $^{13}$C$\alpha$ was achieved using band-selective SPECIFIC-CP [55]. In the intra-residue NCACX experiment, the C$\alpha$ magnetization was selected by a 17-kHz $^{15}$N spin-lock field and a 25-kHz $^{13}$C spin lock field on resonance with C$\alpha$. In the inter-residue NCOCX experiment, the spin-lock fields were 27 kHz for $^{15}$N and 30 kHz for $^{13}$C on resonance with C$\alpha$. In the latter case, since the CO offset was 18 kHz, the effective field of the CO resonances was ~35 kHz, thus satisfying the sideband matching condition with $^{15}$N. After SPECIFIC-CP, $^{13}$C-$^{13}$C
magnetization transfer occurred during a DARR mixing time, which was 40 ms for NCACX and 60 ms for NCOCX.

For the NCOCX experiment, the $^{15}$N spectral width was 4 kHz and the maximum $t_1$ evolution time was 4.75 ms, corresponding to 38 $t_1$ points. The $^{13}$CO spectral width was 2 kHz and the number of $t_2$ points was 22, giving a maximum $t_2$ of 5.56 ms. The experimental time was about 2.5 days. For the NCACX experiment, the $^{15}$N spectral width was the same as that of NCOCX, while the $^{13}$C$\alpha$ ($\omega_2$) spectral width was 4.5 kHz and the acquisition time was 5.56 ms, corresponding to 50 $t_2$ points. The experimental time was 2 days.

The 2D $^{15}$N-$^{15}$N correlation experiment was conducted using $^1$H-driven spin diffusion (PDSD) under 6 kHz MAS at 253 K. The mixing time was 3 s to obtain inter-residue $^{15}$N-$^{15}$N correlation peaks. The indirect dimension has a spectral width of 16 kHz and an acquisition time of 6.88 ms, corresponding to 220 $t_1$ points. The number of scan per $t_1$ slice was 144 and the recycle delay was 1.5 s, giving an experimental time of 40 hours.

The N-H doubled HNCH experiment to determine $\phi$ torsion angles [36,47] were carried out on the 2.5 mm rotor sample under 8 kHz MAS at 263 K. $^1$H homonuclear decoupling during the N-H and C-H evolution periods were achieved using the FSLG sequence [56]. The C$\alpha$ peaks of the chemical shift dimension were assigned based on the 2D and 3D correlation spectra. Although resonance overlap was unavoidable, most residues have consistent $\phi$ angles between the TALOS and the crystal structure results, thus allowing the residues of interest to be measured. For sensitivity reasons, only the intensities at 0 and half a rotor period of dipolar evolution times were measured. Simulations (Figure S4) indicate that the intensity at half a rotor period is the most sensitive to $\phi$ angle variations.

Structure calculation
The $^{13}$C$_\alpha$, $^{13}$C$_\beta$, $^{13}$CO and $^{15}$N chemical shifts were inputted into the TALOS program [46] to obtain ($\phi$, $\psi$) torsion angles. Structure calculations were performed using the simulated annealing protocol of the XPLOR-NIH program [57,58]. Input constraints included 60 torsion angles, 9 distances associated with the disulfide bonds, 46 distances between sequential residues, and 24 non-sequential inter-residue distances. An ensemble of 200 structures were first calculated by performing molecular dynamics at 3500 K for 40 ps, followed by slow cooling from 3500 K to 25 K in 12.5 K steps. At each temperature 0.4 ps of dynamics was performed using a soft square NOE potential. Each structure was then refined with the same protocol but with 10 ps of initial annealing using a hard square potential with the $k_{\text{NOE}}$ force constant held at 30 kcal. After refinement, the 10 lowest-energy structures were chosen to represent the final structure of HNP-1. All structures were visualized in Chimera (UCSF).

Accession numbers

The HNP-1 SSNMR structure coordinates have been deposited in the Protein Data Bank (PDB code: 2KHT). The chemical shifts and distance restraints have been deposited in the Biological Magnetic Resonance Bank (accession number: resb101138).

Acknowledgement

The authors thank Chih-Chia Su for assistance with microcrystal visualization, Professor Chad Rienstra for sharing the design of microcrystal transfer tools, Professor Robert Griffin, Dr. Jozef Lewandowski, Dr. Gaël De Paëpe and Dr. Tony Bielecki for assistance with the 900 MHz NMR experiments, and Dr. Andrew Severin and Dr. Shenhui Li for help with the XPLOR-NIH structure calculations.

This work is supported by the NIH grant GM66976 to M.H. and P41-EB-002026 for the 900 MHz NMR time at MIT.

References


Supporting Information

Figure S1. Summary of the $^{13}$C chemical shifts of the aromatic sidechains of HNP-1.

Figure S2. Assignment of C10 $^{13}$C chemical shifts. (a) F2-F3 plane of the NCACX spectrum at the $^{15}$N chemical shift of 115.6 ppm. The $^{15}$N, $^{13}$C and $^{13}$C$\beta$ chemical shifts of the C10 spin system are identified. (b) F2-F3 plane of the NCOCX spectrum at 115.6-ppm $^{15}$N chemical shift. C10-A9 cross peaks are clearly seen. (c) 100 ms 2D $^{13}$C-$^{13}$C DARR correlation spectrum. The C10 and A9 connectivities (dashed lines) are indicated. Sequential cross peaks between A9 and C10 are highlighted in yellow and assigned in blue, and confirm the C10 C$\beta$ assignment to be 33.4 ppm on the TMS scale.
Figure S3. Cα and Cβ chemical shifts of the six cysteines in HNP-1 overlaid with literature values as a function of oxidation state. The HNP-1 cysteine chemical shifts are referenced to DSS to be consistent with the solution-NMR chemical shifts. Two literature statistical analyses of cysteine chemical shifts are shown. (a) Statistical study by Montelione and coworkers. Green: oxidized cysteines. Blue: reduced cysteines. Red: Zn-ligated cysteines. The HNP-1 cysteine chemical shifts are shown as red circles along with the disulfide bonds in the native protein. (b) Statistical study by Sharma and Rajarathnam. The oxidized cysteines (S-S) in the β-sheet conformation (B, filled circles) lie in the lower right corner of the figure. Both comparisons verify that the six HNP-1 cysteines, including C10, are disulfide bonded.
Figure S4. Simulated $\phi$ angle dependent HNCH curves for key backbone C$\alpha$ peaks to confirm chemical-shift based torsion angles. The residues at each chemical shift frequency and their $\phi_1$ angles are shown. For most residues the $\phi_1$ angles agree well between the TALOS predictions and the crystal structure to within ±10°. For C10 and Y22, highlighted in red, significant disagreement exists. Simulations for both cases are shown. Simulations were carried out using the experimental condition of 8 kHz MAS, FSLG $^1$H homonuclear decoupling, and doubling of the N-H dipolar coupling.$^{3,4}$ (a) 54.5 ppm, using the NMR $\phi_1$ angle for Y22. (b) 54.5 ppm, using the crystal structure $\phi_1$ angle for Y22. (c) 52.6 ppm. (d) 52.0 ppm. (e) 51.2 ppm. (f) 50.8 ppm, where the C10 $\phi_1$ angle is 173° for NMR and 132° for X-ray.
**Figure S5.** $^{13}$C variable-temperature CP-MAS spectrum of microcrystalline HNP-1. Spectra were measured on the 4 mm MAS rotor sample from 288 K to 253 K. The resolution is not affected until 253 K, where sidechain line broadening is observed. The intensities are similarly high in this temperature range, indicating the protein is already largely immobilized at 288 K.
Figure S6. Conformational variability of the loop region of HNP-1 in the membrane-bound state. Site-specifically labeled HNP-1 bound to DMPC/DMPG (3 : 1) membranes was measured. (a) The Ala_{12} chemical shifts are the same as in the microcrystalline state. (b) The Ala_{12} chemical shifts are more helical than in the microcrystalline state.
Table S1: HNP-1 ($\phi$, $\psi$) angles obtained from SSNMR chemical shifts versus crystal structure torsion angles of HNP-3. Residues with significant angular deviations between the two structures are shaded in gray.

<table>
<thead>
<tr>
<th>Residues</th>
<th>NMR (HNP-1)</th>
<th>X-ray (HNP-3)</th>
<th>$\phi$ from TALOS</th>
<th>$\psi$ from TALOS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\phi$</td>
<td>$\psi$</td>
<td>$\phi$</td>
<td>$\psi$</td>
</tr>
<tr>
<td>A2</td>
<td>- -</td>
<td>- -</td>
<td>19.3</td>
<td>16.3</td>
</tr>
<tr>
<td>C3</td>
<td>-131.0</td>
<td>154.9</td>
<td>-112.6</td>
<td>148.0</td>
</tr>
<tr>
<td>Y4</td>
<td>-123.2</td>
<td>129.8</td>
<td>-134.1</td>
<td>157.9</td>
</tr>
<tr>
<td>C5</td>
<td>-119.7</td>
<td>142.7</td>
<td>-100.0</td>
<td>123.0</td>
</tr>
<tr>
<td>R6</td>
<td>-131.9</td>
<td>141.6</td>
<td>-133.4</td>
<td>150.0</td>
</tr>
<tr>
<td>I7</td>
<td>-135.0</td>
<td>147.7</td>
<td>-136.1</td>
<td>134.9</td>
</tr>
<tr>
<td>P8</td>
<td>-135.0</td>
<td>144.4</td>
<td>-86.7</td>
<td>-11.5</td>
</tr>
<tr>
<td>A9</td>
<td>-141.1</td>
<td>151.6</td>
<td>-163.1</td>
<td>173.2</td>
</tr>
<tr>
<td>C10</td>
<td>-127.1</td>
<td>150.2</td>
<td>-71.7</td>
<td>152.2</td>
</tr>
<tr>
<td>I11</td>
<td>-120.4</td>
<td>158.9</td>
<td>-101.6</td>
<td>179.7</td>
</tr>
<tr>
<td>A12</td>
<td>-71.8</td>
<td>147.0</td>
<td>-54.6</td>
<td>131.9</td>
</tr>
<tr>
<td>G13</td>
<td>83.2</td>
<td>22.6</td>
<td>87.2</td>
<td>8.9</td>
</tr>
<tr>
<td>E14</td>
<td>-132.1</td>
<td>152.3</td>
<td>-125.1</td>
<td>158.4</td>
</tr>
<tr>
<td>R15</td>
<td>-134.0</td>
<td>147.2</td>
<td>-130.2</td>
<td>150.2</td>
</tr>
<tr>
<td>R16</td>
<td>-99.0</td>
<td>124.2</td>
<td>-82.7</td>
<td>129.9</td>
</tr>
<tr>
<td>Y17</td>
<td>-116.2</td>
<td>143.2</td>
<td>-116.0</td>
<td>-6.8</td>
</tr>
<tr>
<td>G18</td>
<td>-120.6</td>
<td>134.9</td>
<td>-179.8</td>
<td>-157.7</td>
</tr>
<tr>
<td>T19</td>
<td>-109.5</td>
<td>133.8</td>
<td>-130.3</td>
<td>146.9</td>
</tr>
<tr>
<td>C20</td>
<td>-131.6</td>
<td>149.1</td>
<td>-117.7</td>
<td>152.8</td>
</tr>
<tr>
<td>I21</td>
<td>-98.8</td>
<td>127.5</td>
<td>-116.9</td>
<td>123.2</td>
</tr>
<tr>
<td>Y22</td>
<td>-109.4</td>
<td>116.6</td>
<td>-165.2</td>
<td>134.2</td>
</tr>
<tr>
<td>Q23</td>
<td>50.8</td>
<td>43.0</td>
<td>54.0</td>
<td>41.8</td>
</tr>
<tr>
<td>G24</td>
<td>79.3</td>
<td>3.3</td>
<td>66.3</td>
<td>20.5</td>
</tr>
<tr>
<td>R25</td>
<td>-123.2</td>
<td>151.4</td>
<td>-131.5</td>
<td>164.9</td>
</tr>
<tr>
<td>L26</td>
<td>-107.6</td>
<td>131.4</td>
<td>-95.3</td>
<td>136.6</td>
</tr>
<tr>
<td>W27</td>
<td>-106.0</td>
<td>139.2</td>
<td>-127.5</td>
<td>152.5</td>
</tr>
<tr>
<td>A28</td>
<td>-89.8</td>
<td>125.9</td>
<td>-81.2</td>
<td>138.1</td>
</tr>
<tr>
<td>F29</td>
<td>-111.7</td>
<td>134.7</td>
<td>-95.2</td>
<td>120.8</td>
</tr>
<tr>
<td>C30</td>
<td>-126.1</td>
<td>154.4</td>
<td>-123.3</td>
<td>140.3</td>
</tr>
<tr>
<td>C31</td>
<td>- -</td>
<td>- -</td>
<td>-138.8</td>
<td>-</td>
</tr>
</tbody>
</table>
Table S2. XPLOR-NIH NOE input files for microcrystalline HNP-1 from solid-state NMR data. Note the residue number starts with Ala₁ in the NOE calculation, but the rest of the paper uses the convention of Ala₂ as the first residue to align with the other HNP proteins.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Distance</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>assign (resid 2 and name CA) (resid 30 and name CA)</td>
<td>4.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 2 and name CB) (resid 30 and name CB)</td>
<td>4.1</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 2 and name SG) (resid 30 and name SG)</td>
<td>1.9</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 4 and name CA) (resid 19 and name CA)</td>
<td>6.0</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 4 and name CB) (resid 19 and name CB)</td>
<td>3.7</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 4 and name SG) (resid 19 and name SG)</td>
<td>1.9</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 9 and name CA) (resid 29 and name CA)</td>
<td>5.6</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 9 and name CA) (resid 29 and name CA)</td>
<td>3.7</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 9 and name SG) (resid 29 and name SG)</td>
<td>1.9</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 1 and name CA) (resid 2 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 4 and name CA) (resid 5 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 5 and name CA) (resid 6 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 8 and name CA) (resid 9 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 11 and name CA) (resid 12 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 23 and name CA) (resid 24 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 28 and name CA) (resid 29 and name CA)</td>
<td>3.8</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>assign (resid 9 and name CA) (resid 13 and name CB)</td>
<td>3.8</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>assign (resid 13 and name CG) (resid 29 and name CA)</td>
<td>3.8</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>assign (resid 20 and name CG1) (resid 21 and name CB)</td>
<td>3.8</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>assign (resid 24 and name CB) (resid 25 and name CG)</td>
<td>3.8</td>
<td>1.3</td>
<td>1.0</td>
</tr>
<tr>
<td>assign (resid 1 and name CA) (resid 29 and name CA)</td>
<td>4.6</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>assign (resid 3 and name CA) (resid 4 and name CB)</td>
<td>4.6</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>assign (resid 3 and name CB) (resid 4 and name CA)</td>
<td>4.6</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>assign (resid 3 and name CA) (resid 29 and name CB)</td>
<td>4.6</td>
<td>2.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>
assign ( resid 4 and name CA ) ( resid 5 and name CB ) 4.6 2.1 0.8 !
assign ( resid 5 and name CA ) ( resid 27 and name CA ) 4.6 2.1 0.8 !
assign ( resid 6 and name CG1 ) ( resid 25 and name CB ) 4.6 2.1 0.8 !
assign ( resid 6 and name CG1 ) ( resid 26 and name CA ) 4.6 2.1 0.8 !
assign ( resid 7 and name CB ) ( resid 8 and name CA ) 4.6 2.1 0.8 !
assign ( resid 7 and name CB ) ( resid 8 and name CB ) 4.6 2.1 0.8 !
assign ( resid 8 and name CB ) ( resid 9 and name CA ) 4.6 2.1 0.8 !
assign ( resid 9 and name CB ) ( resid 10 and name CG2 ) 4.6 2.1 0.8 !
assign ( resid 9 and name CB ) ( resid 15 and name CD ) 4.6 2.1 0.8 !
assign ( resid 11 and name CB ) ( resid 12 and name CA ) 4.6 2.1 0.8 !
assign ( resid 15 and name CA ) ( resid 27 and name CB ) 4.6 2.1 0.8 !
assign ( resid 19 and name CB ) ( resid 20 and name CG2 ) 4.6 2.1 0.8 !
assign ( resid 20 and name CA ) ( resid 25 and name CA ) 4.6 2.1 0.8 !
assign ( resid 22 and name CB ) ( resid 23 and name CA ) 4.6 2.1 0.8 !
assign ( resid 5 and name CD ) ( resid 29 and name CB ) 5.8 3.3 0.5 !
assign ( resid 10 and name CB ) ( resid 11 and name CA ) 5.8 3.3 0.5 !
assign ( resid 15 and name CB ) ( resid 28 and name CA ) 5.8 3.3 0.5 !
assign ( resid 19 and name CA ) ( resid 20 and name CD1 ) 5.8 3.3 0.5 !
assign ( resid 21 and name CB ) ( resid 22 and name CA ) 5.8 3.3 0.5 !
assign ( resid 4 and name CB ) ( resid 28 and name CA ) 7.5 5.0 0.5 !
assign ( resid 5 and name CA ) ( resid 6 and name CB ) 7.5 5.0 0.5 !
assign ( resid 5 and name CG ) ( resid 27 and name CA ) 7.5 5.0 0.5 !
assign ( resid 6 and name CA ) ( resid 7 and name CA ) 7.5 5.0 0.5 !
assign ( resid 8 and name CB ) ( resid 9 and name CA ) 7.5 5.0 0.5 !
assign ( resid 9 and name CB ) ( resid 10 and name CG1 ) 7.5 5.0 0.5 !
assign ( resid 3 and name N ) ( resid 4 and name N ) 3.5 1.0 0.5 !
assign ( resid 4 and name N ) ( resid 5 and name N ) 3.5 1.0 0.5 !
assign ( resid 6 and name N ) ( resid 7 and name N ) 3.5 1.0 0.5 !
assign ( resid 8 and name N ) ( resid 9 and name N ) 3.5 1.0 0.5 !
assign ( resid 16 and name N ) ( resid 17 and name N ) 3.5 1.0 0.5 !
assign ( resid 17 and name N ) ( resid 18 and name N ) 3.5 1.0 0.5 !
assign ( resid 21 and name N ) ( resid 22 and name N ) 3.5 1.0 0.5 !
assign ( resid 22 and name N ) ( resid 23 and name N ) 3.5 1.0 0.5 !
assign ( resid 23 and name N ) ( resid 24 and name N ) 3.5 1.0 0.5 !
assign ( resid 26 and name N ) ( resid 27 and name N ) 3.5 1.0 0.5 !
assign ( resid 4 and name N ) ( resid 18 and name N ) 6.0 3.0 0.5 !
assign ( resid 7 and name N ) ( resid 26 and name N ) 6.0 3.0 0.5 !
assign ( resid 7 and name N ) ( resid 27 and name N ) 6.0 3.0 0.5 !
assign ( resid 9 and name N ) ( resid 12 and name N ) 6.0 3.0 0.5 !
assign ( resid 16 and name N ) ( resid 29 and name N ) 6.0 3.0 0.5 !
assign ( resid 17 and name N ) ( resid 26 and name N ) 6.0 3.0 0.5 !
assign ( resid 21 and name N ) ( resid 24 and name N ) 6.0 3.0 0.5 !
assign ( resid 23 and name N ) ( resid 25 and name N ) 6.0 3.0 0.5 !

