NMR STRUCTURAL STUDIES OF LUNG SURFACTANT PROTEIN B (SP-B) PEPTIDES

MUZADDID SARKER

NMR Structural Studies of Lung Surfactant Protein B (SP-B) Peptides

By

Muzaddid Sarker

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirements for
the Degree of Master of Science

Department of Physics and Physical Oceanography
Memorial University of Newfoundland
St. John's, Newfoundland, Canada.

Your file Votre référence
ISBN: 978-0-494-19395-2
Our file Notre référence
ISBN: 978-0-494-19395-2

Abstract

Mammalian lungs are composed of millions of tiny air sacs called alveoli where gas exchange takes place. The inner surface of alveoli is coated by an aqueous layer to prevent it from drying up. However, because the attraction between water molecules is stronger than the force between water and air, a surface tension is created at the air-water interface. The tension tends to collapse the alveoli and increase the work of breathing. Lung surfactant is a material that counteracts these tendencies by reducing the surface tension to extremely low values and thus prevents the alveolar collapse and also eases the work of breathing. The lung surfactant material is a mixture of lipids and proteins. Surfactant Protein B (SP-B) is an essential component of the surfactant and is thought to function by facilitating large-scale rearrangement of the lipid molecules and stabilizing the structures. However, neither the structural basis for this ability nor the physiological ramifications of lipid rearrangements are yet understood. SP-B is a lipid-associated hydrophobic protein, which makes it difficult to address with X-ray or conventional solution NMR structural techniques. These difficulties have been addressed in NMR structural studies by the use of fluorinated organic solvents and lipid micelles to solubilize the proteins. The present work has focused on SP-B$_{CTERM}$ and Mini-B, two peptide fragments of SP-B that retain significant biological activity of the full-length protein. The structural features of these peptides were studied using high-resolution solution NMR. Firstly, the conformational features of SP-B$_{CTERM}$ and its interactions with lipids were investigated in micelles mimicking the lipid environment in the lungs. The peptide exhibited the ability to cause aggregation of micelles formed from lipids similar to those found in lungs. This was indicative of the large-scale lipid rearrangement and stabilization of structures facilitated by SP-B in natural surfactant. In the second and main phase of current research, the structure of Mini-B was determined in the structure-inducing fluorinated organic solvent hexafluoroisopropanol (HFIP). The peptide was found to consist of two α-helices at the termini connected by an unstructured loop at the middle. These studies help to define the structural properties that underlie SP-B's function and provide a platform to probe the lipid-protein interactions that are responsible for the ability of lung surfactant to dramatically lower the surface tension at the air-water interface in alveoli.

Acknowledgements

I would like to express, with great pleasure, my deepest gratitude and thanks to my supervisor Dr. Valerie Booth for her continuous guidance, precious suggestions and heartfelt cooperation at each and every step of my program including this research. It was, indeed, a wonderful experience for me to obtain an opportunity to work with her.

I acknowledge the supportive and truly encouraging roles of my co-supervisor Dr. Mike Morrow throughout the whole program. I also acknowledge the great contribution of my other co-supervisor Dr. Kevin Keough for providing the necessary fund to perform this research.

I thank Dr. David Heeley for his kind attitude towards me and also for helping me in conducting CD experiments in his Lab. I thank Dr. Alan Waring and colleagues for preparing the peptide samples used in this work. I also thank Dr. Brian Dawson and colleagues for performing some crucial NMR experiments and Dr. Qi Wang and colleagues for conducting DLS experiments.

Jennifer, Doyle, Sheila and June were always very cooperative and helpful and it was a nice experience for me to work in the same Lab with them.

Last, but not the least, I must acknowledge the contributions of my family who are integral parts of my life.

Contents

Chapter 3 : SP-BCTERM – Lipid Interactions

Chapter 4 : Structure of Mini-B

Chapter 5 : Conclusion

List of Tables

List of Figures

List of Abbreviations

CD	: Circular Dichroism
CMC	: critical micellar concentration
CSI	: chemical shift index
DHPC	: dihexanoylphosphatidylcholine
DLS	: Dynamic Light Scattering
DPC	: dodecylphosphocholine
DPPC	: dipalmitoylphosphatidylcholine
DSS	: sodium 2,2-dimethyl-2-silapentane-5-sulfonate
FID	: free induction decay
HFIP	: hexafluoroisopropanol
HSQC	: Heteronuclear Single Quantum Correlation
LPA	: lysophosphatidic acid
LPG	: lysophosphatidylglycerol
NMR	: Nuclear Magnetic Resonance
NOE	: Nuclear Overhauser Effect
NOESY	: Nuclear Overhauser-Effect Spectroscopy
PC	: phosphatidylcholine
PG	: phosphatidylglycerol
POPC	: palmitoyloleoylphosphatidylcholine
ppm	: parts per million
RDS	: respiratory distress syndrome
SDS	: sodium dodecyl sulphate
SP-B	: Surfactant Protein B
TMS	: tetramethylsilane
TOCSY	: Total Correlation Spectroscopy

Chapter 1

Introduction

1.1 Lung Surfactant

Living organisms need a continuous supply of oxygen to perform various physiological activities and also need to get rid of carbon dioxide produced in metabolic reactions. For air-breathing vertebrates, blood carries both of these gases with the exchange taking place in the lungs. This function of lungs is accomplished through airways ending in tiny air sacs known as alveoli. Adult human lungs contain approximately 2400 kilometers of airways and 300 million alveoli, having a total surface area of about 140 m^2. Alveoli are surrounded by capillaries carrying blood and the barrier in between is extremely thin, allowing for rapid O_2/CO_2 exchange (Figure 1.1). The inner surface of the alveoli is coated by an aqueous layer to prevent it from drying up due to its constant contact with air.

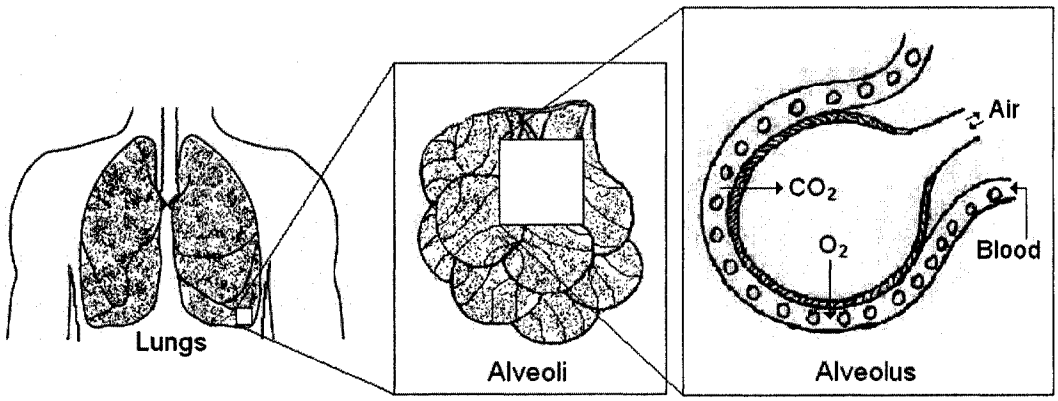


Figure 1.1 : Schematic diagram of human respiratory system.

The lungs, and thereby the alveoli, inflate and deflate cyclically with inhalation and exhalation. These fluid coated air sacs are roughly spherical in shape. Maintaining their structure during cyclical changes in lung volume is critical for normal respiration. Since the attraction between water molecules is stronger than that of water and air, a surface tension is created at the air-water interface in the alveoli. This is largely governed by LaPlace's law that relates the fluid pressure, P, inside a spherical vessel of radius R with the surface tension, T;

$$T = RP/2$$

The surface tension, unless countered, would promote lung-collapse and increase the work required to re-inflate the lung at the end of expiration [1]. However, it is countered by a material called lung surfactant (also known as pulmonary surfactant). Lung surfactant is a mixture of lipids and proteins that lines the air-water interface in the alveoli. Pure water has a surface tension of about 70 mN/m. By interfering with the interaction of the surface water molecules, the surfactant lowers it down to an equilibrium value of 25 mN/m in the lungs, which is further reduced to almost 0 mN/m during expiration [2]. This prevents alveolar collapse at the end of exhalation and also eases the work of breathing during subsequent inhalation. This behavior of the surfactant has been termed as the surface activity.

Lung surfactant was first discovered by von Neergaard in the late-1920s [3]. However, the significance of the discovery was not appreciated at that time. In the mid-1950s, surfactant's ability to reduce surface tension led to its rediscovery by Pattle [4] and Clements [5]. Since then, extensive research has been conducted to explore the molecular mechanisms that underlie its physiological function. Over the years, important advancements have been made. However, a complete molecular description of the structures of the components and their interactions during the respiratory dynamics are still unavailable.

1.2 Surfactant Source and Composition

The alveolar fluid coating has a thickness of about 2 μm. Lung surfactant is synthesized and secreted into the alveolar fluid by some of the boundary-forming cells of the alveoli, known as Type II cells (Figure 1.2). These cells are cuboidal in shape and contain specialized secretory intermediaries known as lamellar bodies (LB) [6]. An LB is the storage form of the surfactant. It consists of a core composed of multilamellar membranes surrounded by another membrane [7]. It is secreted in response to local

biochemical factors [8], signals from the automatic nervous system [9] and deep breathing [10] or stretch [11]. In the aqueous environment, the LB contents become hydrated and can unravel into tubular myelin (TM). The TM is an ordered array of tubules of nearly rectangular cross section. The walls of the TM have bilayer form and the corners are either intersecting or very closely apposed bilayers with very high curvature. Some models for surfactant formation assume TM as the immediate precursor of the surface film. Surfactant components are released from it and form a surface active film at the air-water interface [12, 13]. The surfactant film has long been considered as a simple monolayer. However, recent electron-microscopic and surface activity studies suggest that actually several layers are closely packed at the interface [14 - 16].

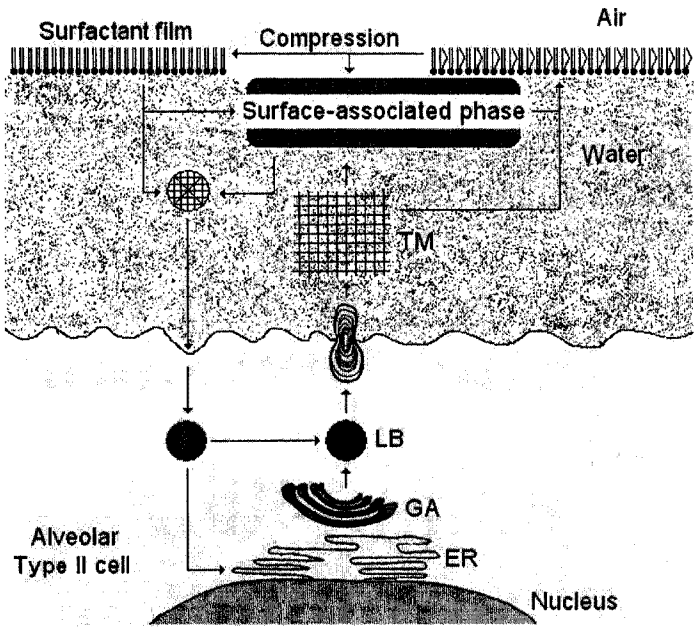


Figure 1.2 : One model describing the formation of lung surfactant [13, 16]. Components are synthesized in the endoplasmic reticulum (ER), transported to golgi apparatus (GA) and packed in lamellar bodies (LB). LB, being secreted into the liquid layer (Hypophase), swells and unravels to form tubular myelin (TM). TM supplies surfactant lipids and proteins to the surface film.

Lung surfactant is a complex mixture of lipids (~ 90% by weight) and proteins (~ 10% by weight). Phospholipids make up the bulk of the surfactant materials (~ 80%) but there are some neutral lipids (~ 10%) as well. Composition varies significantly between different mammalian and nonmammalian species and also throughout the physiological development of a particular species [17]. Table 1.1 lists the names of the components of human lung surfactant with the percentage of share by weight [17, 18].

Table 1.1 : Components of human lung surfactant [17, 18]

Phospholipids (~80%)	1. Phosphatidylcholine (PC) ~ 70% (~35% dipalmitoylphosphatidylcholine (DPPC), ~15% other disaturated species and ~20% unsaturated forms) 2. Phosphatidylglycerol (PG) ~ 9% 3. minor amount of Phosphatidylinositol (PI) 4. minor amount of Phosphatidylethanolamine (PE) 5. minor amount of Phosphatidylserine (PS) 6. minor amount of Sphingomyelin (SM)
Neutral lipids (~10%)	1. mainly Cholesterol ~ 10% 2. minor amounts of mono, di and triglycerides 3. minor amount of free fatty acids
Proteins (~10%)	1. Surfactant Protein A (SP-A) ~ 6% 2. Surfactant Protein B (SP-B) ~ 1.5% 3. Surfactant Protein C (SP-C) ~ 1.5% 4. Surfactant Protein D (SP-D) ~ 1%

1.3 Surfactant Biophysical Activity

An effective surfactant exhibits two critical properties. Firstly, it adsorbs very rapidly to the air-water interface, producing equilibrium surface tensions of around 25 mN/m within seconds [2, 16]. This ensures that the surfactant reaches the respiratory interface before a newborn completes the first inspiration, so that alveolar collapse can be avoided by reducing the surface tension at the end of the subsequent expiration. Secondly, once adsorbed at the interface, the surfactant must reduce the surface tension to values close to 0 mN/m upon compression [5]. This, in addition to avoiding lung collapse at the end of expiration, eases the work of next inspiration.

Since its rediscovery in the mid-1950s, lung surfactant has been subjected to various analytical and experimental techniques, but the molecular mechanisms underpinning its functions are not yet known in detail. There are lots of unanswered

questions regarding the structures, roles and mechanisms adopted by each individual components at different stages. However, based on various *in vitro* biophysical studies, a few models have been proposed. The models suggest that the phospholipids, due to their ability to form oriented interfacial monolayers, act as the principal surface active agents in the surfactant. However, the two properties mentioned above seemed to be mutually exclusive in single phospholipid systems. Phospholipids exhibiting rapid adsorption and spreading properties (e.g., POPC) are not capable of lowering surface tensions when the formed monolayers are compressed [19]. Conversely, phospholipids achieving extremely low surface tensions at monolayers following compression (e.g., DPPC) adsorb very slowly from the hypophase [18]. Thus, the solution to obtaining optimal performance appears to be a mixture of suitable phospholipids. During inspiration, some of the lipid components ensure rapid adsorption, formation and replenishment of the surface-active film. Then in subsequent expiration, other lipid components provide the low compressibility necessary to achieve stable low surface tension.

The protein components are also essential for the surfactant to function properly. The proteins are primarily considered as modulating agents, evolved to optimize the surface activity of the phospholipids during the respiratory cycle. However, one of the proteins, namely SP-B, plays such crucial roles that in its absence the surfactant loses its surface activity and fails to function. SP-B and some other proteins also produce dramatic acceleration of the interfacial adsorption and speed the process further under certain circumstances [8]. Thus, the mixture of lipids and proteins has been evolutionally optimized in lung surfactant to perform all the required tasks simultaneously [18].

1.4 Surfactant Disorder - Disease and Treatment

Deficiency or absence of functional lung surfactant leads to potentially fatal health conditions such as respiratory distress syndrome (RDS). This is a common

problem in premature newborns, as well as in adults with acute injury or illness. The development of lung surfactant replacement treatments in the early 1990s has greatly improved the outlook for newborns with RDS [19]. However, use of surfactant replacement techniques in addressing RDS for adults is still under development [20]. The need for more efficacious therapies is thus obvious. Clinical trials have demonstrated that artificial surfactants which include the proteins SP-B and SP-C are much more effective than protein-free surfactant preparations [22].

1.5 Surfactant Lipids

The roles of the major lipid components of the lung surfactant in surface activity have been identified but those of the minor lipid species are less-well characterized. Also to be noted is that, most of the experimental focus has been on surface tension reducing properties of the surfactant. The lipid components, however, could also be involved in other metabolic aspects of the surfactant life cycle [23].

During expiration, surfactant film at the alveolar air-water interface is compressed and the surface tension must be reduced to the order of 0-1 mN/m [5]. This means that the surfactant film sustains surface pressure close to 70 mN/m at the interface. Experiments on monolayers composed of the major surfactant phospholipids led to the conclusion that only pure DPPC film is capable of withstanding such high pressure [24]. This is because the saturated acyl chains of DPPC can be packed tightly enough to sustain the highest pressure. Thus DPPC acts as the most vital component of the surfactant. However, this phospholipid adsorbs very poorly by itself into the interface, especially at temperatures below its gel-to-liquid crystalline phase transition temperature of 41 ^{0}C [25].

Mixing with more fluid phospholipids, such as unsaturated PCs, PG and/or PI, improves the adsorption of DPPC to a large extent [23]. However, lipids alone do not make fully functional surfactant. The hydrophobic proteins SP-B and SP-C boost the adsorption substantially, but the mechanisms involved are yet to be understood. Nevertheless, most of the components of surfactant other than DPPC are considered to facilitate both its adsorption to the interface and its re-spreading in the successive respiratory cycles [18].

Another lipid element of the surfactant, present in mentionable amount, is cholesterol. The ability of cholesterol to alter essential properties of membranes is well-established but no clear data are available to date to evaluate its actual role in lung surfactant.

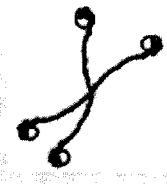
1.6 Surfactant Proteins

The protein portion of lung surfactant comprises four proteins SP-A, SP-B, SP-C and SP-D. The names of the proteins were given with respect to their chronological order of discovery. SP-A is the most abundant among the proteins in surfactant. Its main role is thought to be in defending the lungs against germs. Experiments showed that the mice born without SP-A could breathe normally but were extremely prone to developing infections [26]. *In vitro* and *in vivo* studies also showed that SP-A plays important role in the formation of tubular myelin [27]. It has further been observed that SP-A can improve the surface activity of surfactant under challenging conditions such as low surfactant concentrations and presence of inhibitory plasma proteins or oxidants [18]. SP-B fulfils a crucial role since hereditary deficiency of this protein is lethal in humans [28] and in knockout mice [29]. Deactivation of SP-B by antibodies also causes RDS *in vivo* [30]. SP-C deficiency is not lethal at birth but its absence modifies biophysical properties of monolayers [31]. It is associated with familial interstitial lung disease [32]. The exact

roles played by SP-B and SP-C in surfactant function and the mechanisms by which they act are not yet known. However, experiments showed that the presence of SP-B and SP-C provide necessary enhancement in adsorption of phospholipids at the alveolar air-water interface [33]. These proteins also seem to be important to prevent detaching of the folded structures from the water layer [18]. SP-D is not associated with surface active lipid-protein complexes and is not considered to play any significant roles in mechanisms of reducing surface tension at air-water interface [18].

The main structural properties of the surfactant proteins are listed in Table 1.2 [16]. It should be noted that the overall 3D structures of these proteins have not been determined thus far and the schematic models shown here are based on preliminary assumptions.

Table 1.2 : Structural features of the lung surfactant proteins [16]

Protein	SP-A	SP-B	SP-C	SP-D
Model structure				
Nature	Hydrophilic	Hydrophobic	Hydrophobic	Hydrophilic
Quat. structure	Octadecamer	Dimer	Monomer	Dodecamer
MW/Monomer	26-38 kDa	8.7 kDa	4.2 kDa	43 kDa
Phospholipid binding	DPPC	Anionic/zwitterionic phospholipids	Anionic phospholipids	PI

1.7 Surfactant Protein B (SP-B)

SP-B is the most important among the protein components of the surfactant as it is essential for survival. In mammalian lung, it is found as a covalently linked homodimer

having molecular mass of about 18 kDa [34]. Human SP-B monomer is 79 residues (amino acids) in length [35]. However, the SP-B gene is first transcribed and translated into a significantly larger precursor consisting of 381 residues. The mature form of SP-B consists of residues 201-279 of the precursor. This is cleaved in compartments that lie between the trans-golgi and the lamellar body [36]. Table 1.3 lists the amino acid sequence of human SP-B.

Table 1.3 : Amino acid sequence of human surfactant protein B

1-16	Phe-Pro-Ile-Pro-Leu-Pro-Tyr-Cys-Trp-Leu-Cys-Arg-Ala-Leu-Ile-Lys-
17-32	Arg-Ile-Gln-Ala-Met-Ile-Pro-Lys-Gly-Ala-Leu-Ala-Val-Ala-Val-Ala-
33-48	Gln-Val-Cys-Arg-Val-Val-Pro-leu-Val-Ala-Gly-Gly-Ile-Cys-Gln-Cys-
49-64	Leu-Ala-Glu-Arg-Tyr-Ser-Val-Ile-Leu-Leu-Asp-Thr-Leu-Leu-Gly-Arg-
65-79	Met-Leu-Pro-Gln-Leu-Val-Cys-Arg-Leu-Val-Leu-Arg-Cys-Ser-Met

Compared to water soluble proteins, a large portion (about 60%) of SP-B's amino acids are hydrophobic, which are alanine, valine, leucine, isoleucine, methionine, cysteine, phenyalanine and tryptophan. There are seven cysteines, six of which form three intra-molecular disulfide bridges and the remaining one forms an inter-molecular bond responsible for dimerization. SP-B also exhibits a strong cationic profile as it contains nine positively charged (arginine & lysine) and two negatively charged (aspartate & glutamate) residues yielding a net charge of + 7 at neutral pH.

The properties of SP-B which have been identified *in vitro* include, membrane binding, membrane lysis, membrane fusion, promotion of lipid adsorption to air-water surfaces, stabilization of monomolecular surface films and re-spreading of films from collapsed phases [36]. All these activities point to SP-B's ability to make large-scale lipid rearrangements that are assumed to be crucial in surfactant function. However, the details of SP-B's essential roles *in vivo* are still far from being understood.

The tertiary (overall 3D) and even the secondary (α-helix/β-sheet/loop/turn/coil) structure of the full-length SP-B has not yet been determined. However, some inference can be drawn from proteins with similar amino acid sequence whose structures have already been determined. SP-B belongs to the Saposin protein superfamily. Experimentally derived structures of four members of this superfamily, namely Saposin B (PDB entry 1N69 [37]), Saposin C (PDB entry 1SN6 [38]), NK-Lysin (PDB entry 1NKL [39]) and Amoebapore (PDB entry 1OF9 [40]), display four (or five) dominant helical regions although the way the helices are packed together can differ. Sequence homology with these proteins suggests that SP-B likely consists of four or five α-helices connected by loops (Figure 1.3 (A)).

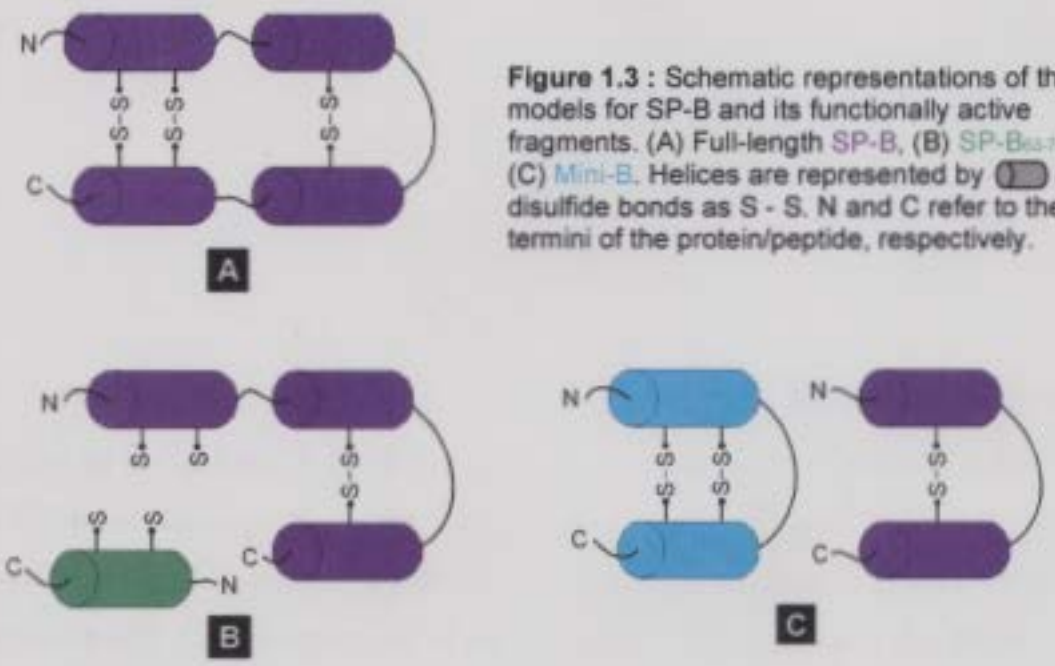


Figure 1.3 : Schematic representations of the models for SP-B and its functionally active fragments. (A) Full-length SP-B, (B) SP-B$_{8-78}$ and (C) Mini-B. Helices are represented by ⬭ and disulfide bonds as S - S. N and C refer to the two termini of the protein/peptide, respectively.

1.8 SP-B Fragments and Their Activity

Fragments of SP-B containing individual helices or pairs of helices were shown to retain significant biological activity of the full-length protein. SP-B$_{CTERM}$ (also known as

SP-B63-78) is composed of amino acids 63-78 of SP-B and contains the C-terminal helix (Figure 1.3 (B)). Mini-B is made up of amino acids 8-25 and 63-78 of the native protein and contains both the N- and C-terminal helices linked by a loop (Figure 1.3 (C)). Mini-B can be produced with and without the disulfide bonds linking the two helices.

Measurements of blood oxygen content of surfactant deficient rats treated with various artificial surfactants have demonstrated that SP-B peptides retain significant biological activity (Figure 1.4) [personal communication from Alan Waring]. The artificial surfactants were composed from lipids alone and lipids plus SP-B, Mini-B or SP-BCTERM. Treatment with lipids alone did not improve the oxygen content in blood. Lipids + SP-BCTERM improved it almost by half of that of lipids + SP-B. Lipids + Mini-B performed as well as the lipids + SP-B. The findings indicate that these peptides, especially Mini-B, retain vital functional properties of native SP-B. This points to the likelihood that Mini-B contains functionally active regions of the full-length protein.

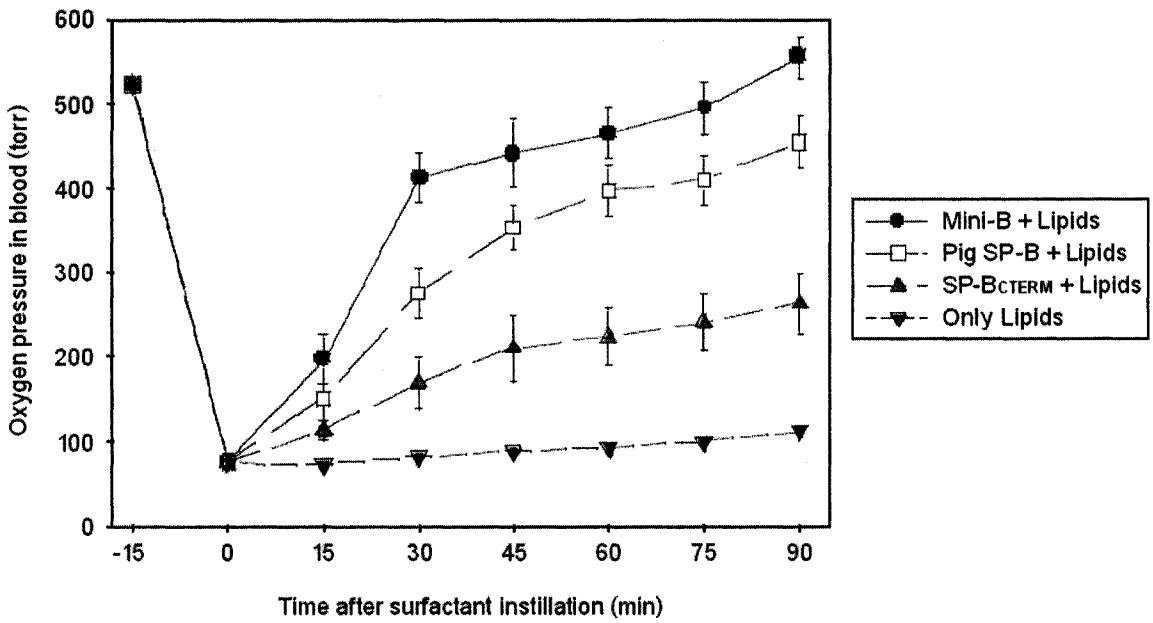


Figure 1.4 Relative performance of different artificial lung surfactants applied on surfactant deficient rats. Treatment with only lipids did not improve the oxygen content in blood. However, lipids with SP-B peptides, especially Mini-B, showed significant improvement.

1.9 Previous Structural Studies of SP-B Peptides

Structures of two SP-B peptides containing the two terminal regions of the full-length protein have previously been determined by NMR. The structure of the 15 residue N-terminal peptide, SP-B_{11-25}, has been determined in organic solvent methanol by Kurtz *et al.* (PDB entry 1KMR [41]). The structures of the 16 residue C-terminal peptide, SP-B_{63-78} (called earlier and later as SP-B_{CTERM}), have been determined in micelles formed from Sodium Dodecyl Sulphate (SDS) as well as in organic solvent hexafluoroisopropanol (HFIP) by Booth *et al.* (PDB entries 1RG3 and 1RG4 [42]).

Kurtz *et al.* determined the structure of SP-B_{11-25} in methanol using solution NMR [41]. The structure showed that the peptide consists of an α-helix (Figure 1.5), running through amino acids 13 to 21. The C-terminus of SP-B_{11-25} was found to be disordered.

Figure 1.5 : Ensemble of 10 lowest energy ribbon structures of SP-B_{11-25} in methanol [41]. [Diagram taken from Protein Data Bank, ID 1KMR]. The peptide contains an α-helix from amino acids 13 to 21 with a disordered C-terminus.

Booth *et al.* determined the structures of SP-B_{63-78} in two different environments [42]. The organic solvent sample was prepared by dissolving 2 mM SP-B_{63-78} in 40% HFIP, 50% H_2O and 10% D_2O with 0.2 mM DSS. The sample had pH of 3.0 and the NMR data were acquired at 5 ^{0}C. The lipid micelle sample was prepared by mixing 2 mM peptide, 200 mM deuterated SDS and 0.2 mM DSS in 90% H_2O and 10% D_2O. The sample had pH of 6.9 and the NMR experiments were performed at 25 ^{0}C. The studies showed that the structure of the peptide is mostly α-helical in both the environments, the α-helix runs through amino acids 68 to 78 in HFIP and 69 to 78 in SDS (Figure 1.6). However, the N-terminal region of SP-B_{63-78} was found disordered in HFIP.

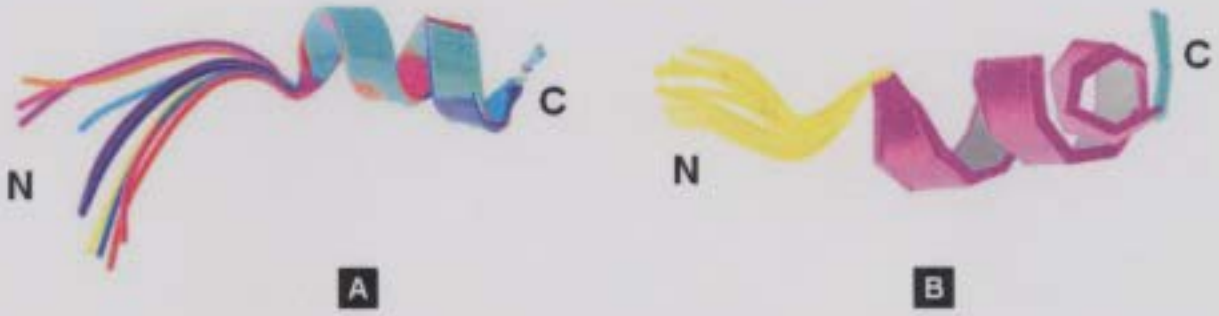


Figure 1.6 : Ensembles of 10 lowest energy ribbon structures of SP-B_{63-78} in (A) HFIP and (B) SDS micelle [42]. [Diagrams taken from Protein Data Bank, IDs 1RG4 and 1RG3]. The peptide contains an α-helix from amino acids 68 to 78 in HFIP and from amino acids 69 to 78 in SDS. The N-terminal region of the peptide is disordered in HFIP .

1.10 Objectives of the Present Work

The present NMR structural studies were conducted to

(i) probe the interactions of SP-B_{CTERM} with the micelle-forming analogs of the physiological surfactant phospholipids, and

(ii) determine the structure of Mini-B in structure-promoting organic solvent HFIP.

A protein's biological function comes about due to the organization of its amino acid chain into a specific, well-folded, 3D structure. Hence, determination of the structure of SP-B will provide a basis for understanding its function. However, SP-B is a lipid-associated and exceptionally hydrophobic protein, making it difficult to address with conventional X-ray or solution NMR structural techniques. Besides, it is very hard to express the full-length SP-B in bacteria or synthesize it chemically. Therefore, initial structural studies were carried out with two chemically synthesized segments (peptides), SP-B_{CTERM} and Mini-B, that retain significant biological activity comparable to the full-length protein. The structural features of these peptides provided important indications about how the full-length SP-B works in native conditions.

Chapter 2

Protein NMR Spectroscopy

2.1 Protein Structure

A protein is a polymer composed of a long chain of chemical subunits called amino acids (also known as residues). There are 20 different types of amino acids that are commonly found in proteins [43]. All of them have the same backbone chemical structure (– N – C_α – C –). Three of the substituents of C_α are identical for every amino acid. The difference lies in the fourth substituent, the sidechain (or R group). (Figure 2.1 (A)). Individual atoms in an amino acid are referred to as H_N, C_α, H_α, C_β, H_β, etc. Amino acids are linked into a linear sequence by peptide bonds between the carboxyl C of one residue and the amino N of the next residue (Figure 2.1 (B)). The first and the last residues in the chain are termed as the N- and C-terminal residues, respectively.

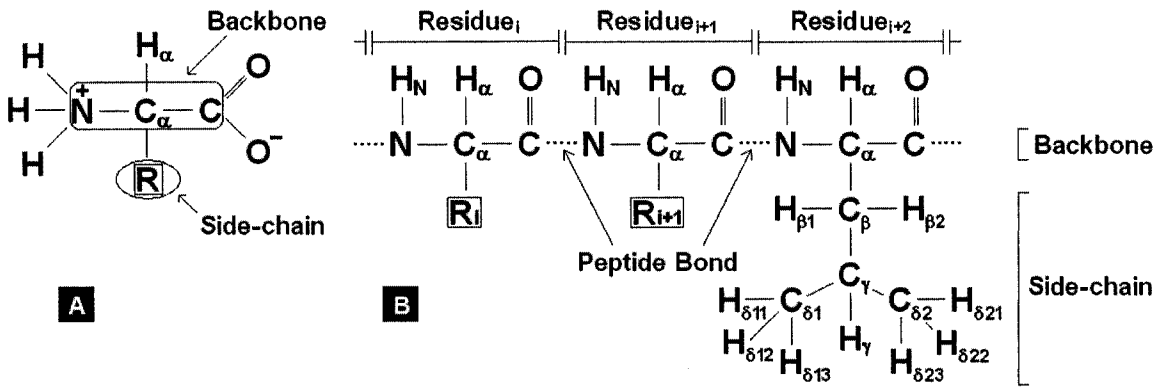


Figure 2.1 : Chemical structure of a protein. (A) A single amino acid. All 20 amino acids have same backbone chemical structure (– N – C_α – C –) but different side-chains (Rs). (B) Amino acids are connected by peptide bonds to form a full-length protein. Residue$_{i+2}$ is shown with the names of the side-chain nuclei as an example.

There are four levels of protein structure; primary, secondary, tertiary and quaternary. The amino acid sequence of the polypeptide chain is called the primary structure. Different regions of the chain form local regular secondary structures, such as α-helices and β-strands. The tertiary structure is formed by packing such structural elements into one or more compact globular units (domains). The final protein may contain several polypeptide chains arranged in a quaternary structure. It is the overall 3D structure of a protein that leads to function [44].

2.2 Protein Structure Determination

Proteins play numerous crucial physiological roles in living organisms. Since the function of a protein depends on its 3D structure, knowledge on the structural features is a precondition to understand the functional properties. Only two methods are available for determining the structure of a protein at atomic resolution; X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy [45].

X-ray crystallography requires the protein to be grown as a well-ordered crystal [46]. A pattern of X-rays diffracted by the protein crystal is recorded. An electron density map is then generated into which the known sequence is fitted. This method calculates the atomic coordinates precisely. However, it can only provide information about one of the low-energy conformations that the protein may adopt but nothing regarding the dynamics. Therefore, the structure may sometime be misleading and may lack physiological relevance. Furthermore, for lipid-associated membrane proteins, it is virtually impossible (with rare exceptions) to grow diffraction-quality crystals due to the problems caused by aggregation.

NMR spectroscopy has only been developed as a technique for protein structure determination in the last two and a half decades. The first complete 3D structure solved using this technique was presented in 1986. Although solid-state NMR can be used to obtain some structural information, high-resolution protein structures are determined by NMR mainly in solution state. NMR method exploits the quantum mechanical property of spin 1/2 nuclei. The sample is placed in a strong static magnetic field and the spin 1/2 nuclei (^{1}H, ^{13}C, ^{15}N, etc.) of the protein are perturbed by specific radiofrequency (RF) pulses. High-resolution structural data are generated by manipulating the magnetization through scalar and dipolar couplings. The data contain information on inter-nuclear distances and backbone dihedral angles which are used to compute the structural model of the protein.

In addition to providing structural data, high-resolution solution NMR can also be used to study the flexibility of proteins and their dynamics over a wide range of time scales. This method also enables the study of membrane proteins in lipid micelles which mimick the physiological environment. However, the inherent properties of solution NMR impose a limitation on the size of the protein that can be studied. For standard triple resonance techniques, the size is limited typically to 200 residues.

2.3 Basic Principles of NMR

Nuclear magnetic resonance (NMR) is a phenomenon which occurs when a group of nuclei with nonzero spin angular momentum is immersed in a static magnetic field and exposed to RF pulses of certain specific energy. Nuclei with zero spin (e.g., ^{12}C or ^{16}O) are thus not observable by NMR. Nuclei with spin 1 (e.g., ^{2}H or ^{14}N) are also not generally useful for structural studies in solution. They possess quadrupole moments and the lifetimes of their magnetic states are very short resulting in broad resonance lines which are difficult to observe. Nuclei with spin 1/2 (such as ^{1}H) are the most useful ones for high-resolution solution NMR [45]. For biomolecular structural studies, ^{12}C and/or ^{14}N nuclei of the molecule are often replaced by their isotopes ^{13}C and/or ^{15}N, respectively, which are spin 1/2 nuclei.

In the absence of any external magnetic field, the spin 1/2 nuclei of a sample, which may be considered as tiny bar magnets, do not have any preferred orientation. However, if the sample is placed in a large static external magnetic field, quantum mechanical requirements mean that all the nuclei take one of the two allowed states, given by the magnetic quantum number $m = \pm 1/2$. The two states have different energy levels known as Zeeman levels (Figure 2.2). In thermal equilibrium, these states are unequally populated, having some more nuclei in the lower energy state than the higher, as determined by the Boltzmann distribution. The small nuclear magnets may

spontaneously flip from one orientation (energy state) to the other but the rate is extremely slow. However, if energy equal to that of the difference between the two states is applied, much more flipping is induced. This is an example of resonance phenomena and hence given the name Nuclear Magnetic Resonance (NMR) [47].

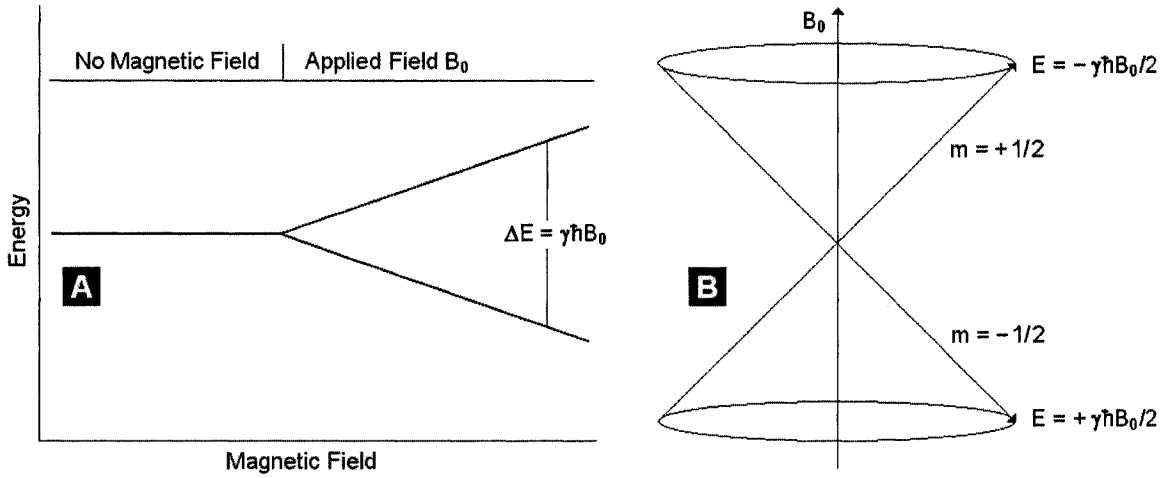


Figure 2.2 : Energy levels of a spin 1/2 nucleus in an external magnetic field. (A) The applied field B_0 causes the energy state of the nucleus to be splitted into two levels (Zeeman levels). (B) The nucleus assumes one of the two quantum mechanically allowed energy values and transition occurs by the absorption or emission of energy equal to the difference of the two states.

A nucleus with spin angular momentum **I** also possesses nuclear magnetic moment $\boldsymbol{\mu}$, and the relation between the two is

$$\boldsymbol{\mu} = \gamma \mathbf{I}$$

Here γ is the proportionality constant known as the gyromagnetic ratio of the nucleus. The magnitude of γ, in part, determines the receptivity of a nucleus in NMR spectroscopy. The larger the gyromagnetic ratio the stronger the NMR signal is. Now, the magnitude of the spin angular momentum is

$$I = m\hbar$$

Here $\hbar = h/2\pi$ and h is Planck's constant. Therefore, the nuclear magnetic moment is

$$\mu = \gamma m\hbar$$

Again, in an external magnetic field **B**, the spin states of a nucleus have energies given by

$$E = -\boldsymbol{\mu}.\mathbf{B}$$

The energy of a particular spin state m is then

$$E_m = -m\hbar\gamma B$$

Thus the energy difference between the two Zeeman levels of a spin 1/2 nucleus placed in an external static magnetic field $\mathbf{B_0}$ is

$$\Delta E = \gamma\hbar B_0$$

This is also the irradiation energy required to cause Zeeman transition between the two states. Now the Bohr condition

$$\Delta E = h\nu$$

enables the frequency of the nuclear transition, known as Larmor frequency, to be written as $\nu_0 = (1/2\pi)\gamma B_0$ or $\omega_0 = \gamma B_0$

in units of Hz or rad/s, respectively. Hence, the irradiation energy depends on both γ and B_0. In practice, it is in the radiofrequency (RF) range and is typically applied as a short pulse [47].

Table 2.1 lists the gyromagnetic ratios and the Larmor frequencies of selected NMR-active nuclei in an 11.74 T field [45, 48].

Table 2.1 : Properties of nuclei commonly used in NMR [45, 48]

Nucleus	Natural abundance (%)	I	γ $(T\cdot s)^{-1}$	ω_0 at B_0 = 11.74 T (MHz)
^1H	99.98	1/2	2.68×10^8	500
^2H	0.02	1	4.11×10^7	77
^{13}C	1.11	1/2	6.73×10^7	125.7
^{15}N	0.36	1/2	-2.72×10^7	50.7

The Zeeman splitting of the magnetic states of the spin 1/2 nuclei in an external magnetic field also causes a net bulk magnetization $\mathbf{M_0}$ to evolve as the two states are unequally populated. It is this net magnetization that determines the NMR signal, not the magnetic moment of an individual nucleus. Conventionally the direction of the applied field $\mathbf{B_0}$ is taken as the z-axis of the coordinate system, so, $\mathbf{M_0}$ also evolves along z-axis.

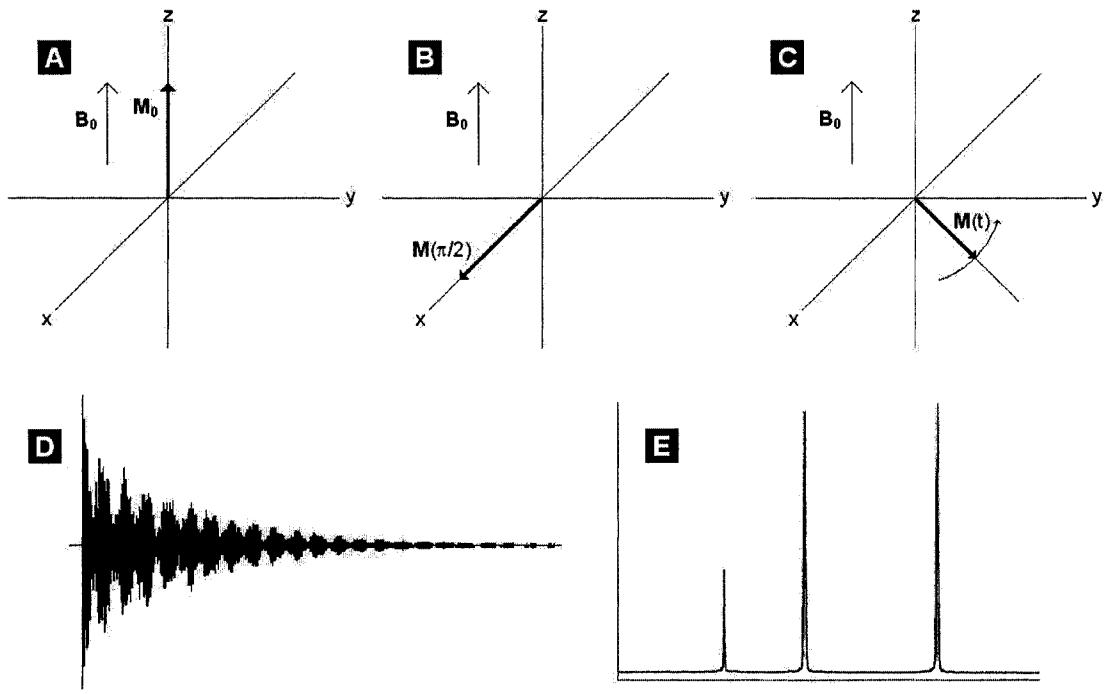


Figure 2.3 : Steps of a simple one-pulse NMR experiment. In a static external field $\mathbf{B_0}$, the net bulk magnetization $\mathbf{M_0}$ lies along z-axis (A). A 90^0 pulse applied along the y-axis rotates $\mathbf{M_0}$ to the x-axis making it $\mathbf{M}(\pi/2)$ having the same magnitude (B). Relaxation starts immediately as $\mathbf{M}(\pi/2)$ precesses in the xy-plane under the influence of $\mathbf{B_0}$. The time-varying magnetic field $\mathbf{M}(t)$ induces a current in the spectrometer coil (C). The free induction decay (FID) is recorded during the acquisition period that starts at the end of the pulse (D). However, the FID is a time-domain signal. The real component of the complex frequency domain NMR spectrum is obtained by Fourier transformation of the FID (E).

During NMR experiments (Figure 2.3), the net magnetization $\mathbf{M_0}$ is manipulated by an RF pulse, which is a time varying magnetic field $\mathbf{B_1}$, oscillating at the same Larmor frequency ω_0. Often, $\mathbf{B_1}$ is a 90^0 pulse and applied along y-axis. This applies a torque on the bulk magnetization and tips it along x-axis into the xy-plane. In this situation, the bulk magnetization rotates about the z-axis with frequency ω_0 under the influence of $\mathbf{B_0}$ and induces an electric current in the detection coil of the spectrometer. In the absence of any further perturbing RF pulses, relaxation processes eventually return the spin system to thermal equilibrium with the magnetization oriented back along the z-axis. During this period, the transverse magnetization decays with time. The corresponding free induction decay (FID) is recorded for a certain period. However, the FID is a time-domain signal with contributions typically from many different nuclei having slightly different

frequency. The usual frequency-domain spectrum is obtained by computing the Fourier transform of the FID.

2.4 Basic NMR Parameters

Relaxation Times (T_1 and T_2) : Relaxation is the process by which a nuclear spin system, perturbed by absorbing an RF energy, returns back to thermal equilibrium. There are two relaxation processes; spin-lattice or longitudinal relaxation, characterized by the time T_1 and spin-spin or transverse relaxation, characterized by the time T_2 [49]. At equilibrium, the net magnetization vector, $\mathbf{M_0}$, lies along the z-axis, so there is no transverse component and the longitudinal component represents the total magnetization, i.e., $M_Z = M_0$ and $M_X = M_Y = 0$ (Figure 2.4 (A)). A 90^0 RF pulse with appropriate energy and phase rotates the magnetization and over time (of the order of 10 μs) diminishes M_Z to zero (Figure 2.4 (B)). After the pulse, the spin system gradually (of the order of 1 s for a typical protein) returns back to the equilibrium and M_Z regains its original value. The time constant which describes how M_Z returns to its equilibrium value is called the spin-lattice relaxation time (T_1). The equation governing this behavior as a function of time t after the displacement is

$$M_Z(t) = M_0 (1 - e^{-t/T_1})$$

Again, when the net magnetization is rotated to the XY-plane by the application of a 90^0 RF pulse, it evolves as transverse magnetization and precesses in the plane under the influence of the static external field. However, immediately it starts to dephase (fan out) (Figure 2.4 (C)). Fluctuating fields which perturb the energy levels of the spin states cause the transverse magnetization to dephase. The longer the elapsed time, the greater the phase difference is. The time constant that describes how the transverse magnetization, M_{XY}, returns to its equilibrium value, M_{XY_0}, is called the spin-spin relaxation time (T_2). This is governed by

$$M_{XY}(t) = M_{XY_0} e^{-t/T_2}$$

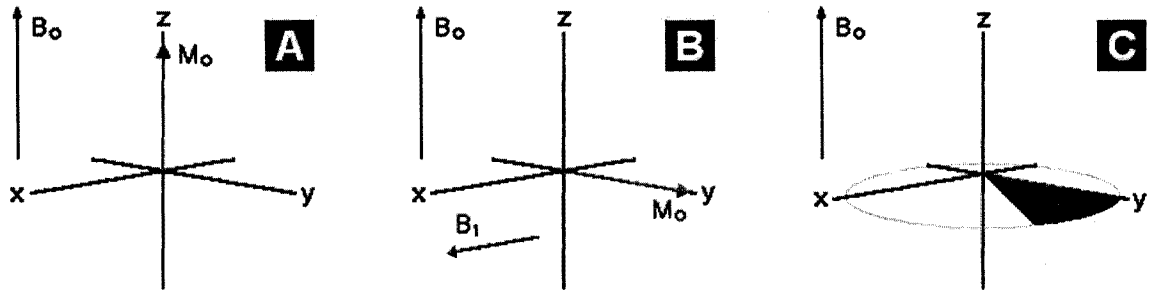


Figure 2.4 : The net magnetization $\mathbf{M_0}$ at different stages. (A) In static magnetic field $\mathbf{B_0}$, $M_Z = M_0$. (B) A 90^0 RF pulse $\mathbf{B_1}$ rotates the magnetization into the xy-plane. (C) The magnetization starts to diphase immediately due to fluctuations in the magnetic fields.

In reality both the T_1 and T_2 processes occur simultaneously, the only restriction being that T_2 is always less than or equal to T_1. Following a pulse sequence that perturbs the spin system from equilibrium, T_2 governs the length of time during which the FID can be observed and T_1 governs the minimum time required for equilibrium to be restored. T_2 plays a crucial role in protein NMR. It is inversely proportional to the overall rotational correlation time (τ_c) of the protein and thus depends on the mass and shape of the molecule. There are actually two factors that contribute to the decay of transverse magnetization; mutual exchange of spin energies (said to lead to a pure T_2 molecular effect) and variations in the static external field $\mathbf{B_0}$ (said to lead to an inhomogeneous T_2 effect). The combination of these two factors is what actually results in the decay of transverse magnetization. The combined time constant is called T_2^* and is expressed as

$$1/T_2^* = 1/T_2 + 1/T_{2\text{inhomo}}$$

Typical values of T_1, T_2 and τ_c for proteins in solution are 1, 0.05 and 10^{-9} s, respectively.

Chemical Shift : In NMR spectroscopy, nuclei of different elements, having different gyromagnetic ratios, yield signals at different frequencies when placed in a particular external magnetic field (Table 2.1). However, the observed resonance frequencies also depend on the local chemical environments that the individual nuclei find themselves in and differ slightly from the frequencies predicted by

$$\omega_0 = \gamma B_0$$

The difference between the predicted and observed resonance frequencies is referred to as a chemical shift. It offers the possibility of distinguishing between nuclei in different chemical environments. This phenomenon arises because motions of electrons induced by the external magnetic field generate secondary fields. The net local magnetic field experienced by a nucleus results from the combination of the external and the secondary fields. The effect of the secondary field, called nuclear shielding, can enhance or oppose the main field. Depending on the chemical environment around the nucleus, each nucleus experiences a slightly different degree of shielding resulting in the variation in local magnetic fields [45, 50]. Consequently, the resonance condition is modified to

$$\omega = \gamma B_{local} = \gamma B_0(1-\sigma)$$

where ω is the effective resonance frequency due to the actual local field B_{local} experienced by the nucleus and σ is a nondimensional screening or shielding constant. However, it is difficult to determine the resonance frequency of a nucleus with absolute accuracy. It is rather easier to determine the difference between the resonance frequencies of a nucleus of interest and a reference [45]. Since the induced secondary field that causes the chemical shift is about a million times weaker than the applied main static field, the chemical shift is expressed in terms of parts per million (ppm). This also removes the dependence of the chemical shift on applied magnetic field strength and makes it a dimensionless number δ (Figure 2.5) given by

$$\delta = [(\omega_{int} - \omega_{ref})/\omega_0] \text{ X } 10^6$$

in which ω_{int} and ω_{ref} are the frequencies of the signal of interest and the reference signal, respectively, and ω_0 is the operating frequency of the spectrometer. In practice the reference signal is set to 0 ppm so that $\omega_{ref} = \omega_0$. The most commonly used reference compounds are tetramethylsilane (TMS) and sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). TMS is used as an external reference while DSS is used as an internal reference [45].

Scalar Coupling : Scalar interactions (through-bond) between nuclei linked via a small number of covalent bonds in a chemical structure result in mutual splitting of the NMR

signal from each nucleus into multiplets (Figure 2.5) [47]. It occurs because the nuclei are coupled through interactions mediated by the electrons forming the chemical bonds between the nuclei. This coupling is ordinarily not important beyond three bonds. The two-bond coupling is termed as geminal (e.g., H-C-H) while the three-bond one as vicinal (e.g., H-C-C-H). To a first approximation, the relative intensities of the multiplets are given by binomial coefficients; 1:1 for a doublet (i.e., a nucleus interacting with another nucleus), 1:2:1 for a triplet (i.e., a nucleus interacting with two other equivalent nuclei) and 1:3:3:1 for a quartet (i.e., a nuclei interacting with three other equivalent nuclei). The difference between any two adjacent components of a multiplet is the same and yields the value of the spin-spin coupling constant J (in Hz). This is independent of the magnetic field strength. To simplify a protein spectrum and to improve the signal to noise (S/N) ratio, decoupling is often employed by irradiating the nuclei coupled to the observed nuclei with a weak RF field.

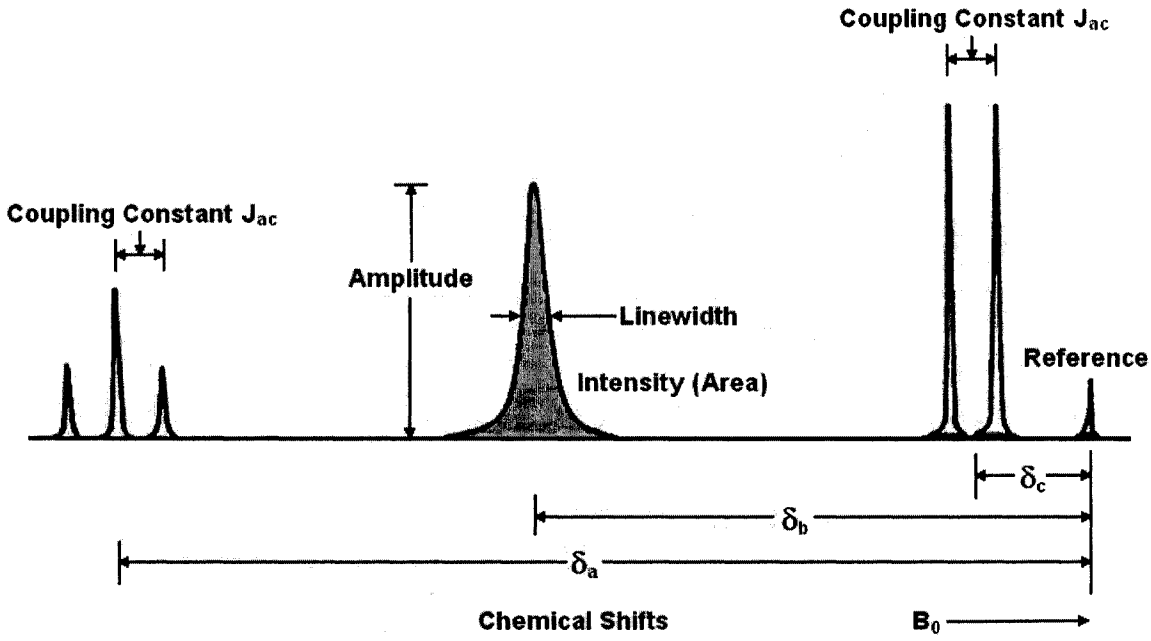


Figure 2.5 : Basic NMR spectral parameters. The reference signal is assigned a chemical shift of 0 ppm. The chemical shifts of the other signals are measured from the reference and named as δ_a and so on. The height of a signal represents its amplitude and the area under the curve represents its intensity which is proportional to the number of nuclei contributing to the signal. The coupling constant represents the difference between two adjacent multiplets which results from scalar coupling.