References

Chapter 4

3D $^{13}$C-$^{13}$C-$^{13}$C Correlation NMR for De Novo Distance Determination of Solid Proteins and Application to a Human Alpha Defensin

_A paper published in J. Magn. Reson._


Shenhui Li, Yuan Zhang, and Mei Hong*

Abstract

The _de novo_ structure of an antimicrobial protein, human $\alpha$-defensin 1 (HNP-1), is determined by combining a 3D $^{13}$C-$^{13}$C-$^{13}$C (CCC) magic-angle spinning (MAS) correlation experiment with standard resonance assignment experiments. Using a short spin diffusion mixing time to assign intra-residue cross peaks and a long mixing time to detect inter-residue correlation peaks, we show that the 3D CCC experiment not only reduces the ambiguity of resonance assignment, but more importantly yields two orders of magnitude more long-range distances without recourse to existing crystal structures. Most of these distance constraints could not be obtained in a _de novo_ fashion from 2D correlation spectra due to significant resonance overlap. Combining the distance constraints from the 3D CCC experiment and the chemical-shift-derived torsion angles, we obtained a _de novo_ high-resolution NMR structure of HNP-1, with a heavy-atom RMSD of 3.4 Å from the crystal structure of the analogous HNP-3. The average energy of the minimum-energy ensemble is less than of 40 kcal/mol. Thus, the 3D CCC experiment provides a reliable means of restraining the three-dimensional structure of insoluble proteins with unknown conformations.
Introduction

Magic-angle spinning (MAS) solid-state NMR (SSNMR) spectroscopy has become a complementary method to solution NMR and X-ray crystallography for structure determination of proteins. Techniques for measuring a large number of conformational constraints, such as chemical shifts [1-9], torsion angles [10-12], and distances [13-16], have been recently developed. For uniformly $^{13}$C-labeled proteins, distance constraints are usually obtained from 2D $^{13}$C-$^{13}$C correlation spectra such as PDSD, DARR, and CHHC [17-20]. To obtain long-distance constraints, long spin diffusion mixing times are usually required, resulting in a large number of cross peaks that are difficult to resolve and assign, since even the subset of short-range cross peaks already required 3D heteronuclear correlation experiments to fully resolve. Recently, a Proton Assisted Recoupling (PAR) experiment [21,22] was introduced to selectively enhance the intensities of weakly coupled long-range cross peaks in 2D spectra. Despite this improvement, 2D $^{13}$C-$^{13}$C correlation approaches remain of limited utility for distance extraction, and often existing crystal structures or structural models were relied on for assigning the cross peaks in these congested 2D spectra.

Three-dimensional $^{13}$C-$^{13}$C-$^{13}$C (CCC) correlation experiments have been proposed to facilitate resonance assignment [23,24]. Baldus and co-workers [23] demonstrated a CCC experiment that contains a double-quantum (DQ) indirect dimension, which avoids the cubic diagonal in the 3D spectrum and identifies the intra-residue cross peaks through the DQ signals. Rienstra and coworkers [24] proposed an alternative 3D CCC experiment that uses soft 180° and 90° pulses to selectively detect the aliphatic region of the $^{13}$C spectrum in the F1 and F2 dimensions, thus shortening experimental time. However, both studies focused on increasing intra-residue resonance assignment rather than extracting long-range distances. In this work, we show that the 3D CCC experiment not only improves resonance assignment, but allows distance extraction in a de novo fashion, which is essential for restraining the three-dimensional fold of structurally unknown proteins.

Human neutrophil α-defensins (HNPs) are cationic antimicrobial proteins of the
human innate immune system to protect against microbial infections [25]. Six human α-defensins are known. They contain 29 - 33 amino acid residues and six conserved cysteins that form three intramolecular disulfide bonds [26-29]. HNPs kill a broad range of pathogens, including bacteria, fungi and certain enveloped viruses [25,30]. Similar to most antimicrobial peptides (AMPs), the main mode of action of HNPs is the disruption of the cell membrane of the invading pathogen [31,32]. Several crystal structures of HNPs have been reported [33-36]; however, all these structures were solved in the absence of any membranes or membrane-mimetic solvents, thus their relevance for the membrane-bound structure and mechanism of HNPs remains unknown. Understanding the detailed mechanism of human α-defensins thus requires high-resolution structures of HNPs in the lipid membrane.

We have initiated a solid-state NMR investigation of the three-dimensional structure of HNP-1 in lipid membranes. As a first step, the resonance assignment of the 30-residue HNP-1 in a microcrystalline state has been completed using various 2D 13C-13C and 3D 13C-15N correlation techniques [37]. However, the long-distance correlation peaks observed in the 2D CC spectra could not be assigned without using information from the crystal structure of the analogous HNP-3 [33]. In this work, we show that the 3D CCC experiment allows a large number of inter-residue distances to be identified, thus permitting a de novo structure determination of HNP-1. We also quantify the increased information content of the 3D CCC experiment compared to the 2D CC experiment.

Materials and Methods

Sample preparation

Uniformly 13C, 15N-labeled HNP-1 in microcrystalline form was prepared as described in a separate report [37]. Briefly, a fusion protein containing the HNP-1 sequence, GST-proHNP1, was expressed in E. coli, folded, and cleaved to yield the HNP-1 precursor protein proHNP-1. ProHNP-1 was purified by reversed-phase HPLC and cleaved by cyanogen bromide to give folded HNP-1. Microcrystalline HNP-1 was prepared by precipitation from a pH 6.5 polyethyleneglycol-400 solution. The hydrated
protein microcrystals were centrifuged into the MAS rotors. 3-5 mg of microcrystals were packed into a 4 mm rotor and a 2.5 mm rotor.

Solid-state NMR

Solid-state NMR experiments were carried out on a Bruker AVANCE-600 (14.1 Tesla) spectrometer. The 2D $^{13}$C-$^{13}$C DARR spectrum with a mixing time of 100 ms was measured at 253 K under 12 kHz MAS using a 2.5 mm rotor sample. The spectral widths of the direct (F2) and indirect (F1) dimensions were 50 kHz and 27 kHz, respectively. A total of 300 t$_1$ slices and 96 scans per slice were measured, resulting in an experimental time of 17 hours.

Figure 4.1. Pulse sequence for the 3D CCC MAS correlation experiment. Phase cycles are: $\phi_1 = 0$ 2, $\phi_2 = 1$, $\phi_3 = 00112233$, $\phi_4 = 11223300$, $\phi_5 = 33001122$, $\phi_6 = 11223300$, $\phi_7 = 33001122$, and receiver = 10323210, 32101032. Here 0 = +x, 1 = +y, 2 = -x, and 3 = -y.

The 3D CCC pulse sequence is shown in Figure 4.1. The experiment was carried out at 253 K under 8 kHz MAS using a 4 mm rotor sample. The first and second DARR mixing times were 20 ms and 100 ms, respectively. The F1 and F2 spectral widths were 13 kHz (86.7 ppm). The $^{13}$C carrier was set to 39.6 ppm, thus the F1 and F2 spectral windows ranged from -3.5 to 82.7 ppm. This spectral range folded the Arg C$\zeta$ region in 157-169 ppm to 70.8-82.8 ppm, and the CO region at 170 -180 ppm to -2.4-7.6 ppm. Both folded regions avoided resonance overlap with the aliphatic carbons. The 3D spectrum was measured with 130 t$_1$ slices, 130 t$_2$ slices, and 8 scans per slice, resulting in an experimental time of 71 hours.
Typical 90° pulse lengths were 4-5 µs for both $^1$H and $^{13}$C channels. The $^1$H-$^{13}$C cross polarization contact time was 700 µs and the $^1$H decoupling field strength was 71 kHz in the experiments. The $^{13}$C chemical shifts were referenced externally to the $\alpha$-Gly $^{13}$C′ signal at 176.49 ppm on the TMS scale.

Structure calculation

The 2D and 3D $^{13}$C correlation spectra were assigned using the program SPARKY. Structure calculation was carried out using the XPLOR-NIH program [38,39]. Distance constraints were inputted similar to that of Castellani et al [13], where sequential Cα-Cα distances were fixed to 3.8±0.5 Å, and other inter-residue distances were given the range of 2.5-7.5 Å for the 100 ms DARR mixing time used in the 3D CCC experiment [13,40]. In addition to unique distance constraints, ambiguous distances with a degeneracy of less than 5 were included in the input file. Dihedral angle constraints obtained from the assigned $^{13}$C and $^{15}$N chemical shifts [41] were also included [37]. Structure calculations proceeded in two stages. For the annealing process, an ensemble of 200 structures was calculated by performing molecular dynamics at 3500 K for 40 ps, followed by slow cooling from 3500 K to 25 K in 12.5 K increments, with 0.4 ps of dynamics at each temperature using a soft square NOE potential. In the second stage, each structure was refined by repeating the annealing protocol with only 10 ps of initial annealing using a hard square NOE potential with the $k_{\text{NOE}}$ force constant held at 30 kcal. After structure refinement, the ensemble of 10 lowest energy structures was selected to represent the final structure of HNP-1. The final structure has been deposited to the Protein Data Bank (Accession code: 2KHT).

Results and Discussion

2D CC spectrum at long mixing times
Figure 4.2. 2D $^{13}$C-$^{13}$C correlation spectrum of microcrystalline HNP-1 at a DARR mixing time of 100 ms. The spectrum was measured under 12 kHz MAS at 253 K. Inter-residue cross peaks are annotated in blue, and some intra-residue peaks are also assigned, in black.

Before describing the analysis of the 3D CCC experiment, we first present the 100 ms 2D DARR spectrum (Figure 4.2), which was measured to obtain long-range correlation peaks in a conventional manner [37]. All inter-residue cross peaks in the 2D spectrum are annotated in blue, and selected intra-residue cross peaks are annotated in black. For most inter-residue cross peaks, the assignments are ambiguous, with a
degeneracy of 2 or higher. Only one of the assignment possibilities is listed in Figure 4.2. For example, the peak at (F1 = 33.7 ppm, F2 = 42.5 ppm) can be assigned either to R6Cβ-R16Cδ or C10Cβ-R16Cδ, corresponding to an assignment degeneracy of 2. From this 100 ms 2D CC spectrum, a total of 24 inter-residue cross peaks with an assignment degeneracy of 1 – 8 were found. Among these peaks, 15 peaks were assigned to sequential correlations (|i-j|=1), 1 peak was assigned to medium correlations (1<|i-j|≤4), and 8 peaks were assigned to long-range correlations (|i-j|>4).

3D CCC experiment

Figure 4.3. F2-F3 planes of the 3D CCC spectrum at various F1 frequencies. (a) F1=54.0 ppm, (b) F1=49.0 ppm, (c) F1=33.8 ppm, (d) F1=27.1 ppm. The amino acid spin systems are connected by dashed lines. All inter-residue cross peaks are assigned. The non-sequential inter-residue cross peaks are shown in brown while the sequential cross peaks are in black.
The 3D CCC experiment was carried out using a DARR mixing time of 20 ms between F1 and F2 and 100 ms between F2 and F3 dimensions. The short mixing time allows polarization transfer mostly within the same residue while the long mixing time establishes inter-residue correlations. Figure 4.3 shows 4 representative 2D planes at F1 = 54.0, 49.0, 33.8, and 27.1 ppm. Compared to the 2D spectrum, each F2-F3 plane has a much smaller number of peaks, thus facilitating resonance assignment. For example, based on the previous assignment [37], the F2-F3 plane at F1 = 54.0 ppm (Figure 4.3a) can be attributed to five Cα sites: R6Cα (53.9 ppm), R16Cα (54.5 ppm), Y22Cα (54.2 ppm), Q23Cα (54.2 ppm) and C31Cα (54.0 ppm). Indeed, five networks of resonances indicated by dashed lines were observed, representing the spin systems of R6, R16, C31, Q23, and Y22. For example, in the R6 spin system (blue dashed lines), R6Cα (53.9 ppm), R6Cβ (33.7 ppm), and R6Cγ (26.7 ppm) are detected in the F2 dimension due to spin diffusion from R6Cα during the 20 ms mixing time. Intra-residue cross peaks such as R6Cα-Cβ-Cδ (54.0 ppm, 34.3 ppm, 41.2 ppm) and R6Cα-Cγ-Cδ (54.0 ppm, 26.9 ppm, 41.2 ppm) are readily identified in the blue square.

More important than the intra-residue cross peaks are inter-residue cross peaks in these 2D planes, which are annotated in brown for non-sequential correlation peaks and black for sequential cross peaks. In assigning these inter-residue peaks, we use the following principles. First, due to the short mixing time between the F1 and F2 dimensions the F1-F2 assignment usually does not extend beyond sequential correlations, with the exception of disulfide bonds and a few very high intensity peaks such as the Ala methyl carbons. Second, for peaks lying at diagonal planes of the 3D spectrum, where either the F1 and F2 frequencies are identical or the F1 and F3 frequencies are identical, the degenerate assignment is resolved by looking for peaks in other planes that belong to the spin systems of the degenerate sites. For example, for two spins A and A’ with similar chemical shifts ωA, the observation of a cross peak at (F1, F2, F3) = (ωA, ωA, ωC) is two-fold degenerate. But if A and A’ have different frequencies for their directly bonded spins B and B’, then the observation of a cross peak at (ωB, ωA, ωC) but not at (ωB’, ωA, ωC) justifies the assignment of (ωA, ωA, ωC) to a long-range correlation between A and C and not between A’ and C.
We illustrate the above principles with several examples. First, the cross peak at 
(F1, F2, F3) = (54.0 ppm, 33.8 ppm, 52.8 ppm), which does not lie in any diagonal plane, 
can be uniquely assigned to R6Cα-Cβ-C5Cα, which establishes a sequential constraint. 
Second, the cross peak at (54.0 ppm, 39.4 ppm, 34.2 ppm) can be uniquely assigned to 
C31Cα-C30Cβ-E14Cγ, indicating spatial proximity between C30 and E14. Here the 
other four possibilities for the F1 frequency of 54.0 ppm are ruled out because the short 
mixing time prohibits long-distance polarization transfer. This non-diagonal cross peak 
also allowed the peak at (54.0 ppm, 54.0 ppm, 31.0 ppm) to be assigned to C31Cα-
C31Cα-E14Cβ with reasonable confidence, even though it appears on a diagonal plane. 
Another example of a peak with degenerate assignment is the peak at (54.0 ppm, 54.0 
ppm, 16.7 ppm). It can be assigned to XCα-Cα-A28Cβ, where X can be one of the five 
Cα sites at 54.0 ppm. However, one cross peak at (33.8 ppm, 54.0 ppm, 51.3 ppm) in F1 
= 33.8 ppm plane (Figure 4.3c) and another cross peak at (27.1 ppm, 42.5 ppm, 16.5 ppm) 
in Figure 4.3d were detected that can be uniquely assigned to R6Cβ-R6Cα-A28α and 
R16Cγ-R16Cδ-A28Cβ, respectively. These R6-A28 and R16-A28 correlations indicate 
that the (54.0 ppm, 54.0 ppm, 16.7 ppm) peak can be assigned to both R16Cα-R16Cα-
A28Cβ and R6Ca-R6Ca-A28Cβ. In general, non-sequential inter-residue correlations 
manifest as multiple peaks in different planes of the 3D spectrum, thus increasing the 
confidence of assignment.

Figure 4.3b shows the F1 = 49.0 ppm plane. Two squares corresponding to the 
A2 (Cα at 49.5 ppm) and A9 (Cα at 48.8 ppm) spin systems are well resolved. Five 
sequential cross peaks between A2 and C3, A9 and C10 can be uniquely assigned. In 
addition, two sequential cross peaks with a degeneracy of 2 were detected that correspond 
to A9-P8 or A9-C10 correlations. In this spectrum, one non-sequential inter-residue cross 
peak is observed that indicates proximity between P8 and C5 sidechains.

Figure 4.3c shows the F2-F3 plane at F1 = 33.8 ppm, with R6, C10 and E14 spin 
systems, and Figure 4.3d displays the F1 = 27.1 ppm plane with the R6 and R16 spin 
systems. Most non-sequential inter-residue cross peaks fall outside the diagonal planes in 
these two planes, simplifying assignment. Altogether, 26 non-sequential inter-residue
constraints are identified in Figure 4.3, greatly exceeding the 9 non-sequential cross peaks identified in the 100 ms 2D \(^{13}\text{C}-^{13}\text{C}\) DARR spectrum. Most importantly, the 3D assigned long-range correlations are \textit{de novo} rather than back-confirmed with the crystal structure, thus this approach can be applied to proteins with no pre-existing structure or structural models.

\textbf{Distance statistics from 3D CCC and 2D CC spectra and assignment map}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Number of assigned peaks and their degeneracies from 2D and 3D \(^{13}\text{C}\) NMR. (a) Statistics of the 100 ms 2D CC spectrum. (b) Statistics of the 3D CCC spectrum, measured with mixing times of 20 ms and 100 ms. The 3D spectrum contains two orders of magnitude more resolved peaks than the 2D spectrum, with lower assignment degeneracies.}
\end{figure}

In the two previously published papers on the 3D CCC experiments [23,24], no analysis was given about the long-range distances that can be obtained. We now quantify the number of long-range distance constraints as well as the completeness of assignment by the 3D CCC experiment on microcrystalline HNP-1. \textbf{Figure 4.4} shows the number of assigned peaks as a function of assignment ambiguity from the 100 ms 2D CC spectrum versus the 3D CCC spectrum. A total of 24 inter-residue correlation peaks was observed in the 100 ms 2D CC spectrum, including 15 sequential, 1 medium-range and 8 long-range correlations. Among the 24 inter-residue peaks, only 9 were uniquely assigned. The average assignment degeneracy is 2.71 with a standard deviation of 0.41.

In comparison, the 3D CCC spectrum yielded 586 inter-residue correlations peaks, among which 450 were uniquely assigned. Among these unique cross peaks 220 are
sequential and 230 are non-sequential cross peaks. The average assignment degeneracy decreased to 1.30 with a standard deviation of 0.04. Therefore, the 3D CCC experiment not only increased the number of inter-residue cross peaks by two orders of magnitude but also decreased the number of ambiguous assignments significantly. Another benefit of the 3D CCC spectrum is that by having a large number of inter-residue correlations, erroneous assignment is better tolerated compared to the case when a much smaller number of constraints are available.

**Figure 4.5.** 2D contour map of the number of inter-residue cross peaks in HNP-1. A red diagonal line (red) indicating auto-correlation is drawn to distinguish from sequential correlation peaks.

Based on the 3D CCC spectrum, **Figure 4.5** plots the number of uniquely assigned inter-residue correlation peaks as a function of residue number. The number of distance constraints between two residues can be directly read off the map. For example, the numbers of correlations for the (R6, A28) pair and the (A28, R6) pair are 5 and 3, respectively, which result from eight observed cross peaks in the 3D spectrum, five of which are unique: R6Cγ-A28Cα, R6Cα-A28Cα, R6Cβ-A28Cβ, R6Cα-A28Cβ, R6Cβ-A28Cα, and A28Cα-R6Cβ, A28Cβ-R6Cβ, A28Cβ-R6Cα. These five unique constraints were inputted as NOE constraints for structure calculation. **Figure 4.5** also shows that the
correlations C3-C31, Y4-C30, and R6-A28 indicate two antiparallel β-strands β1 and β3, while the E14-C31, R15-F29, R16-A28, I21-L26, Y22-R25 cross peaks indicate another two antiparallel β-strands, β2 and β3. These antiparallel β-strands are consistent with those of the crystal structure (Figure 4.6d). Moreover, three disulfide bond correlations, C3-C31, C5-C20 and C10-C30, which are absent in the 2D spectrum, are now clearly identified, providing another set of important constraints for the HNP-1 structure.

**Structure calculation**

![Figure 4.6](image)

Figure 4.6. Three-dimensional structures of HNP-1 from SSNMR and HNP-3 from crystallography. (a) Average minimum-energy SSNMR structure obtained from long-distance constraints detected in the 2D CC spectrum. (b) Ensemble of 10 minimum-energy SSNMR structures obtained from long-distance constraints detected in the 3D CCC spectrum. (c) Average SSNMR structure of (b). All NMR structures shared the same torsion angle constraints obtained from chemical shifts. (d) Crystal structure of HNP-3. Yellow, purple, and red ribbons represent the three β-strands β1, β2, and β3. Black indicates loops. The three disulfide bonds are shown in orange.