Linewidth : In an NMR spectrum a series of absorption peaks or signals, representing resonance frequencies of the nuclei in different chemical environments, are observed. The area of a signal (not the height) is directly proportional to the number of nuclei contributing to the signal under suitable experimental conditions. The linewidth ($W_{1/2}$) is defined as the full-width at half-height of the resonance line shape (Figure 2.5) [47]. It is a primary factor affecting both resolution and S/N ratio of the spectrum. It is directly related to the spin-spin or transverse relaxation time T_2 by

$$W_{1/2} = 1/\pi T_2$$

The relaxation time, however, is inversely proportional to the overall rotational correlation time (τ_c) of the molecule in solution. Again, τ_c depends on mass and shape of the molecule. Thus, observed linewidths significantly larger than expected primarily indicate increase in rotational correlation time caused by molecular aggregation. If aggregates are large enough, the lines become too broad to observe.

Nuclear Overhauser Effect (NOE) : The NOE is the fractional change in intensity of one NMR line when another resonance is irradiated in a double irradiation experiment [51]. It is customarily quoted in percent of the unperturbed resonance intensity. The NOEs are observed because of the dipolar cross-relaxation (through-space) between pairs of nuclei (protons) with sufficiently close spatial proximity (usually < 5 Å). NOE intensity is proportional to the inverse sixth power of the distance between the two nuclei and given by

$$\text{NOE} = f(\text{dynamics}) \times 1/r^6$$

Thus the NOEs observed between pairs of protons in a protein characterize inter-proton distances that are invaluable for the 3D structure determination. Another application of this effect is the information on protein dynamics since the NOE is a function of the dynamics too. The dynamics arise from the conformational changes in proteins while performing their specific functions. With the NOE experiment for each amino acid, the rigidity and flexibility of the local backbone conformation of a protein can be determined in a quantitative manner [52].

2.5 Protein NMR Experiments

Complex multi-dimensional, as well as simple one-dimensional, NMR experiments are performed for the purpose of protein structure determination. Proton (^{1}H) is the most useful nucleus in this context because of its large natural abundance and large gyromagnetic ratio (Table 2.1). However, ^{13}C or ^{15}N nuclei are also used in protein NMR by replacement of their natural isotopes (^{12}C and ^{14}N). The structure determination process begins with identifying which resonance frequency belongs to which magnetic nucleus in the chemical structure of the protein. Afterwards, NOEs are used to identify pairs of protons that are close in space. The NOE data is combined with molecular dynamics simulation to provide the 3D atomic coordinates of the protein.

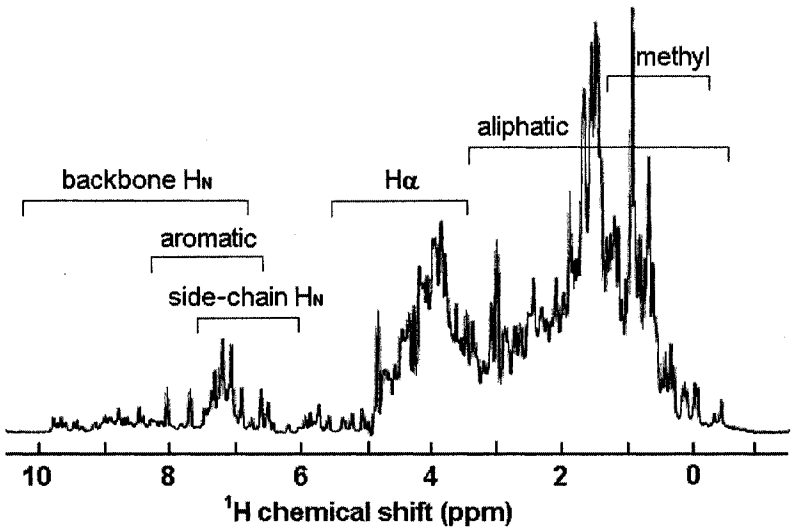


Figure 2.6 : Schematic representation of 1D ^{1}H NMR spectrum of a protein. Chemical shift ranges observed for the various types of proton resonances are shown.

In a simple 1D ^{1}H experiment, the net magnetization that evolves along the z-axis is rotated into the xy-plane by a 90^0 RF pulse of sufficient power for a few microseconds. At the end of the pulse, the transverse magnetization decays and the data acquisition begins by recording the FID. The time domain FID is then Fourier-transformed to yield the frequency domain spectrum. Because the resonance frequency is modulated by the chemical environment that a ^{1}H finds itself in, the 1D spectrum shows a dispersion of signals obtained from protons at different positions in various amino acids. However, a

group of protons at similar positions in all amino acids (e.g., amide protons or α-protons) has a characteristic range of frequencies and hence the chemical shifts (Figure 2.6) [45].

The second basic experiment, the proton-detected 2D Heteronuclear Single Quantum Correlation (HSQC), uses a magnetization transfer mechanism. For this experiment, at least some the ^{12}C and/or ^{14}N nuclei must have been replaced by their spin 1/2 isotopes ^{13}C and/or ^{15}N during the protein production. ^{13}C will not be mentioned any further as only ^{15}N labels were used for present studies. The magnetization is transferred from covalently linked ^{1}H to ^{15}N during the mixing period that precedes the acquisition. The peaks in the spectrum correlate the frequencies of ^{15}N-^{1}H pairs of nuclei (Figure 2.7).

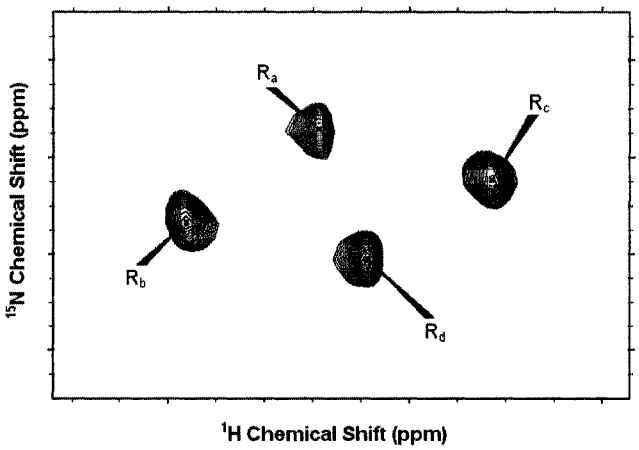


Figure 2.7 : Schematic representation of 2D ^{15}N-^{1}H HSQC spectrum for a protein with selected ^{15}N-labeled residues. The peaks correlate the resonance frequencies of covalently linked ^{15}N - ^{1}H pairs of nuclei. R_a, R_b, R_c and R_d represent peaks from four different ^{15}N-labeled amino acids.

The 1D ^{1}H and 2D HSQC experiments are routinely done to optimize the sample conditions (concentration, pH, temperature, etc.) and to check the status of the sample before and after any complex multidimensional experiments. The spectra also indicate the homogeneity of the protein's structure, molecular aggregation, etc. Preliminary hints on secondary structure (α-helix/β-sheet) may also be obtained as a particular proton group (especially the backbone amide protons) exhibits characteristic shifts in frequency with respect to that of random coil configuration.

To determine a protein's structure, the resonance frequency of each magnetic nucleus must be determined. For that, a number of multidimensional homonuclear and

heteronuclear experiments are performed which eventually lead to the complete resonance assignments.

2D Total Correlation Spectroscopy (TOCSY) and Nuclear Overhauser-Effect Spectroscopy (NOESY) are the standard homonuclear experiments for identifying spin-spin coupling connectivities between pairs of protons. These experiments rely on magnetization transfer from one proton to the next, which takes place during a part of the pulse sequence called the mixing time. In TOCSY, the magnetization is transferred from one proton to another through covalent bonds using scalar coupling (up to 3 bonds). In NOESY, the magnetization transfer is achieved through space using dipolar coupling (up to ~ 5 Å). The diagonal peaks of these 2D spectra reproduce the complete 1D ^{1}H spectrum. The off-diagonal cross-peaks, on the other hand, correlate pairs of protons between which the magnetization has been transferred (Figure 2.8, only selected few peaks are shown for clarity).

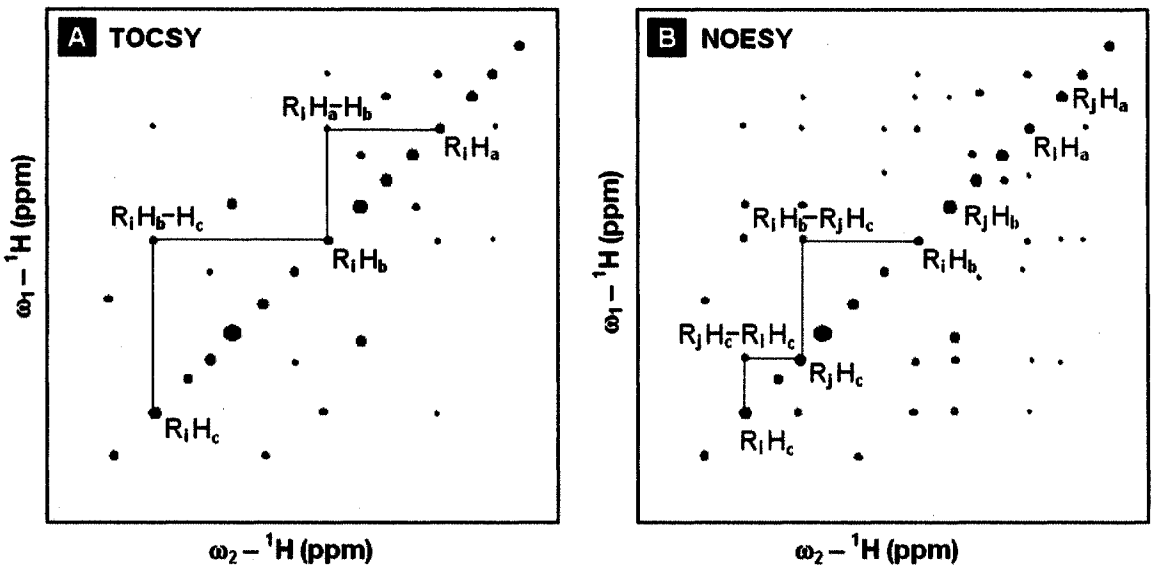


Figure 2.8 : Schematic representations of 2D homonuclear NMR spectra for a protein. (A) TOCSY, in which the cross-peaks (red) result from the magnetization transfer between protons through covalent bonds (up to 3 bonds). (B) NOESY, in which the cross-peaks result from the magnetization transfer between protons through space (up to ~ 5 Å). NOESY generates peaks from both intra-residue (red) as well as sufficiently close inter-residue (green) proton-pairs. The diagonal peaks (blue) represent complete 1D ^{1}H spectrum in both. The spectra are symmetrical with respect to the diagonal.

The NOEs (observed as the cross-peaks in NOESY spectrum) act as the primary source of structural information in NMR, as they indicate the spatial proximity of protons within the molecule. The intensities of the NOEs vary approximately as the inverse sixth power of the inter-proton distances. However, there are difficulties in using the NOE intensity quantitatively since the intensity is also affected by spin diffusion, intra-protein motions and artifacts. Therefore, all the NOE intensities are classified into three categories (strong, medium and weak) and a set of distance restraints is developed. This set of distance restraints act as the most crucial input in the protein structure calculation.

Additional restraints are derived from experiments that are sensitive to backbone dihedral angles. Double Quantum Correlated Spectroscopy (DQCOSY) measures the scalar coupling between the H_N and H_α, the strength of which depends on the dihedral angle ϕ. The frequencies of the resonances (and hence the chemical shifts) are sensitive to the secondary structure and can be used in combination with databases of structures and chemical shifts to predict backbone dihedral angles [50, 53]. Information on secondary structure obtained by other methods, such as Fourier Transform Infrared (FTIR) Spectroscopy, may also be used to generate H-bond restraints between different residues.

As 2D homonuclear spectra are crowded with cross-peaks (especially the NOESY), overlapping of peaks often provide significant hindrance to resonance assignments. 3D and 4D heteronuclear-edited NMR experiments resolve overlapped cross-peaks according to the chemical shift of the heteronuclei (e.g., ^{15}N) bonded directly to the ^{1}H spins [54, 55]. In 3D experiments, the spectrum is separated along a third dimension (^{15}N or ^{13}C), so that the signals are distributed in a cube instead of a plane. This spreading-out is achieved by combining an HSQC at the end of a TOCSY/NOESY in a single experiment. The acquisition starts after the HSQC instead of at the end of the TOCSY/NOESY [56]. A 4D experiment consists of the concatenation of a NOESY pulse sequence and two HSQC building blocks. The peaks in the 4D spectra are first separated with respect to one heteronuclei to generate 3D cubes which are then sliced further with respect to the other heteronuclei attached to ^{1}H spins.

2.6 Protein Structure Calculation

Protein structure is calculated using computer programs that convert proton-proton distance and dihedral angle data (restraints) into 3D molecular coordinates. The calculated structure reflects both the experimental data, as well as restrictions on bond lengths, angles, etc. that are known from the chemical structure of the protein. Both the NMR data and chemical structure data are implemented as empirical energy functions. In practice, an extended initial structure with ideal covalent geometry is generated first, using the known amino acid sequence. The computer program then attempts to refold the initial structure in a way such that the experimentally determined inter-proton distances are satisfied. To achieve this, each known parameter is assigned an energy potential. The program calculates a structure by minimizing the overall energy [57].

Without the experimentally determined distance and dihedral angle restraints from the NMR spectra, the protein can adopt a huge number of conformations that are still consistent with the covalent geometry imposed by its chemistry. It is therefore important to identify as many restraints as possible from the NMR spectra to restrict the conformational space to the subset of space actually sampled by the protein.

Simulated annealing, combined with the molecular dynamics simulation, is the most commonly used method for calculating NMR structure of a protein in solution. In this method, an initial structure is heated to a high temperature in a simulation giving the atoms a high thermal mobility. During the subsequent cooling steps, the initial structure can evolve towards the energetically favorable final structure under the influence of a force field derived from the restraints [58]. The total energy is classified into two categories and given by,

$$E_{Total} = E_{Empirical} + E_{Experimental}$$

with $$E_{Empirical} = E_{Bond} + E_{Angle} + E_{van\text{-}der\text{-}Waals} + E_{Electrostatic}$$

and $$E_{Experimental} = E_{NOE} + E_{Dihedral} + E_{H\text{-}bond}$$

The structures are calculated using these energy functions and the simulated annealing protocol which finds the sets of coordinates that minimize the total energy [48].

E_{NOE}, the pseudo energy term corresponding to the NOEs, plays crucial role in protein structure calculation. However, the NOE-derived inter-proton distances are imprecise. Therefore, many closely related structural models are consistent with the observations. Hence, NMR structures of proteins are usually reported as ensembles of atomic coordinates and the conformational space sampled by the ensemble reflects the conformational space sampled by the protein in solution.

2.7 Strategies for Hydrophobic Proteins

Hydrophobic proteins are not water soluble. SP-B is a lipid-associated hydrophobic protein and hence presents unique challenges to structure determination by solution NMR. However, in solution NMR, detergent or lipid micelles can be used to mimic the lipid environment in which SP-B functions in the lungs (Figure 2.9). A micelle is an aggregate of lipid molecules, usually of globular shape, in a polar solvent (mostly water). These lipid molecules contain a polar (hydrophilic) head group and one long or two short nonpolar (hydrophobic) tail(s) (acyl chain(s)). When the lipid concentration in water is greater than a certain value, known as the critical micellar concentration (CMC), spherical structures called micelles are formed spontaneously by exposing the hydrophilic head groups to water molecules and burying the hydrophobic tails inside.

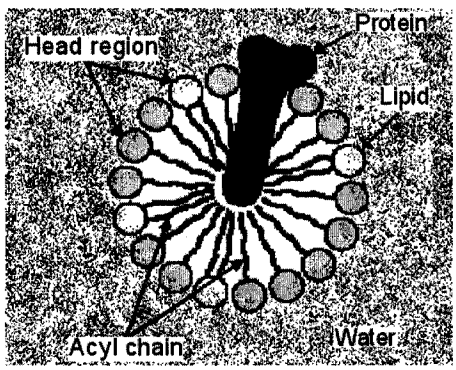


Figure 2.9 : Schematic representation of lipid micelle and protein/peptide. When mixed in water in excess of CMC, lipids with single acyl chain or short double acyl chains form spherical micelles. The membrane or lipid-associated proteins/peptides settle on the micelles which may resemble the physiological situations.

However, the lipid-protein complex is significantly bigger than the size of the protein itself. The larger complex tumbles more slowly which leads to more efficient transverse relaxation. Faster transverse relaxation rates lead in turn to broader lines in the NMR spectra and consequent decreases in the S/N ratio, as well as increases in peak overlap. The relaxation slows down because of the higher rotational correlation time which in turn broadens the resonance linewidths affecting both the resolution and the S/N ratio. As a result, the micelle data of protein/peptide is time consuming to analyze compared to data acquired in solvent.

Organic solvents provide an alternative means to solubilize hydrophobic proteins such as SP-B for solution NMR studies. However, the structure of the protein in organic solvent is unlikely to resemble its lipid-associated structure. Hence, both organic solvent as well as micelles formed from different lipids have been employed for the structural studies of SP-B peptides (SP-BCTERM and Mini-B).

In the first phase of this research, the conformational properties of SP-BCTERM and its ability to bind micelles together were studied in four different lipid micelle systems. The spectral features were compared with that obtained by Booth *et al.* while determining the structures of SP-BCTERM in hexafluoroisopropanol (HFIP) and sodium dodecyl sulphate (SDS) [42]. The lipids used were dodecylphosphocholine (DPC), lysophosphatidylglycerol (LPG), lysophosphatidic acid (LPA) and dihexanoylphosphatidylcholine (DHPC), which are micelle-forming analogs of natural lung surfactant lipids PC and PG (Figure 2.10). PCs are the most abundant phospholipids in the surfactant. They have two acyl chains connected to a zwitterionic head that contains both positively and negatively charged chemical groups. DPC (Figure 2.10A) has a similar zwitterionic head region as PC but it has one acyl chain. LPG (Figure 2.10B) contains similar anionic (negatively charged) head group as PG but again has one acyl chain. LPA (Figure 2.10C) also possesses an anionic head group and a single acyl chain. The head group of LPA is not similar to that of any surfactant lipids, but this was still chosen to see the response of the cationic (positively charged) SP-B_{CTERM} in an anionic environment additional to LPG.

DHPC (Figure 2.10D) has two short acyl chains but the head group is completely analogous to that of PC.

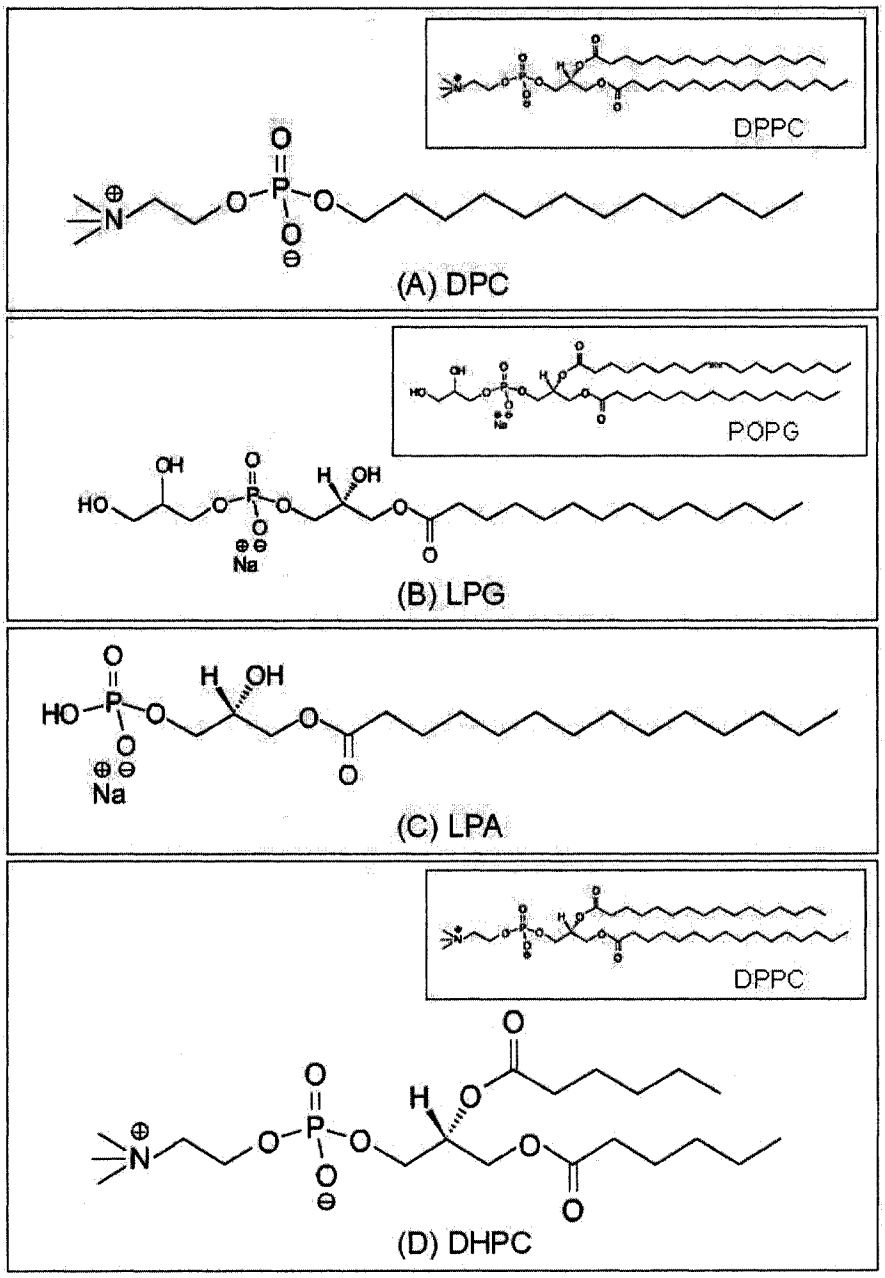


Figure 2.10 : Micelle-forming analogs of the lung surfactant phospholipids (shown in the insets) used to study the conformational features of SP-B$_{CTERM}$. (A) dodecylphosphocoline (DPC), (B) lysophosphatidylglycerol (LPG), (C) lysophosphatidic acid and (LPA) and dihexanoylphosphatidylcholine (DHPC).

While preparing the samples, each of the lipids were mixed in H_2O/D_2O in amount above the CMC to ensure the spontaneous formation of micelles. Table 2.2 lists some basic properties of the micelles formed from the lipids used in present work including that of SDS, which is the most commonly used detergent to form micelles for NMR structural studies of proteins [59 - 65].

Table 2.2 : Properties of the micelle-forming lipids used to study SP-B$_{CTERM}$ [59 - 65]

Lipid	SDS	DPC	LPG	LPA	DHPC
CMC (mM)	1.57	1	0.018	0.54	14
Micelle Diameter (nm)	4	4	5.58	Not available	3.3
Molecules/Micelle	70	56	125	Not available	40

In the second and main phase of the research, the structure of the other SP-B peptide, Mini-B, was determined in deuterated organic solvent HFIP (Figure 2.11). Two fluorinated alcohols, tri-fluoro-ethanol and methanol, have been the most common organic solvents used for protein structure determination. However, HFIP has been found to offer better homogeneity of the protein conformation and the sample remains stable over longer periods. Therefore, HFIP was chosen for Mini-B structure determination. This research provided the first detailed structural characterization of Mini-B. This will also work as a primary database for any future NMR structural studies of Mini-B in lipid micelle systems.

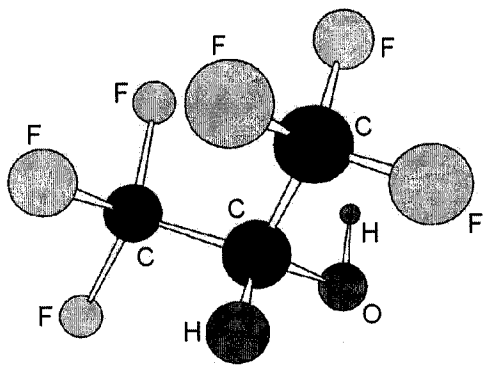


Figure 2.11 : Molecular structure of organic solvent hexafluoroisopropanol (HFIP) used to determine the structure of Mini-B.

Chapter 3

SP-BCTERM – Lipid Interactions

3.1 Overview of Structural Studies of SP-B$_{CTERM}$

Structures of SP-B$_{CTERM}$ were determined earlier by Booth *et al.* [42] in both organic solvent HFIP and lipid micelles formed from SDS. The first phase of the present research was originally intended to investigate the conformational features of this peptide in more physiologically relevant environments than HFIP and SDS. Four different lipids (DPC, LPG, LPA and DHPC) were chosen for the purpose (see Section 2.7, Figure 2.10). These lipids form micelles in water and are single acyl chain or short double acyl chains analogs of phospholipids found in the natural lung surfactant.

The investigations began with simple NMR experiments (1D ^{1}H and 2D ^{15}N-^{1}H HSQC) of SP-B$_{CTERM}$ in DPC, LPG, LPA and DHPC micelles. The spectra indicated aggregation of micelles, resulting in large complex sizes not amenable to study by solution NMR. To assess the basic structural properties of the peptide in presence of the aggregation, Circular Dichroism (CD) was used, in collaboration with David Heeley (Department of Biochemistry, MUN). In order to directly observe the aggregation implied by the NMR spectra, Dynamic Light Scattering (DLS) data were collected, in collaboration with Qi Wang (Agriculture and Agri-Food Canada). Overall, the results indicated that SP-B$_{CTERM}$ induced huge aggregation of LPG and LPA micelles, significant aggregation of DPC micelles and maybe some aggregation of DHPC micelles. This result is very relevant to SP-B's function in lungs where it is thought to be important in reorganizing lipid materials and stabilizing the structures at different stages of surfactant's life cycle.

To underline the uniqueness of SP-B$_{CTERM}$'s interactions with lipids, similar studies were also performed with the peptide RP-1. RP-1 possesses similar topological features to SP-B. It consists of a single helix which is hydrophobic on one face and hydrophilic on the other, and has several positively charged residues. At least in DPC, RP-1 did not induce the aggregation of micelles that was observed with SP-B$_{CTERM}$.

3.2 Materials and Methods

SP-BCTERM is a 16 residue peptide that contains most of the C-teminal region of human SP-B. It consists of amino acids 63 – 78 of the full-length protein. The amino acid sequence is :

Gly-Arg-Met-Leu-Pro-Gln-Leu-Val-Cys-Arg-Leu-Val-Leu-Arg-Cys-Ser

1 5 10 15

The peptide had six ^{15}N labels at leucines 4, 7, 11 and 13 and valines 8 and 12, i.e., the natural isotope ^{14}N of those residues at the backbone was replaced by ^{15}N which is a spin 1/2 nucleus. The peptide has two cysteines and three positively charged Arginines.