Figure 4.6(a-c) shows the calculated HNP-1 structures where the long-range distance constraints were obtained from the 2D CC experiment and 3D CCC experiment. Both structures are compared with the crystal structure of HNP-3 (Figure 4.6d). As listed in Table 1, a total of 60 distance constraints were obtained from 2D $^{13}$C-$^{13}$C and $^{15}$N-$^{15}$N correlation experiments, including $^{13}$C-$^{13}$C DARR experiments with 20 ms, 40 ms, 100 ms and 200 ms mixing, a CHHC experiment with 100 μs mixing, and a $^{15}$N-$^{15}$N
correlation experiment with 3.0 s mixing [37]. Among these constraints, 36 were sequential \(|i-j|=1\), 4 were medium-range \((1<|i-j|\leq 4)\), and another 20 constraints were long-range restraints \((|i-j|>4)\). Although the long-range cross peaks were assigned by comparing with the crystal structure, due to the small number of distances, the heavy-atom RMSD of the 2D-based NMR structure from the crystal structure is still 3.8 Å (Figure 4.6a). These differences include 1) a bulge in the \(\beta_2\) strand of the NMR structure, which is absent in the crystal structure, 2) a different orientation of the \(\beta_3\) strand, and 3) a different orientation of the loop connecting the \(\beta_1\) and \(\beta_2\) strands. The lack of \textit{de novo} distance determination made it difficult to assess whether the deviation between the NMR structure and the crystal structure results from the low number of distance restraints or from a true conformational distinction between HNP-1 and HNP-3.

Table 4.1 SSNMR structure statistics of HNP-1.

<table>
<thead>
<tr>
<th>Distance constraints and calculation results</th>
<th>From 2D NMR</th>
<th>From 3D CCC NMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential ((</td>
<td>i-j</td>
<td>=1))</td>
</tr>
<tr>
<td>Medium-range ((1&lt;</td>
<td>i-j</td>
<td>\leq 4))</td>
</tr>
<tr>
<td>Long-range ((</td>
<td>i-j</td>
<td>&gt;4))</td>
</tr>
<tr>
<td>Unique</td>
<td>60</td>
<td>213</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>270</td>
</tr>
</tbody>
</table>

| Total \((\phi, \psi)\) torsion angle constraints | 56 | 56 |
| Total energy \(E_{\text{global}}\) (kcal/mol) | 264 | 38.5 |
| Backbone RMSD from crystal structure (Å) | 2.7 | 2.4 |
| Heavy-atom RMSD from crystal structure (Å) | 3.8 | 3.4 |

From the 3D CCC spectrum, we obtained a total of 213 unique distances and 57 ambiguous distances from the 450 uniquely assigned cross peaks. Some atomic pairs exhibit multiple cross peaks in the 3D spectrum, thus increasing the confidence of assignment. Among these distances, 129 are sequential, 45 are medium-range, and 96 are
long-range constraints (Table 4.1). As shown in Figure 4.6b, the ten minimum-energy structures are well clustered, and the average structure of the ensemble (Figure 4.6c) is more similar to the crystal structure (Figure 4.6d) [33] with a heavy-atom RMSD of 3.4 Å. Specifically, the β2 bulge is significantly attenuated in the new NMR structure, and the orientation of the β3 strand is also much closer to that of the HNP-3 structure. The remaining difference is the position of the I7-E14 loop, which can now be more confidently attributed to real conformational differences between the two proteins [37]. We hypothesize that this remaining conformational difference results from the sequence difference between HNP-1 and HNP-3 and is correlated with the antimicrobial activity difference between the two proteins. On average the HNP-1 antimicrobial activity is about 2-fold stronger than HNP-3 [42]. Since the three core β-strands are well constrained by the disulfide bonds and a tight turn, the only conformationally variable region of the protein is the non hydrogen-bonded loop between β1 and β2 strands. By subtly changing the conformation of the protein, this loop may thus influence the protein’s interaction with the target lipid membrane and hence the antimicrobial activity [34].

How robust is the 3D CCC experiment for protein samples with limited sensitivity? If we increase the lowest contour level by a factor of 1.6, effectively removing the weakest cross peaks, then we obtain fewer (323 inter-residue) cross peaks and fewer (145) distance constraints. But even with a smaller subset of constraints, the backbone RMSD between the resulting average NMR structure and the crystal structure remains reasonably low, 2.6 Å. Thus, the information content of the 3D CCC experiment is sufficient to permit de novo structure determination of proteins with lower sensitivities than available here. Moreover, the present distance analysis only focuses on the correlations between aliphatic carbons. If we include carbonyl and aromatic correlations with the aliphatic carbons, then the number of long-range distance constraints will increase further and ease structure determination.
Figure 4.7. Effects of linewidths on the number of resolved cross peaks from 3D CCC and 2D DARR spectra. (a) F1=54.0 ppm plane of the 3D CCC spectrum, processed with Gaussian line broadening values of 0.5 ppm (left), 1.0 ppm (middle), and 2.0 ppm (right). The numbers of cross peaks are 47, 33, and 35, respectively. (b) 2D 100 ms DARR spectrum, processed with Gaussian line broadening values of 0.5 ppm, 1.0 ppm, and 2.0 ppm, from the left to the right. The numbers of cross peaks are 194, 126, and 84, respectively.

It is also interesting to assess how the number of distance constraints from the 3D CCC experiment is affected by linewidths, considering membrane proteins and other non-crystalline proteins typically have broader lines than microcrystalline proteins. As an example, we processed the F1 = 54.0 ppm plane of the 3D spectrum (Figure 4.3a) with different Gaussian line broadening values of 0.5, 1.0 and 2.0 ppm. The number of cross peaks in the aliphatic region (10 – 62 ppm) was found to be 47, 33, and 35, respectively (Figure 4.7a). Thus, at linewidths typical of non-crystalline proteins (1-2 ppm), the number of resolved peaks decreases only modestly, to about 75% of the maximum.
number. Interestingly, the peak number does not decrease from 1.0 ppm to 2.0 ppm linewidths, suggesting that each plane of the 3D spectrum is sufficiently uncongested that it is insensitive to changes in this range of linewidths. In comparison, the number of aliphatic cross peaks in the 100 ms 2D DARR spectrum is 194, 126, and 84 for 0.5, 1.0, and 2.0 ppm line broadening (Figure 4.7b). Thus, the resolved peaks decreased to only 43% of the maximum value. Moreover, the peak number is significantly smaller under 2.0 ppm line broadening than under 1.0 ppm broadening, reflecting the intrinsic spectral congestion. Thus, the 3D CCC spectra are less affected by larger linewidths than 2D CC spectra, making the 3D CCC experiment also robust in terms of resolution.

**Conclusion**

We have shown that the 3D CCC experiment is very useful for resolving and extracting a large number of inter-residue distances that are crucial for *de novo* determination of the three-dimensional structure of proteins by solid-state NMR. By reducing peak congestion, the 3D CCC experiment reduces assignment ambiguity and significantly increases the number of long-range distance constraints. The information content is much larger than several 2D correlation experiments combined. This 3D CCC approach can be readily combined with techniques that preferentially detect long-range distances such as the PAR experiment. Similarly, one can combine the 3D CCC experiments with alternative isotopic labeling schemes to further facilitate long distance extraction.

**Acknowledgment**

This work is supported by an NIH grant GM66976 to M.H and an NSF instrumentation grant DBI421374 for the 600 MHz NMR at Iowa State University.

**References**


[12]  Hong, M.; Gross, J. D.; Griffin, R. G., Site-resolved determination of peptide
torsion angle phi from the relative orientations of backbone N-H and C-H bonds by solid-

[13]  Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat,
H., Structure of a protein determined by solid-state magic-angle-spinning NMR

M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M., Dipole tensor-based atomic-resolution

Structural constraints for the Crh protein from solid-state NMR experiments, J. Bio.

Malliavin, T.; Boeckmann, A., 3D structure determination of the Crh protein from highly

[17]  Grommek, A.; Meier, B. H.; Ernst, M., Distance information from proton-driven

Ziliox, M.; Khorana, H. G.; Sheves, M.; Smith, S. O., Dipolar assisted rotational
resonance NMR of tryptophan and tyrosine in rhodopsin, J. Biomol. NMR 29 (2004)
11-20.

[19]  Lange, A.; Luca, S.; Baldus, M., Structural constraints from proton-mediated rare-
spin correlation spectroscopy in rotating solids, J. Am. Chem. Soc. 124 (2002) 9704-
9705.

transfer dynamics in rotating solids and their use for 3D structure determination, J. Am.

[21]  De Paepe, G.; Lewandowski, J. R.; Loquet, A.; Bockmann, A.; Griffin, R. G.,
245101.


[33] Hill, C. P.; Yee, J.; Selsted, M. E.; Eisenberg, D., Crystal-structure of defensin HNP-3, an amphiphilic dimer - mechanisms of membrane permeabilization, Science 251


Chapter 5

Membrane-Bound Structure and Topology of a Human Alpha Defensin Indicates A Dimer Pore Mechanism for Membrane Disruption

A paper published in Biochemistry

2010, vol. 49, pp. 9770-9782

Yuan Zhang, Wuyuan Lu, and Mei Hong

Abstract

Defensins are cationic and disulfide-bonded host defense proteins of many animals that target microbial cell membranes. Elucidating the three-dimensional structure, dynamics and topology of these proteins in phospholipid bilayers is important for understanding their mechanisms of action. Using solid-state NMR spectroscopy, we have now determined the conformation, dynamics, oligomeric state and topology of a human α-defensin, HNP-1, in DMPC/DMPG bilayers. 2D correlation spectra show that membrane-bound HNP-1 exhibits a similar conformation to the water-soluble state, except for the turn connecting the β2 and β3 strands, whose sidechains exhibit immobilization and conformational perturbation upon membrane binding. At high protein/lipid ratios, rapid $^1$H spin diffusion from the lipid chains to the protein was observed, indicating that HNP-1 was well inserted into the hydrocarbon core of the bilayer. Arg Cζ - lipid $^{31}$P distances indicate that only one of the four Arg residues forms tight hydrogen-bonded guanidinium-phosphate complexes. The protein is predominantly dimerized at high protein/lipid molar ratios, as shown by $^{19}$F spin diffusion experiments. The presence of a small fraction of monomers and the shallower insertion at lower protein concentrations suggest that HNP-1 exhibits concentration-dependent oligomerization and membrane-bound topology. These data strongly support a “dimer pore” topology of HNP-1 in which the polar top of the dimer lines an aqueous pore while
the hydrophobic bottom faces the lipid chains. In this structure R25 lies closest to the membrane surface among the four Arg residues. The resulting pore does not have large lipid disorder, in contrast to the case of protegrin-1, a two-stranded β-hairpin antimicrobial peptide. These results provide the first glimpse into the membrane-bound structure and mechanism of action of human α-defensins.

Introduction

Defensins are host-defense antimicrobial proteins present in many animals and plants. These cationic polypeptides contain three intramolecular disulfide bonds, whose linkage patterns distinguish the α-, β-, and θ-defensins [1]. Humans have six α-defensins of about 30 residues long [2]. The human neutrophil peptides (HNP) 1-4 were found in the azurophilic granules of neutrophils [3-5], while HD-5 and HD-6 are present in intestinal epithelial cells [6,7]. Except for HD-6, the human α-defensins show wide-spectrum antimicrobial activities in the microgram per milliliter range [8,9]. Similar to many other antimicrobial peptides (AMPs), the general mechanism of antimicrobial action is believed to be permeabilization of the microbial cell membrane [10,11].

![Figure 5.1](image)

**Figure 5.1.** Different structural models of HNPs in lipid bilayers. The basic oligomeric state of the protein is a hydrogen-bonded dimer shaped like a basket. (a) Structure of the HNP monomer [12] with Arg residues indicated. (b) Surface-bound wedge model with the hydrophobic basket
bottom inserted into the bilayer and the polar basket top in contact with water. (c) Membrane-inserted wedge model. (d) Dimer pore model, where the hydrophobic basket bottom contacts the lipids while the polar top faces the water pore. (e) General pore model, where the dimers are rotated by 90˚ from the dimer pore orientation. (f) Multimeric pore model, where six dimers form a pore with a ~25 Å inner diameter. The height of the multimer, measured between two Pro8 residues, is 27 Å. The approximate dimensions of the various oligomeric assemblies are indicated in (b-f).

High-resolution X-ray structures of all six human α-defensins in the absence of membrane-mimetic solvents have been determined [12,13]. The structures are very similar despite significant sequence differences between HNP 1-3 and the other three defensins. The proteins consist of three antiparallel β-strands (β1, β2, β3) connected by turns and a longer loop (Figure 5.1a). The proteins are dimerized in the crystal through intermolecular H-bonds between the two β2 strands, thus extending the triple-stranded β-sheet to a six-stranded β-sheet. The dimer has a basket shape and is amphipathic, with a polar top and an apolar base.

Based on the crystal structure of HNP-3 [12], several mechanistic models had been proposed to account for the antimicrobial activities of HNPs. The wedge model [12] suggests that the dimer inserts into the bilayer with the hydrophobic bottom contacting the hydrocarbon region of the membrane, thus disrupting lipid packing (Figure 5.1b, c). The depth of the wedge is unknown, so the dimer may be partly or completely immersed in the lipid membrane. Alternatively, HNPs might form membrane-spanning pores with the hydrophobic base facing the lipids. In the so-called “dimer pore” model [12] (Figure 5.1d), the polar tops of two dimers are oriented to favor transmembrane (TM) orientation of a small solvent channel observed at the dimer interface in the crystal structure [12]. In the “general pore” model [12], the dimers are rotated by 90˚ around the horizontal axis (Figure 5.1e), so that the side view, parallel to the membrane normal, shows the basket shape of the dimer. A multimeric pore model was also proposed based on vesicle leakage and dextran permeability experiments. Here 6-8 HNP dimers may form a large pore with an inner diameter of 20-25 Å [14]. The dimers would be oriented with the long axis of the basket top at approximately 45˚ from the bilayer normal, so that the Arg residues in the
protein are located in two rings separated by about 16 Å along the bilayer normal, promoting favorable electrostatic interactions with the lipid phosphates (Figure 5.1f).

Despite abundant crystal structure information of HNPs, no high-resolution structure information for membrane-bound HNPs has been reported. To elucidate the mechanism of action of human defensins, we have initiated a solid-state NMR study of recombinant HNP-1 uniformly labeled with $^{13}\text{C}$ and $^{15}\text{N}$. Since membrane-bound proteins typically have broad linewidths than proteins outside the membrane, due largely to the conformational disorder of the lipids, we found it necessary to first conduct resonance assignment of the protein in the more ordered water-soluble state, in the absence of lipids. Unlike the bactericidal channel-forming colicins [15-17], which exhibit abundant membrane-induced flexibility, HNPs cannot undergo large conformational rearrangements due to their multiple disulfide bonds. We thus first determined the NMR structure of lipid-free microcrystalline HNP-1 using 2D and 3D magic-angle-spinning (MAS) $^{13}\text{C}$-$^{13}\text{C}$ and $^{13}\text{C}$-$^{15}\text{N}$ correlation experiments [18,19]. Together with distance constraints, these chemical shifts led to the first solid-state NMR structure of a human $\alpha$-defensin, which confirmed the 3D fold seen in the crystal structures (PDB code: 2KHT). This NMR structure determination yielded the complete $^{13}\text{C}$ and $^{15}\text{N}$ chemical shifts of the protein [19], with which the membrane-bound HNP-1 chemical shifts can now be compared. In this work, we investigate the conformation and dynamics of HNP-1 in the membrane, its oligomeric state [20,21], and most importantly, its interactions with lipids and with water. These results yielded the global topology of HNP-1 in the lipid membrane, enabling us to rule out most structural models and allowing us to propose the membrane-disruptive mechanism of HNP-1.

**Materials and Methods**

**HNP-1 expression and synthesis**

Recombinant $^{13}\text{C}, ^{15}\text{N}$-labeled HNP-1 (ACYCRIPACIAGERRYGTCIYGQRLWAFC) (Figure 5.3a) was obtained as a cleavage product from its precursor protein, proHNP1. The residues are numbered from
A2 to C31 due to sequence alignment with other mammalian defensins, many of which contain an additional N-terminal residue before A2 [12]. Using this numbering system also facilitates comparison with the structure of HNP-3, which was the first structure determined in the HNP family. Briefly, the expression protocol starts with expression of proHNP1 as a GST-fusion protein in *E. coli* using uniformly $^{13}$C, $^{15}$N-labeled Spectra 9 media (Cambridge Isotope Laboratories) [19,22]. GST-proHNP1 was folded and then cleaved using thrombin, producing proHNP1. ProHNP1 was purified by reversed-phase HPLC, then cleaved using cyanogen bromide to yield correctly folded HNP-1.

Antimicrobial assays confirmed the activity of the protein. For example, 100% killing of *S. aureus* is reached at 64 µg/ml HNP-1 [19,22]. Crude HNP-1 was purified by reversed-phase HPLC. About 1.5 mg HNP-1 (molecular weight: 3634 Da) was purified from 1 L culture. $^{19}$F-labeled HNP-1 for CODEX experiments was synthesized using t-Boc chemistry [23], where Tyr4 was replaced by 4-$^{19}$F-Phg. The use of phenylglycine removes the possibility of sidechain $\chi_1$ motion that could complicate the interpretation of the CODEX data [24,25]. The choice of Tyr4 as the $^{19}$F labeled site was based on the crystal structures of various HNPs, all showing close inter-subunit contact of this residue [12,13,26]. Previous studies of 4-[$^{19}$F]Phg mutants of the related β-hairpin antimicrobial peptide, PG-1 [24], suggest that the mutation generally does not perturb the antimicrobial activities of the peptides.

**Membrane sample preparation**

DMPC and DMPG lipids (Avanti Polar Lipids, Alabaster, AL) were codissolved in chloroform at a molar ratio of 3:1, dried under a stream of nitrogen gas, suspended in cyclohexane and lyophilized overnight. The resulting homogeneous lipid powder was suspended in a pH 7 phosphate buffer, vortexed and then freeze-thawed five times. The vesicle solution was sequentially extruded through polycarbonate filters with pore sizes of 400 nm and 100 nm to obtain large unilamellar vesicles. HNP-1 was dissolved in 1 mL phosphate buffer, added to the lipid vesicle solution and dialyzed overnight. The mixture was centrifuged at 200,000 g for 3 hours at 4°C to obtain a membrane pellet. The pellet was pipette-transferred into 4 mm MAS rotors for NMR experiments.
Three samples were prepared: two U-\textsuperscript{13}C, \textsuperscript{15}N-labeled HNP-1 samples at protein : lipid (P/L) molar ratios of 1:18 and 1:40, and one \textsuperscript{19}F-labeled HNP-1 sample at P/L = 1:18. The 1:18 molar ratio corresponds to a mass ratio of 1:3.3, thus there is sufficient lipid to ensure appropriate protein-lipid interactions.

\textit{Solid-state NMR spectroscopy}

Solid-state NMR experiments were carried out on a Bruker (Karlsruhe, Germany) AVANCE-600 (14.1 Tesla) spectrometer operating at a \textsuperscript{13}C Larmor frequency of 150.92 MHz for most resonance assignment experiments, and a Bruker DSX-400 (9.4 Tesla) spectrometer for \textsuperscript{13}C-\textsuperscript{31}P REDOR experiments, \textsuperscript{1}H spin diffusion and \textsuperscript{13}C-\textsuperscript{1}H dipolar coupling measurements. Triple-resonance MAS probes with 4-mm rotors were used in all experiments. \textsuperscript{13}C and \textsuperscript{31}P chemical shifts were referenced externally to the \(\alpha\)-glycine \textsuperscript{13}CO signal at 176.49 ppm on the TMS scale and the hydroxyapatite \textsuperscript{31}P signal at 2.73 ppm on the phosphoric acid scale, respectively. \textsuperscript{15}N chemical shifts were referenced to the \textsuperscript{15}N signal of N-acetyl-valine at 122.0 ppm on the liquid ammonia scale.

\textsuperscript{13}C-\textsuperscript{13}C DARR correlation spectra [27] were measured at 273 K under 5 kHz MAS with a \textsuperscript{13}C spin diffusion time of 40 ms. 2D N(CO)CX and N(CA)CX correlation spectra [28,29] were measured at 273 K under 7.5 kHz MAS with a \textsuperscript{13}C mixing time of 40 ms for NCACX and 60 ms for NCOCX. The \(\textsuperscript{15}N-\textsuperscript{13}C\) SPECIFIC [30] cross-polarization (CP) contact times were 3 ms. In the intra-residue NCACX experiment, \(\textsuperscript{15}N-\textsuperscript{13}C\alpha\) magnetization transfer was achieved using a \textsuperscript{15}N spin-lock field of 17 kHz and an on-resonance \textsuperscript{13}C\alpha spin-lock field of 25 kHz. In the inter-residue NCOCX experiment, \(\textsuperscript{15}N-\textsuperscript{13}CO\) magnetization transfer was accomplished with a \textsuperscript{15}N on-resonance spin-lock field of 27 kHz and a tilted \textsuperscript{13}CO spin-lock field strength of 35 kHz, which was the result of a 30 kHz applied pulse on resonance at C\alpha and an 18 kHz offset for CO.

\textsuperscript{13}C-\textsuperscript{1}H dipolar couplings were measured at 303 K using the 2D dipolar chemical-shift (DIPSHIFT) correlation experiment [31]. The sample was spun at 4.5 kHz, and the MREV-8 sequence [32] was used for \textsuperscript{1}H homonuclear decoupling. The t\textsubscript{1}-curves were fit using a Fortran program, and the fit values were divided by the MREV-8 scaling factor to
obtain the true couplings. To account for uncertainties in the MREV-8 scaling factor, we measured the rigid-limit C-H couplings using the crystalline model peptide formyl-Met-Leu-Phe (f-MLF) [33,34], where Leu Cα, Met Cβ, and Leu Cδ signals represented CH, CH₂, and CH₃ groups, respectively. The CH and CH₂ peaks gave C-H rigid-limit dipolar couplings of 11.2 kHz (fit values), which corresponded to true couplings of 23.8 kHz when the theoretical scaling factor of 0.47 for MREV-8 was used [35]. For the Leu CH₃ signal, the fit value was 3.5 kHz. The ratio of the measured HNP-1 couplings with the rigid-limit f-MLF couplings yielded the order parameter S_{CH}.

To determine the depth of insertion of HNP-1, we carried out 2D ¹³C-detected ¹H spin diffusion experiments [36,37] at 303 K or 308 K under 5 kHz MAS. For the P/L = 1:18 sample, ¹H spin diffusion mixing times (tₘ) of 2.25 – 225 ms were applied to transfer the ¹H magnetization of mobile lipids and water to HNP-1. For the P/L=1:40 sample, two mixing times of 100 ms and 225 ms were measured since the sensitivity was too low to measure a complete buildup curve. The ¹H magnetization transfer was detected through the protein ¹³C signals. To ensure that only the mobile lipid and water polarization served as the ¹H magnetization source, we suppressed the ¹H polarization of the rigid protein by a ¹H T₂ filter of 0.4 x 2 ms before the t₁ evolution period. For the 2.25 ms spectrum of the 1:18 sample, the water ¹H cross section was extracted to analyze the water-proximal residues in the protein.

A frequency-selective rotational-echo double-resonance (REDOR) experiment [38] was used to measure protein ¹³C distances to lipid ³¹P. The experiments were performed at 233 K under 4.5 kHz MAS to suppress lipid motions as well as protein sidechain motions. A rotor-synchronized ¹³C Gaussian 180° pulse of 444 µs was applied in the middle of the REDOR period to suppress ¹³C-¹³C J-couplings between the on-resonance ¹³C and its directly bonded spins. ³¹P 180° pulses of 9 µs were applied every half a rotor period to recouple the ¹³C-³¹P dipolar coupling. The time-dependent REDOR intensities were fit using two-spin simulations, which had been shown [39] to correspond to the vertical distance between the ¹³C spin and the ³¹P plane for long distances of greater than ~6 Å.
19F CODEX experiments [20,40,41] for determining the oligomeric state of HNP-1 were carried out at 233 K under 7 kHz MAS. The 19F chemical shift anisotropy (CSA) was recoupled by two rotor periods of 180° pulses. During the 19F mixing time, spin diffusion changes the 19F CSA and prevents complete refocusing of a stimulated echo. A z-filter after the second 180°-pulse train allows the correction of 19F T1 relaxation effects by conducting two experiments for each tm, a control experiment (S0) in which the short z-filter period (10 µs) was applied between the two 180°-pulse trains while the long tm occurred after, and a dephasing experiment (S) where the long tm occurs between the two 180°-pulse trains. The intensity ratio, S/S0, at long mixing times gives the number of 19F spins in close proximity. Two mixing times, 500 ms and 1 s, were measured.

Results

Conformation and dynamics of HNP-1 in the lipid membrane

Figure 5.2. 1D 31P and 13C spectra of membrane-bound HNP-1. (a) 31P static spectra of DMPC/DMPG membranes with HNP-1 at P/L = 1:18 at 303 K (thick line) and without the protein (thin line). The protein-bound spectrum exhibits no isotropic peak and is identical to the spectrum of the non-protein-containing membrane. (b) 31P spectrum of HNP-1-bound DMPC/DMPG membrane at 253 K. (c) 13C MAS spectrum of membrane-bound HNP-1 at 298 K. (d) 13C spectrum of membrane-bound HNP-1 at 273 K. (e) 13C spectrum of membrane-bound HNP-1 at 268 K. The protein intensities and linewidths are
little affected by temperature, suggesting that HNP-1 is immobilized in the liquid-crystalline membrane. (e) $^{13}$C spectrum of microcrystalline HNP-1 without the lipids at 268 K. The intensity distribution of the membrane-bound HNP-1 is similar to that of the microcrystalline protein, indicating similar conformations.

**Figure 5.2a,b** shows the $^{31}$P spectrum of the DMPC/DMPG (3:1) membrane with bound HNP-1 above (303 K) and below (253 K) the membrane phase transition temperature. At 303 K the $^{31}$P spectrum has the classical uniaxial lineshape indicative of lamellar bilayers and is identical to the protein-free lipid spectrum, indicating that HNP-1 does not cause significant orientation disorder to the lamellar bilayer. Below the phase transition temperature (23°C) the $^{31}$P spectrum exhibited the expected broadening, with a rigid-limit span of 210 ppm at 253 K.

We examined the membrane-bound conformation of HNP-1 by comparing the $^{13}$C spectra of the membrane-bound and microcrystalline states. **Figure 5.2c,d** show the 1D $^{13}$C CP-MAS spectra of HNP-1 in DMPC/DMPG bilayers at 298 K and 273 K. The intensity and linewidths were largely unaffected by temperature, suggesting that HNP-1 was immobilized at ambient temperature. Compared to the microcrystalline protein spectrum (**Figure 5.2e**), the membrane-bound sample exhibited similar intensity distributions except for the additional lipid peaks. Thus, HNP-1 overall adopts a similar average conformation in the membrane as in the microcrystalline state, as expected for this disulfide-bonded protein.
Figure 5.3. Resonance assignment of membrane-bound HNP-1 and conformational comparison with the microcrystalline protein. (a) Amino acid sequence of HNP-1. (b) 2D $^{13}$C-$^{13}$C DARR spectrum of membrane-bound HNP-1 at 273 K with a mixing time of 40 ms. (c-e) Regions that show chemical shift changes between the membrane-bound (black) and microcrystalline (red) states. (c) I21 sidechain. (d) R25 sidechain. (e) R25 and R16 sidechains. Note the significant increase of the R25 sidechain intensities in the membrane-bound state in (d) and (e).

To better resolve the resonances, we measured 2D $^{13}$C-$^{13}$C and $^{15}$N-$^{13}$C correlation spectra. The 2D $^{13}$C-$^{13}$C DARR spectrum at 40 ms mixing (Figure 5.3b) showed mostly intra-residue cross peaks, which could be readily assigned by comparison with the microcrystalline spectra [19]. 2D N(CO)CX and N(CA)CX spectra (Figure S1) further corroborated these assignments. Most residues exhibited similar chemical shifts in the membrane as in the microcrystalline form except for the sidechain peaks of I21 and R25. The I21 $C\gamma$ and $C\varepsilon$ peaks moved downfield from the microcrystalline state: the $C\gamma$ chemical shift increased from 24.6 ppm to 25.5 ppm while the $C\varepsilon$ peak moved from 11.5 ppm to 12.4 ppm (Figure 5.3c). In the microcrystalline state, the R25 sidechain resonances were not visible in the 2D $^{13}$C-$^{13}$C spectrum and were much weaker than the
R6 and R16 signals in the 3D NCC spectra [19]. In the membrane-bound state, however, the R25 sidechain resonances became much stronger, with clear C\(\beta\)-C\(\gamma\) and C\(\delta\)-C\(\beta\) cross peaks (Figure 5.3d, e). I21 and R25 are both located in the \(\beta_2-\beta_3\) turn at the hydrophobic base of the dimer (Figure 5.1a). This turn was seen in the HNP-3 crystal structure to deviate significantly from crystallographic symmetry and was thought to exhibit flexing motions [12]. Thus, the membrane-induced conformational and dynamical changes of I21 and R25 sidechains are likely significant, and may reflect site-specific interactions of this \(\beta_2-\beta_3\) turn with lipids (see below). Figure 5.3f shows the C\(\alpha\) and C\(\beta\) secondary chemical shifts [42] of HNP-1 for both the membrane-bound and microcrystalline states, confirming the three \(\beta\)-strand motifs of the protein and the similarity of the protein confirmation without and with the lipids. The Y17-G18 junction in the middle of the \(\beta_2\) strand deviates from the ideal \(\beta\)-sheet conformation, as noted before [19], and may reflect the role of G18 as a hinge for the \(\beta_2-\beta_3\) hairpin [13].

**Figure 5.4.** \(^{13}\)C-\(^1\)H dipolar couplings of DMPC/DMPG-bound HNP-1 at 303 K under 4.5 kHz MAS. (a) \(^{13}\)C chemical shift dimension of the 2D spectrum. (b) C-H dipolar coupling curves of two C\(\alpha\) cross sections and their best-fit, compared with the Leu C\(\alpha\) data of f-MLF. (c) C-H dipolar coupling curves of two CH\(_2\) peaks and their best-fit, compared with the Met C\(\beta\) data of f-MLF. (d) C-H dipolar coupling curves of two CH\(_3\) cross sections and their best-fit simulations at 3.0 and 3.2 kHz, compared to the Leu C\(\delta\) signal of f-MLF. All coupling indicated are fit values,
and convert to true couplings after they are divided by the homonuclear decoupling scaling factor.

To evaluate the mobility of membrane-bound HNP-1 more quantitatively, we measured the $^{13}$C-$^1$H dipolar couplings using the 2D DIPSHIFT experiment. Figure 5.4 shows the $^{13}$C chemical shift dimension of the 2D spectrum, with partial resolution of the sites. Several dipolar cross sections representing the backbone Cα-Hα, sidechain CH$_2$ and CH$_3$ groups are shown. We found rigid-limit $S_{CH}$’s of 0.95–1.05 for Cα and methylene groups, while the methyl groups of Ile and Ala exhibited $S_{CH}$ values of 0.26-0.28. Since methyl three-site jumps alone give an order parameter of 0.33, the small degree of additional scaling (0.78-0.85) indicates that other torsional degrees of freedom in the aliphatic sidechains of these residues are insignificant. Thus the DIPSHIFT data is consistent with the temperature-independent $^{13}$C spectra in supporting that HNP-1 is largely immobilized in the liquid-crystalline phase of the bilayer, with only a moderate amount of sidechain motion. This finding is not only consistent with the conformational rigidity of this disulfide-bonded protein, but more importantly indicates that HNP-1 is not monomeric in the membrane. If the protein were monomeric, the small molecular weight would give rise to fast whole-body uniaxial diffusion of the protein in the membrane, as observed in many small to medium membrane proteins with molecule weights up to several tens of kDa [43-47]. Since all HNPs form dimers in the crystal [12,13,48], the immobilization suggests that when bound to the membrane, HNP-1 is at least dimeric and may form even higher-order oligomers.

**Depth of insertion of HNP-1 from lipid-protein $^1$H spin diffusion**

To determine the insertion motif of HNP-1, we carried out a 2D $^{13}$C-detected $^1$H spin diffusion experiment [36]. The experiment measures inter-proton distances from the mobile water and lipid acyl chains to the rigid protein through distance-dependent $^1$H magnetization transfer. TM proteins in contact with both the membrane surface and the hydrocarbon core exhibit fast $^1$H spin diffusion from both lipid chains and water [37], while surface-bound proteins have much slower $^1$H spin diffusion from the lipid chains.
Figure 5.5. $^1$H spin diffusion from lipid chains and water to HNP-1 in DMPC/DMPG bilayers. (a) Representative 2D spectrum for P/L = 1:18 at 303 K, at a $^1$H mixing time of 100 ms. (b) Integrated $^1$H cross sections from $^{13}$C chemical shifts of 63 - 37 ppm as a function of $^1$H mixing times, compared with the 1D $^1$H spectrum at the top. (c) $^1$H spin diffusion buildup curves for lipid CH$_2$ and water at P/L=1:18. (d) $^1$H spin diffusion intensities for lipid CH$_2$ and water at P/L=1:40. The CH$_2$ buildup is much slower at the lower protein concentration.

Figure 5.5 shows the $^1$H spin diffusion data of membrane-bound HNP-1 (P/L = 1:18) at 305 K. A representative 2D spectrum ($^1$H mixing time 100 ms) shows the characteristic cross peaks between the protein $^{13}$C signals (63 – 37 ppm) and the lipid CH$_2$ protons at 1.3 ppm, indicating that HNP-1 is within spin diffusion reach of the acyl chains. Strong water-protein cross peaks were also observed. To quantify the distances, we measured the 2D spectra as a function of $^1$H spin diffusion mixing time. Figure 5.5b shows the integrated $^1$H cross sections for $^{13}$C chemical shifts of 63–37 ppm for mixing times of 9 to 225 ms. After correcting for $^1$H $T_1$ relaxation of the water and lipid CH$_2$ protons, we plotted the cross-peak intensities as a function of the square root of the mixing time. Figure 5.5c shows that both the water and lipid cross peaks with the protein reached a plateau by 100 ms, qualitatively indicating that HNP-1 is in close contact with both the hydrophobic core of the membrane and water. The equilibration allows both the water and lipid cross peaks to be normalized with respect to their own maximum intensities at 100 ms.

We simulated the CH$_2$ buildup curve using a 1D lattice model to obtain the minimum lipid-protein separation. In this model, $^1$H magnetization ($M_i$) transfer was modeled as a discrete process along the bilayer normal [49]. The transfer rate, $\Omega = D/a^2$,
depends on the lattice spacing $a$, fixed at 2 Å, and the diffusion coefficients of the various membrane components. Based on previous calibrations we used diffusion coefficients of 0.012 nm$^2$/ms for lipids ($D_L$), 0.03 nm$^2$/ms for water ($D_W$) and 0.3 nm$^2$/ms for the protein ($D_P$) [24,36,50]. For the interfacial $D_{LP}$ between the lipid and the protein, we used a value of 0.0025 nm$^2$/ms, which was within the range reported before [24,36,50]. The simulation yields the minimum lipid-protein or water-protein distances because once $^1$H magnetization crosses the interface from the soft lipid matrix (or water) to the nearest protein residues, it is rapidly equilibrated within the immobilized protein. The lack of site specificity allowed us to use the integrated intensities for the protein $^{13}$C signals to obtain global information about the depth of insertion of HNP-1. Figure 5.5c shows that the CH$_2$ buildup curve for the P/L=1:18 sample was best fit to a distance of 2 Å, indicating that HNP-1 was well inserted into the hydrophobic region of the DMPC/DMPG bilayer, in immediate contact with the acyl chains. This close proximity rules out the surface-bound wedge model (Figure 5.1b): if the basket-shaped dimer is exposed to the aqueous phase to any significant extent, for example one third of its 18 Å height, then the membrane-immersed part of the protein would be ~12 Å, which would make the protein mostly outside the hydrophobic region of the bilayer.

Some antimicrobial peptides have been found to exhibit concentration-dependent interactions with the lipid bilayer: at low concentrations they bind to the surface of the lipid membrane while at high concentrations they insert into the membrane [51,52]. To determine whether the depth of insertion of HNP-1 changes with its concentration, we measured the $^1$H spin diffusion spectra of a P/L = 1:40 sample. We measured two mixing times, 100 ms and 225 ms, which were sufficiently long to give equilibrium intensities for the 1:18 sample. The lipid-protein cross peaks at 225 ms were clearly higher than at 100 ms after taking $^1$H T$_1$ relaxation into account, indicating that the CH$_2$ cross peaks had not plateaued by 225 ms. To obtain the correct normalization, we scaled the CH$_2$ cross-peak intensities with the water cross-peak intensities in the 100 ms spectrum, since the water intensities were already equilibrated at 100 ms (Figure 5.5d), consistent with all membrane proteins studied so far [53-55]. This ratio was further scaled by the equilibrium intensity ratio (0.61) between the CH$_2$ peak and the water peak in the 1D $^1$H
spectrum:

\[
I_{CH_2}^{\text{norm}}(t_m) = \frac{I_{CH_2}^{\text{obs}}(t_m) e^{+t_m/T_1,CH_2}}{I_{H_2O}^{\text{obs},100ms} \cdot e^{+t_m/T_1,H_2O}} \left( \frac{I_{eq,CH_2}}{I_{eq,H_2O}} \right) \tag{1}
\]

The normalized CH\textsubscript{2} intensities at P/L = 1:40 are significantly lower than those of the 1:18 sample, indicating that HNP-1 is more shallowly inserted at lower protein concentrations. Thus, HNP-1 has similar concentration-dependent insertion motifs to several other antimicrobial peptides such as protegrin-1 [51,52].

**Membrane topology of HNP-1 from \textsuperscript{13}C-\textsuperscript{31}P distances**

**Figure 5.6.** Arg \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P distances of HNP-1 in DMPC/DMPG bilayers. (a) \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P REDOR S/S\textsubscript{0} values as a function of mixing time and representative S\textsubscript{0} and S spectra. Best-fit simulation used a 1:3 combination of a 4 Å distance and a 7 Å distance. (b)-(e) Solid lines represent REDOR simulations using \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P distances based on the vertical distance differences among the four Arg \textsuperscript{\zeta} \textsuperscript{13}C atoms. Dashed lines represent simulations based on the Arg \textsuperscript{\alpha} vertical distance differences. (b) Membrane-immersed wedge model and the best-fit Arg \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P REDOR curve. Simulated curves decay faster than the experimental data. (c) General pore model and the best-fit \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P curves, which decay faster than the measured data. (d-e) Simulation of the Arg \textsuperscript{\zeta} \textsuperscript{13}C-\textsuperscript{31}P REDOR data
using the dimer pore model. (d) Dimer pore model based on the crystal structure of HNP-3 and a 35 Å bilayer thickness. The simulated Cζ-P REDOR curves decay slower than the experimental data. (e) Dimer pore model after adjusting the Arg sidechain rotameric states (Table 1) and using a 30-Å membrane thickness. The calculated Cζ-P curve fits the data very well.

We also investigated the membrane topology of HNP-1 by measuring 13C-31P distances between the Arg residues and the lipid headgroups. The well resolved Arg Cζ signal at 157 ppm provides a site-specific probe of the interaction of the four Arg’s with the lipid phosphates. Figure 5.6a shows the normalized Cζ intensities, S/S₀, as a function of REDOR mixing time. The intensities decayed quickly to 0.67 in the first 10 ms, then more slowly to ~0.5 by 18 ms. The bimodal decay cannot be fit to a single distance, but to a short distance of 4 Å for the initial fast regime and three longer distances of ~7 Å for the long-time regime. The weighting factor of 1 : 3 reflects the intensity plateau at 0.67 between 8 ms and 14 ms. Thus, the data suggests that one out of four guanidinium ions forms tight H-bonded complexes with the lipid phosphates, while the other Arg’s are more distant.