The peptide was synthesized chemically by Alan Waring (UCLA), following the solid phase peptide synthesis method [42]. Materials (organic solvents and other reagents) used for the synthesis and purification of SP-BCTERM were high performance liquid chromatography (HPLC) grade or better (Fisher Scientific, Aldrich Chemical). O-fluorenylmethyl-oxycarbonyl (Fmoc) chemistry was employed for synthesizing the peptide. Fmoc amino acids were supplied by AnaSpec. The ^{15}N-labeled amino acids were obtained from Cambridge Isotope Laboratories and converted to their Fmoc derivatives also by AnaSpec. After synthesis, the molecular weight of the peptide was confirmed by fast atom bombardment or MALDI-TOF mass spectrometry and the purity (> 95%) was determined by analytical HPLC. The purified peptide was lyophilized (powdered by cooling in liquid nitrogen) and stored at 4 ^0C.

All the four NMR samples of SP-BCTERM in lipid micelles were prepared by mixing 1.5 mM peptide, 150 mM lipid and 0.4 mM DSS in 90% H_2O and 10% D_2O. The peptide was weighed out in dry form assuming it was 100% pure. DSS was used to obtain the proton reference signal at 0 ppm and D_2O was used to provide the lock nucleus (deuterium) to the field-frequency lock system of the spectrometer. The volume of the sample was 600 μL for each. To avoid any undesired folding of the peptide resulting from disulfide bonds between cysteines (both intra- and inter-chain), the pH of the

samples were set below 7. The pH was measured by an accumet Electrode bought from Fisher Scientific.

The lipids were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL) and the DSS and D_2O were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA).

The NMR experiments (1D ^1H and 2D ^{15}N-^1H HSQC) for SP-BCTERM were performed on a Bruker Avance 500 MHz spectrometer of Centre for Chemical Analysis, Research and Training (C-CART) at Memorial University of Newfoundland (MUN). The spectrometer was equipped with z-gradients and triple-resonance TXI probes. Presaturation and flip-back water suppression methods were used for 1D ^1H and 2D HSQC experiments, respectively. The spectra were processed using the software NMRPipe 2.2 [66].

The CD spectra for two SB-BCTERM samples (in DPC and LPG) were obtained using a Jasco J-810 spectropolarimeter of Heeley Lab at MUN. The CD samples were prepared by taking portions of NMR samples and diluting these about 10^3 times. Experiments were done at 25 ^0C using light path of 2 nm taking 2 accumulations with scanning speed of 100 nm/sec. The DLS spectra for DPC samples (with and without the peptide) were acquired by a DLS system of Agriculture and Agri-Food Canada, Guelph. The DLS samples maintained the same material concentrations and experimental conditions as NMR.

3.3 NMR Spectra of SP-B$_{CTERM}$ in DPC Micelles

The first lipid micelle system used to study the conformational and micelle-binding features of SP-B$_{CTERM}$ was formed by the zwitterionic lipid DPC (a single chain analog of DPPC). The sample had a pH of 5.5 and the initial NMR experiments were done at 25 ^{0}C.

The 1D ^{1}H spectrum of SP-B$_{CTERM}$ in DPC is shown in Figure 3.1. Corresponding spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are shown in the insets. Comparison of the backbone H_N regions (~ 6.5 – 8.5 ppm) indicate that the spectrum of SP-B$_{CTERM}$ in DPC did not show dispersion of resonance signals at a level comparable to the spectra of SP-B$_{CTERM}$ in HFIP and SDS. The signals were also weaker as shown by the numbers corresponding to the intensities of the highest signals in H_N regions normalized by the number of scans. The intensity of the signal in DPC is over 5 times weaker than that in HFIP and about 4 times weaker than that in SDS.

The 2D ^{15}N-^{1}H HSQC spectra of SP-B$_{CTERM}$ obtained in DPC is shown in Figure 3.2. Corresponding spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are also shown. The peptide had six ^{15}N-labeled amino acids. The spectra in HFIP and SDS showed the expected six peaks corresponding to those amino acids. However, the spectrum in DPC showed more than six peaks. Also the peaks in DPC were weaker by factors of about 50 and 5, respectively, than what were observed in HFIP and SDS, as shown by the numbers representing the contour levels (CL).

The experiments were repeated at 45 ^{0}C but no significant changes were noticed. The 1D ^{1}H proton signals seemed a little more dispersed but still weaker. The 2D HSQC spectrum continued to show more and weaker peaks. Change in pH of the sample also did not produce any significant alteration in spectral features.

The most likely explanation for the reduction in peak intensity is that the peptide induced aggregation of the DPC micelles resulting in large complexes. The intensities reduced because the resonance linewidths were broadened as a consequence of more efficient transverse relaxation caused by the large complexes. The extra peaks in 2D HSQC indicated that multiple conformations of the peptide were present in DPC environment.

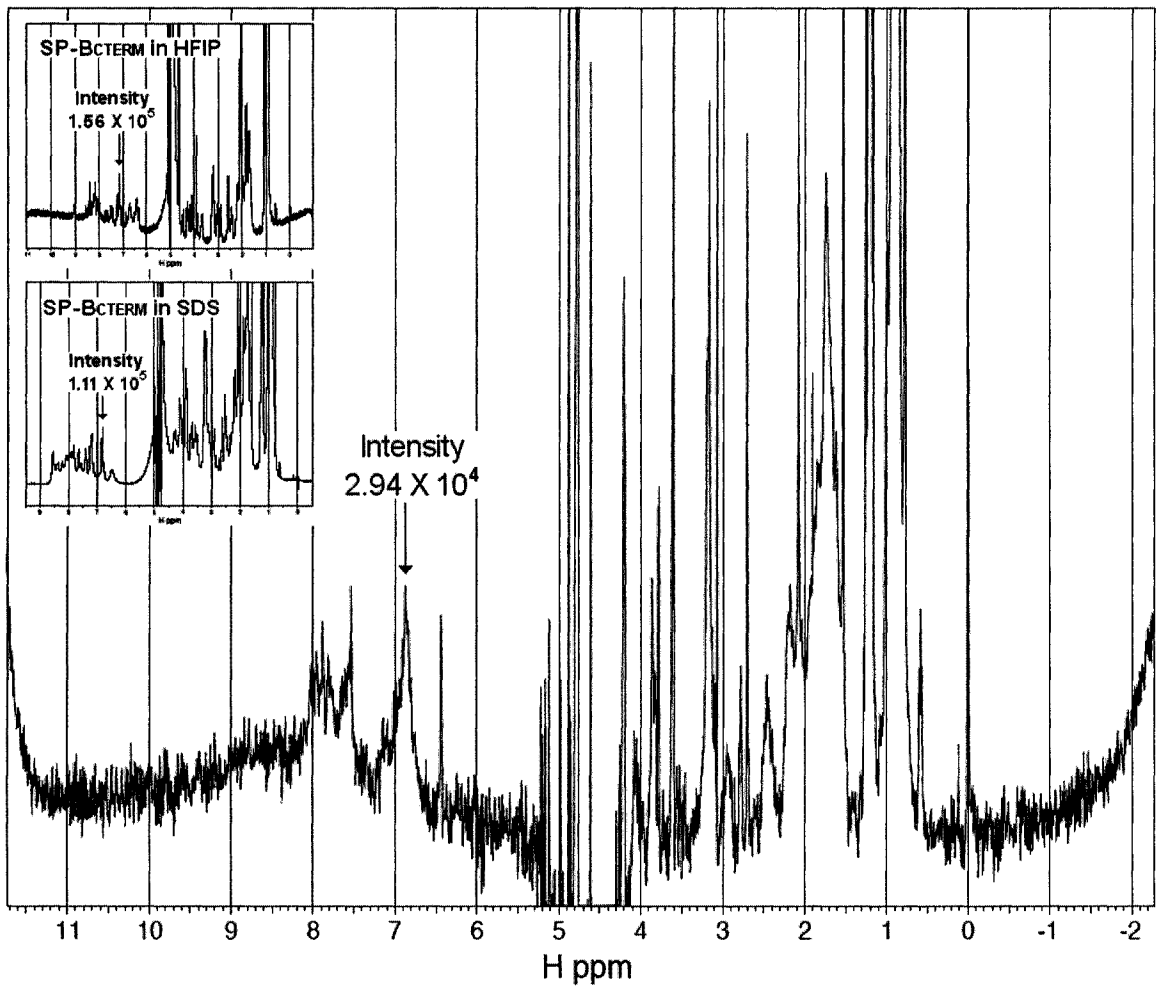


Figure 3.1 : 1D ^{1}H spectrum of SP-BCTERM in DPC micelles. All the regions including that for the backbone amide protons (~ 6.5 – 8.5 ppm) show much less dispersion of chemical shift signals in comparison with what were observed earlier by Booth *et al.* in HFIP and SDS (shown in the insets). The intensity of the signals in DPC are also weaker than that in HFIP and SDS as shown by the corresponding numbers.

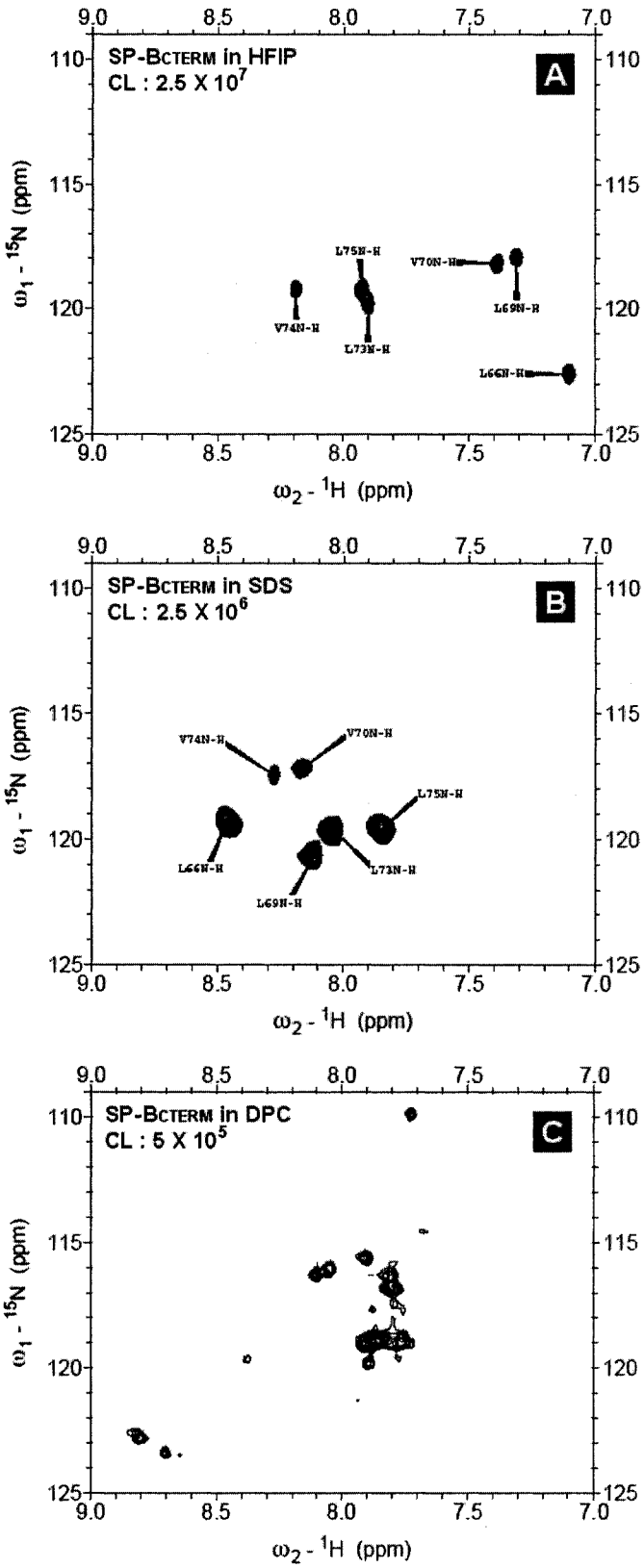


Figure 3.2 : 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM in HFIP (A) and SDS micelles (B) obtained earlier by Booth *et al.* and in DPC micelles (C). The peptide had six ^{15}N-labeled amino acids. More peaks are seen in DPC spectrum, although the spectra of HFIP and SDS samples had six peaks. Also, the DPC peaks are weaker than that of HFIP and SDS as indicated by the contour levels (CL).

3.4 NMR Spectra of SP-BCTERM in LPG Micelles

Next, the conformational and micelle-binding features of SP-BCTERM were studied in micelles formed by the anionic lipid LPG (a single chain analog of PG). The pH of the sample was set at 4.5 and the initial experiments were performed at 25 ^{0}C.

The 1D ^{1}H spectrum of SP-BCTERM in LPG is shown in Figure 3.3. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are shown in the insets. Comparison of the backbone H_N regions (~ 6.5 – 8.5 ppm) indicate that the spectrum of SP-BCTERM in LPG showed only a few very weak resonance signals that were almost lost in the noise. Signals were too weak to comment on dispersion. The numbers corresponding to the intensities of the highest signals in H_N regions, normalized by the number of scans, show that the intensity of the signal in DPC is over 72 times weaker than that in HFIP and about 52 times weaker than that in SDS.

The 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM obtained in LPG is shown in Figure 3.4. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are also shown. The spectra in HFIP and SDS showed the expected six peaks corresponding to the six ^{15}N-labeled amino acids. However, the spectrum in DPC showed three weak peaks and several very weak peaks. The peaks in LPG were weaker by factors of about 100 and 10, respectively, than what were observed in HFIP and SDS, as shown by the numbers representing the contour levels (CL).

The experiments were repeated at 45 ^{0}C. Almost no change was noticed in the 1D spectrum. The 2D HSQC spectrum showed somewhat six peaks this time but those were still very weak. Change in pH also did not produce significant changes in spectral features.

The reduction in peak intensity, again, was likely due to the peptide induced aggregation of the LPG micelles forming large complexes. Reduction of the peak intensity resulted from the broadenning of the resonance linewidths due to more efficient transverse relaxation caused by the large complexes. The less number of peaks with similar intensity indicate that the petide didnot have proper secondary folding either. Several very weak peaks in 2D HSQC indicate that the peptide took multiple conformations as well.

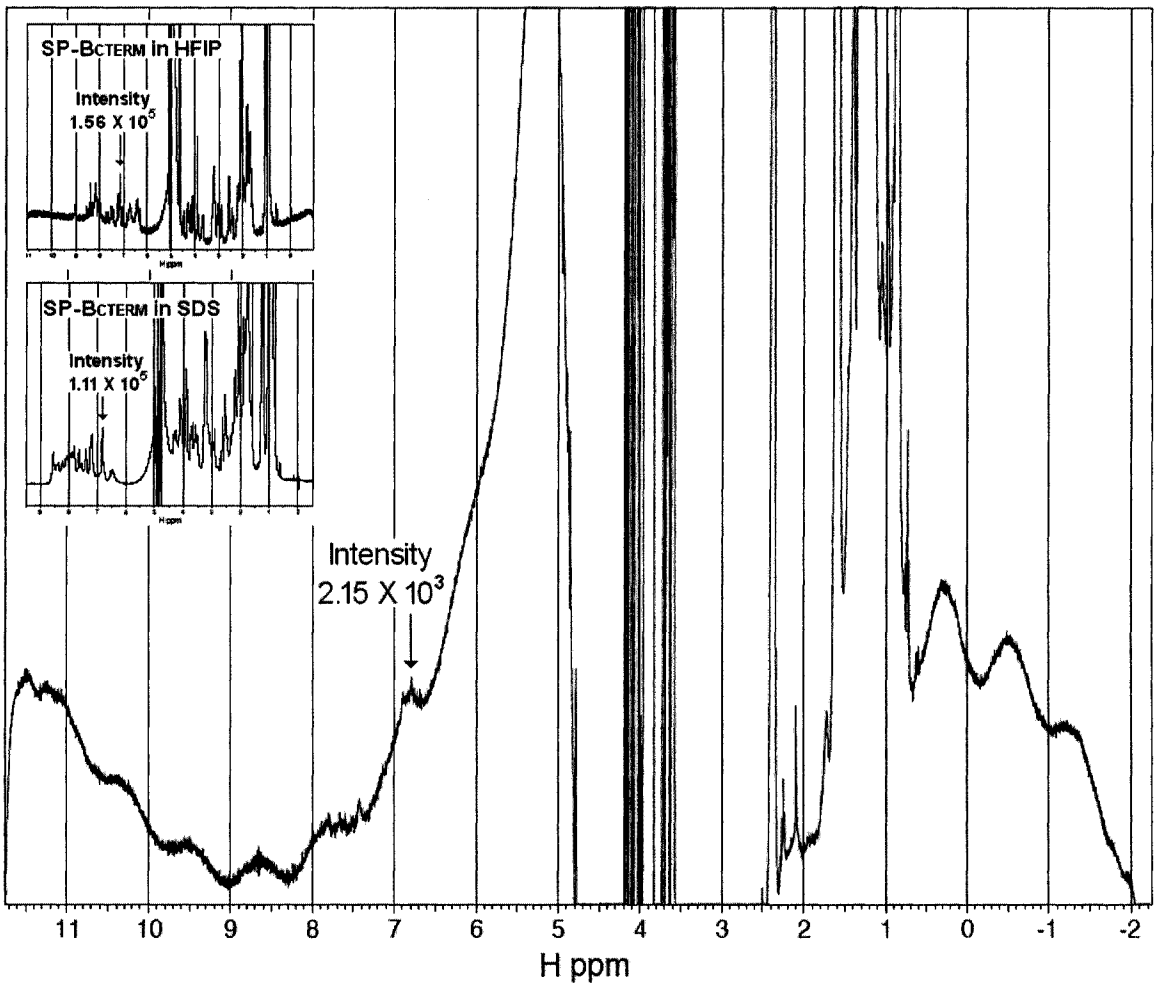


Figure 3.3 : 1D ^{1}H spectrum of SP-BCTERM in LPG micelles. All the regions including that for the backbone amide protons (~ 6.5 – 8.5 ppm) show very little signals. Similar spectra obtained in HFIP and SDS (shown in the insets) obtained earlier by Booth *et al.* showed good disersion. The intensity of the signals in LPG were also much weeker than that observed in HFIP and SDS as shown by the corresponding numbers.

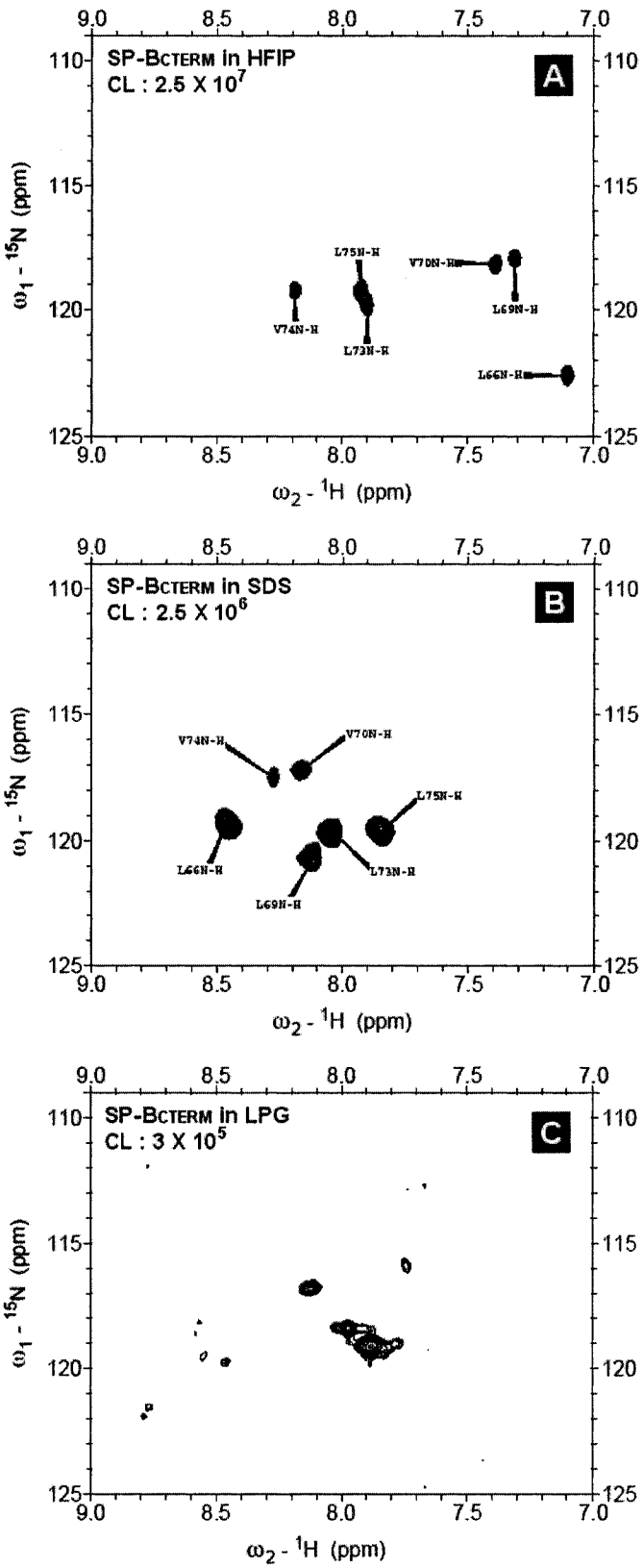


Figure 3.4 : 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM in HFIP (A) and SDS micelles (B) obtained earlier by Booth *et al.* and in LPG micelles (C). The peptide had six ^{15}N-labeled amino acids but the expected number of peaks are not seen in LPG spectrum, although the spectra of HFIP and SDS samples had six peaks. Also, the LPG peaks are much weaker than that of HFIP and SDS as indicated by the contour levels (CL).

3.5 NMR Spectra of SP-B$_{CTERM}$ in LPA Micelles

The cationic protein SP-B has been suggested to bind to the anionic phospholipids of the native surfactant [16]. Hence NMR studies were continued to investigate the conformational and micelle-binding features of SP-B$_{CTERM}$ in another micelle system formed by the lipid LPA. This lipid possesses a single acyl chain and an anionic headgroup. The sample had pH of 3.3 and the data were acquired at 25 ^{0}C.

The 1D ^{1}H spectrum of SP-B$_{CTERM}$ in LPA is shown in Figure 3.5. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are shown in the insets. The backbone H_N region (~ 6.5 – 8.5 ppm) of the spectrum in LPA showed only a few peaks with very poor intensity and very little dispersion. Comparison with the spectra in HFIP and SDS, considering the intensities of the highest signals in H_N regions normalized by the number of scans, showed that the intensity of the signal in LPA is over 21 times weaker than that in HFIP and over 15 times weaker than that in SDS.

The 2D ^{15}N-^{1}H HSQC spectra of SP-B$_{CTERM}$ obtained in LPA is shown in Figure 3.6. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are also shown. The peptide had six ^{15}N-labeled amino acids. The spectra in HFIP and SDS showed the expected six peaks corresponding to those amino acids. However, the spectrum in LPA showed no peaks at all indicating that the intensity of the peptide signals were similar or smaller than that of the noise.

The most likely explanation for the reduction of intensity of the 1D ^{1}H signals, again, is that the peptide induced aggregation of the LPA micelles producing complexes bigger than the NMR accessible sizes. However, no comments on the folding or conformational homogeiniety can be made due to the absence of any peaks in the 2D HSQC spectrum.

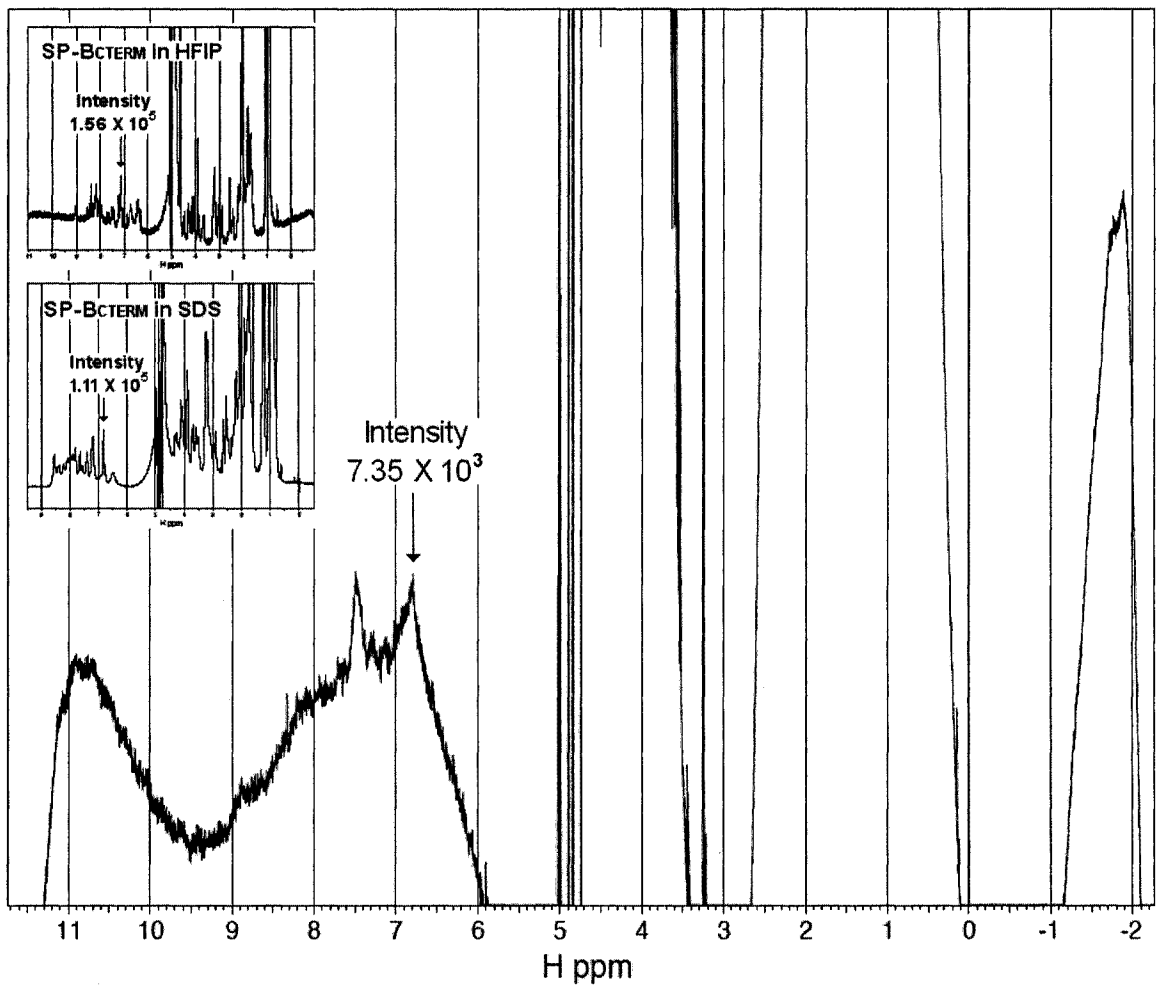


Figure 3.5 : 1D ^{1}H spectrum of SP-BCTERM in LPA micelles. All the regions including that for the backbone amide protons (~ 6.5 – 8.5 ppm) show only a few very weak chemical shift signals in comparison with what were obtained earlier by Booth *et al.* in HFIP and SDS (shown in the insets). The intensity of the signals in LPG are much weaker than that in HFIP and SDS as shown by the corresponding numbers.