Table 5.1. Arg Cζ distances to lipid 31P (RCP, Å) in various topology models of HNPs in lipid bilayers.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Original HNP-3 structure 2</th>
<th>35 Å bilayer thickness</th>
<th>Arg-modified structure, 30 Å bilayer thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rotamer RCP, wedge</td>
<td>RCP, general pore</td>
<td>RCP, dimer pore</td>
</tr>
<tr>
<td>R6</td>
<td>mmt180 ᵃ</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>mtt-85</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>R15</td>
<td>mtt180</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>R16</td>
<td>ttt180</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>R25</td>
<td>mtt-85</td>
<td>18</td>
<td>15</td>
</tr>
</tbody>
</table>

1. Each residue in principle has four distances to the membrane surfaces due to the dimer state and two bilayer surfaces. Tabulated here is the shortest distance, which is the one relevant for the 13C-31P distance experiment.
The crystal structure of HNP-3 was used to model the dimer pore [12], since the NMR structure is for the HNP-1 monomer and has lower resolution for the sidechains.

The symbols m, t, p indicate -60°, 180°, and +60°, respectively, for the consecutive χ torsion angles from the backbone [56].

We now consider the compatibility of the various membrane topological models of HNP-1 with the Arg Cζ-P distances. In the fully immersed wedge model (Figure 5.6b, 1c), R6 and R15 are closest to the membrane surface, with their Cζ depths differing by only ~2 Å when the axis of the basket-shaped dimer is oriented parallel to the bilayer normal (Table 5.1) [12]. Thus, if R6 Cζ is 4 Å away from 31P, then R15 Cζ should be about 6 Å away. Using this approach, we estimated all four Cζ-P distances for the immersed wedge model (Table 5.1) and simulated the resulting REDOR curve. Figure 5.6b shows that the simulated REDOR curve decays faster than the experimental intensities, especially in the 8 – 14 ms regime.

Since Arg residues have multiple rotameric states with similar energetic stabilities [56] and the lipid bilayer is expected to exert a strong influence to Arg sidechain conformation through salt bridge interaction [39,57], the Arg sidechains may adopt different conformations in the membrane than found in the crystal. However, any rotamer differences should shorten the Cζ-P distances and speed up the dipolar dephasing, rather than slowing down the dephasing, which would be required to fit the experimental data. When the four Cζ-P distances were estimated based on the relative positions of the four Arg backbone Cα atoms to eliminate sidechain conformational effects, the simulated REDOR curve (dashed line) (Figure 5.6b) still decayed too fast compared to the observed data.

In the general pore model (Figure 5.6c, Figure 5.1e), more than one Arg sidechain was close to the 31P atoms. Specifically, R6, R15, and R16, all have similar Cζ depths based on the crystal structure (Table 5.1) [12]. Keeping the least inserted R15 Cζ-P distance at 4 Å, as constrained by the REDOR data, we obtain R16 and R6 Cζ-P distances of ~6 and 7 Å. The resulting REDOR curve again decays faster than the measured data, and cannot be remedied by changing the sidechain conformation.
Finally, we considered the dimer pore model, where the dimer orientation was rotated by 90° from that of the general pore model (Figure 5.6d, Figure 5.1d). Here the situation is qualitatively different: R25 is now closest to the membrane surface, with the next nearest Arg, R15 in the other monomer, ~4 Å deeper in the membrane. Using the rotamers in the HNP-3 structure, we found the simulated curve to now decay more slowly than the measured data (Figure 5.6d), thus allowing us to modify the sidechain conformations to minimize the Cζ-P distances. In addition, $^{13}$CO-$^{31}$P and $^{13}$Cα-$^{31}$P REDOR data suggested a slight membrane thinning from 35 Å to 30 Å (Figure 5.7).

Combining these two changes, we obtained Arg rotamers that gave Cζ-P distances of 4, 6.5, 9, and 10 Å (Table 5.1), which gave excellent fit to the experimental data (Figure 5.6e). Moreover, R25, the residue with the shortest Cζ-P distance, is also the residue showing significant immobilization upon membrane binding (Figure 5.3). Thus, the conformational changes also support the dimer pore topology of HNP-1.

Backbone CO and Cα distances to $^{31}$P are consistent with the Arg Cζ REDOR data. Both signals decay slowly, with average S/S₀ values of ~0.8 at 14 ms. Simulations yielded best-fit average distances of 7.6 Å for CO and 7.2 Å for Cα. Since no site resolution was attempted, we added the REDOR dephasing of every CO and Cα atoms in each structural model based on their distances to the membrane surface. Figure S2 shows the distance distribution for the dimer pore model, where the shortest distance to $^{31}$P was fixed at 4 Å. The simulated REDOR curves fit the measured dephasing well. In reaching this agreement, it was necessary to thin the membrane to 30 Å, suggesting that HNP-1
insertion slightly reduces the bilayer thickness. Although the backbone $^{13}$C-$^{31}$P REDOR did not exclude the other structural models, it was consistent with the dimer pore structure. In addition, the $^{13}$CO-$^{31}$P REDOR data verified that HNP-1 was inserted into the gel-phase membrane, since simulations using a surface-bound wedge model were inconsistent with the experimental result (**Figure S3**).

**Residue-specific water-protein distances**

**Figure 5.8.** Water-protein $^1$H spin diffusion to determine the membrane topology of HNP-1. (a) $^{13}$C spectrum after 2.25 ms $^1$H spin diffusion from water (red), superimposed with the $^{13}$C CP spectrum (black). Simulated stick spectrum based on the dimer pore model is shown in blue. (b) $^{13}$C CP spectrum and simulation to fix the CP weighting factors. (c) Difference between the experimental and simulated water-edited spectrum from (a). (d) Dimer pore model, where residues close to water are shown in blue, interfacial residues in gray, and residues far from water in red. The respective intensity weighting factors are 1, 0.55 and 0.1.

To complement the topology information from the lipid-to-protein spin diffusion experiment, we examined the water-to-protein $^1$H spin diffusion profile at the short
mixing time of 2.2 ms, when the $^1$H spin diffusion from water to the protein has not fully equilibrated in the protein, thus giving site-specific information about residue distances to water. **Figure 5.8a** shows the water-edited $^{13}$C spectrum (red) extracted from the water $^1$H cross section of the 2D spectrum. This water-edited spectrum was superimposed with the $^{13}$C CP spectrum, which reflects the equilibrium $^{13}$C magnetization. The two spectra were scaled such that the water-edited spectral intensities at most equaled but not exceeded the CP intensities. It can be seen that the intensity distribution of the water-edited spectrum is clearly different from that of the CP spectrum due to the selectivity of the water-to-protein spin diffusion at this short time. For example, the lipid chain signals from 32.5 ppm to 22 ppm were much attenuated, as expected because of the immiscibility of water with the hydrocarbon core. The protein signals in the 53-40 ppm region, which mainly result from $\text{C}_\alpha$ and $\text{C}_\beta$, were also reduced relative to the $\text{C}_\alpha$ intensities in the 60-55 ppm range.

To simulate the water-edited $^{13}$C spectrum, we first reproduced the equilibrium CP spectrum by using appropriate weighting factors for each carbon. These weighting factors accounted for different CP dynamics of CH$_n$ (n=1, 2,3) groups and the different mobilities between the backbone and sidechains (**Figure 5.8b**). Next, we assigned each carbon a spin-diffusion weighting factor based on the proximity of each site to water on the membrane surface and water in the pore. Both the crystal structure of HNP-3 [12] and the solid-state NMR structure of HNP-1 [19] were considered in estimating the carbon-water distances, and the simulation was insensitive to the exact starting structure. Three categories of spin diffusion weighting factors were used. Carbons close to water were given a weighting factor of 1, while those far from water a weighting factor of 0.1. Residues in the middle were assigned a weighting factor of 0.55. Combining the water-proximity weighting factors with the CP weighting factors, we simulated the water-edited spectrum expected for the dimer pore topology. **Figure 5.8a** shows that the simulated stick spectrum is in good agreement with the measured spectrum. The RMSD difference between the two was 0.22 as normalized to the 54-ppm peak intensity (**Figure 5.8c**), comparable to the experimental noise of 0.18, indicating that the measured spectrum is consistent with the dimer pore topology. We did not consider the 33 – 22 ppm region in
the simulation due to significant overlap between the lipid and protein signals in this region. The use of the solid-state NMR structure for the choice of spin diffusion weighting factors did not affect the good fit (Figure S4).

**Oligomeric structure of HNP-1 in the lipid membrane**

![Figure 5.9](image)

**Figure 5.9.** $^{19}$F CODEX control ($S_0$) and dephased (S) spectra of 4-$^{19}$F-Phg$_4$-labeled HNP-1 in DMPC/DMPG (3:1) membranes at P/L = 1:18. The mixing time was 1 s. The S/$S_0$ value was 0.66 ± 0.06.

So far we have assumed that HNP-1 assembles into a basic unit of dimers in the membrane, upon which loose higher-order oligomers may form. Although the immobilization result supports this assumption, we further verified this hypothesis using the $^{19}$F CODEX experiment. The experiment detects $^{19}$F magnetization exchange between orientationally different singly fluorinated proteins. The oligomeric number $n$ is obtained from the CODEX echo intensity at long mixing times, which is $1/n$. We substituted Tyr4 in the $\beta$1 strand by 4-$^{19}$F-Phg, since crystal structures suggest this residue to lie in the dimer interface, with inter-monomer distances of less than 10 Å (PDB codes: 1DFN, 3HJ2). Figure 5.9 shows the $^{19}$F CODEX control ($S_0$) and exchange (S) spectra of HNP-1 in DMPC/DMPG membranes (P/L = 1:18) at a mixing time of 1 s. The S/$S_0$ value was 0.66 ± 0.06 and was the same between 1 s and 500 ms within experimental uncertainty, indicating that intermolecular spin diffusion has equilibrated. The slightly larger than 0.5 final value indicates that the majority (66%) of $^{19}$F-labeled Tyr$_4$ exists in a dimer state ($n$
while the rest (33\%) has $^{19}$F-$^{19}$F distances greater than \(~15\) Å. Thus, at a protein concentration of about 5 mol\% in the membrane, the majority of HNP-1 is tightly dimerized, while the rest either exists in a monomer state or forms loose dimers with Tyr4 separations greater than the upper limit of this $^{19}$F spin diffusion technique.

**Discussion**

*Evidence of pore formation by HNPs in lipid membranes from biochemical data*

Since the discovery of human $\alpha$-defensins [3,4], many antimicrobial assays, lipid vesicle experiments, and high-resolution structures have been reported to understand the mechanisms of action of these host-defense proteins [1,58]. Similar to smaller $\beta$-hairpin antimicrobial peptides, bactericidal activity by HNPs followed inner membrane permeabilization [59]. Dye release and fluorescence spectroscopy experiments found that HNP-1 caused both fusion and lysis of negatively charged lipid vesicles through electrostatic interactions [60]. For rabbit neutrophil defensins, vesicle leakage depends on the membrane composition: it is all-or-none for whole *E. coli* lipids but graded for POPG vesicles [61]. Electron micrographs of human parasite cells *Trypanosoma cruzi* in the presence of micromolar concentrations of HNP-1 showed distinct 25-nm sized pores in the cellular and flagellar membranes [62], through which HNP-1 appears to enter the trypanosome cells, causing subsequent DNA fragmentation and cell destruction. Vesicle leakage experiments also showed that HNP-induced pores increase with the concentration of the anionic lipid [63]. These biochemical studies all indicated that human $\alpha$-defensins cause pore formation in anionic lipid membranes. However, the molecular structure and topology of HNP-1 at the pore and the type of lipid disorder has remained elusive.

*Structural constraints on HNP-1-induced pores in anionic lipid membranes*

The $^1$H spin diffusion, $^{13}$C-$^{31}$P REDOR, and $^{19}$F CODEX results combined to give the following constraints to HNP-1 structure in the anionic lipid membrane. At high protein concentrations of P/L = 1:18, HNP-1 fully spans the membrane and contacts the hydrophobic chains. Among the four Arg’s, one Arg forms hydrogen-bonded complexes with the lipid phosphates. This Arg is most likely R25 in the $\beta2$-$\beta3$ turn, because its
signal was enhanced in the 2D correlation spectra by lipid-induced immobilization. The DMPC/DMPG bilayer is thinned slightly upon HNP-1 binding. Seen at Tyr$_4$, the majority of the protein is at least dimerized at P/L = 1:18. These observations support the dimer pore model for HNP-1 in the membrane. The fully immersed wedge model and the general pore model would place R25 to be the furthest Arg residue from the membrane surface, which is inconsistent with the 2D $^{13}$C spectra, because the center of the hydrocarbon region is the most fluid region of the bilayer and should not immobilize R25. The depth distribution of the four Arg sidechains in the fully immersed wedge model and general pore model also do not agree with the stoichiometry that only one Arg C$_\zeta$ is H-bonded to the lipid phosphates. The data also rule out a surface-bound wedge model where the polar basket top is significantly exposed to water, since it does not agree with the fast $^1$H spin diffusion from the lipid chains to the protein. In the multimeric pore model, the Arg backbones lie at roughly the same depths in each lipid leaflet, with the two C$_\alpha$ rings separated by ~16 Å. Thus, the C$_\zeta$ atoms are expected to be similarly close to the membrane surface $^{31}$P, which does not agree well with the distance distribution of the C$_\zeta$-P REDOR data. At lower protein concentrations (P/L=1:40), HNP-1 is less inserted into the membrane, as seen in the slower lipid-protein $^1$H spin diffusion. Although sensitivity limitations preclude the determination of the exact topology, it is reasonable to expect the surface-bound wedge model to be the most likely scenario.

In model-specific fitting of the various experimental data, we primarily used the crystal structure of HNP-3 (PDB: 1DFN) because the solid-state NMR structure [19] (PDB: 2KHT) is that of the monomer, with no direct intermolecular constraints for the dimer. Nevertheless, Figure S4 shows that the hypothetical dimer NMR structure would give similar conclusions as the HNP-3 crystal structure for the water-edited $^{13}$C spectrum, suggesting that the exact input structure does not affect the conclusion of the global topology of HNP-1.
Figure 5.10. Membrane topology of HNP-1 in anionic lipid bilayers. At high protein concentrations HNP-1 is mostly dimerized and spans the membrane, lining a central water pore. The R25 sidechain lies closest to the membrane surface interacting with the phosphate groups. The lipid chains near the pore do not exhibit significant orientational disorder. (a) Side view. (b) Side view with the pore behind one dimer. (c) Top view. The distances between the two dimers and the total oligomeric number of the assembly are not probed by the experiments reported here.

Taken together, the solid-state NMR data support the dimer pore topology, where HNP-1 dimers span the membrane with the R25-containing β2-β3 turn pointing at the membrane surfaces (Figure 5.10). When water on the membrane surface and in the pore is both considered, the dimer pore topology reproduces the observed water-edited protein $^{13}$C spectrum. The orientation and depth of each dimer is well constrained by the NMR data, and the $^{19}$F CODEX data supports the dimerization of the majority of the protein. The small fraction of monomers is likely related to the observed shallower insertion of the protein at lower concentrations, and suggests that concentration-dependent oligomerization may be important for the membrane-disruptive activity of HNP-1.

Does the dimer pore mechanism of HNP-1 apply to other human α-defensins? Comparisons of the activities and specificities of the six human α-defensins indicate that their interactions with lipid membranes are diverse. Against Gram-positive S. aureus, the relative potencies were HNP-2 > HNP-1 > HNP-3 > HNP-4, while the relative potencies against the Gram-negative E. coli were HNP-4 > HNP-2 > HNP-1 = HNP-3 [8]. Among HNP 1, 2, and 3, whose amino acid sequences differ only in their N-terminal residue, HNP-3, which possesses an N-terminal polar residue Asp, has weaker activities than
HNP-1 and HNP-2, which have a hydrophobic residue at the N-terminus [5,8]. Consistently, acetylation and amidation of HNP-2 to remove the terminal charges modulated the protein’s antimicrobial activity and vesicle leakage [64]. Thus, the positive charge density and charge distribution have a significant effect on the membrane interaction of HNPs. In comparison, HD5 and HD6, which are found in intestinal epithelial cells, have much lower sequence homologies to HNPs [6,7]. HD6 has nearly two orders of magnitude weaker activities than the HNPs [8], while HD5, while localized on the cell membrane, was suggested to possibly interact with the cells in a receptor-mediated fashion [65]. Based on the high sequence homology and functional similarity between HNP-1 and HNP-2, we speculate that the dimer pore mechanism may apply to HNP-2 and should be relevant to HNP-3 as well, but the other defensins in this family may adopt different orientations and depths of insertion in the lipid membrane.

**Comparison of HNP-1 with small β-hairpin antimicrobial peptides**

Our previous studies of the two-stranded disulfide-bonded β-hairpin antimicrobial peptide, PG-1, indicated that the strong interactions between Arg guanidinium ions and the lipid phosphate groups drove the formation of toroidal pore defects in the membrane, where PG-1 lines the pore as a TM β-barrel [24,39,55]. When the guanidinium ions were dimethylated, the mutant showed 3-fold weaker antimicrobial activities and no longer formed large β-barrels in the membrane [66]. A PG-1 mutant with only half the number of Arg residues inserted only partially into the anionic membrane and exhibited much weaker interactions with the lipid headgroups [50]. Guanidinium-phosphate interactions were also observed in Arg-rich cell-penetrating peptides, penetratin [67] and the HIV TAT peptide [68]. In HNP-1, although only one guanidinium group is within H-bonding distance to the lipid phosphates, the average distance for the other three Arg residues is 7 Å, which is short compared to the bilayer thickness (Figure 5.6a). Lu and coworkers have shown by mutagenesis that when three of the four Arg’s were converted to Lys, HNP-1 activity was significantly weakened, and the effect was more pronounced against the Gram-positive *S. aureus* than the Gram-negative *E. coli* [69].