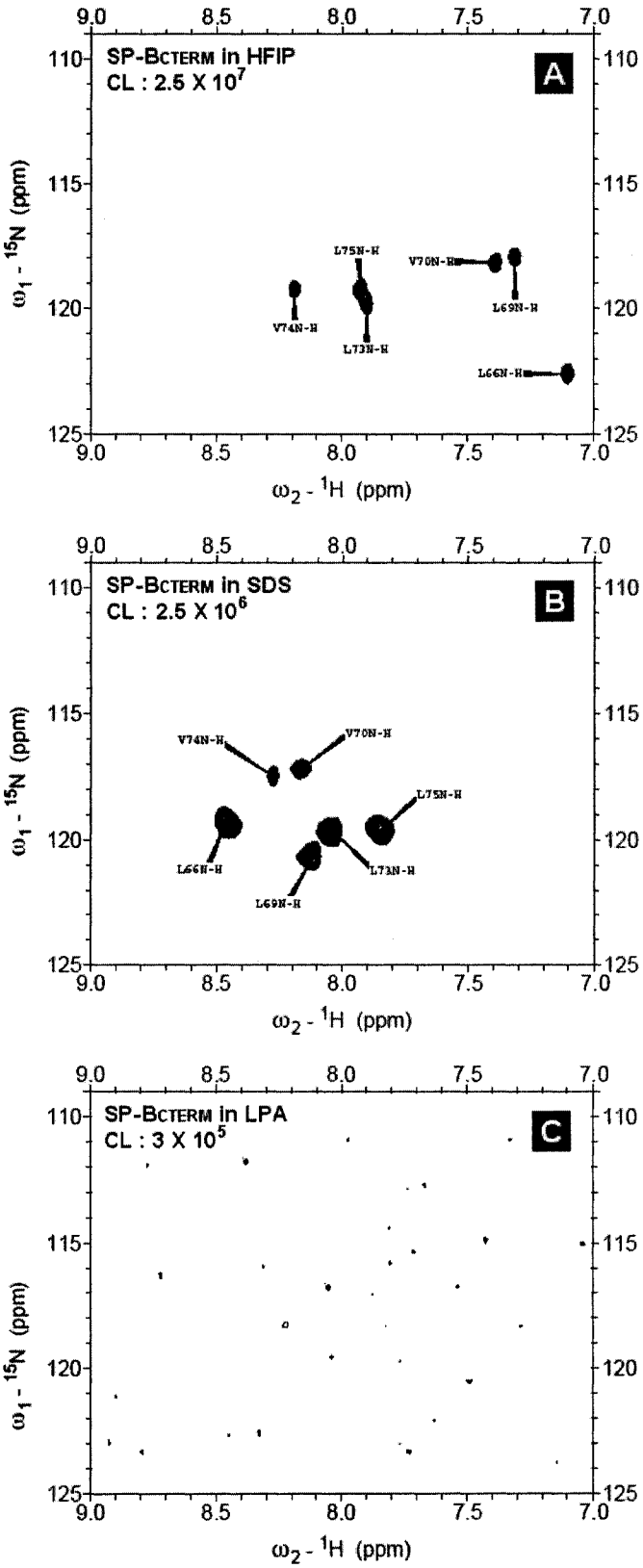


Figure 3.6 : 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM in HFIP (A) and SDS micelles (B) obtained earlier by Booth *et al.* and in LPA micelles (C). The peptide had six ^{15}N-labeled amino acids but no peak is seen for the LPA spectrum, although the spectra of HFIP and SDS samples had six peaks. The peaks for LPA are completely lost in the noise.

3.6 NMR Spectra of SP-BCTERM in DHPC Micelles

Lastly, the conformational and micelle-binding features of SP-BCTERM were investigated in micelles formed by another zwitterionic lipid DHPC (an analog of PC). This lipid contains two acyl chains but those are short enough so that it can form micelles in water and can be used for protein NMR studies. The pH of the sample was 3.3 and the data were recorded at 25 ^{0}C.

The 1D ^{1}H spectrum of SP-BCTERM in LPA is shown in Figure 3.7. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are shown in the insets. As seen from the backbone H_N regions (~ 6.5 – 8.5 ppm), the spectrum in DHPC exhibited a better dispersion of proton resonance signals in comparison with that of the previous three lipid micelles systems (Figure 3.11). However, the signals in DHPC were still weaker than that in HFIP and SDS. The intensities of the highest signals in H_N regions, normalized by the number of scans, showed that the intensity of the signal in DHPC is over 6 times weaker than that in HFIP and over 4 times weaker than that in SDS.

The 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM obtained in DHPC is shown in Figure 3.8. Similar spectra obtained earlier by Booth *et al.* for the peptide in HFIP and SDS are also shown. The spectra in DHPC, unlike that for the other three micelle systems, showed six distinct peaks corresponding to the six ^{15}N-labeled amino acids although two of them were almost overlapped. However, the peaks in DHPC were weaker by factors of 25 and 2.5, respectively, than what were observed in HFIP and SDS, as shown by the numbers representing the contour levels (CL).

Some reduction of the intensity of signals in DHPC was seen in comparison to that in HFIP and DPC. It is, therefore, possible that the peptide induced aggregation of the micelles but the complexes were still within the NMR accessible sizes. However, SP-BCTERM folded into a specific secondary structure and did not take multiple

conformations in DHPC environment, as indicated by the spectral features of the 1D ^{1}H and 2D HSQC spectrum.

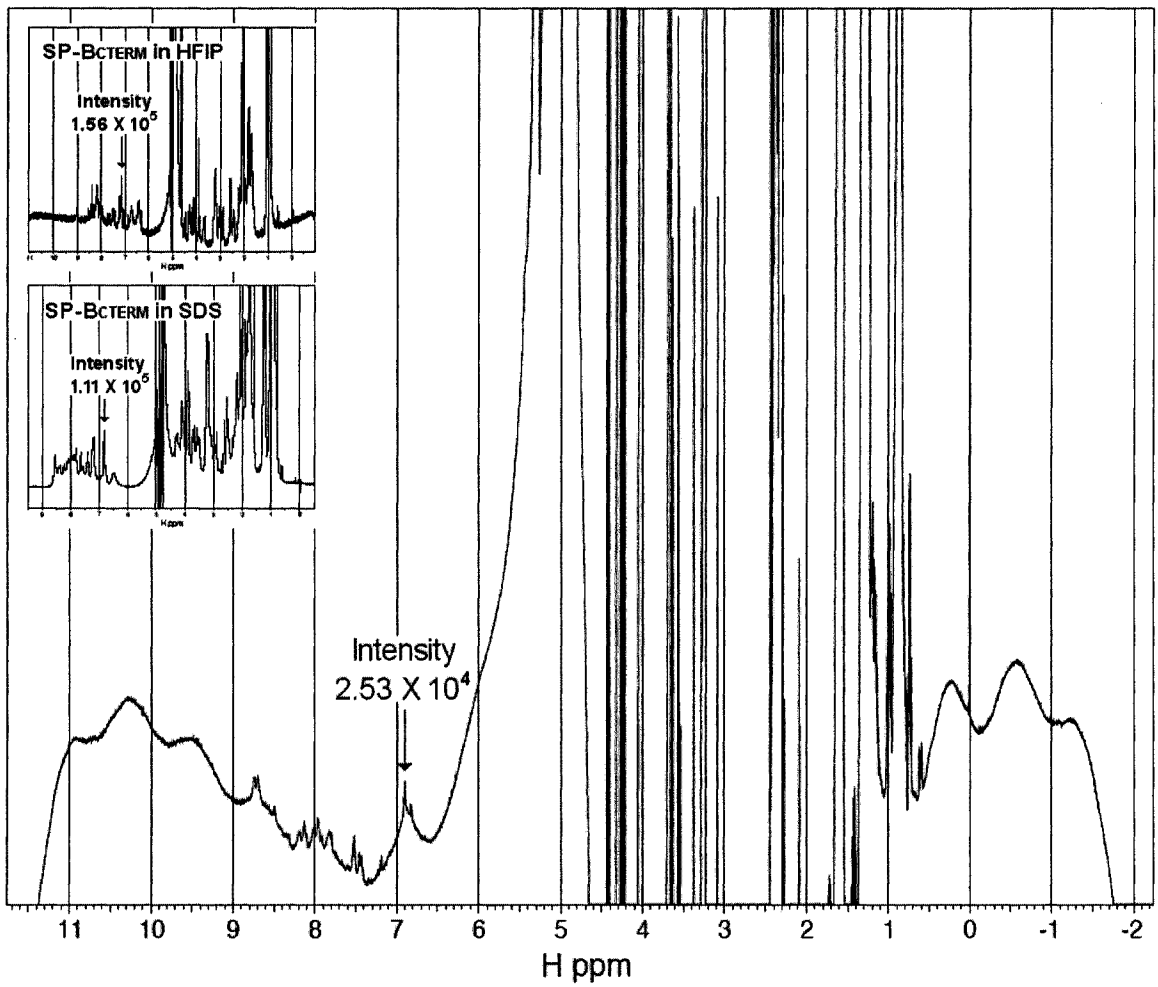


Figure 3.7 : 1D ^{1}H spectrum of SP-BCTERM in DHPC micelles. All the regions including that for the backbone amide protons (~ 6.5 – 8.5 ppm) show reasonable dispersion of chemical shift signals. However, the intensity of the signals was weaker than that observed earlier by Booth *et al.* in HFIP and SDS (shown in the insets).

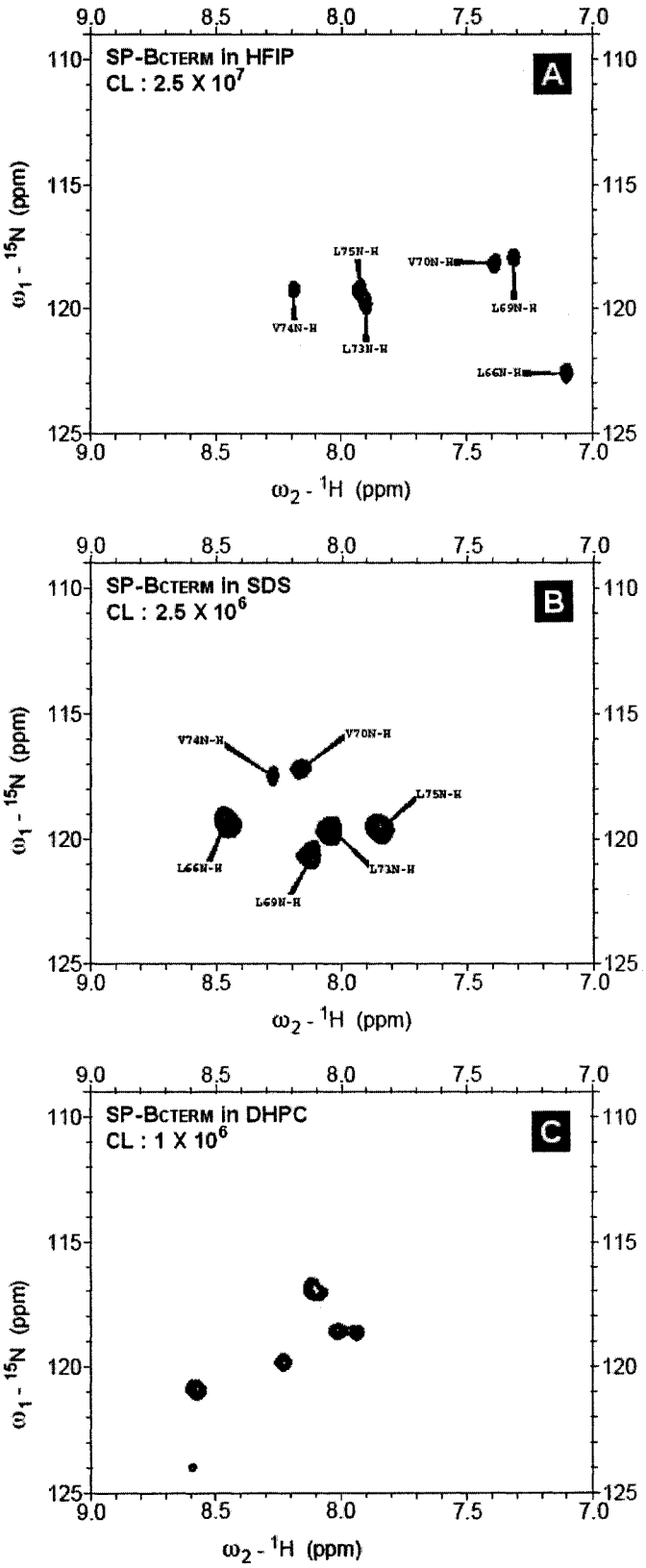


Figure 3.8 : 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM in HFIP (A) and SDS micelles (B) obtained earlier by Booth *et al.* and in DHPC micelles (C). The peptide had six ^{15}N-labeled amino acids. There are six peaks seen in DHPC spectrum with two peaks almost overlapped. The spectra of HFIP and SDS samples also had six peaks. The intensity of the peaks in DHPC was a bit weaker than that in HFIP and SDS as seen from the contour levels (CL).

3.7 CD Spectra of SP-BCTERM in DPC and LPG Micelles

The NMR results for SP-BCTERM in DPC and LPG micelles indicated peptide-induced aggregation of micelles. Since SP-BCTERM had been proposed to form β-sheet aggregates, circular dichroism (CD) was used to assess the nature of the secondary structure of the peptide within the aggregated complexes. Unlike the solution NMR, CD can indicate the secondary structure, even in the aggregated samples.

Circular dichroism is a technique that employs polarised light. The basis of CD is the differential absorption of two kinds of circularly polarised light, right (electric field having a clockwise spiral) and left (anti-clockwise spiral), by a chiral chromophore (such as a protein). A molecule which can not be superimposed on its mirror image is known as chiral and the chemical group that gives color to a molecule is called chromophore. When a sample containing a chiral chromophore is illuminated by a beam comprising the two kinds of circularly polarised light, the amplitude of the electric field vector of one is reduced compared to the other and the transmitted light traces an elliptical path as it moves forward.

The backbone of a protein/peptide shows different optical activity depending upon wheather it is in helical (α) or extended (β) type configuration. Therefore, CD can provide some information about the secondary structure. α-helical peptide bonds yield two negative peaks (minima) at around 208 and 222 nm and a positive peak (maximum) at around 190 nm. The β-sheet conformation is characterised by a minimum at 216 nm and a maximum around 195 nm. However, the spectral extremes of disordered polypeptides are similar to those of β-structure and can thus interfere. For this reason CD is considered to work better for detecting α-type structures than β-type.

Both of the CD spectra obtained for SP-BCTERM in DPC and LPG micelles showed two minima at around 208 and 222 nm and one maximum close to 190 nm (Figures 3.9

and 3.10) which were indicative of perfectly α-helical structure of the peptide. Hence, despite of inducing aggregation, the peptide remained largely α-helical and did not take on β-sheet or unstructured conformations. However, the NMR samples were diluted by about 1000 times to obtain the CD spectra. Since, the secondary structure of a peptide

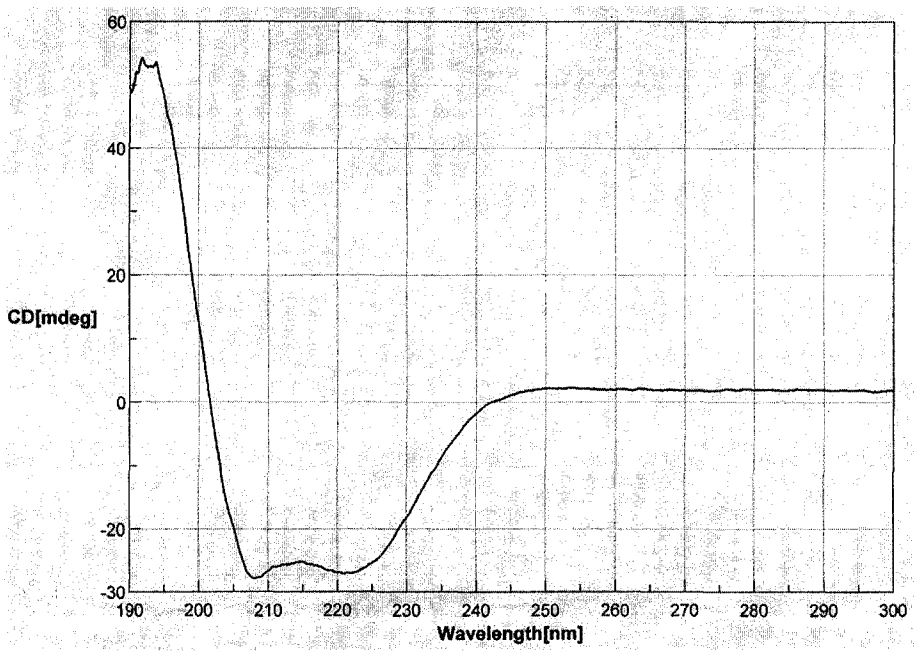


Figure 3.9 : Far-UV CD spectrum of SP-BCTERM in DPC micelles. [Peptide concentration ~ 5 μm, light path 2 nm, number of accumulations 2, temperature 25 ^{0}C and pH 5.5]. The curve shows two minima close to 208 nm and 222 nm and one maximum close to 190 nm. These are characteristics of α-helical peptide.

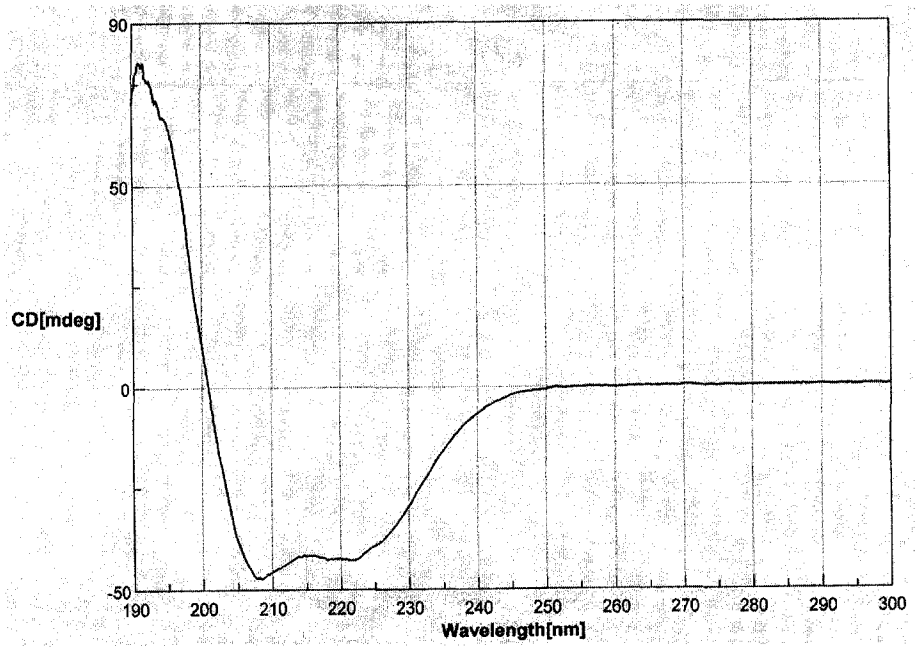


Figure 3.10 : Far-UV CD spectrum of SP-BCTERM in LPG micelles. [Peptide concentration ~ 25 μm, light path 2 nm, number of accumulations 2, temperature 25 ^{0}C and pH 4.5]. The curve shows two minima close to 208 nm and 222 nm and one maximum close to 190 nm. These are characteristics of α-helical peptide.

depends on concentration also, it could not be concluded without doubt that SP-BCTERM had uniform single helical conformations in the original NMR samples in DPC and LPG.

3.8 Dynamic Light Scattering of SP-BCTERM in DPC Micelles

The NMR spectra of SP-BCTERM in DPC and LPG micelles indicated aggregation of lipid molecules induced by the peptide. The CD spectra, on the other hand, indicated that despite aggregation, the structure of the peptide itself remained largely α-helical. At this stage, in order to directly determine the sizes of SP-BCTERM-micelle aggregates in the DPC sample, dynamic light scattering (DLS) experiments were performed. This was done in collaboration with Qi Wang of Agriculture and Agri-Food Canada.

DLS is one of the most popular methods used to determine the size of particles in solution. Shining a monochromatic light beam, such as a laser, onto a solution with spherical particles in Brownian motion, causes a Doppler shift when the light hits the moving particle. This is due to the change in wavelength of the incoming light. The change, however, is related to the size of the particle. It is thus possible to compute the sphere size distribution and give a description of the particle's motion in the medium, by measuring the diffusion coefficient of the particle and using the autocorrelation function.

To determine the size of the peptide-micelles aggregates, DLS experiments were done with the original NMR sample, i.e., 1.5 mM SP-BCTERM, 150 mM DPC and 0.4 mM DSS, mixed in 90% H_2O and 10% D_2O. For the purpose of comparison, data were also collected for a sample with 150 mM DPC but no peptide.

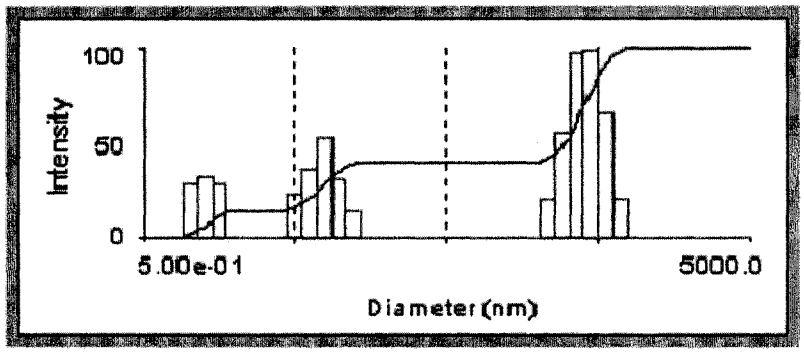
The spectrum of the DPC-only sample showed a distribution of particle sizes (diameters) in three different regions (Figure 3.11); one at ~ 1.25 nm with relative intensity 33 which likely represented dust, one at ~ 7 nm with relative intensity 54 which

represented micelles in NMR spectroscopic observation range and the other at ~ 440 nm with relative intensity 100 that represented aggregated lipids beyond the solution NMR observable range.

The spectrum of DPC + SP-BCTERM sample showed more distribution of particle sizes (Figure 3.12). The first group of particles had diameters at ~ 1.33 nm with relative intensity 12 which again probably represented dust, the second group had diameters of ~ 7 nm with relative intensity 3 that represented lipid-peptide complexes observable by NMR and the third group had distribution at ~ 420 nm with relative intensity 100 that represented aggregated micelles not observable by NMR because of the bigger size. There had been a fourth group of particles as well of diameters around 6000 nm and over with relative intensity 2 that were even further from the NMR observable range.

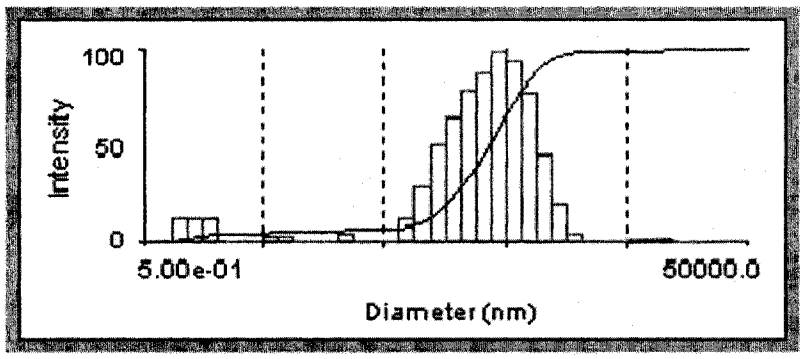
The expected minimum size of a single DPC micelle (i.e., without any aggregation) is 4 nm [Table 2.2] which is consistent with the population of particles observed at 5-7 nm. Particles much larger than this would not be expected to be observable in solution NMR spectra. Addition of SP-BCTERM to DPC had the effect of increasing the ratio of very large aggregates (~ 400 nm) to single micelles (~ 5 nm). In the presence of the peptide, there also appeared a greater range in the aggregate size as seen by ~ 200-400 nm in DPC alone and ~ 75-1800 nm in DPC + SP-BCTERM. In presence of the peptide, a small population of intermediate sized particles, ~ 20 nm, also appeared.

The DLS findings imply that when only DPC was added to water, it formed micelles of different sizes, both suitable for NMR detection as well as bigger than that. However, when SP-BCTERM was mixed with DPC, NMR observable complexes were almost lost and mostly the bigger complexes were present with a much higher distribution range. This had been a clear indication of SP-BCTERM induced aggregation of DPC micelles which was predicted earlier from NMR results.



Diameter (nm)	Intensity	Diameter (nm)	Intensity	Diameter (nm)	Intensity
1.00	30	11.95	15	142.81	0
1.25	33	14.97	0	178.94	0
1.57	30	18.76	0	224.21	22
1.87	0	23.51	0	260.83	56
2.46	0	29.46	0	352.00	99
3.09	0	36.91	0	441.05	100
3.87	0	46.24	0	552.62	67
4.85	24	57.94	0	682.43	22
6.08	37	72.80	0	867.60	0
7.61	54	90.97	0	1087.08	0
9.54	32	113.98	0	1362.09	0

Figure 3.11 : DLS data of only DPC (150 mM DPC in 90% H_2O and 10% D_2O). The particles with diameters ~ 1.25 nm probably represent dusts. The particles with diameters ~ 7.61 nm represent micelles formed by DPC molecules that can be observed in NMR. The particles with diameters ~ 441.05 nm also represent DPC micelles but are too big to be detected in NMR spectroscopy.



Diameter (nm)	Intensity	Diameter (nm)	Intensity	Diameter (nm)	Intensity
1.00	12	23.71	5	562.34	95
1.33	12	31.62	0	749.89	78
1.78	12	42.17	0	1000.00	47
2.37	0	56.23	0	1333.52	21
3.16	0	74.99	13	1778.28	4
4.22	0	100.00	29	2371.37	0
5.62	3	133.35	51	3162.28	0
7.50	3	177.83	65	4216.96	0
10.00	0	237.14	80	5623.41	2
13.34	0	316.23	89	7498.94	2
17.78	0	421.70	100	10000.00	2

Figure 3.12 : DLS data of DPC + SP-B_{CTERM} (150 mM DPC and 1.5 mM SP-B_{CTERM} in 90% H_2O and 10% D_2O). The particles with diameters ~ 1.33 nm possibly represent dusts. The particles with diameters 5.62 and 7.50 nm represent peptide-micelle complexes that are observable by NMR but the intensity is very low. The particles with diameters ~ 421.70 nm represent bigger complexes that can not be detected by NMR although the intensity is very high.

3.9 NMR Spectra of RP-1 in DPC Micelles

The ability to induce micellular aggregation of another amphiphilic and positively charged peptide, RP-1, was investigated to probe the uniqueness of this ability in SP-BCTERM. The DPC micelle system was used for this purpose. RP-1 is modeled on human proteins with antimicrobial activities called kinocidins. It is an 18 residue peptide with 8 positively charged amino acids. Its structure were determined in SDS micelles which showed that the peptide possesses an α-helical conformation. The conformational features of RP-1 were investigated further in DPC micelle system. The results were then compared to that of SP-BCTERM.

The SDS sample of RP-1 consisted of 1.5 mM peptide, 150 mM SDS and 0.4 mM DSS in 90% H_2O and 10% D_2O. The sample had pH of 6.9 and the data were collected at different temperatures from 25 to 45 ^{0}C. The DPC sample of RP-1 was prepared with analogous conditions to the SP-BCTERM sample, i.e., 1.5 mM peptide, 150 mM DPC and 0.4 mM DSS were mixed to 90% H_2O and 10% D_2O. The pH of the sample was 6.8 and the experiments were done at 25 ^{0}C.

The 1D ^{1}H spectrum of RP-1 in DPC is shown in Figure 3.13. Similar spectra obtained for the peptide in SDS at 25 and 40 ^{0}C are shown in the insets. All the 1D ^{1}H spectra showed reasonable dispersion of H_N and H_α signals. The intensity of signals in DPC was also comparable to that in SDS. The 2D ^{15}N-^{1}H HSQC spectrum of RP-1 in DPC is shown in Figure 3.14. Similar spectra obtained for the peptide in SDS at 25 and 40 ^{0}C are also shown. The peptide had five ^{15}N-labeled amino acids. The spectrum in DPC showed the expected five peaks although two of them were almost overlapped. The intensity of the peaks were same for both SDS and DPC samples. These spectral features indicated that RP-1 took on a well-structured and homogeneous conformation in both the lipid systems without causing any significant micelle aggregation.

Comparison of the results for RP-1 and SP-BCTERM in DPC micelles draws clear conclusion that SP-BCTERM caused aggregation of the micelles while RP-1 did not.