It is interesting that no isotropic peak was observed in the $^{31}$P NMR spectra,
indicating that high-curvature defects such as micelles or toroidal pores were absent in the HNP-bound DMPC/DMPG membrane (Figure 5.2). Thus, in the DMPC/DMPG membrane, the HNP-1 dimer pores exist in regular lamellar bilayers, consistent with the classical barrel-stave model. The retention of the bilayer integrity contrasts with the behavior of PG-1, which caused substantial membrane disorder to phosphocholine (PC)/phosphatidylglycerol (PG) membranes as well as phosphatidylethanolamine (PE)/PG membranes. The HNP-1 interaction with the DMPC/DMPG membrane is more akin to that of tachyplesins [70] and a synthetic antimicrobial arylamide [71], which did not disrupt PC/PG membranes. However, tachyplesin-1 caused clear disruption to the PE/PG membrane [70]. The extent of membrane disorder depends on both the distribution of Arg residues in the protein sequence and the composition of the lipid membrane. It is possible that HNP-lipid interactions are sensitive to the membrane composition, similar to tachyplesins, so that while HNP-1 does not disrupt PC/PG membranes, it may disrupt PE/PG membranes. Similarly, an increased percentage of the negatively charged PG lipids may cause a large amount of membrane disorder. A study of rabbit neutrophil defensins suggested that the presence of PE and cardiolipin lipids increased the membrane disruption [61]. Given the increased complexity of HNPs over two-stranded β-hairpin antimicrobial peptides, structural investigations as a function of membrane composition will be useful to further elucidate the membrane interaction and mechanism of action of this class of defensins. Finally, the precise antimicrobial mechanism of HNP-1 in vivo may depend on factors other than HNP-phospholipid interaction. It was recently reported that HNP-1 activity correlates with the amount of lipid II, a bacterial cell wall precursor [63]: inhibition of lipid II synthesis weakened the HNP-1 antibacterial activity. This result suggests that interaction of cell wall components with HNP to be involved in the antimicrobial action of HNPs, and may explain why D-amino-acid analogs of HNPs were found to have weaker bactericidal activities but similar membrane-disruptive abilities compared to their L-amino acid counterparts. Thus, the mechanism of action of HNPs, similar to nisin [72], may involve multiple targets in the bacteria. The dual mechanisms may also contribute to the lack of significant membrane disorder seen in the $^{31}$P NMR spectra.
References


[12] Hill, C. P.; Yee, J.; Selsted, M. E.; Eisenberg, D., Crystal structure of defensin


[34] Rienstra, C.; Tucker-Kellogg, L.; Jaroniec, C.; Hohg, M.; Reif, B.; McMahon, M.; Tidor, B.; Lozano-Pérez, T.; Griffin, R., De novo determination of peptide structure with


[56] Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C., The penultimate


Supporting Information

Figure S1. 2D $^{15}\text{N}$-$^{13}\text{C}$ correlation spectra of HNP-1 bound to DMPC/DMPG membranes to confirm resonance assignment. The Arg and L26 resonances are shown in red. (a) N(CA)CX spectrum. (b) N(CO)CX spectrum. R15 signals are not observed in these spectra, indicating it is dynamically disordered at the temperature of the experiment, 273 K.
Figure S2. REDOR distance simulations using the dimer pore model. (a) $^{13}\text{C}O^{13}\text{P}$ distance distribution and the resulting REDOR curve, superimposed with the experimental data. (b) $^{13}\text{C}\alpha^{13}\text{P}$ distance distribution and the resulting REDOR curve, superimposed with the experimental data.
Figure S3. Simulations of the $^{13}$CO-$^{31}$P REDOR distances for the surface-bound wedge model to verify that HNP-1 is inserted into the membrane at low temperature. (a) If the HNP-1 wedge is partly immersed into the membrane with a quarter of its full height (4.5 Å), then the simulated CO-P REDOR dephasing disagrees completely with the measured data. (b) If the wedge has its bottom on the membrane surface, then the simulated dephasing still contradicts with the data. (c) If the protein is 2 Å above the membrane surface, then the simulated dephasing will agree with the data. However, this scenario of the protein “floating” above the membrane surface contradicts the Arg Cζ-P REDOR data. Therefore, HNP-1 is well inserted into the membrane at low temperature.
Figure S4. Simulation of the water-to-protein $^1$H spin diffusion spectrum at 2.25 ms mixing using a dimer pore model where the solid-state NMR-determined HNP-1 structure was used as input. (a) Experimental $^{13}$C spectrum superimposed with the stick spectrum (blue) simulated based on the proximity of residues to water in the dimer pore model. (b) Difference between the experimental and simulated spectra in (a). (c) Simulation (blue) of the experimental $^{13}$C CP spectrum. (d) Dimer pore model, where residues close to water are shown in blue, residues far from water shown in red, and the interface residues in gray.

Reference

Chapter 6

Bicelle Alignment and Orientation Simulation of HNP-1 in Bicelle Systems

Introduction

In solid-state NMR of membrane proteins, aligned membrane samples are often used to obtain information about peptide orientation in the lipid membrane and membrane disruption by the peptide. In the presence of water, different lipids and combination of lipids self-assemble to form phospholipid bilayers that enable the proteins to maintain their native structures and functions. The proteins can be uniaxially aligned along with the phospholipids in the magnetic field of the NMR spectrometer [1]. Two of the most common oriented lipid systems are lipids mechanically aligned on glass plates [2,3] and lipid bicelles aligned magnetically [4-7]. Mechanically aligned lipid bilayers have the advantage that they allow wider choices of lipid combinations. However, they suffer from low sample stability due to the difficulty of maintaining sample hydration in the presence of high-power radiofrequency irradiations [8]. Bicelle samples have the disadvantage of limited choices of lipid compositions, since not all lipids will align in a magnetic field, and restricted temperature range for each lipid composition. But they have the benefit of being easier and quicker to make and allowing measurements in an MAS probe.

The bicelle membrane is intermediate between micelles and bilayers [9,10]. It consists of a mixture of long-chain phosphatidylcholine (PC) lipids (between 12 and 18 carbons) and short chain PC lipids (6 to 8 carbons). The long-chain lipids form the planar bilayer region of the bicelle while the short-chain lipids ‘cap’ the rim of the bilayer [11]. Bicelles are characterized by q, the molar ratio of long- to short-chain lipids. Small ‘isotropic’ bicelles (q<0.5) reorient rapidly with no preferred direction of alignment
[7]. We are interested in larger bicelles with $q > 2.5$, which are aligned by the strong magnetic field, forming a liquid crystalline phase with the bilayer normal perpendicular to the direction of the magnetic field [12]. This orientation is driven by the cooperative diamagnetic susceptibilities of the lipid molecules. A glycerophospholipid with saturated chains has the parallel diamagnetic susceptibility, $\chi_\parallel$, (along the lipid long axis) smaller than the perpendicular $\chi_\perp$, resulting in a negative anisotropy, $\Delta \chi = \chi_\parallel - \chi_\perp$ [13]. Bicelles can also be ‘flipped’ 90° by the addition of lanthanide ions [14]. Because such lanthanides possess a strong positive $\Delta \chi$, they can turn the bilayer normal to be uniaxially aligned parallel to the magnetic field [15]. Protein structures can be determined in either perpendicular or parallel bicelles [1,16].

The common short-chain lipids used for bicelles are DHPC, 6-O-PC, and DCPC (Yuan, include any other before the introduction of Triton). Recently Opella and coworkers reported bicelles with enhanced alignment order and improved stability [17]. This system consists of long-chain phosphatidylcholine lipids mixed with the short-chain lipid Triton X-100 with a molar ratio of 5:1 ($q = 5$). They prepare the sample by adding a solution containing Triton X-100 to the proteoliposome pellet with a lipid concentration of 20% (w/v). Upon formation of the magnetically aligned bilayer phase, the translucent proteoliposome pellet turns into a clear, non-viscous solution when placed in ice. This bicelle sample exhibits narrow line widths of $\sim 0.5$ ppm and remains well-aligned over a broad range of temperature. It is also stable for long periods of time (months).

In this chapter, we study the alignment of bicelles with different lipid combinations, long- to short-chain lipid ratios and hydration levels. The alignment of bicelles with peptides was also studied and a phase diagram was generated. We show that the orientation of HNP-1, a dimeric protein with 3 $\beta$-strands in each monomer, can determined in principle by 2D correlation spectra of $^{15}$N anisotropic chemical shift and N-H dipolar coupling. We show simulated 2D spectra employing site-specifically labeled residues.
**Procedure for preparing bicelles**

This procedure roughly follows the protocol described by De Angelis and Opella [11], while additional papers are referenced for comparing the sample preparation conditions [18-20]. First, the desired amount of the short-chain lipid such as 6-O-PC is weighed out. It is important for the short-chain lipid to be non-hydrolyzable to increase the long-term stability of the sample. In addition, the lipid mass needs to be accurate to obtain the desired q. Thus, it is a good practice to dry the lipid in a lyopholizer for 2-4 hours and then reweigh to confirm the mass. The dry mass is then used to calculate the appropriate amount of DMPC (the long-chain lipid) required to reach the desired q value. We usually use q = 3.2. Dissolve 6-O-PC in sufficient water or buffer to give a lipid/water ratio of 35% (w/v). The micelle solution should be clear and homogeneous at this stage. Add this micelle solution to the long-chain lipid powder. The resulting mixture at this step is heterogeneous, with water on top of the lipid. To create a homogeneous mixture, the sample is heated to 42 °C with hot water, and cooled to 0 °C in an ice bath, then vortexed for 30 seconds. Repeat the cool-thaw cycle three more times until the solution becomes homogeneous and the entire amount of long-chain lipid is dissolved. Even if there are no visible particles floating in the solution, the sample should be placed in a 4ºC refrigerator overnight to allow the lipid to dissolve and the sample to equilibrate.

The next day, the bicelle solution should be viscous at room temperature and become non-viscous and fluidic at 0°C. If this viscosity test is passed, then the bicelle sample can be checked by $^{31}$P NMR for degree of alignment. To transfer the sample into the rotor, we place it in the refrigerator for ~5 minutes to cool it to the solution phase, together with the pipet tip we are going to use. Once the sample becomes fluid, we use the pipet tip to take out the solution and transfer it into the rotor. This step needs to be done quickly to avoid the sample returning to the gel phase at room temperature in the pipet tip. The $^{31}$P single-pulse experiment is carried out on the 600 MHz NMR spectrometer (14.1 T) using a static $^{31}$P probe containing a solenoid coil whose axis is perpendicular to the magnetic field. $^{31}$P chemical shifts are referenced to 85% H$_3$PO$_4$ at 0 ppm. The temperature of alignment can be different from expected, since it depends
sensitively on the the long- to short-chain lipid ratio and hydration level. Thus a range of temperatures (298-320 K) should be scanned before identifying the temperature for the best bicelle alignment. Normally a 3 K difference is sufficient to cause a different lineshape.

After a well aligned bicelle sample is prepared and characterized, it can be added directly to the peptide using a precooled pipet tip. At the beginning, the peptide may not be completely dissolved. This problem can be overcome by the cool-and-thaw cycles. A clean needle is also useful to physically stir the mixture and speed up dissolution. After the peptide is added to the bicelle solution, the resulting mixture should be allowed to sit overnight to equilibrate. With a high peptide concentration, the solution can be viscous at all temperatures. The sample is now ready for $^{31}$P NMR.

**Factors Affecting Bicelle Alignment**

![Figure 5.1](image)

**Figure 5.1.** $^{31}$P spectra of different phases of lipids. (a) Magnetically-oriented bicelle phase. The sample is DMPC/6-O-PC with 65% hydration at 310 K. (b) Aligned bicelle coexisting with isotropic bicelles (arrow). The sample is DMPC/6-O-PC in the presence of RC100b with 65% hydration at 314 K. (c) Isotropic bicelles. The sample here is DMPC/6-O-PC containing RC100b, with 70% hydration at 314 K after pulsing overnight. (d) Unoriented bilayers mixed with isotropic bicelles. The sample here is DMPC/6-O-PC with 70% hydration at 303 K and the spectra is collected right after the sample was put into the magnetic field.

Static $^{31}$P spectra are used to characterize the phase of the bicelle lipid mixture. The lipid mixtures can form different phases under different temperature-composition-
hydration conditions, resulting in different static $^{31}$P lineshapes (Figure 5.1). For the successfully oriented bicelle, the $^{31}$P spectrum shows two single sharp peaks for the long-chain lipids and short-chain lipids at anisotropic chemical shift frequencies. The intensities of the two peaks should reflect the q value. For oriented bicelles coexisting with isotropic bicelles or micelles, the spectrum shows a sharp peak at 0 ppm representing the isotropic phase, in addition to the two anisotropic chemical shift peaks for the oriented bicelles. When the lipids are fully in the isotropic phase, the spectrum will have only one peak at 0 ppm and the two bicelle peaks disappear. When the lipids are unaligned but do not form small isotropic vesicles, then a uniaxial powder pattern is observed, representing unoriented lamellar bilayers.

**Lipid Ratio**

**Figure 5.2.** Phase diagrams of two bicelles as a function of temperature and q value. (a) Phase diagram of DMPC/DCPC with 80% hydration [21]. (b) Phase diagram of DMPC/DCPC with 75% hydration [22]. The phase diagrams are summarized based on $^{31}$P NMR spectra reported in the literature. The upward triangle represents a single phase of bicelles oriented along the magnetic field. The downward triangle stands for oriented bicelles coexisting with isotropic or unoriented phases. The opened circle denotes unoriented and isotropic phases. The open rectangle denotes the isotropic phase. The open diamond stands for the lamellar phase. [21] [22]

The lipid ratio q, which is defined as the molar ratio of long- to short-chain lipids, is one of the major factors affecting bicelle alignment. Raffard et al. [21] and Triba et al. [22] have examined the bicelle alignment with different q ratios at fixed hydration levels by $^{31}$P NMR (Figure 5.2). For the DMPC (14:0 chains) /DCPC (1,2-dicaproyl-sn-glycero-
3-phosphocholine, 6:0 chains) system [21] with q ratios less than 2, the lipids exhibit a single sharp line centered at -0.4 ppm, which is the isotropic chemical shift of phosphatidylcholines. This reflects the formation of micelles or small isotropic bicelles at all temperatures. When the q ratio is larger than 9, the spectrum is composed of a single axially symmetric powder pattern, reflecting an unoriented lamellar phase at all temperatures. The region in which bicelles self-orient in the magnetic field is roughly delineated by $q = 2.5-7$ and $T = 298 – 318$ K. Similarly, for DMPC/DHPC (1,2-diheptanoyl-sn-glycero-3-phosphocholine, 7:0 chains) systems [22], aligned bicelle phase occurs between $q = 2-5$ and $T = 308 – 323$ K. Thus, for our DMPC/6-O-PC sample, we choose the q ratio of 3.2, which is in the middle of the q range that is supposed to give oriented bicelles in the magnetic field.

**Figure 5.3.** Chemical structures of DMPC, DHPC, DCPC and 6-O-PC.

People have found that increasing the chain length of the short-chain and long-chain lipids shrinks the zone where the oriented bicelles are observed [22]. For instance, the DMPC/DCPC system has a bicelle phase ranges from $q = 2$ to $q = 7$, while the DMPC/DHPC system’s bicelle phase shrinks to $q = 2-5$. We have measured the $^{31}$P spectra for both DMPC/6-O-PC and DMPC/DHPC systems at different temperatures.
The sample containing 6-O-PC, which has a shorter aliphatic chain, gives the narrowest FWHM of 0.9 ppm and the line shape is narrow from 297 K to 316 K. In contrast, for the DMPC/DHPC system the narrowest line width is 1.2 ppm and the lines start to broaden significantly (~2.5 ppm) below 297 K.

**Figure 5.4.** Factors affecting bicelle alignment. For all samples shown here, q = 3.2. (a) DMPC/6-O-PC at 310 K with 65% hydration (top) and DMPC/DHPC at 300 K with 65% hydration (bottom). (b) DMPC/6-O-PC with 65% hydration. The top spectrum is prepared in 25 mM HEPES buffer and aligned at 310 K and the lower one is prepared in 10 mM HEPES buffer and aligned at 306 K. (c) DMPC/DHPC mixture with different hydration levels: 70% at 310 K, 65% at 300 K and 60% at 320 K. (d) $^{31}$P spectra of DMPC/6-O-PC with (at 310 K) and without the peptide RC100b (at 314 K).
**Bicelle Alignment in the presence of salt**

Previously, people have found that the temperature range and the center of the bicellar domain do not change with the addition of salt (ref[??]). In other words the salt does not appear to be necessary for bicelle formation. However, it did shift the bicellar domain in the direction of increasing short-chain lipids, in other words the bicelles form at lower q in the presence of salt than in the absence of salt [23].

We measured the $^{31}$P spectra of DMPC/6-O-PC with different concentrations of HEPES buffer, 10 mM and 25 mM. The line width of the sample prepared with 10 mM HEPES buffer is broader than at 25 mM, indicating that salt facilitates the alignment of the bicelle in the magnetic field. The reason for this better alignment may be that the addition of ions may increase the absolute value of $\Delta \chi$, thereby influencing the lipid alignment with respect to the bilayer normal.

**Bicelle Alignment as a Function of Hydration**

Hydration level is another factor that affects bicelle alignment. We have prepared several DMPC/DHPC samples with hydration levels of 60%, 65% and 70%. Among them, the 65% hydrated sample gave the narrowest line width (0.9 ppm). From the literature, people have found that the sample with lower hydration level tends to be more stable over a wider temperature range and aligns in the field more easily [21]. This is probably because of tighter packing of the bicelle, which leads to less swelling. Our later investigation of bicelles with peptides also shows that DMPC/6-O-PC containing retrocyclin has a better alignment at 65% hydration than at 70% hydration. This result further supports the statement that moderate dehydration leads to better bicelle alignment. The reason why 60% hydration has a broader FWHM than the 65% is probably because the bicelles are too crowded to have the free space to orient.

**Bicelle with Peptides**

Addition of peptides to bicelles can interfere with alignment and lead to broader line widths. For example, the best FWHM we can obtain for DMPC/6-O-PC bicelle alone is 0.9 ppm at 65% hydration. After adding the retrocyclin into the bicelle at a
peptide/lipid molar ratio of 1/40, the $^{31}$P line width increased to 1.5 ppm. The alignment temperature range also changes by the addition of peptide. Without retrocyclin, the alignment temperature range is 300 K – 310 K. With retrocyclin, the best alignment temperature increased to 315 K. Decreasing the peptide concentration should help the bicelle alignment.

In conclusion, the alignment of bicelles is sensitive to several factors such as lipid composition, temperature and hydration levels. The bicellar domains are found for lipid molar composition between 63% and 85% for long-chain lipid (q = 2-5), for a temperature range of 25-45 ºC and for hydration levels from 40% to 95% (w/w). Increasing the chain length for both short-chain and long-chain lipids, for example, going from DMPC/DCPC to DPPC/DHPC, reduces the zone where the oriented discs are observed. The presence of salts can increase the bicelle domain and sometimes improves the bicelle orientation in the magnetic field. Moderate dehydration can also favor alignment due to tighter packing of the bicelles [21].

**Orientation Simulation of HNP-1**

The orientation of a membrane protein can be determined from the two-dimensional spectra correlating $^1$H-$^{15}$N dipolar coupling and $^{15}$N chemical shift. A common method for measuring such spectra is PISEMA [24,25]. Previous studies showed that the PISEMA spectra of membrane proteins in oriented lipid bilayers are an sensitive indicator of the $\alpha$-helical structure and topology. The resonance frequencies in the PISEMA spectra depend on the helix orientation, as well as backbone dihedral angles, the magnitudes and orientations of the principal element of the amide $^{15}$N chemical shift tensor. Because the $^{15}$N chemical shift tensor is reasonably well characterized for protein backbone amides, it is possible to calculate solid-state NMR spectra for specific structural models of proteins.

For helical proteins, the characteristic ‘wheel-like’ patterns of resonances observed in these spectra reflect the helical wheel projections of residues and therefore provide direct information on the secondary structure and topology of membrane proteins in phospholipid bilayers. These patterns are referred as PISA (Polarity Index Slant Angle)
wheels [24,25]. In this study, we calculate the 2D $^{15}\text{N}-^{1}\text{H}/^{15}\text{N}$ correlation spectra for ideal $\alpha$-helices and $\beta$-sheets using a FORTRAN program, then simulate the spectrum expected for the $\beta$-strand rich HNP-1. The Fortran program defines a molecule-fixed coordinate system that reflects the $\alpha$-helix axis and geometry and calculates the anisotropic frequencies based on the orientation of the magnetic field ($B_0$) in this coordinate system. For $\alpha$-helices, the z axis of this reference system is the $\alpha$-helix axis, and is obtained as the average orientation of an even number of consecutive N-H bonds in the helix. The x-y plane is defined as the plane perpendicular to the $\alpha$-helix axis. There are two parameters to define the orientation of the helices: the tilt angle $\tau$, which is between the $\alpha$-helix axis and the $B_0$ field, and the rotation angle $\rho$, which is defined as the angle between the projection of $B_0$ onto the x-y plane and the radial axis of the helix on the x-y plane that passes through the C$\alpha$ carbon of the second residue. The z axis of the $^{15}\text{N}$ chemical shift tensor is 17º from the N-H bond, whereas the x axis is 25º from the peptide plane. The rigid limit N-H dipolar coupling is 10 kHz, corresponding to a bond length of 1.07 Å. Literature $^{15}\text{N}$ chemical shift principal values of (64, 77, 217) ppm were used to simulate the orientation dependent spectra. The $^{15}\text{N}$ anisotropic chemical shift frequencies $\delta_{||}$, which is the frequency when the protein is uniformly aligned and the alignment axis parallel to the magnetic field (0º -aligned sample), can be calculated by

$$\delta_{||} = \omega_{\text{aligned}} = \frac{1}{2} \delta (3\cos^2\tau - 1 - \eta \sin^2\tau \cos 2\rho) + \omega_{\text{iso}}$$

Here $\delta = \delta_{zz} - \delta_{\text{iso}}$ and $\eta = (\delta_{yy} - \delta_{xx})/\delta$ are the anisotropy and asymmetry parameters, respectively, of the rigid-limit interaction tensor. The N-H dipolar coupling can be derived by the equation below:

$$\omega_{NH} = \frac{1}{2} (3\cos^2\tau - 1) \omega_{NH,\text{rigid}}$$

Here $\omega_{NH,\text{rigid}}$ is the rigid limit N-H dipolar coupling.
Figure 5.5. Simulated PISEMA spectra for the full range of possible orientations of an ideal α-helix ($\phi = -65^\circ$, $\psi = -40^\circ$) aligned by glass plates (b). (a) Literature reported simulations [25]. (b) Our own simulations using the Fortran program. The two sets of simulations show perfect agreement.