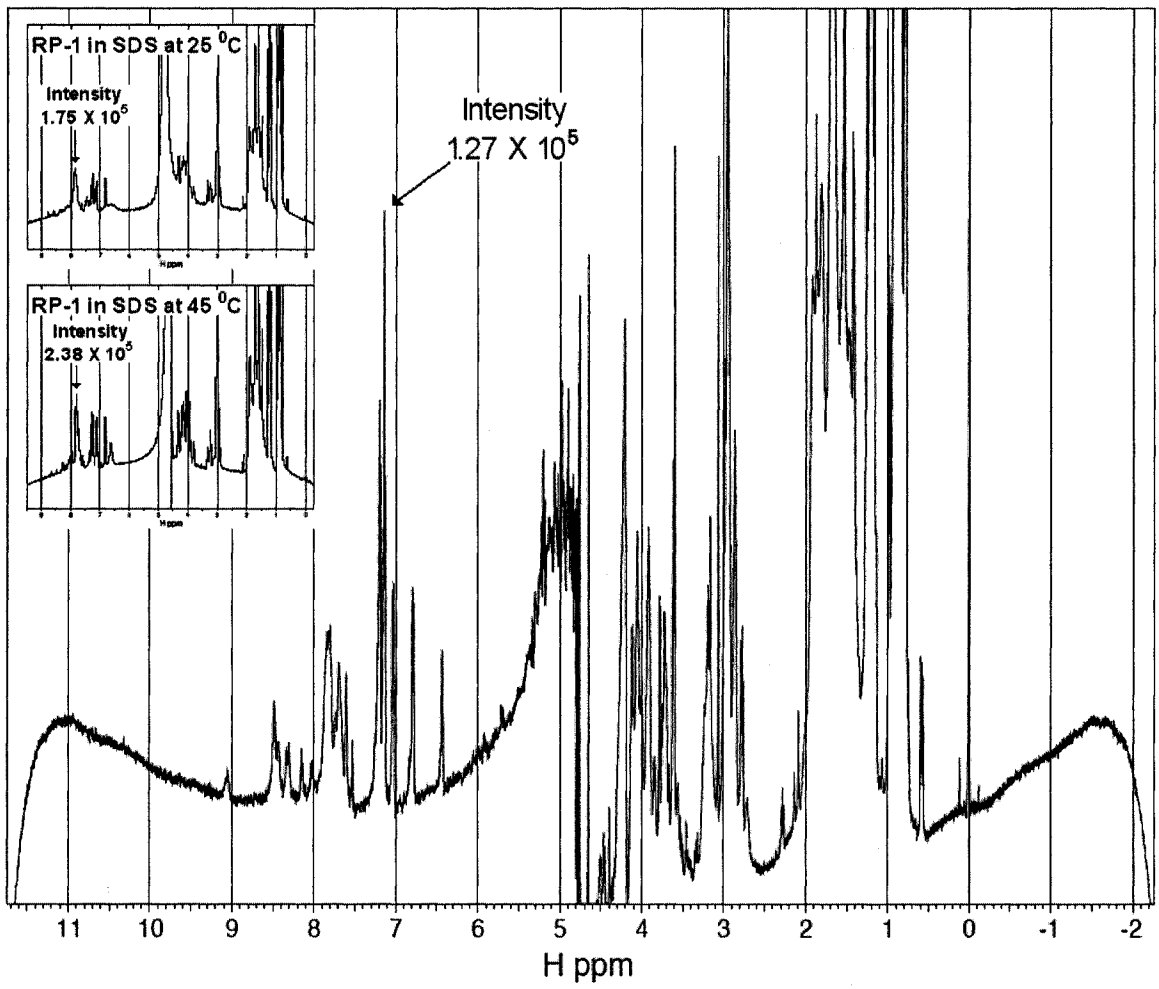


Figure 3.13 : 1D 'H spectrum of RP-1 in DPC micelles. All the regions including that for the backbone H_N and H_α regions show good dispersion of chemical shift signals. The dispersion is comparable to that obtained earlier for RP-1 in SDS micelles at 25 and 40 ^{0}C (shown in the insets). The intensity of the signals are comparable too as shown by the corresponding numbers.

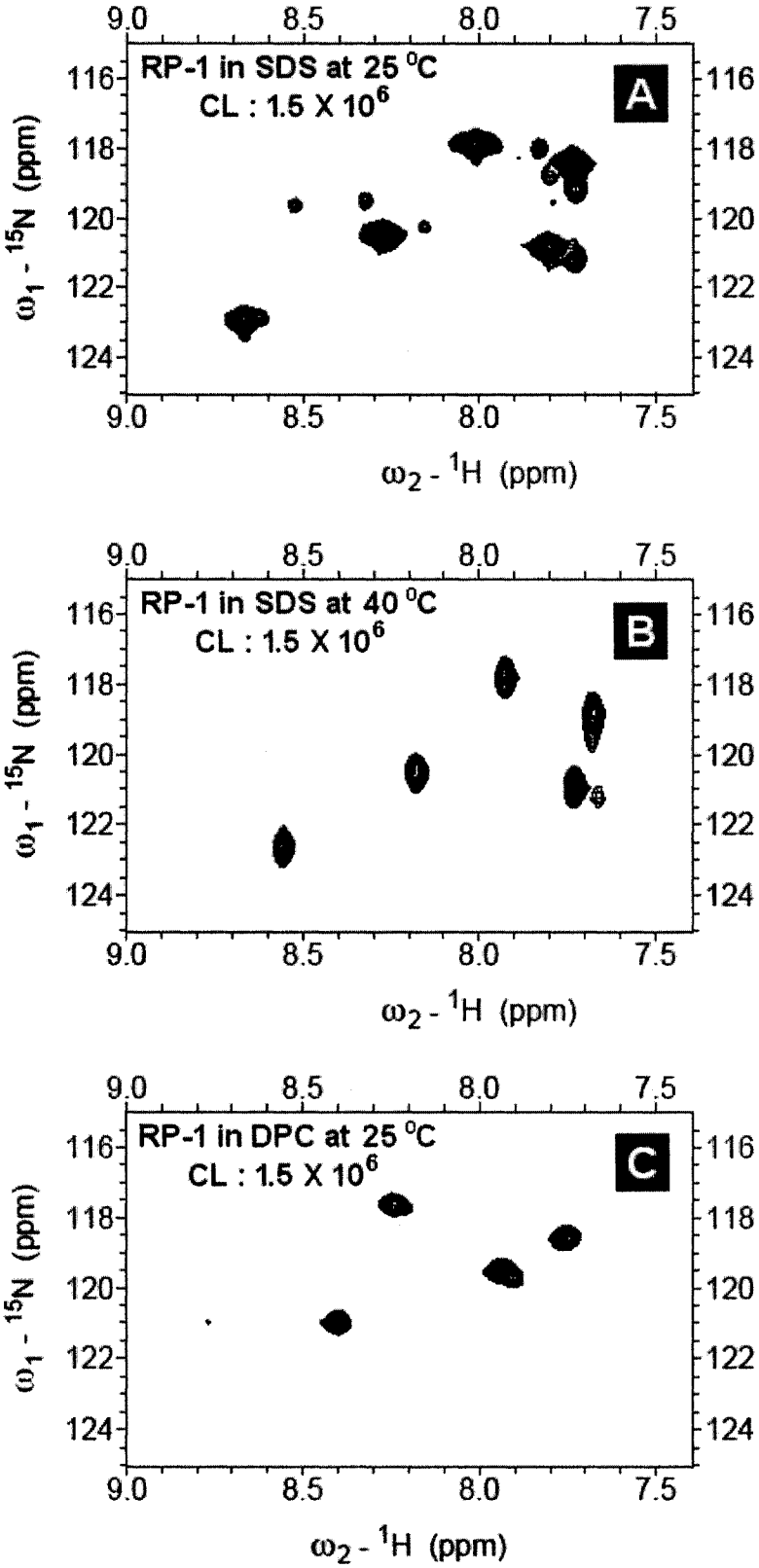


Figure 3.14 : 2D ^{15}N-^{1}H HSQC spectra of RP-1 in SDS micelles at 25 ^{0}C (A) and 40 ^{0}C (B) and in DPC micelles (C). The peptide had five ^{15}N-labeled amino acids. The spectra for SDS show more peaks corresponding to additional minor conformations which is more pronounced at 25 ^{0}C. The spectrum for DHPC show five peaks with two peaks almost overlapped. The intensities of the peaks in all three are same as shown by the contour levels (CL).

3.10 Discussion

Lung surfactant prevents alveolar collapse and eases the work of breathing by reducing the surface tension to extremely low values at air-water interface. SP-B is an essential component of the surfactant. Absence of SP-B is lethal in newborns [28]. Mice with the SP-B gene knocked-out also die early after birth [29]. The basis for SP-B's critical role is still not clear but it appears to originate somehow from SP-B's ability to reorganize lipid structures. Biophysical studies have yielded a long list of *in vitro* activities that may be important in SP-B's *in vivo* roles, including membrane binding, membrane lysis, membrane fusion, promotion of lipid adsorption to air-water surfaces, stabilization of monomolecular surface films and respreading of films from collapsed phases [36]. The studies presented here add additional insight by probing SP-B's ability to modulate the structures of highly curved lipid micelles.

The amino acid distribution along the sequence of SP-B supports formation of several amphiphilic α-helical segments which are hydrophobic on one side and hydrophilic on the other [67]. This suggests a relatively superficial disposition of the protein in lipid bilayers and monolayers [16, 68] with the charged and polar amino acids interacting with the lipid headgroups and the hydrophobic amino acids interacting with the lipid acyl chains. SP-B contains nine positively charged amino acids, with a net charge of +7. This gives the protein a cationic character that suggests a preferential interaction of the protein with anionic phospholipids, such as PG, both in bilayers [69] and monolayers [70]. In particular, selective interaction of SP-B with anionic phospholipids could be important for the formation of nonbilayer intermediates required in bilayer-monolayer transitions [71]. SP-B – PG interactions have also been proposed to promote enrichment of the interfacial films with the most surface-active phospholipids, during compression [16, 72]. SP-B also plays a critical role in the formation of SP-A – DPPC tubular myelin structures that are important in transporting phospholipids from the aqueous subphase to the surface monolayer [73, 74]. All of these observations point to

SP-B's essential contributions in maintaining the molecular structure and continuity of the surfactant monolayer during breathing as well as facilitating the incorporation of phospholipids from aqueous subphase to air-water interface [42]. Large-scale rearrangement of phospholipids and stabilization of the structures by SP-B through binding underlie all these functions. Models have been proposed with several types of SP-B – lipid combined structures, with SP-B attached to monolayer, bilayer, in between monolayer and bilayer, within bilayer, inside curve, outside curve, in bilayer-break, etc. [16, 18]. Figure 3.15 shows the interactions of SP-B and phospholipids.

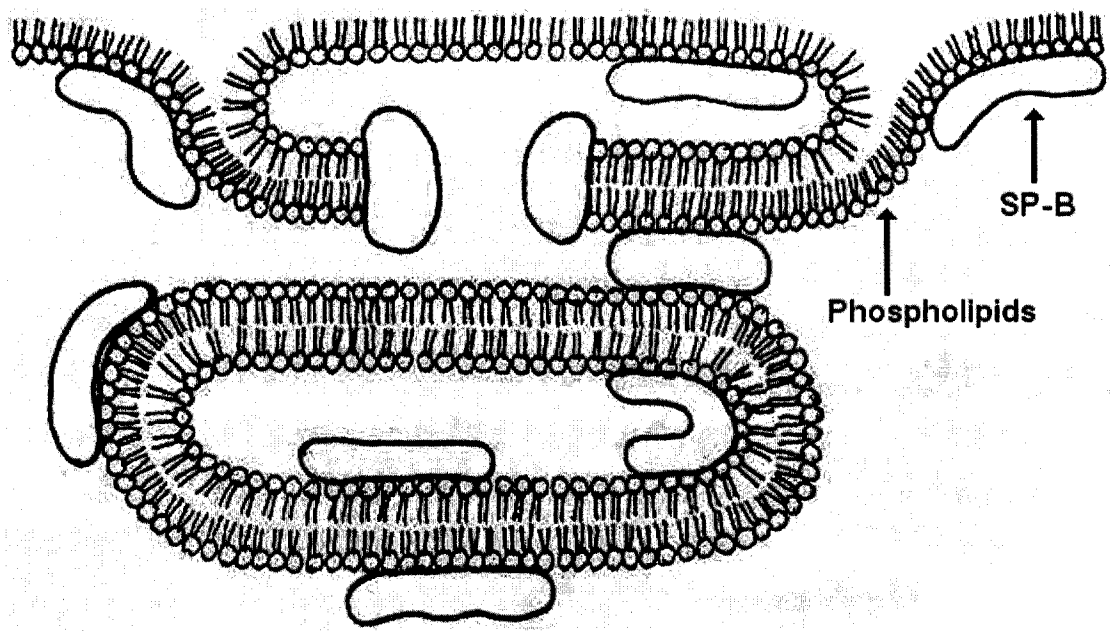


Figure 3.15 : A model showing SP-B – phospholipids interactions in surfactant layers.

As an active peptide segment of the full-length SP-B, the SP-BCTERM was expected to take on α-helical conformation and to exhibit ability of binding lipids analogous to the physiological phospholipids. The results from NMR studies and additional information obtained from CD and DLS experiments supported these hypotheses. SP-BCTERM clearly demonstrated its ability to facilitate large-scale lipid rearrangement through binding lipid micelles.

The 1D ^{1}H spectra of SP-BCTERM in all the four lipid micelle systems, composed from DPC, LPG, LPA and DHPC, showed significant reduction of resonance signal intensity in comparison with that in HFIP and SDS, studied earlier by Booth *et al.* [42]. More reduction of intensity was noticed for the micelle systems formed by LPA and LPG. Moreover, the chemical shifts of the amide proton signals, which are usually the best resolved resonances in protein NMR, showed far less dispersion than expected with some exceptions for that in DHPC. In comparison with previously obtained spectra in HFIP and SDS, the 2D ^{15}N-^{1}H HSQC spectra of SP-BCTERM also showed much weaker intensity of peaks in DPC and LPG. The peaks in LPA were even totally lost in the noise. Also, in DPC and LPG, the peptide exhibited multiple conformations indicated by more peaks than expected. In DHPC, the spectra showed six expected peaks but the signals were still weaker. These observations lead to the following conclusions regarding the conformational features of SP-BCTERM in DPC, LPG, LPA and DHPC and its ability to bind the micelles formed from these lipids.

Firstly, the reduction of peak intensity in both 1D and 2D spectra was a direct consequence of bigger peptide-lipid complexes caused by peptide induced aggregation of lipid micelles. In solution NMR, increasing the complex size leads to slower tumbling and hence more efficient relaxation. This, in turn, leads to broader and less-intense peaks in the NMR spectra. Comparison of the spectral features in four lipids showed that the intensity reduction was more prominent for LPA and LPG, the lipids containing anionic headgroups, than for DPC and DHPC, the other two lipids containing zwitterionic headgroups. Figure 3.16 shows SP-BCTERM and micelles drawn with appropriate relative size scaling and presents a model of how the peptide may initiate aggregation of micelles. The model is consistent with the charge distribution on SP-BCTERM shown in Figure 3.17. The peptide is capable of interacting with two or more lipid headgroups at each of its positively charged patches. These patches are distributed over the surface of SP-BCTERM in such a way that electrostatic interactions with lipids may occur on two nearly opposite faces of the peptide as well as at the positively charged end. While SP-BCTERM was most effective in inducing aggregation of anionic phospholipids, it also induced significant

aggregation of the zwitterionic phospholipids. Additionally, earlier work with the anionic detergent SDS showed no signs of SP-Bcterm induced aggregation of SDS micelles. Clearly, the interaction is more complex than just a simple charge-charge interaction.

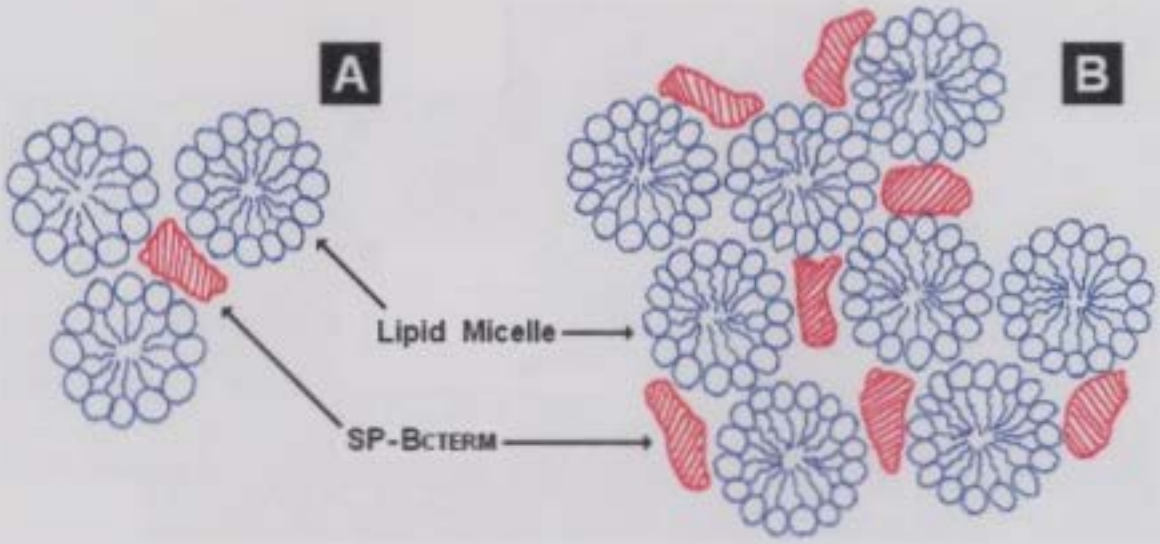


Figure 3.16 : Possible models for aggregation of lipid micelles induced by a single SP-Bcterm (A) or by a number of SP-Bcterm molecules (B).

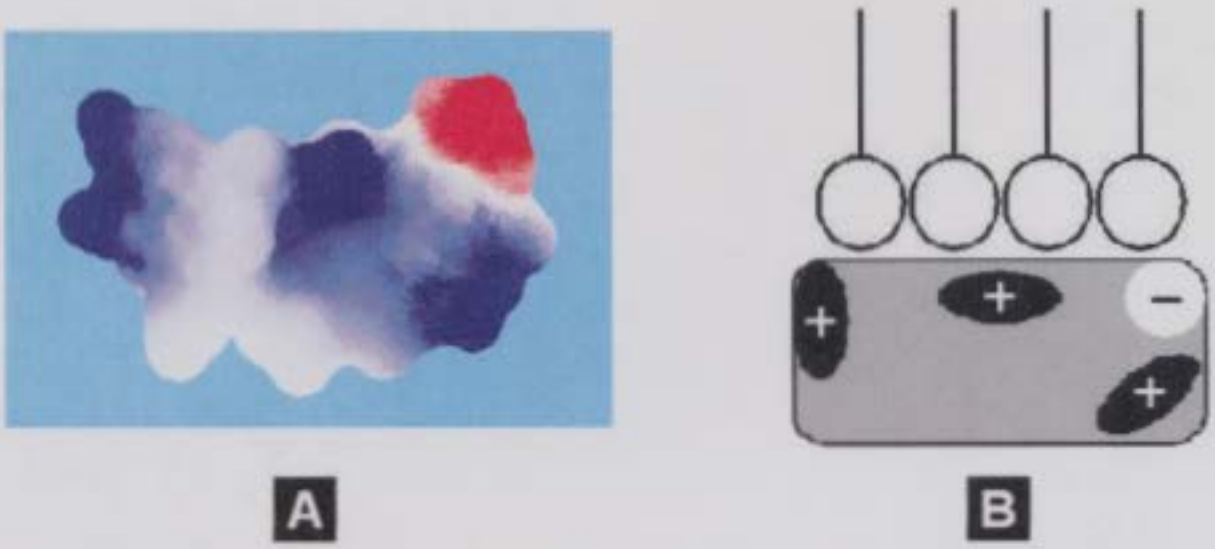


Figure 3.17 : Interaction of SP-Bcterm with phospholipids. (A) Surface charge plot of SP-Bcterm obtained from the peptide's structure in SDS [40]. Color code : blue – positively charged surface and red – negatively charged surface. (B) Simplified model showing that four phospholipid molecules can interact from one side of the peptide at once.

Secondly, the lack of chemical shift dispersion in 1D spectra and extra peaks in 2D spectra indicated that the SP-BCTERM did not fold properly into a single 3D structure in DPC, LPG and LPA. The spectra indicated that the peptide took on multiple conformations possibly as required to bind the lipid micelles together from different positions (Figure 3.16B). This view is not contradictory with the observations from CD spectra in DPC and LPG, which demonstrated the structures of the peptide as mostly α-helical. The CD is only able to indicate if a structure is helical, not if there are multiple helical structures present.

Thirdly, the NMR spectra indicated that SP-BCTERM induced aggregation of micelles with a variety of diameters. The diameters of the micelles are 4 nm for DPC, 5.6 nm for LPG and 3.3 nm for DHPC (Table 2.2). As the number of lipids that form each micelle is known (except for LPA), the number of micelles per peptide can be calculated. It is 2.7 for DPC, 1.2 for LPG and 3.75 for DHPC. SP-BCTERM is thus able to induce aggregation of a variety of micelle/peptide ratios (Figure 3.16).

To probe the uniqueness of the ability of SP-BCTERM in inducing aggregation of micelles, another cationic amphiphilic peptide with known helical structure, RP-1, was studied in DPC. The 1D ^{1}H spectrum obtained for this peptide showed signals with intensity levels and chemical shift dispersion expected for a single helix conformation bound to a non-aggregated single micelle. The 2D ^{15}N-^{1}H HSQC spectrum showed the expected number of peaks corresponding to the ^{15}N-labeled amino acids also with good intensity. Therefore, RP-1 took on a well-defined homogeneous conformation in DPC and did not produce any bigger peptide-micelle complexes.

Hence, the same lipid system that underwent aggregation in the presence of SP-BCTERM, did not aggregate in the presence of RP-1. This further emphasized the uniqueness of SP-BCTERM's ability to induce aggregation of certain phospholipid micelles.

These studies show that a single helix from SP-B can induce aggregation of anionic, and to a lesser extent, zwitterionic phospholipids. This observation constrains the models of interactions of full-length SP-B with lipids presented in Figure 3.15. Since the single helix of SP-B$_{CTERM}$ retains about half the activity of full-length SP-B (Figure 1.4), interactions that require SP-B to cover a large distance (e.g., to traverse the bilayer) cannot be of utmost importance. These studies show that a single helix of SP-B is capable of inducing strong interactions between regions of phospholipids with "outside" curvature – most similar to the interactions shown at the far left of Figure 3.15. The high degree of biological functionality of SP-B$_{CTERM}$ implies that these types of interactions are of high biological consequence when compared to myriad of lipid-binding activities that have been proposed for SP-B.

Chapter 4

Structure of Mini-B

4.1 Overview of Structural Studies of Mini-B

SP-B is essential for normal lung surfactant function but the molecular basis for its activity is not yet understood. This is because the 3D structure of this protein has not been determined so far. Theoretical models predicted that the N- and C-terminal domains of SP-B fold as charged amphiphilic (hydrophobic in one side and hydrophilic at the other) α-helices connected by disulfide bonds and suggested that these adjacent helices participate in critical surfactant activity [75]. The peptide fragment of SP-B containing these two terminal regions was termed as Mini-B. Mini-B in model surfactant lipid mixtures exhibited marked *in vitro* activity, similar to that observed with native SP-B. *In vivo*, too, Mini-B showed oxygenation and dynamic compliance that compared favorably with that of full-length SP-B. These studies concluded that the combination of structure and charged patches for the amphiphilic α-helical N- and C-terminal domains of SP-B (i.e., Mini-B) are key to its function [76]. However, to understand the mechanisms of function, a detailed 3D structure of Mini-B is essential. Hence, using solution NMR, the high-resolution structure of Mini-B was determined in structure-promoting organic solvent HFIP as the main part of present research.

4.2 Materials and Methods

Mini-B is a synthetic 34 residue peptide consisting of a fusion of residues 8-25 and 63-78 of human SP-B. Its amino acid sequence is shown below.

Cys-Trp-Leu-Cys-Arg-Ala-Leu-Ile-Lys-Arg-Ile-Gln-Ala-Met-Ile-Pro-Lys-Gly-Gly-Arg-
1 5 10 15 20
Met-Leu-Pro-Gln-Leu-Val-Cys-Arg-Leu-Val-Leu-Arg-Cys-Ser
21 25 30 34

Mini-B was synthesized chemically by Alan Waring (UCLA) following the same procedure as for SP-BCTERM (Section 3.2) [42]. Nine of the residues, namely leucines 3, 7, 22, 25, 29 and 31, alanines 6 and 13 and glycine 18 were ^{15}N labeled, i.e., the natural isotope ^{14}N at the backbone of these residues were replaced by spin 1/2 ^{15}N nuclei. The peptide contains four cysteines but it was synthesized without any disulfide bonds between any pairs of cysteines.

The NMR sample of Mini-B in fluorinated organic solvent was prepared by dissolving 1.5 mM peptide in 40% HFIP, 50% H_2O and 10% D_2O with 0.4 mM DSS. The peptide was weighed out in dry form assuming it was 100% pure. HFIP (> 99% deuterated), DSS and D_2O were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). DSS was used to obtain the proton reference signal at 0 ppm and D_2O was used to provide the lock nucleus (deuterium) to the field-frequency lock system of the spectrometer. The volume of the sample was 500 μL. To avoid any undesired folding of the peptide resulting from disulfide bonds between cysteines (both intra- and inter-chain), the pH of the sample was set at 5. The pH was measured by an accumet Electrode bought from Fisher Scientific.

Initial 1D ^{1}H and 2D ^{15}N-^{1}H HSQC experiments were done at three different temperatures (15, 25 and 35 ^{0}C) to evaluate the temperature dependence of the peptide conformation. It was observed that the peptide exhibited improved conformational homogeneity at 15 ^{0}C. Therefore, all subsequent complex multidimensional NMR experiments were conducted at 15 ^{0}C. The experiments included two 2D TOCSY (with mixing times of 45 and 80 ms), three 2D NOESY (with mixing times of 100, 200 and 300 ms) and one 3D ^{15}N-edited NOESY (with mixing time of 200 ms). Lastly, two additional 2D TOCSY and NOESY experiments (with mixing times of 80 and 200 ms, respectively) were performed with a freshly prepared sample of Mini-B in 40% HFIP and 60% D_2O by lyophilizing the peptide from the previous sample.

The NMR data acquisition began on Bruker Avance 500 MHz spectrometer of the Centre for Chemical Analysis, Research and Training (C-CART) at Memorial University of Newfoundland (MUN). However, during the initial acquirements of 1D ^{1}H and 2D ^{15}N-^{1}H HSQC data, the RF probe of the spectrometer failed. While the probe was being repaired, the 1D ^{1}H and 2D HSQC, TOCSY and NOESY spectra were collected using a Bruker Avance 700 MHz spectrometer of Health Canada, Ottawa. The sample was returned to MUN for acquisition of the 3D NOESY and the last 2D TOCSY and NOESY (with lyophilized Mini-B in HFIP/D_2O) spectra. Both of the spectrometers used were equipped with z-gradients and triple-resonance TXI probes. The 1D ^{1}H used presaturation, HSQC used flip-back and TOCSY-NOESY used water-gate water suppression techniques.

The spectra were processed using the software NMRPipe 2.2 [66]. Frequency assignments for all the observable spin 1/2 nuclei of Mini-B were made using Sparky 3.110 [77]. Structures of the peptide were calculated using the simulated annealing algorithm within CNS 1.1 [78]. After each round of calculation, structures were viewed by MOLMOL 2k.2 [79]. At the end of the final round of calculation, an ensemble of the 15 lowest energy structures of Mini-B was retained for analysis. This ensemble of structures has also been deposited in the Protein Data Bank (PDB) [PDB ID 2A2H].

4.3 ^{1}H and HSQC Spectra

Three 1D ^{1}H and 2D ^{15}N-^{1}H HSQC spectra were collected at 15, 25 and 35 ^{0}C to optimize the temperature conditions for further spectral acquisition. 15 ^{0}C was chosen as the optimum temperature as it showed the best H_N dispersion and the least conformational heterogeneity. Later on, 1D ^{1}H and 2D ^{15}N-^{1}H HSQC experiments were always performed routinely before and after any complex multidimensional NMR experiments to monitor for any changes in the sample. Spectra obtained thereby always showed identical response of signals indicating that the peptide conformation remained stable over time.

Figure 4.1 shows a 1D ^{1}H spectrum of Mini-B at 15 ^{0}C. Well-dispersed signals with good intensity are seen which indicate that the protein was folded into a well-defined 3D structure under these conditions. Figure 4.2 shows a 2D ^{15}N-^{1}H HSQC spectrum of the peptide at 15 ^{0}C. The peptide had a total of nine ^{15}N labeled residues and the expected nine peaks corresponding to the covalently bonded ^{15}N – ^{1}H pairs of nuclei are seen here. Absence of any extra peaks indicates that Mini-B took on a single conformation in HFIP. However, the peaks for A13 and L25 are almost overlapped. Note that the ^{15}N frequency of G18 is 107 ppm. Here it appears at 126 ppm due to spectral aliasing since, to optimize the resolution, the spectrum was acquired with a reduced spectral width of 114 to 128 ppm.