Figure 5.5 shows the calculated PISEMA spectra for the full range of possible orientations of an ideal α-helix with identical backbone dihedral angles of ($\phi = -65^\circ$, $\psi = -40^\circ$) for all residues. The peptide is assumed to align with the bilayer normal parallel to the $B_0$ field, as applicable for glass-plate aligned samples. When the helix axis is parallel to the bilayer normal ($\tau = 0^\circ$), all the amides have the same orientation relative to the magnetic field, therefore all the sites give the same N-H dipolar coupling and $^{15}$N chemical shift frequencies. Since the amide backbone N-H bonds in the α-helix are nearly, but not exactly, parallel to the helix axis, the resonance frequencies approach, but do not reach, their maximum values of 10 kHz. Tilting the helix away from the
membrane normal breaks the symmetry, inducing variations in the orientations of the amide N-H bond vectors relative to the field. For all helix orientations other than parallel to the field direction, the PISEMA spectra have characteristic wheel-like patterns whose frequency breadths reflect the extent of the helix tilt. When the helices are tilted to be parallel to the membrane surface, their amide N-H bonds and one of the $^{15}$N chemical shift tensor are nearly orthogonal to the magnetic field and give highly overlapped PISEMA spectra with all resonance frequencies around 5 kHz and 75 ppm. The N-H dipolar coupling values indicated here are not scaled by the homonuclear decoupling scaling factor.

β-strand peptides have very different geometries and are defined in the following way. The z axis of the reference system is the β-strand axis, and is calculated as the average orientation of an even number of consecutive C'_{i-1}-N_{i} bonds. The y-z plane, which is the local β-sheet plane, is defined as the common plane containing the z axis and a specific C=O vector. We used the C=O bond of the second residue to define the y-z plane. The tilt angle $\tau$ is defined as the angle between the β-strand axis and the $B_0$ field, which is the bilayer normal if the sample is aligned by glass plates. The rotation angle $\rho$ is defined as the angle between the y axis and the projection of $B_0$ onto the x-y plane.

Figure 5.6. Simulated PISEMA spectra of β-strands ($\phi = -140^\circ$, $\psi = -135^\circ$) with different orientations relative to the magnetic field aligned by glass plates (b). (a) Literature reported
simulations [26]. (b) Our own simulations using the Fortran program. For each $\tau$ angle, we did several simulations for $\rho$ from 0º to 360º. And the $\rho = 0º$ points were shown in filled circles.

Figure 5.6 shows the calculated PISEMA spectra of $\beta$-strands with several $\tau$ values relative to the magnetic field. The ($\phi$, $\psi$) torsion angles for the strands residues are (-140º, 135º). Same as the $\alpha$-helix, when the strand long axis is oriented parallel to the bilayer normal, all amides have identical orientation relative to the magnetic field, thus all resonances overlap at a single N-H dipolar coupling and $^{15}$N chemical shift value. However, unlike the $\alpha$-helix, of which the resonance frequencies approach the maximum attainable values of 10 kHz, the N-H dipolar couplings of $\beta$-strands approach, but do not reach, a value of 5 kHz. This is because the amide N-H bonds in $\beta$-strands are nearly perpendicular to the strand axis. Therefore, the maximum attainable value is only half of the $\alpha$-helix

$$\omega_{NH} = \frac{1}{2}(3\cos^2 \theta - 1) \times \frac{1}{2}(3\cos^2 \tau - 1) \omega_{NH, rigid} = \left(-\frac{1}{2}\right)\frac{1}{2}(3\cos^2 \tau - 1) \omega_{NH, rigid}.$$ 

When the $\beta$-strand is tilted away from the bilayer normal, the PISEMA spectra trace out butterfly-like resonance patterns. Because $\beta$-strands lack cylindrical symmetry around the main molecular axis, their 2D spectra depend characteristically on both the tilt angle of the strand axis and the rotation angle of the $\beta$-sheet plane, while the PISA wheel patterns of $\alpha$-helix are distinguished mostly by the tilt angle of the helical axis and not by the helix rotation angle.

What we were discussing above is the orientation simulation for proteins in glass-plate oriented bilayers. The bicelle sample is a different case because it is aligned with bilayer normal perpendicular to the magnetic field, while that of the glass-plate sample is parallel to the magnetic field. This difference introduces a scaling factor for both the N-H dipolar coupling and $^{15}$N chemical shift frequencies. For the N-H dipolar coupling:

$$\omega_{NH, bicelle} = \frac{1}{2}(3\cos^2 \theta - 1) \times \frac{1}{2}(3\cos^2 \tau - 1) \omega_{NH, rigid}$$

$$= \frac{1}{2}(3\cos^2 \theta - 1) \omega_{NH, glass-plate}.$$
Here $\theta = 90^\circ$ because of the different orientation between bicelle and glass plate. Hence:

$$\omega_{\text{NH bicelle}} = -\frac{1}{2} \omega_{\text{NH glass plate}}$$

For the $^{15}$N chemical shift frequencies:

$$\delta_{\text{bicelle}} = \frac{1}{2} (3 \cos^2 \theta - 1) \times \frac{1}{2} \delta (3 \cos^2 \tau - 1 - \eta \sin^2 \tau \cos 2\phi) + \omega_{\text{iso}}$$

$$\delta_{\text{bicelle}} = \frac{1}{2} (3 \cos^2 \theta - 1) \times \delta_{\text{ani glass plate}} + \omega_{\text{iso}}$$

Again, with $\theta = 90^\circ$:

$$\delta_{\text{bicelle}} = \left(-\frac{1}{2}\right) \times \delta_{\text{ani glass plate}} + \omega_{\text{iso}}$$

Therefore, the F1 dimension of the PISEMA spectra (N-H dipolar coupling) of the bicelle sample is half of that of the glass plate. The F2 dimension ($^{15}$N chemical shift) of the bicelle is generated by first scaling down the glass plate spectra by a factor of 2 and then taking the mirror image around the $\omega_{\text{iso}}$ (120 ppm). As a result, when $\tau = 0^\circ$, the PISEMA spectrum of $\alpha$-helix in bicelles is a single point at (71 ppm, 5 kHz) and the $\beta$-sheet peak is at (148 ppm, 2.5 kHz).

For a real protein with non-ideal secondary structures, the expected 2D spectra become more complex. The human $\alpha$-defensin HNP-1 consists of three antiparallel $\beta$-strands connected by turns and a long loop. The protein is dimerized through intermolecular H-bonds between the two $\beta_2$ strands. The dimer has a basket shape with a polar top and an apolar base. Since the three $\beta$-strands in this protein are not completely parallel to each other and the relatively long loop region and the dimerization make the structure more complex, we cannot easily calculate the PISEMA spectrum of this protein by simulating the orientation of the three $\beta$-strands separately and then combining them together. Thus we used a different strategy to calculate the PISEMA spectra. We defined a Cartesian coordinate system with x, y, z axes and added the coordinates into the PDB file of the protein. Thus when we rotate these axes, we also rotate the whole body of the
protein. Then we read out the coordinates of the protein and obtain the orientation of the N-H bond vectors and the $^{15}\text{N}$ chemical shift vectors. We place the z axis of the reference coordinate system to be perpendicular to the plane of the top of the basket, the x axis to be parallel to the axis connecting the two R25 Cα sites in the two monomers, and the y axis to be perpendicular to the x-z plane. The tilt angle $\tau$ is defined as the angle between the z axis and the bilayer normal. The rotation angle $\rho$ is defined to be the angle between the projection of the bilayer normal onto the x-y plane and the y axis.

Previously people have proposed several mechanistic models to explain the antimicrobial activity of HNPs [27,28]. The wedge model suggests that the dimer inserts into the bilayer with the hydrophobic bottom contacting the hydrocarbon region of the membrane, which makes the $\tau = 0^\circ$. The dimer pore model suggests that the polar top of the dimer is oriented to favor the transmembrane orientation of the solvent channel at the dimer interface, which has $(\tau, \rho)$ angles of $(90^\circ, 90^\circ)$. In the general pore model, the dimer is rotated by $90^\circ$ around the z axis of the dimer pore model. Thus the $(\tau, \rho)$ angle is $(90^\circ, 0^\circ)$ for the general pore model.
Figure 5.7. Simulated $^{15}$N chemical shift – $^1$H-$^{15}$N dipolar coupling 2D correlation spectra for three orientations of HNP-1 in the lipid membrane. (a) Dimer pore model, with $\tau = 90^\circ$ and $\rho = 90^\circ$. (b) Wedge model, with $\tau = 0^\circ$. (c) General pore model, with $\tau = 90^\circ$ and $\rho = 0^\circ$.

We chose three residue types, leucine (Leu26), tyrosine (Tyr4, Tyr17, Tyr22) and arginine (Arg5, Arg15, Arg16, Arg25), as the proposed labeled sites. Figure 5.7 shows the simulated 2D spectra for the three orientational models. In addition to the labeled sites, several residues were also calculated to verify the validity of the simulation. For example, in the dimer pore model, the N-H vector of A12 is almost parallel to the x axis, thus parallel to the bilayer normal. Therefore, the N-H dipolar coupling of A12 should be close to the maximum value 10 kHz. The simulated A12 value is 8.8 kHz, which is close to 10 kHz. In the general pore model, the N-H vector of I7 is parallel to the z axis, which means it is perpendicular to the bilayer normal. As a result, the N-H dipolar coupling of I7 should be around 5 kHz. And in the calculated spectrum, the I7 N-H dipolar coupling is 4.1 kHz, in agree with the prediction.
The calculated spectra above are for glass-plate samples with the bilayer normal 0° with $B_0$. For bicelle samples, we need to consider the $(\frac{1}{2})$ scaling factor because of the different orientation between the bicelle axis and the glass plate alignment axis. As we discussed above, the N-H dipolar coupling of the bicelle sample is half of that of the glass plate. The $^{15}$N chemical shift of the bicelle is generated by first scaling down the glass plate spectra by a factor of 2 and then taking a mirror image around the $\omega_{\text{iso}}$ (120 ppm). Therefore, we can determine the orientation of the HNP-1 in the lipid bilayers by comparing the experimental PISEMA spectra with the calculated 2D spectra for the proposed models.

Reference


Appendix A

Protocol for purification of $^{15}$N, $^{13}$C labeled HNP-1

1. Pasteurize

Prepare four 1 L flask, use tin to seal. Put the flasks into autoclave, set the temperature to be 121 °C, solution 30 min, dry 40 min. After the autoclave complete, take out the flasks and put them onto a clean table (with tin).

Dissolve LB Agar into deionized water (35 g/L), add kanamycin with final concentration of 50 µg/ml (1000 times dilution). Put the solution into autoclave (121 °C, solution 30 min). Then transfer the solution onto several culture dishes. After it cools down, use parafilm to seal and store in 4 °C freezer.

Prepare 1 M IPTG solution (2.38 g for 10 ml), use 0.22 µm membrane to filter the solution into ten 1.5 ml eppendorf tube. Use parafilm to seal, store in -20 °C freezer.

2. Cell culture and protein expression

Take out the E. coli bacteria from 80 °C freezer (using seeding tool). Use flame to burn the seeding tool for half minute, let it sit for half minute for cooling. Take out some bacteria using the seeding tool (just slightly touch). Inoculate the bacteria onto the solid media. Put the culture dish into the 4 °C freezer. Let it sit overnight. (Plasmid : pGEX-2T-Kana + GST_ProHNP-1; Host cell: Ecoli BL21L (pLysS DE3)

Next morning, near the flame, add 250 ml Spectra9 CN media into one 1 L flask. Add 250µl 50 mg/ml kanamycin solution using 1 ml pipet. Use flame to burn the seeding tool for half a minute, let it sit for half a minute for cooling. Take out the culture dish. Pick one separate clone (~2 mm diameter) using the seeding tool. Put the seeding tool into the media to inoculate the cell. Then use flame to burn the tool. Use tin to seal the flask and shake the cell at 250 rpm, 37 °C overnight (from 4 pm to 10 am). Also put the remaining 750 ml media into 37 °C room overnight to warm up.
Next morning, add 750 µl 50 mg/ml kanamycin solution into the remaining 750 ml Spectra9 CN media. Mix the overnight 250ml media with the rest media. Transfer it into 4 x 1 L flasks (250ml each). Continue to shake cell at 250 rpm, 27 °C for 1.5 hr (from 10 am to 11:30 am). Use 1 ml pipet to suck out 750 µl culture and measure the UV-vis. Continue shaking until the UV-vis reaches 1.2 OD (600 nm). Induce cells with final concentration of 3mM IPTG (750 µl to 250 ml media). Continue shaking for 2 hr (from 12 pm to 2 pm).

Spin cells at 7000 g rpm for 15 min at 4 °C in four 500ml spin tubes. Then pour out the supernatant and use 4 x 10 ml PBS buffer to wash the cells and transfer it into 2 x 50 ml tubes by 5ml pipet. Spin the cell at 7000 rpm for 10 min and pour out the supernatant. Get the wet weight of the cells (2 x 2 g). Completely suspend the cell in 2 x 10 ml (5 ml/g) Bugbuster solution. Add 2 x 5 µl Bioprecessing reagent into the two tubes. Shake for 20 min. Centrifuge at 10000 g rpm for 20 min. Pour out the supernatant and wash the inclusion bodies with 2 x 20 ml PBS + 2M urea solution. Spin at 10000 g rpm for 10 min. Pour out the supernatant. Add 1.5 g DTT (in total for the two bottles) and dissolve the inclusion bodies with 200 ml 8 M urea solution.

Use 0.45 µm membrane to filter the solution. Adjust pH to 4–6 before loading into the HPLC column for desalting (column: C18, 10-15 µm). Use 10% B (acetonitrile) for washing (~20 min) and 65% B for elution (the first peak should appear in ~10 min after changing to 65% B). The flow rate is 40ml/min and insert rate is 20 ml/min. The detector is at 214 nm and 280 nm. A is water + 1% TFA. Get ~200 ml solution and freeze it in 1 L round flask with liquid nitrogen. Lyophilize overnight (usually need 2–3 days).

3. **Protein refolding and digestion**

Get ~300 mg lyophilized inclusion body. Dissolve it with 250 ml 8 M urea solution (~1 mg/ml). Add 364 mg cystein (for 1 L) into the solution (final concentration is 3 mM). Stir after the solid is completely dissolved. Dilute the solution with 750 ml 0.2 M NaHCO₃ (three times volume, 16.8 g in 750 ml H₂O). Start the refolding process with adding 72 mg cystine into the solution and stir slowly at room temperature overnight.
The next morning, prepare the cleavage buffer. Dissolve 67.8 g Tris-HCl (20 mM), 245.4 g NaCl (150 mM), 10.3 g CaCl₂ (1.5 mM) into 4 L water. Adjust the pH to be 8.4 (using 37% HCl). Dilute this solution 7 times within the 30 L tank. Transfer the 1 L folded protein solution into four 250 ml dialysis tubes (MWCO=10000). Dialyze the protein solution against the cleavage buffer. Change buffer twice within 24 hr (Start at 10:30 am, change buffer at 6:00 pm).

The next morning, transfer the protein solution into a 1 L bottle. Use dionized water to rinse the tube. Add 33 µl bovine α-thrombin into the 1 L solution and stir slowly overnight.

4. Purification of labeled proHNP-1

The thrombin digestion product needs to be adjust to pH 4~6 (using acetic acid, usually ~4.5) and filtered with 0.45 µm filtration membrane. Set the linear gradient acetonitrile from 25 to 45% within 1 hr. Collect target elution peaks and lyophilize the proHNP-1. The volume is ~30 ml (3~4 tubes).

5. Cleavage of proHNP-1 with CNBr

The CNBr is already prepared as 50 mg/ml solution containing 5% TFA and stored in 4 C fridge. Dissolve the 5 mg proHNP-1 with 10 ml water. Put the solution into the hood. Add 10ml CNBr solution into the peptide with gloves. Stir at room temperature overnight.

Adjust the pH of the solution to be 4~6. Set the linear gradient acetonitrile from 25 to 45% within 1 hr. Collect target elution peaks and lyophilize the HNP-1. The volume is ~30 ml (3~4 tubes).

After lyophilization, we can get 2-3 mg HNP-1 powder for 1 L media.

6. Crystallization of HNP-1

The crystalized HNP-1 was prepared by precipitation of protein with poly(ethylene glycol)400 (PEG-400). A solution of PEG-400 containing 30 mM Cacodylate and 60 mM Li₂SO₄, with pH = 6.5 was slowly added to the stock solution. The final concentration of PEG-400 is ~30%. Two batches were precipitated with stock
solution concentration of 36 mg/ml and 30 mg/ml. The high-concentration stock solution (36 mg/ml) yields microcrystals one hour after mixing with PEG. Figure B.1 shows the picture of the peptide solution after mixing with PEG.

After half an hour:  

![Image of peptide solution after half an hour mixing with PEG.]

After 1 hour:  

![Image of peptide solution after 1 hour mixing with PEG.]

After 1 day:  

![Image of peptide solution after 1 day mixing with PEG.]

After 2 days:  

![Image of peptide solution after 2 days mixing with PEG.]

**Figure B.1.** Crystal picture of HNP-1 after mixing with PEG as a function of time.

The low-concentration stock solution (30 mg/ml) produced microcrystals gradually in a 4-day period, with ~80% of the protein eventually precipitating from the supernatant, which is estimated by the residual absorbance of the supernatant at 280 nm. The hydrated precipitate was centrifuged and transferred to the rotor after removal of supernatant.
Appendix B

Input code for protein structure determination by Xplor-NIH

1. anneal.py

xplor.requireVersion("2.14.4")

# slow cooling protocol in torsion angle space for protein G. Uses
# NOE, RDC, J-coupling restraints.
#
# this script performs annealing from an extended structure
#
# CDS 2005/05/10
#

# this checks for typos on the command-line. User-customized arguments can
# also be specified.
#
(optxs, args) =
xplor.parseArguments(["quick"])  

quick=False
for opt in optxs:
  if opt[0]=="quick": #specify -quick to just test that the script runs
    quick=True
    pass
  pass

# filename for output structures. This string must contain the STRUCTURE
# literal so that each calculated structure has a unique name. The SCRIPT
# literal is replaced by this filename (or stdin if redirected using <),
# but it is optional.
#
outFilename = "SCRIPT_STRUCTURE.sa"
numberOfStructures=20   #usually you want to create at least 20
if quick:
  numberOfStructures=3
  pass

# protocol module has many high-level helper functions.
#
import protocol

protocol.initRandomSeed()   #set random seed - by time

command = xplor.command

# generate PSF data from sequence and initialize the correct parameters.
#
from psfGen import seqToPSF
seqToPSF('HNP1.seq')

# generate random extended initial structure with correct covalent geometry
#
protocol.genExtendedStructure()
# a PotList contains a list of potential terms.
# This is used to specify which
# terms are active during refinement.
# from potList import PotList
fromList = PotList()

# parameters to ramp up during the
# simulated annealing protocol
# from simulationTools import MultRamp,
# StaticRamp, InitialParams
rampedParams=
highTempParams=[

# compare atomic Cartesian rmsd with a reference structure
# backbone and heavy atom RMSDs will be printed in the output
structure files
# from posDiffPotTools import create_PosDiffPot
refRMSD =
create_PosDiffPot("refRMSD","name CA or name C or name N",
                            pdbFile='1DFN.pdb',
                            cmpSel="not name H")

# orientation Tensor - used with the dipolar coupling term
# one for each medium
# For each medium, specify a name, and initial values of Da, Rh.
# from varTensorTools import create_VarTensor
media={}
# medium Da rhombicity
for (medium,DA,Rh) in [(t', -6.5, 0.62),
                        (b', -9.9, 0.23)]:
# oTensor = create_VarTensor(medium)
oTensor.setDa(Da)
oTensor.setRh(Rh)
media[medium] = oTensor

# dipolar coupling restraints for protein amide NH.
# collect all RDCs in the rdcPotList
# RDC scaling. Three possible contributions.
# 1) gamma_A * gamma_B / r_AB^3
# for different expts. in the same medium.
# Sometimes the data is
# prescaled so that this is not needed.
scale_toNH() is used for this.
# Note that if the expt. data has been prescaled, the values for rdc rmsd
# reported in the output will relative to the scaled values- not the expt.
# values.
# 2) expt. error scaling. Used here. A scale factor equal to 1/err^2
# (relative to that for NH) is used.
# 3) sometimes the reciprocal of the Da^2 is used if there is a large
# spread in Da values. Not used here.
# from rdcPotTools import create_RDCPot,
scale_toNH
rdcs = PotList('rdc')
for (medium,expt,file, scale) in [
    ('t','NH' ,'tmv107_nh.tbl' ,1),
    ('t',NCO,'tmv107_nc.tbl' ,.05),
    ('t',HNC,'tmv107_hnc.tbl' ,.108),
    ('b','NH' ,'bicelles_new_nh.tbl' ,1),
    ('b',NCO,'bicelles_new_nc.tbl' ,.05),
    ('b',HNC,'bicelles_new_hnc.tbl' ,.108):
    rdc =
create_RDCPot("%s_%s"%(medium,expt),file,media[medium])
#1) scale prefactor relative to NH
   # see python/rdcPotTools.py for exact
calculation
   # scale_toNH(rdc) - not needed for these
datasets -
   # but non-NH reported
   rmsd values will be wrong.