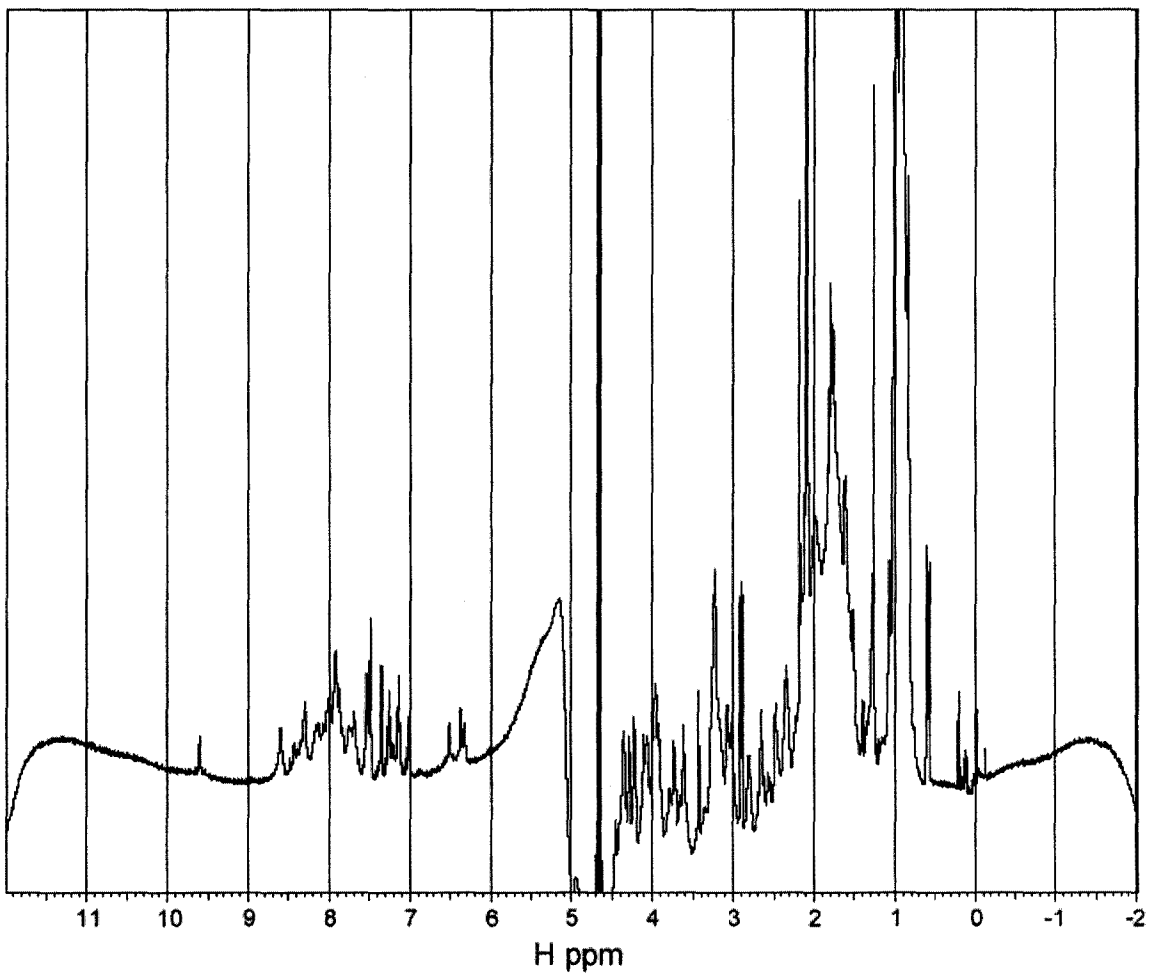


Figure 4.1 : 1D ^{1}H spectrum of Mini-B in HFIP at 15 ^{0}C. The signals correspond to the resonance frequencies (chemical shifts) of the protons in the peptide. The amide protons show good dispersion of signals at ~ 7.2 – 8.6 ppm.

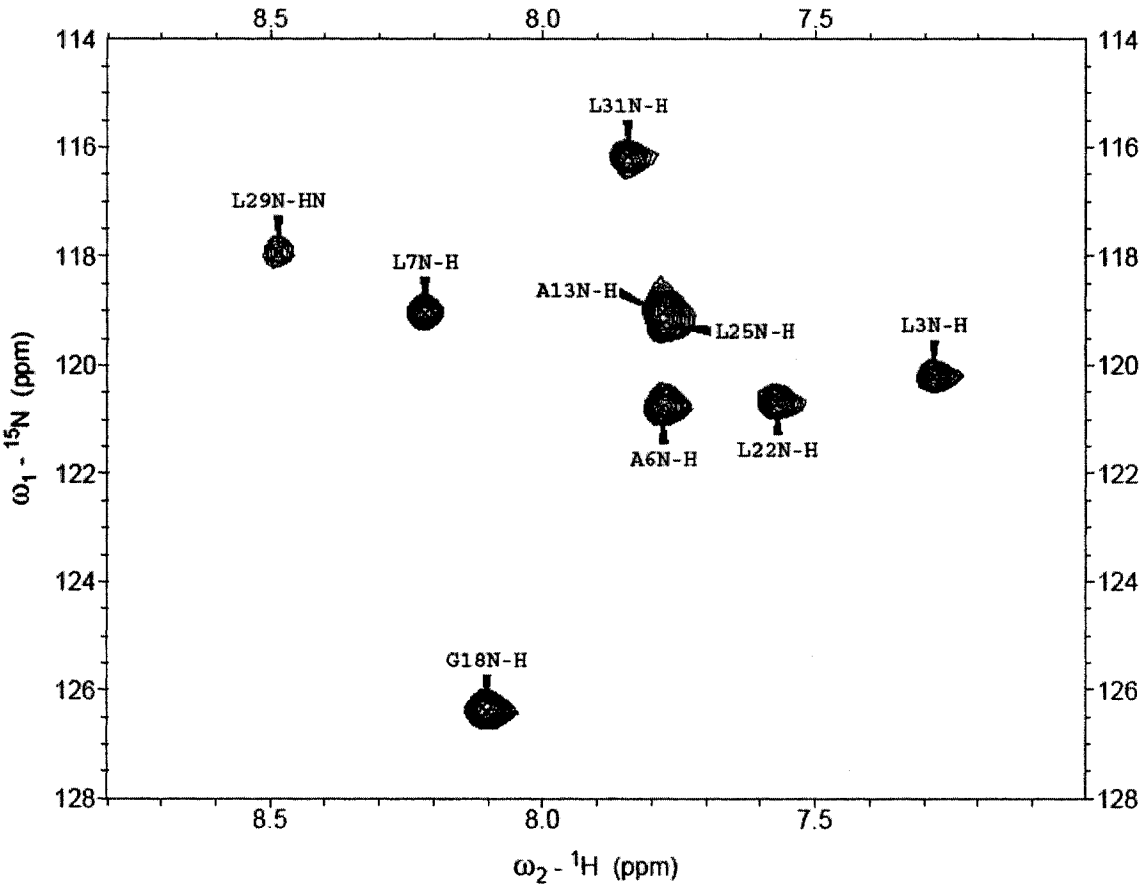


Figure 4.2 : 2D ^{15}N-^{1}H HSQC spectrum of Mini-B in HFIP at 15 ^{0}C. The peptide has nine ^{15}N-labeled amino acids. Nine peaks relating the resonance frequencies of covalently bonded ^{15}N – ^{1}H pairs of nuclei in those amino acides are seen.

4.4 TOCSY and NOESY Spectra

Two 2D TOCSY experiments were performed, with mixing times of 45 and 80 ms, respectively. The spectrum obtained for 45 ms mixing time is shown in Figure 4.3. The off-diagonal cross peaks in the spectrum resulted from the through-bond magnetization transfer between scalar-coupled proton pairs (up to 3 bonds).

Three 2D NOESY experiments were performed with mixing times 100 ms, 200 ms and 300 ms, respectively. The spectrum acquired for 200 ms mixing time is shown in Figure 4.4. The off-diagonal cross-peaks seen here resulted from the through-space magnetization transfer between dipolar-coupled proton-pairs close in space (up to 5 Å).

The higher the mixing time, the greater the magnetization transfer between the proton pairs and the more intense the observed signals in NOESY. However, longer mixing times may allow spin diffusion, when magnetization transfer occurs through a third nucleus. This creates a problem for quantification of the NOEs for structure calculation. Therefore, though all 3 NOESY spectra were used for resonance assignment, only the 200 ms spectrum was used to generate restraints for structure calculation.

The 3D ^{15}N-edited NOESY experiment was done with the mixing time of 200 ms. This spectrum only exhibited NOEs to protons attached to ^{15}N-labeled residues and separated the 2D ^{1}H-^{1}H spectrum by a third dimension (the ^{15}N frequency). The resulting spectral simplification resolved some of the overlapped peaks of the 2D NOESY.

After all other experiments were done, the sample was lyophilized and the peptide powder was dissolved in 40% HFIP and 60% D_2O. 2D TOCSY and 2D NOESY experiments were performed with this freshly prepared new sample with mixing times of 80 and 200 ms, respectively. These two experiments were conducted to check out the presence of any strong H-bonds (i.e., – N – H --- O = C –, between residues) in the

peptide and assign the ring protons of tryptophan. The spectrum of the 2D NOESY is shown in Figure 4.5. Because the amide protons were labile, they were rapidly exchanged with the deuterons and the NOEs that correlated them were lost immediately. Absence of the NOEs implied that Mini-B did not have any strong H-bonds between residues. The quick exchange of amide protons with deuterons was confirmed by a 1D ^{1}H experiment done within 2 minutes of adding D_2O followed by a 2D ^{15}N-^{1}H HSQC experiment and observing no peaks in the spectra corresponding to the amide protons.

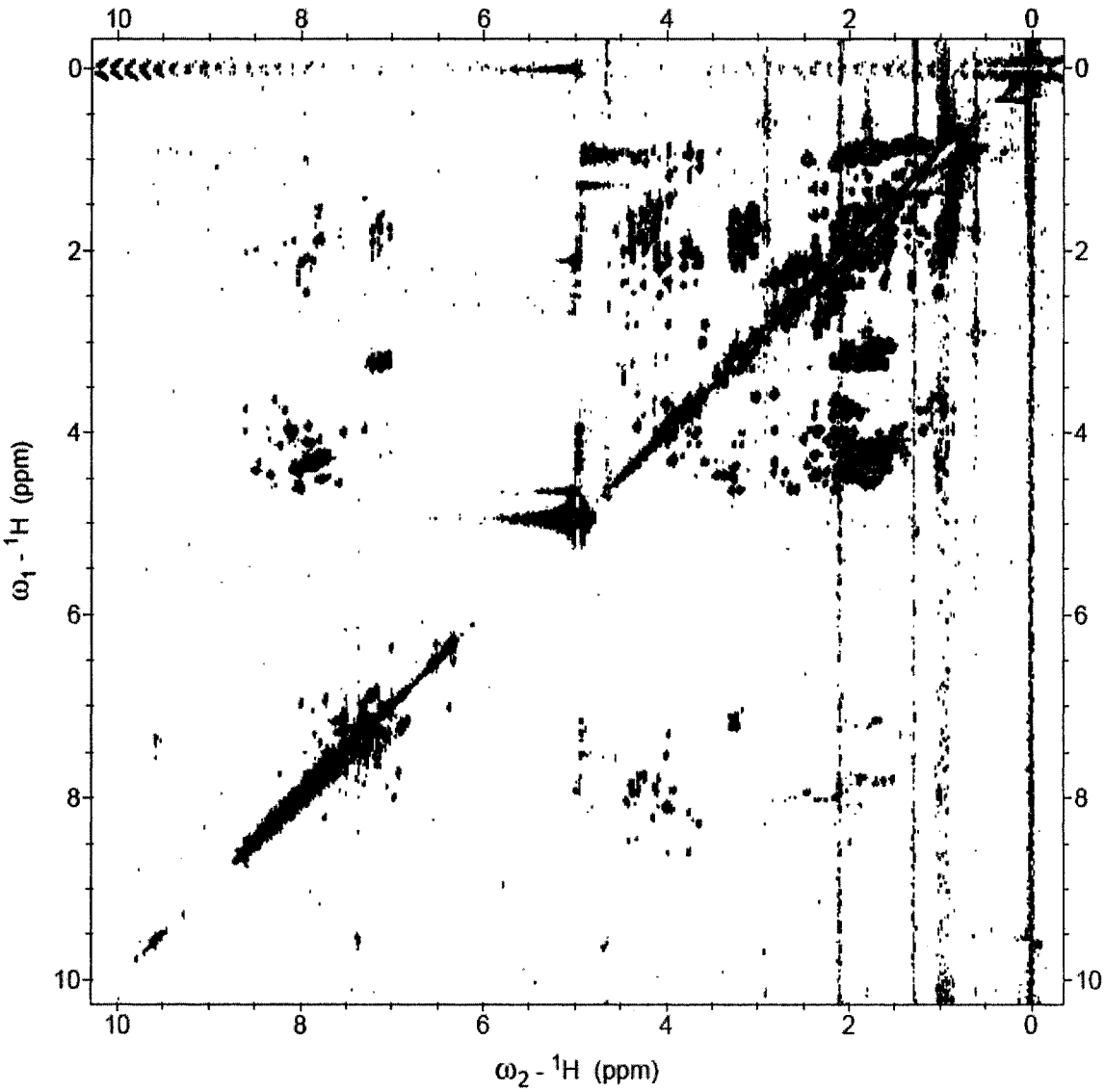


Figure 4.3 : 2D TOCSY spectrum of Mini-B in HFIP with mixing time of 45 ms. The diagonal peaks correspond to the complete 1D ^{1}H spectrum. The off-diagonal cross-peaks represent the correlation resulted from the magnetization transfer between covalently bonded H_N and H_α and then around the side chain in each of the amino acids.

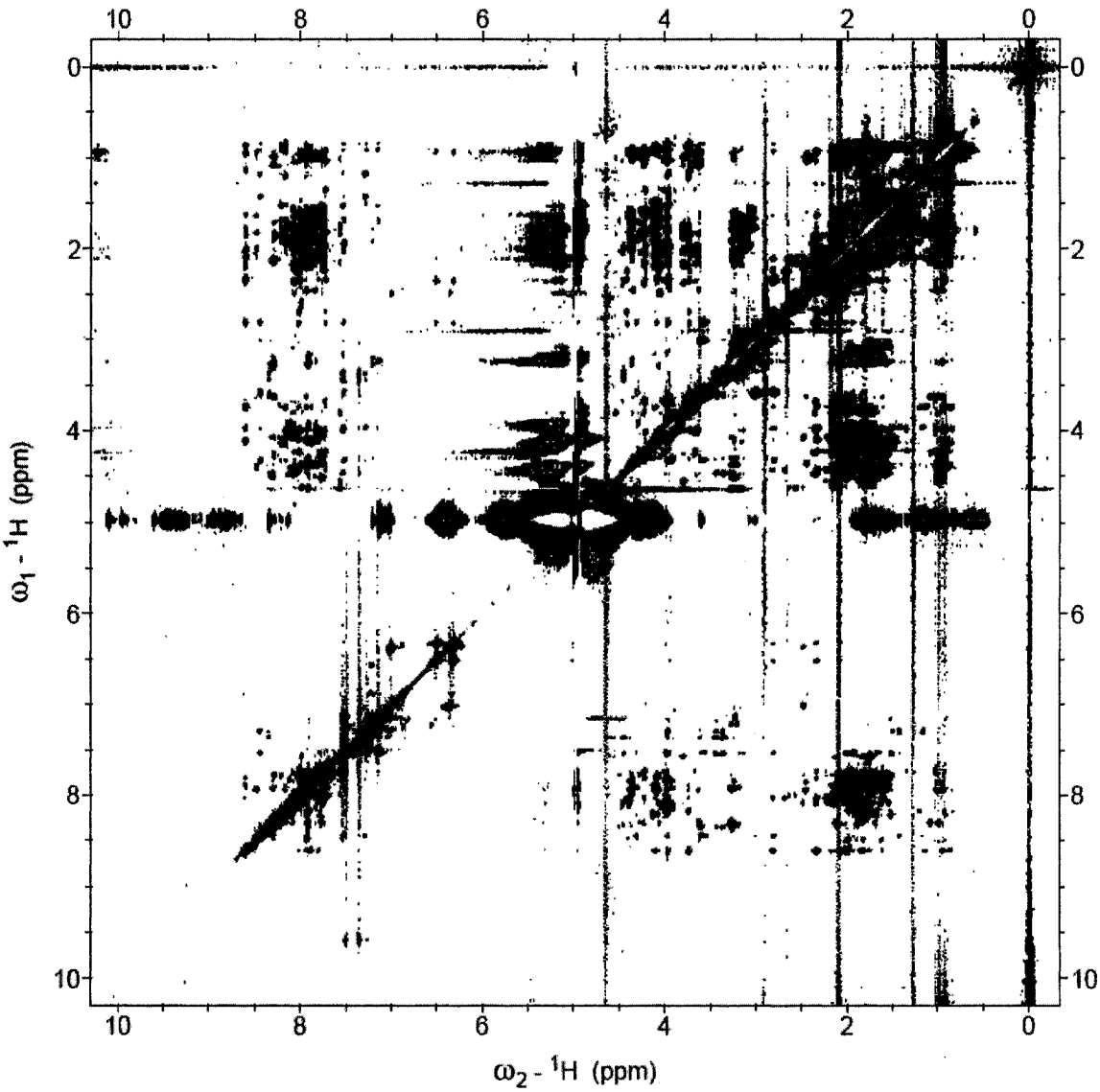


Figure 4.4 : 2D NOESY spectrum of Mini-B in HFIP with mixing time of 200 ms. The diagonal peaks correspond to the complete 1D ^{1}H spectrum. The off-diagonal cross-peaks represent the correlation resulted from the magnetization transfer between proton-pairs that are close in space (less than 5 Å).

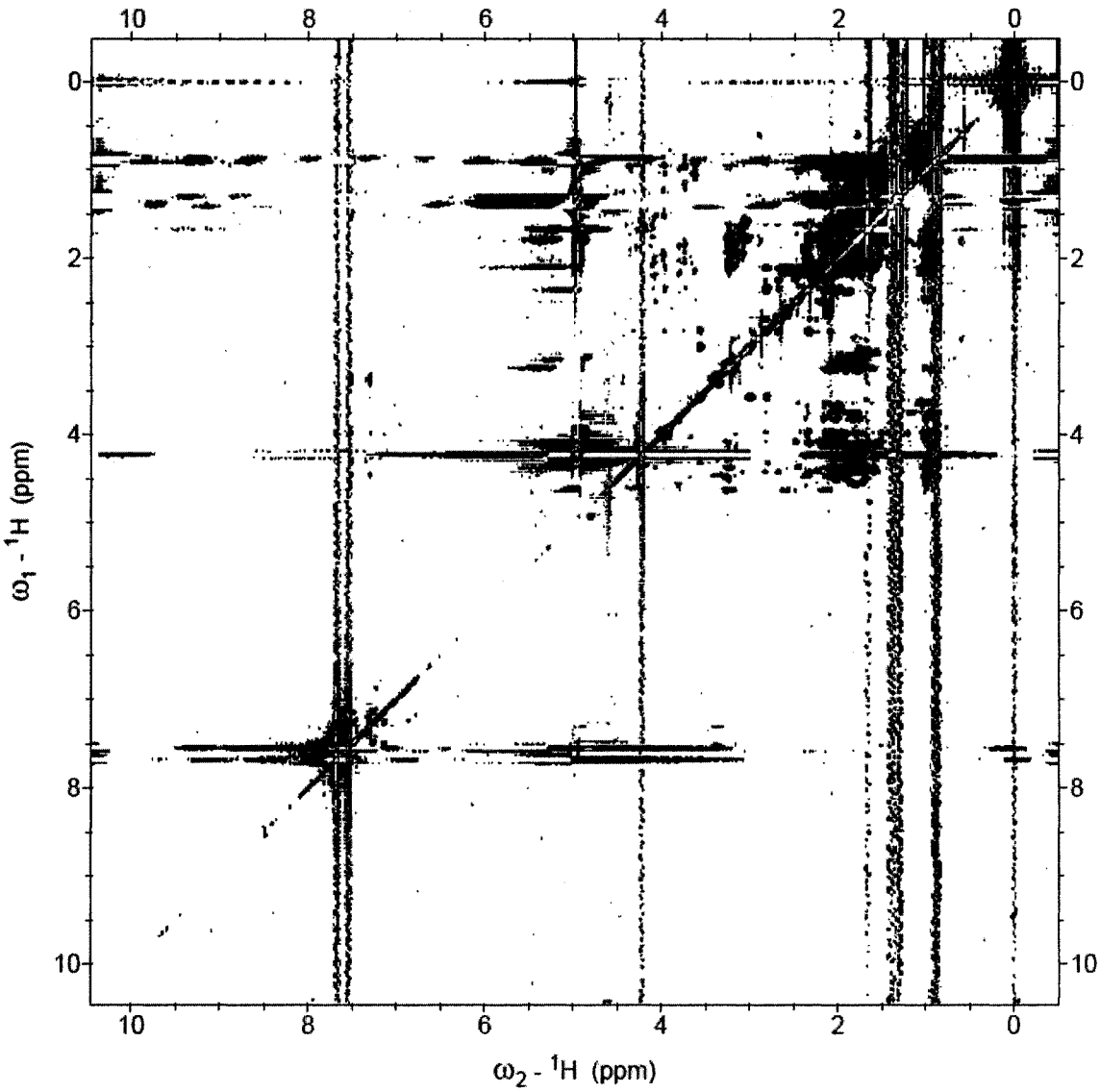


Figure 4.5 : 2D NOESY spectrum of Mini-B in 40% HFIP and 60% D_2O. The amide protons were rapidly exchanged with the deuterons and the peaks correlating them are lost.

4.5 Resonance Assignment

For NMR structural studies, each nuclear magnetic resonance needs to be associated with a specific nucleus in the molecule under investigation [45]. Resonance assignments also have to be sequence-specific, i.e., each resonance must be assigned to a spin in a particular amino acid in the protein sequence. For Mini-B, three types of information were collected from the NMR experiments for the purpose of spectral assignments : through-bond interactions (via scalar coupling, from TOCSY), through-space interactions (via dipolar coupling, from NOESY) and chemical environment (via the chemical shift). Initially, the ^1H resonances were categorized on the basis of their chemical shifts.

The backbone amide protons (^1H_N) typically resonate between 10.0 and 7.0 ppm, the backbone α-protons ($^1H_\alpha$) resonate between 6.0 and 3.5 ppm, the aliphatic side-chain protons resonate between 3.5 and 1.0 ppm and the methyl protons resonate at chemical shifts less than 1.5 ppm [45]. Assignment of individual NMR resonances to individual nuclei in Mini-B started with the backbone amide protons which constituted the best resolved set of resonances. The alpha and side-chain resonances were correlated to the backbone amide protons in TOCSY by the observation of direct or relayed cross-peaks. The H_α-H_N region of the TOCSY spectrum of 45 ms mixing time is shown in Figure 4.6 with the full assignment of the peaks as determined by the procedure discussed below. The $1H_\alpha$-H_N cross-peak is missing as usually the N-terminal amide proton is not observed in TOCSY. Overlapping of peaks is seen in two cases, one for K17 and R28 and the other for R20 and L31.

In NOESY spectra, spin systems were assigned first to particular residues by using through-space dipolar interactions to sequentially connect the spin systems identified from through-bond scalar correlations. Next, the connections between inter-residue protons with sufficient proximity were identified. The majority of short

interproton distances between ^1H_N, $^1H_\alpha$ and $^1H_\beta$ of an α-helical protein/peptide are between residues adjacent in the primary sequence [80]. Thus, identification of intense NOEs from ^1H_N, $^1H_\alpha$ and/or $^1H_\beta$ of one spin system to ^1H_N of a second spin system indicated that the two spin systems are adjacent in the primary sequence, with the first one nearer to the N-terminus. Identification of a series of sequential NOEs placed several spin systems in order and as more spin systems were connected, the sequence eventually matched the primary amino acid sequence of the peptide. The observation of NOEs was not limited only to sequential interactions, since a substantial number of NOEs occurred between nonsequential residues as a result of the secondary structure of the peptide. Those were identified as medium-range NOEs and resulted from proton-pairs from up to five residues apart in the primary chain.

Figures 4.7 and 4.8 show the H_α-H_N and H_N-H_N regions of the NOESY spectrum acquired with mixing time of 200 ms with the full assignment of the peaks. The other portions of the spectrum are shown in Figure 4.9. These spectra showed ambiguity of resonance assignments in many cases, in addition to overlapping of peaks. The ambiguity in assignment occurred when a particular NOE exhibited correlation that might have resulted from more than one pair of inter-residue protons.

The 3D ^{15}N-edited NOESY helped substantially in resolving the overlapped peaks correlated with the residues adjacent to the nine ^{15}N-labeled amino acids. This experiment enabled resolution of the ambiguous protons by generating slices edited by the chemical shifts of the ^{15}N nuclei bounded directly to the ^1H spins. Notably, the resonance assignments for K17, which preceded ^{15}N-labeled G18, were primarily made with the help of this experiment, since the protons of K17 were almost fully overlapped with those of R28 in the 2D homonuclear TOCSY and NOESY. Figure 4.10 shows all the assigned slices of the 3D ^{15}N-edited NOESY.

The majority of the peaks in the NOESY spectrum were assigned with proton resonance frequencies. The remaining unassigned peaks likely originated from peptide impurities or protonated HFIP. A total of 797 NOEs were assigned, out of which 760 were unambiguous while the rest 37 were ambiguous. Ambiguous NOEs were assigned in cases for which up to four different inter-residue assignment possibilities existed. From these assignments, resonance frequencies (and hence the chemical shifts) were obtained for all the magnetic (spin 1/2) nuclei (i.e., ^{1}H and ^{15}N) expected to be observable. Table 4.1 lists the chemical shifts of these nuclei that were identified following the assignments. The few protons left did not produce any observable NOEs and often they remain unassigned in NMR studies. Many of the missing assignments were at the side-chain ends and those neither interact significantly with the rest of the proteins nor possess many NOEs. The chemical shifts of protons at a particular position in the side-chain (e.g., β, γ, etc.) for which assignments could not be made stereospecifically are shown in red and those exhibiting degeneracy are shown in blue.

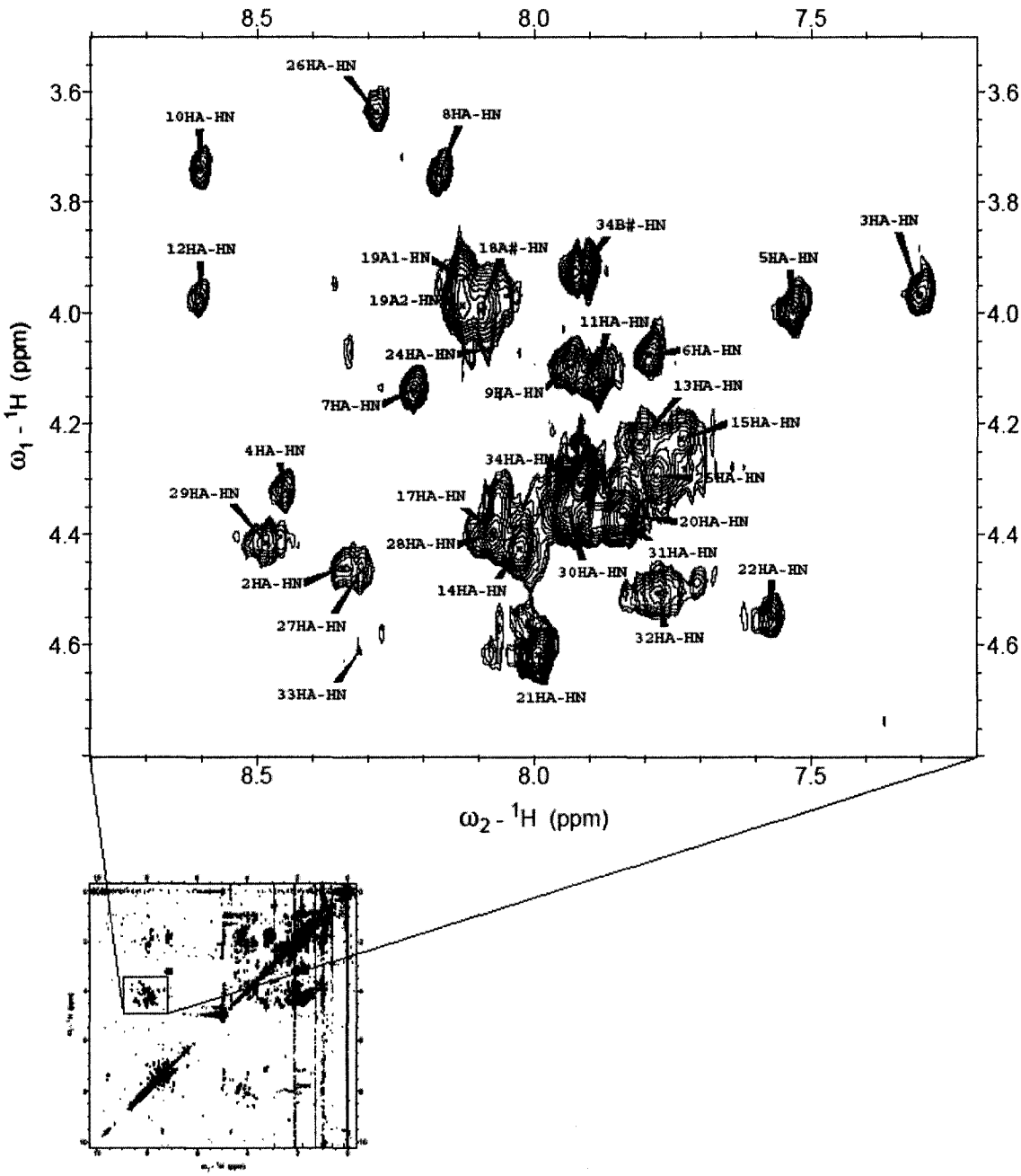


Figure 4.6 : H_α-H_N region of 2D TOCSY spectrum (45 ms mixing time) of Mini-B. The peaks resulted from the correlation (magnetization transfer) between the intra-residue amide and alpha protons and are shown as ....HA-HN. All the 34 residues yielded signals except the N-terminal one (C1). This is because, usually the N-terminal amide proton response is not seen in TOCSY.