   #3) Da rescaling factor (separate
multiplicative factor)
   # scale *= ( 9.9 / rdc.oTensor.Da(0)**2
   # rdc.setScale(scale)
   # rdc.append(rdc)
   # pass
#potList.append(rdc)
#rampedParams.append( MultRamp(0.01,1.
0, "rdc.setScale( VALUE )") )

   # set up NOE potential
   noe=PotList('noe')
potList.append(noe)
from noePotTools import create_NOEPot
for (name,scale,file) in
[('all',1,"HNP1_NOE.tbl"),
   #add entries for additional
tables
   ]:
   pot = create_NOEPot(name,file)
   # pot.setPotType("soft") # if you think
   there may be bad NOEs
   pot.setScale(scale)
   noe.append(pot)
rampedParams.append( MultRamp(2,30,
   "noe.setScale( VALUE )") )

   # setup J coupling - with Karplus
   coefficients
   #from jCoupPotTools import create_JCoupPot
   #jCoup =
create_JCoupPot("jcou","jna_coup.tbl",
   # A=6.98,B=-
   1.38,C=1.72,phase=-60.0)
   #potList.append(jCoup)

   # Set up dihedral angles
   from xplorPot import XplorPot
   protocol.initDihedrals("HNP1_dihe.tbl",
   useDefaults=0)
   potList.append( XplorPot('CDIH') )
   highTempParams.append( StaticRamp("potList[CDIH].setScale(10)") )
   rampedParams.append( StaticRamp("potList
   [CDIH].setScale(200)") )
   # set custom values of threshold values for
   violation calculation
   #
   potList[CDIH].setThreshold( 5 )

   # radius of gyration term
   #
   protocol.initCollapse("resid 1:56",
   Rtarget=10.16)
   potList.append( XplorPot('COLL') )

   # hbda - distance/angle bb hbond term
   #
   #protocol.initHBDA('hbda.tbl')
   #potList.append( XplorPot('HBDA') )

   #Rama torsion angle database
   #
   protocol.initRamaDatabase()
   potList.append( XplorPot('RAMA') )
   rampedParams.append( MultRamp(.002,1,"p
   otList[RAMA].setScale(VALUE)"))

   #
   # setup parameters for atom-atom repulsive
term. (van der Waals-like term)
   #
   #potList.append( XplorPot('VDW') )
   rampedParams.append( StaticRamp("protoc
   ol.initNBond()") )
   rampedParams.append( MultRamp(0.9,0.8,
"command('param nbonds repel VALUE end end')")
"command('param nbonds rcon VALUE end end')")

# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("protocol.initNBond(cutnb=100,
tolerance=45,
onlyCA=1)""")

potList.append( XplorPot("BOND") )
potList.append( XplorPot("ANGL") )

potList['ANGL'].setThreshold(5)
rampedParams.append( MultRamp(0.4,1,"potList['ANGL'].setScale(VALUE)"")
potList.append( XplorPot("IMPR") )
potList['IMPR'].setThreshold(5)
rampedParams.append( MultRamp(0.1,1,"potList['IMPR'].setScale(VALUE)"")

# Give atoms uniform weights, except for the anisotropy axis
# from atomAction import SetProperty
import varTensorTools

AtomSel("not rename ANI").apply( SetProperty("mass",100.) )
varTensorTools.massSetup(media.values(),300)
AtomSel("all ").apply( SetProperty("fric",10.) )

# IVM setup
#  the IVM is used for performing dynamics and minimization in torsion-angle
# space, and in Cartesian space.
# from ivm import IVM
dyn = IVM()

# initialize ivm topology for torsion-angle dynamics
for m in media.values():
    m.setFreedom("fixDa, fixRh")  # fix tensor Rh, Da, vary orientation
    m.setFreedom("varyDa, varyRh")
    protocol.torsionTopology(dyn,oTensors=media.values())

# minc used for final cartesian minimization
# from simulationTools import AnnealIVM
init_t = 3500.  # Need high temp and slow annealing to converge
cool = AnnealIVM(initTemp =init_t, finalTemp=25, tempStep =12.5, ivm=dyn, rampedParams = rampedParams)

# cart_cool is for optional cartesian-space cooling
cart_cool = AnnealIVM(initTemp=init_t,
      finalTemp=25,
      tempStep=12.5,
      ivm=minc,
      rampedParams=rampedParams)

def calcOneStructure(loopInfo):
    # this function calculates a single structure, performs analysis on the structure, and then writes out a pdb file, with remarks.

    # generate a new structure with randomized torsion angles
    #
    from monteCarlo import randomizeTorsions
    randomizeTorsions(dyn)

    # calc. initial tensor orientation
    #
    from varTensorTools import calcTensorOrientation
    for medium in media.values():
        calcTensorOrientation(medium)
        pass

    # initialize parameters for high temp dynamics.
    InitialParams(rampedParams)
    # high-temp dynamics setup - only need to specify parameters which differ from initial values in rampedParams
    InitialParams(highTempParams)

    # high temp dynamics
    #
    protocol.initDynamics(dyn,
                         potList=potList,  # potential terms to use
                         bathTemp=init_t,
                         initVelocities=1,
                         finalTime=800,  # stops at 800ps or 8000 steps
                         numSteps=8000,   # whichever comes first
                         printInterval=100)
    dyn.setETolerance(init_t/100)  # used to det. stepsize. default: t/1000
    dyn.run()

    # initialize parameters for cooling loop
    InitialParams(rampedParams)

    # initialize integrator for simulated annealing
    #
    protocol.initDynamics(dyn,
                         potList=potList,  # potential terms to use
                         numSteps=100,   # at each temp: 100 steps or
                         finalTime=.2,   # .2ps, whichever is less
                         printInterval=100)
    cool.run()  # perform simulated annealing

    # final torsion angle minimization
    #
    protocol.initMinimize(dyn,
                          printInterval=50)
    dyn.run()

    # optional cooling in Cartesian coordinates
    #
    protocol.initDynamics(minc,
                          potList=potList,  # potential terms to use
                          numSteps=100,   # at each temp: 100 steps or
finalTime=.4 ,       # .2ps, whichever is less
printInterval=100)
#cart_cool.run()
# final all- atom minimization
# protocol.initMinimize(minc, potList=potList, dEPred=10)
minc.run()

#do analysis and write structure
loopInfo.writeStructure(potList)
pass

from simulationTools import StructureLoop, FinalParams

2. refine.py

xplor.requireVersion("2.18")

# # slow cooling protocol in torsion angle space for protein G. Uses
# NOE, RDC, J-coupling restraints.
# # this version refines from a reasonable model structure.
# # CDS 2005/05/10
#

xplor.parseArguments() # check for typos on the command-line

outFilename = "SCRIPT_STRUCTURE.sa"
numberOfStructures=3

StructureLoop(numStructures=numberOfStructures, pdbTemplate=outFilename, structLoopAction=calcOneStructure, genViolationStats=1, averageTopFraction=0.5, #report stats on best 50% of structs
averageContext=FinalParams(rampedParams),
averageCrossTerms=refRMSD,
averageSortPots=[potList['BOND'],potList['ANGL'],potList['IMPR'],
noe,rdcs,potList['CDIH']],
averagePotList=potList).run()

# protocol module has many high-level helper functions.
#
import protocol
protocol.initRandomSeed(3421)   #explicitly set random seed

# # annealing settings
#
command = xplor.command

protocol.initParams("protein")

# generate PSF data from sequence and initialize the correct parameters.
#
# from psfGen import seqToPSF
seqToPSF('HNP1.seq')
protocol.initStruct("HNP1.psf") # - or from file

# generate a random extended structure with correct covalent geometry
# saves the generated structure in the indicated file for faster startup
# next time.
#
protocol.genExtendedStructure("HNP1_extended.pdb" %
protocol.initialRandomSeed())

# or read an existing model
#
protocol.loadPDB("1DFN.pdb")
xplor.simulation.deleteAtoms("not known")

protocol.fixupCovalentGeom(maxIters=100, useVDW=1)

# a PotList contains a list of potential terms. This is used to specify which
# terms are active during refinement.
#
from potList import PotList
potList = PotList()

# parameters to ramp up during the simulated annealing protocol
#
from simulationTools import MultRamp, StaticRamp, InitialParams
rampedParams=[]
highTempParams=[]

# compare atomic Cartesian rmsd with a reference structure
# backbone and heavy atom RMSDs will be printed in the output
# structure files

# from posDiffPotTools import create_PosDiffPot
refRMSD =
create_PosDiffPot("refRMSD","name CA or name C or name N",
    pdbFile='1DFN.pdb',
    cmpSel="not name H*")

# orientation Tensor - used with the dipolar coupling term
# one for each medium
# For each medium, specify a name, and initial values of Da, Rh.
#
from varTensorTools import create_VarTensor
media={}

#                        medium  Da   rhombicity
for (medium,Da,Rh) in [('t',   -6.5, 0.62),
    ('b',   -9.9, 0.23 )]:
oTensor = create_VarTensor(medium)
oTensor.setDa(Da)
oTensor.setRh(Rh)
media[medium] = oTensor

# dipolar coupling restraints for protein amide NH.
#
# collect all RDCs in the rdcs PotList
#
# RDC scaling. Three possible contributions.
# 1) gamma_A * gamma_B / r_AB^3 prefactor. So that the same Da can be used
# for different expts. in the same medium. Sometimes the data is
# prescaled so that this is not needed. scale_toNH() is used for this.
# Note that if the expt. data has been prescaled, the values for rdc rmsd
# reported in the output will relative to the scaled values- not the expt.
# values.
# 2) expt. error scaling. Used here. A scale factor equal to 1/err^2
# (relative to that for NH) is used.
# 3) sometimes the reciprocal of the Da^2 is used if there is a large
# spread in Da values. Not used here.
#
# from rdcPotTools import create_RDCPot, scale_toNH
# rdcrs = PotList('rdc')
# for (medium,expt,file, scale) in 
# ['t','NH' ,'tmv107_nh.tbl'       ,1),
#  ('t','NCO','tmv107_nc.tbl'       ,.05),
#  ('t','HNC','tmv107_hnc.tbl'      ,.108),
#  ('b','NH' ,'bicelles_new_nh.tbl' ,1),
#  ('b','NCO','bicelles_new_nc.tbl' ,.05),
#  ('b','HNC','bicelles_new_hnc.tbl',.108) 
# ]:
#    rdc =
create_RDCPot("%s_%s"%(medium,expt),file,media[medium])

# 1) scale prefactor relative to NH
# see python/rdcPotTools.py for exact
calculation
# scale_toNH(rdc) - not needed for these
datasets -
# but non-NH reported
# rmsd values will be wrong.

# 3) Da rescaling factor (separate
# multiplicative factor)
# scale *= ( 1. / rdc.oTensor.Da(0) )**2
# rdc.setScale(scale)
# rdc setShowAllRestraints(1) #all
# restraints are printed during #analysis
# rdc.setThreshold(1.5) # in Hz
# rdc.append(rdc)
# pass
# potList.append(rdc)
#rangedParams.append( MultRamp(0.05,5.0, "rdcs.setScale( VALUE )") )

# calc. initial tensor orientation
# from varTensorTools import
calcTensorOrientation
# for medium in media.values():
#     calcTensorOrientation(medium)
#     pass

# set up NOE potential
# noe=PotList('noe')
# potList.append(noe)
# from noePotTools import create_NOEPot
# for (name,scale,file) in
# ['all',1,"HNP1_NOE.tbl"],
#    #add entries for additional
tables
# ]:
#    pot = create_NOEPot(name,file)
#    # pot.setPotType("soft") - if you think
#    # there may be bad NOEs
#    pot.setScale(scale)
#    noe.append(pot)
# rangedParams.append( MultRamp(2,30, "noe.setScale( VALUE )") )

# set up J coupling - with Karplus
coefficients
# from jCoupPotTools import create_JCoupPot
# jCoup =
create_JCoupPot("jcoup","jna_coup.tbl",
#                        A=6.98,B=-
# 1.38,C=1.72,phase=-60.0)
# potList.append(jCoup)

# Set up dihedral angles
from xplorPot import XplorPot
protocol.initDihedrals("HNP1_dihe.tbl",
useDefaults=0)
potList.append( XplorPot('CDIH') )
highTempParams.append( StaticRamp("potList['CDIH'].setScale(10)") )
rangedParams.append( StaticRamp("potList['CDIH'].setScale(200)") )
# set custom values of threshold values for violation calculation
#
potList['CDIH'].setThreshold(5)  #5 degrees is the default value, though

# gyration volume term
#
from gyrPotTools import create_GyrPot
gyr = create_GyrPot("Vgyr",
   "resid 1:56")  # selection should exclude disordered tails
potList.append(gyr)
#rampedParams.append( MultRamp(.002,1,"gyr.setScale(VALUE)") )

# hbda - distance/angle bb hbond term
#
protocol.initHBDA('hbda.tbl')
potList.append( XplorPot('HBDA') )

# Rama torsion angle database
#
#protocol.initRamaDatabase()
potList.append( XplorPot('RAMA') )
#rampedParams.append( MultRamp(.002,1,"potList['RAMA'].setScale(VALUE)") )

# setup parameters for atom-atom repulsive term. (van der Waals-like term)
#
potList.append( XplorPot('VDW') )
rampedParams.append( StaticRamp("protocol.initNBond()") )
rampedParams.append( MultRamp(0.9,0.8,
   "command('param nbonds rcon VALUE end end')") )
# nonbonded interaction only between CA atoms
highTempParams.append( StaticRamp("protocol.initNBond(cutnb=100,
tolerance=45,
   repel=1.2,
   onlyCA=1)""") )

# Give atoms uniform weights, except for the anisotropy axis
#
from atomAction import SetProperty
import varTensorTools
AtomSel("not resname ANI").apply( SetProperty("mass",100.) )
varTensorTools.massSetup(media.values(),300)
AtomSel("all ").apply( SetProperty("fric",10.) )

# IVM setup
# the IVM is used for performing dynamics and minimization in torsion-angle space, and in Cartesian space.
#
from ivm import IVM
dyn = IVM()  

# initially minimize in Cartesian space with only the covalent constraints.  
#   Note that bonds, angles and many impropers can't change with the  
#   internal torsion-angle dynamics  
#   breaks bonds topologically - doesn't change force field  
#  
#dyn.potList().add( XplorPot("BOND") )  
#dyn.potList().add( XplorPot("ANGL") )  
#dyn.potList().add( XplorPot("IMPR") )  
#  
#dyn.breakAllBondsIn("not resname ANI")  
#import varTensorTools  
#for m in media.values():  
#    m.setFreedom("fix")                 #fix tensor parameters  
#    varTensorTools.topologySetup(dyn,m)  
#setup tensor topology  
#  
#protocol.initMinimize(dyn,numSteps=1000)  
#dyn.run()  

# reset ivm topology for torsion-angle dynamics  
#  
#dyn.reset()  

for m in media.values():  
    m.setFreedom("fixDa, fixRh")        #fix tensor Rh, Da, vary orientation  
    m.setFreedom("varyDa, varyRh")  
#vary tensor Rh, Da, vary orientation  
protocol.torsionTopology(dyn,oTensors=media.values())  

# minc used for final cartesian minimization  
#  
minc = IVM()  
protocol.initMinimize(minc)  

for m in media.values():  
    m.setFreedom("varyDa, varyRh")  
#allow all tensor parameters float here  
    pass  
protocol.cartesianTopology(minc,oTensors=media.values())  

# object which performs simulated annealing  
#  
from simulationTools import AnnealIVM  
init_t = 3000.    # Need high temp and slow annealing to converge  
cool = AnnealIVM(initTemp =init_t,  
               finalTemp=25,  
               tempStep =12.5,  
               ivm=dyn,  
               rampedParams = rampedParams)  

def accept(potList):  
    """  
    return True if current structure meets acceptance criteria  
    """  
    if potList['noe'].violations()>0:  
        return False  
    if potList['rdc'].rms()>1.2: #this might be tightened some  
        return False  
    if potList['CDIH'].violations()>0:  
        return False  
    if potList['BOND'].violations()>1:  
        return False  
    if potList['ANGL'].violations()>0:  
        return False  
    if potList['IMPR'].violations()>1:  
        return False  
    return True  

def calcOneStructure(loopInfo):  
    """ this function calculates a single structure, performs analysis on the
structure, and then writes out a pdb file, with remarks.

```python
# initialize parameters for high temp dynamics.
InitialParams(rampedParams)
# high-temp dynamics setup - only need to specify parameters which differ from initial values in rampedParams
InitialParams(highTempParams)

# high temp dynamics
#
protocol.initDynamics(dyn,
    potList=potList, # potential terms to use
    bathTemp=init_t,
    initVelocities=1,
    finalTime=10,  # stops at 10ps or 5000 steps
    numSteps=5000,  # whichever comes first
    printInterval=100)

dyn.setETolerance(init_t/100)  #used to det. stepsize. default: t/1000
dyn.run()

# initialize parameters for cooling loop
InitialParams(rampedParams)

# initialize integrator for simulated annealing
#
protocol.initDynamics(dyn,
    potList=potList,
    numSteps=100,  #at each temp: 100 steps or
    finalTime=.2,  # .2ps,
    whichever is less
    printInterval=100)

# perform simulated annealing
#
cool.run()

# final torsion angle minimization
#
protocol.initMinimize(dyn,
    printInterval=50)
dyn.run()

# final all-atom minimization
#
protocol.initMinimize(minc,
    potList=potList,
    dEPred=10)
minc.run()

#do analysis and write structure
loopInfo.writeStructure(potList)
pass
```

from simulationTools import StructureLoop, FinalParams
StructureLoop(numStructures=numberOfStructures,
    pdbTemplate=outFilename,
    structLoopAction=calcOneStructure,
    genViolationStats=1,
    averagePotList=potList,
    averageSortPots=[potList['BOND'], potList['ANGL'], potList['IMPR'],
                     noe, rdcs, potList['CDIH']],'averageCrossTerms=refRMSD,' averageTopFraction=0.5, #report only on best 50% of structs
    averageAccept=accept,  #only use structures which pass accept()
averageContext=FinalParams(rampedParams),

averageFilename="SCRIPT_ave.pdb",
#generate regularized ave structure
  averageFitSel="name CA",

  averageCompSel="not resname ANI and not name H*" ).run()