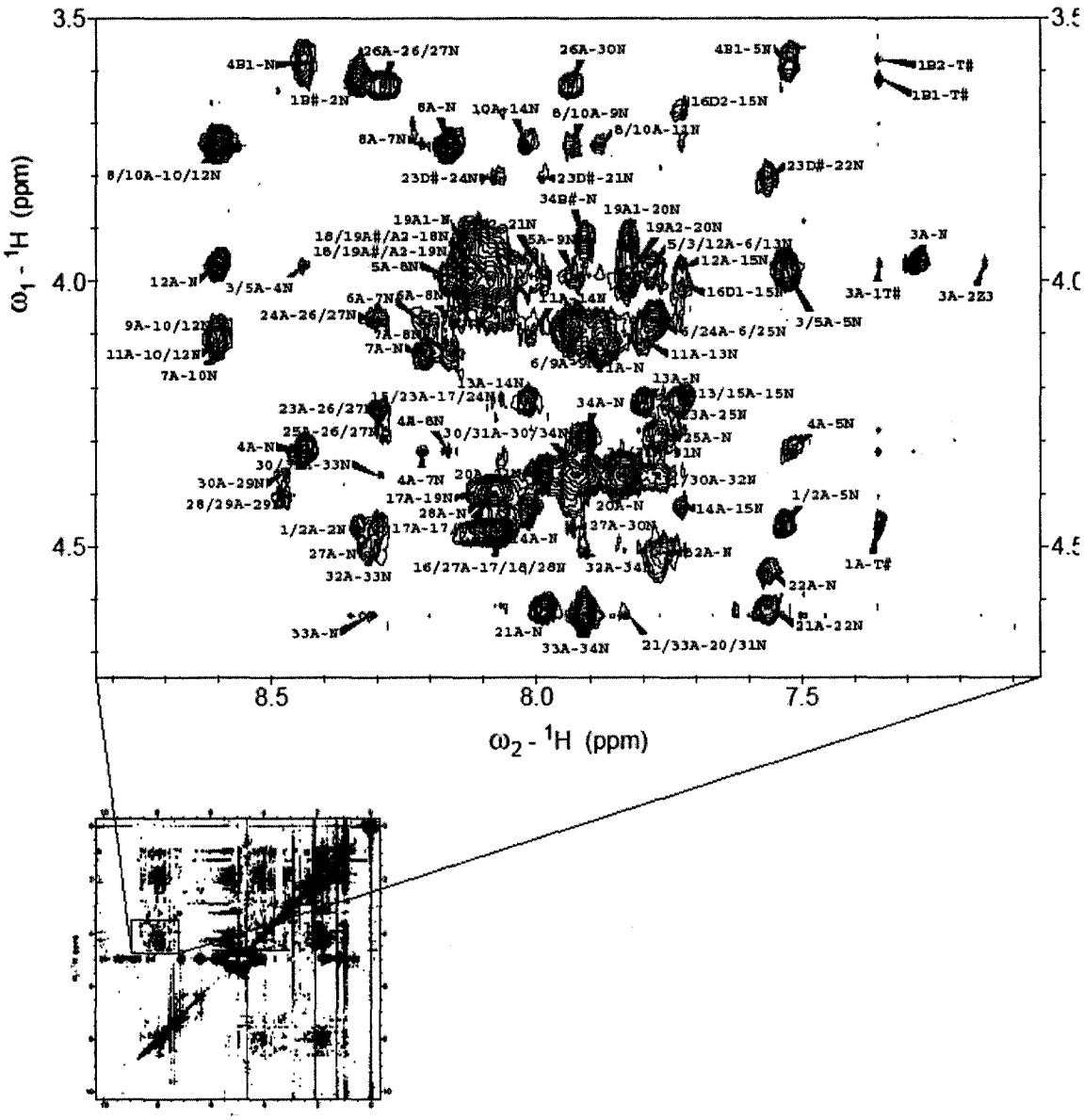


Figure 4.7 : H_α-H_N region of 2D NOESY spectrum (200 ms mixing time) of Mini-B. The NOEs (peaks) resulted from the magnetization transfer between amide and alpha protons (both intra- and inter-residue) those were close in space (less than 5 Å). Ovarlapping as well as ambiguity (possibility of correlating more than one pair of inter-residue protons) are seen for several cases.

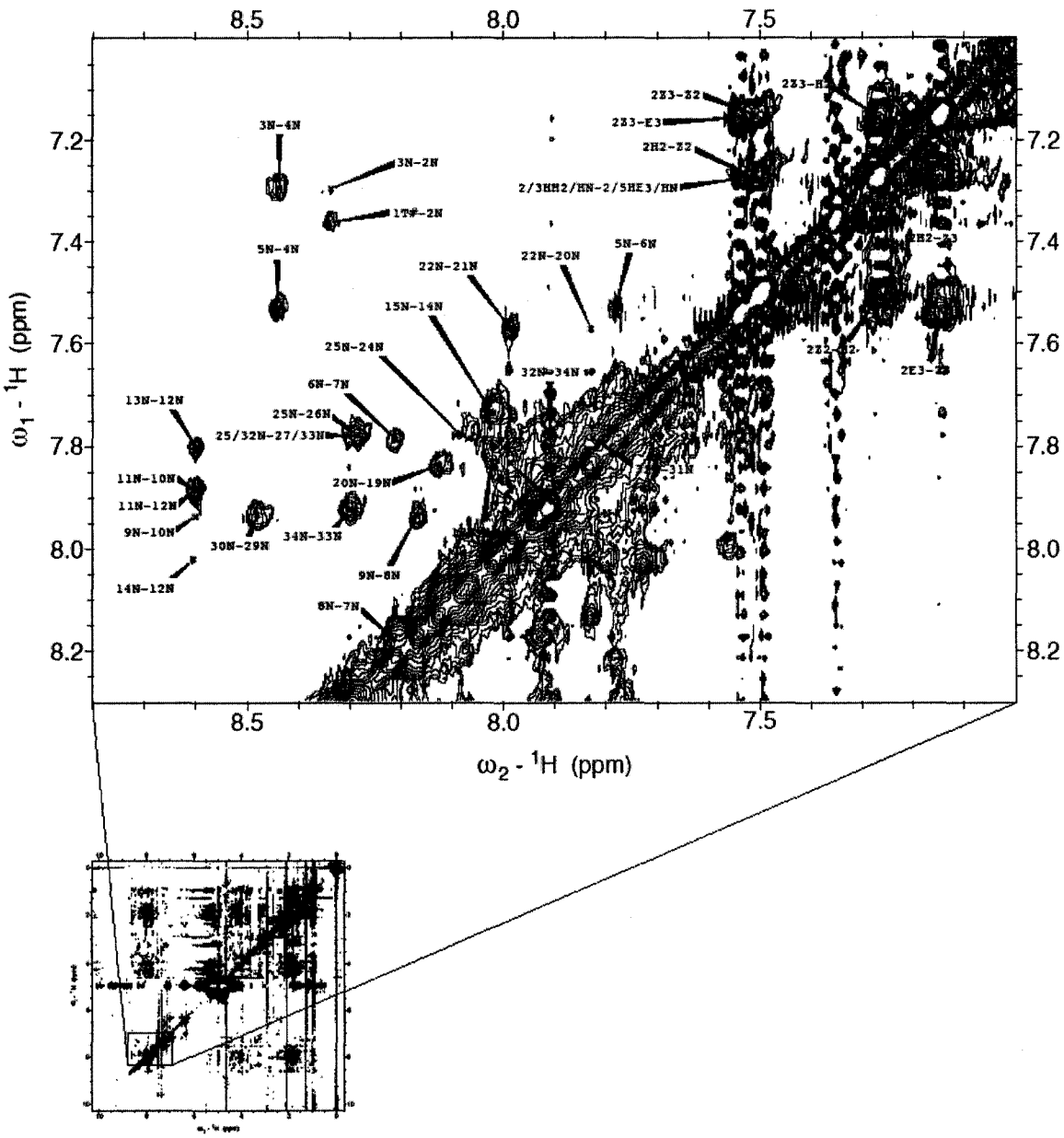


Figure 4.8 : H_N-H_N region of 2D NOESY spectrum (200 ms mixing time) of Mini-B. The NOEs (peaks) resulted from the magnetization transfer between the amide protons (mostly sequential) those were close in space (less than 5 Å).

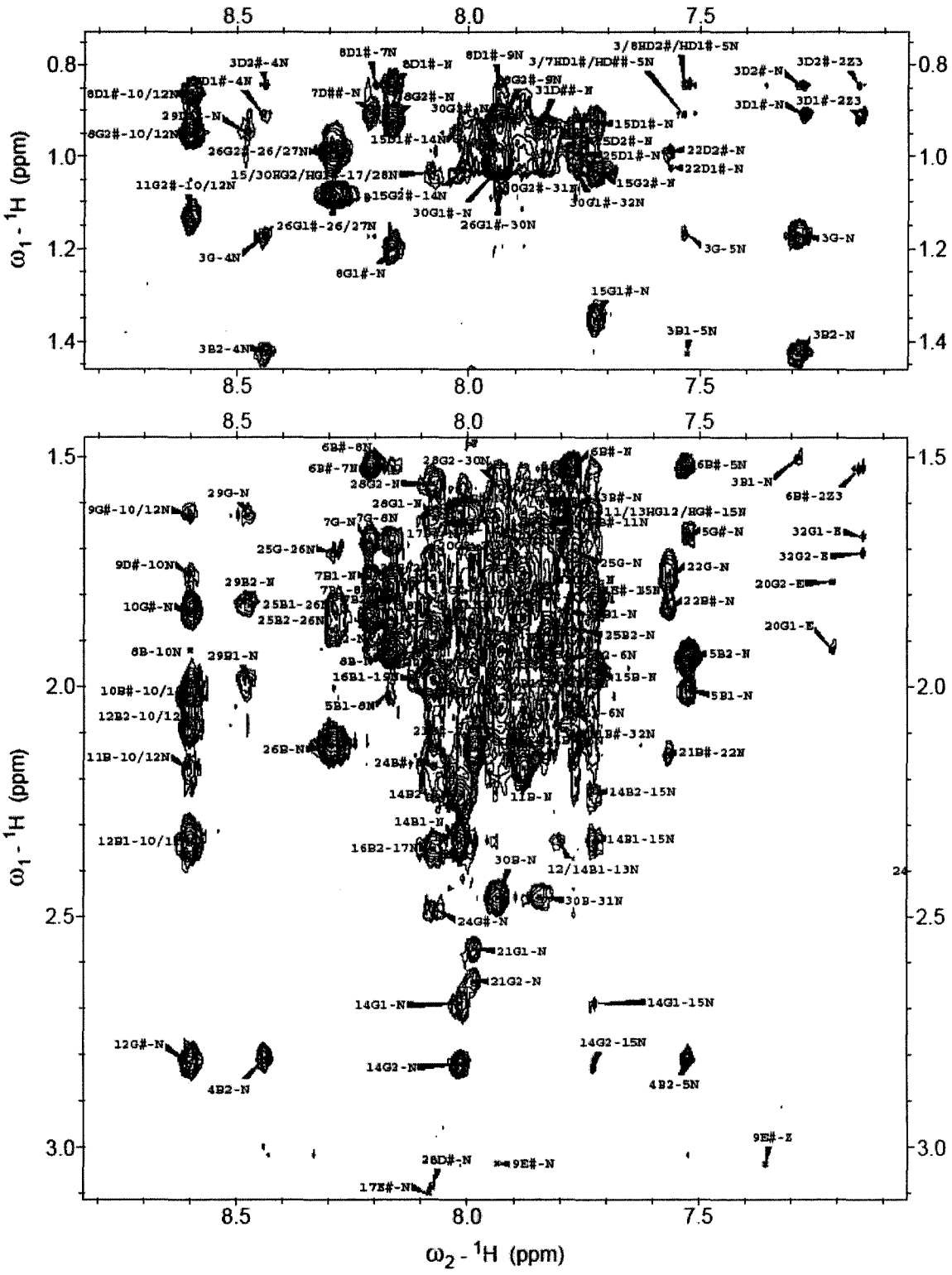


Figure 4.9 : Additional regions of 2D NOESY spectrum (200 ms mixing time) of Mini-B.

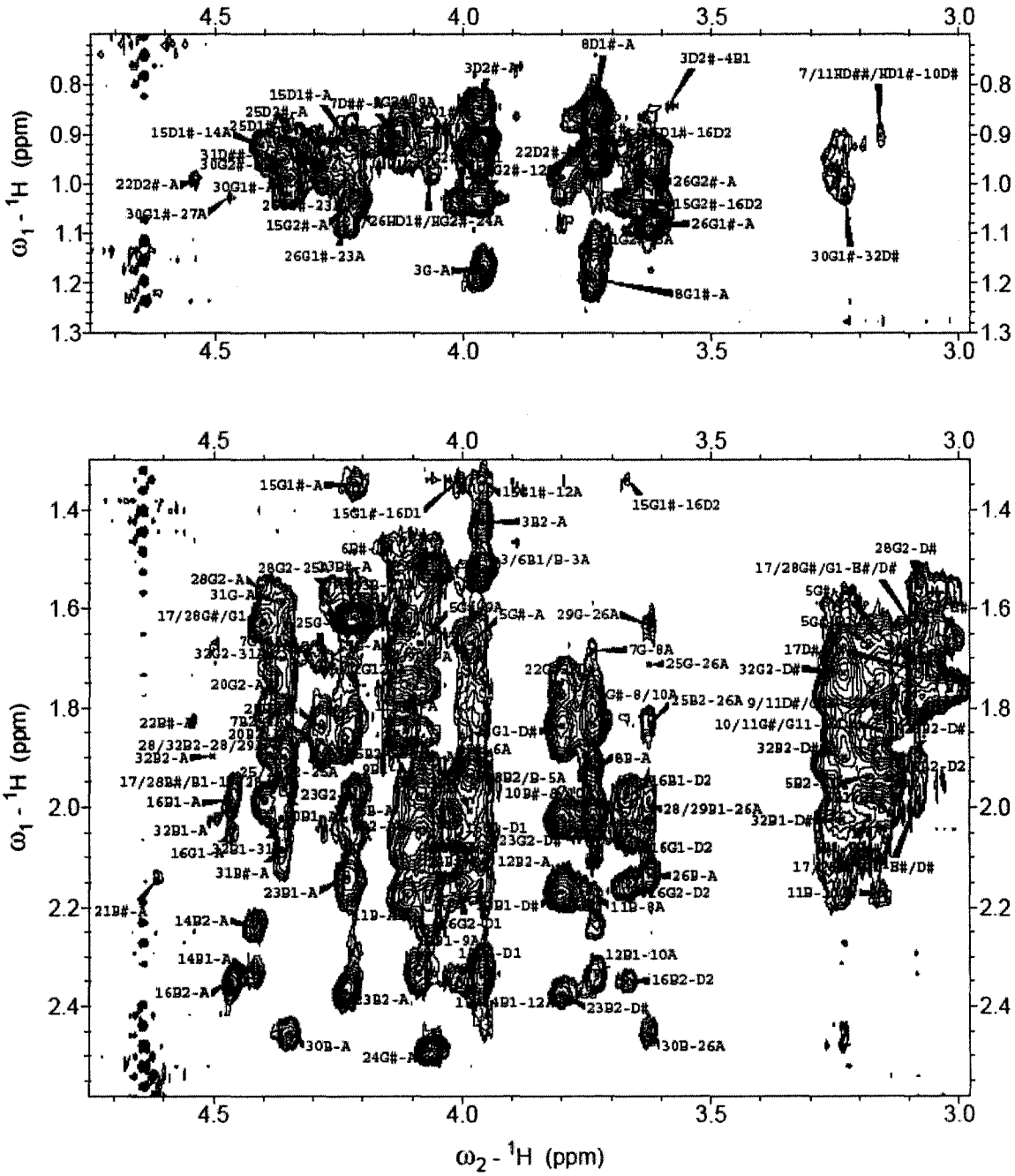


Figure 4.9 (cont.) : Additional regions of 2D NOESY spectrum (200 ms mixing time) of Mini-B.

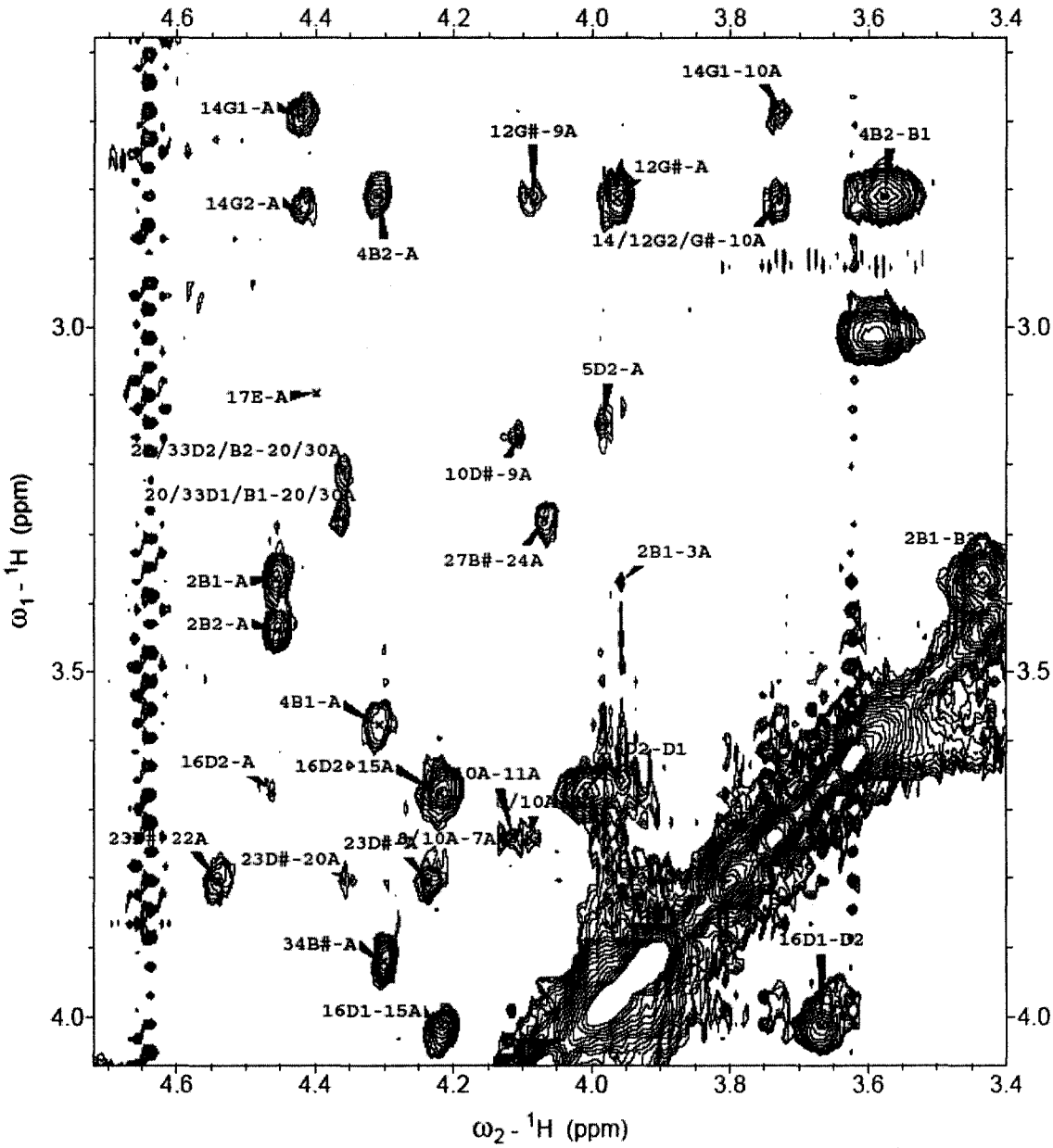


Figure 4.9 (cont.) : Additional regions of 2D NOESY spectrum (200 ms mixing time) of Mini-B.

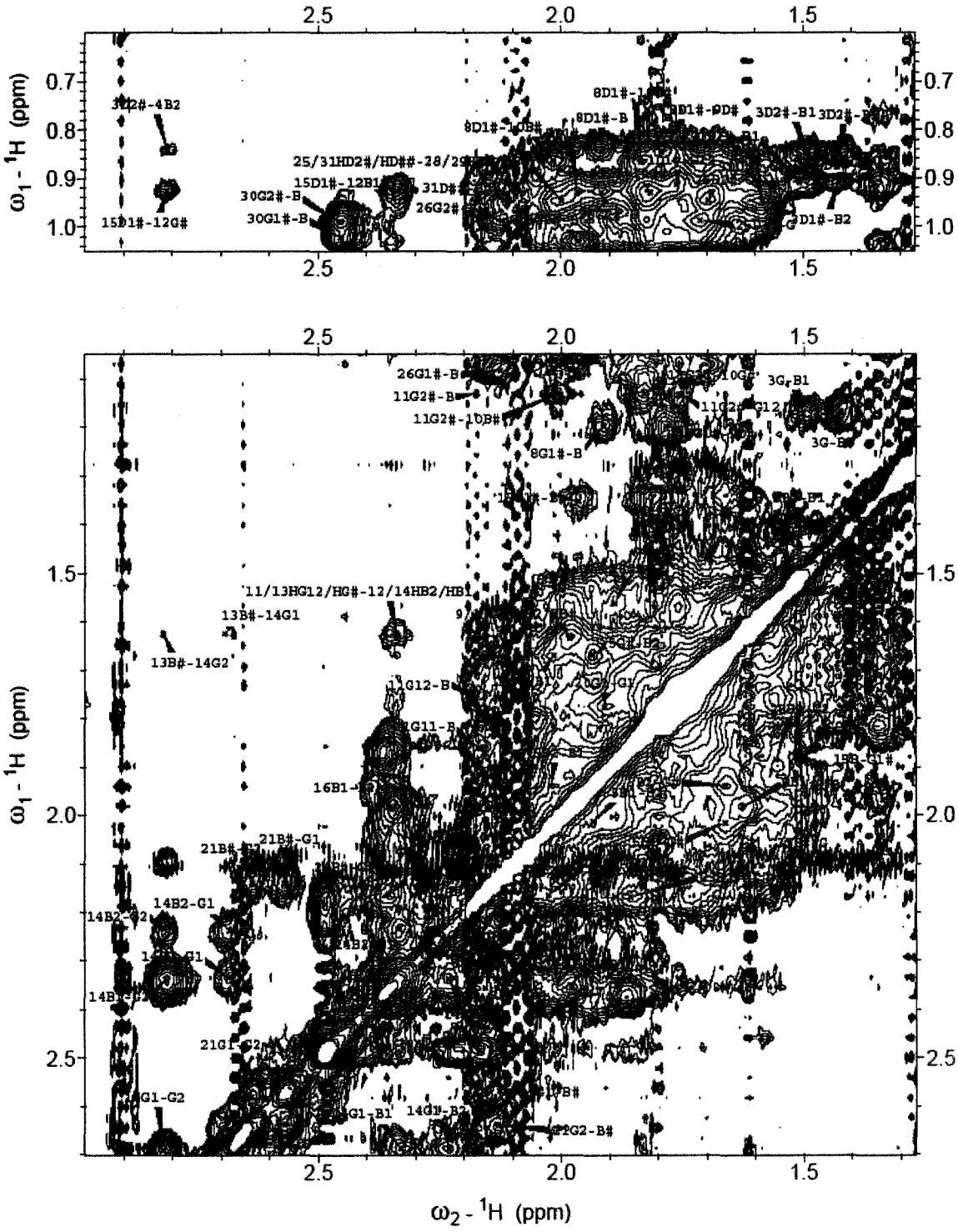


Figure 4.9 (cont.) : Additional regions of 2D NOESY spectrum (200 ms mixing time) of Mini-B.

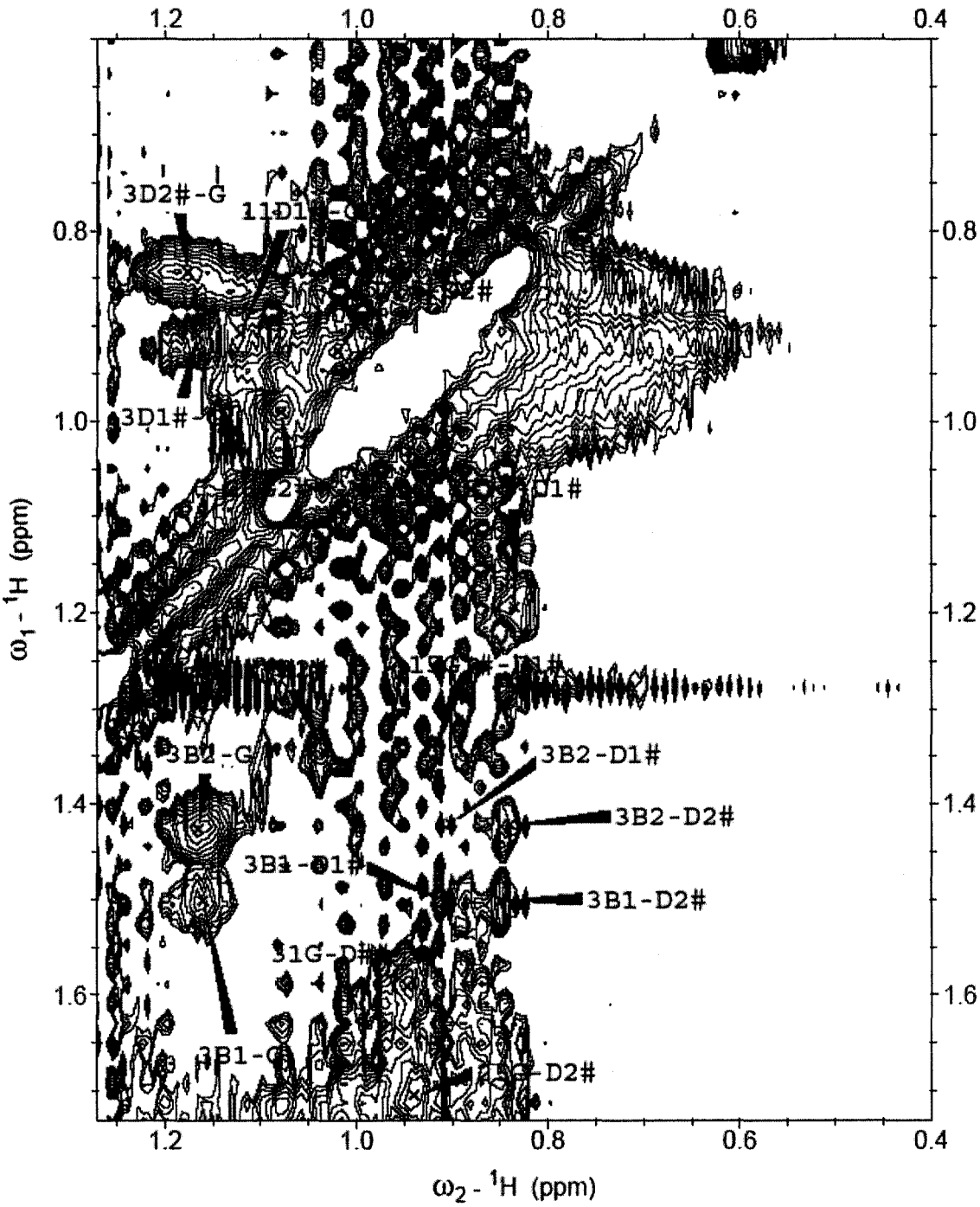


Figure 4.9 (cont.) : Additional regions of 2D NOESY spectrum (200 ms mixing time) of Mini-B.

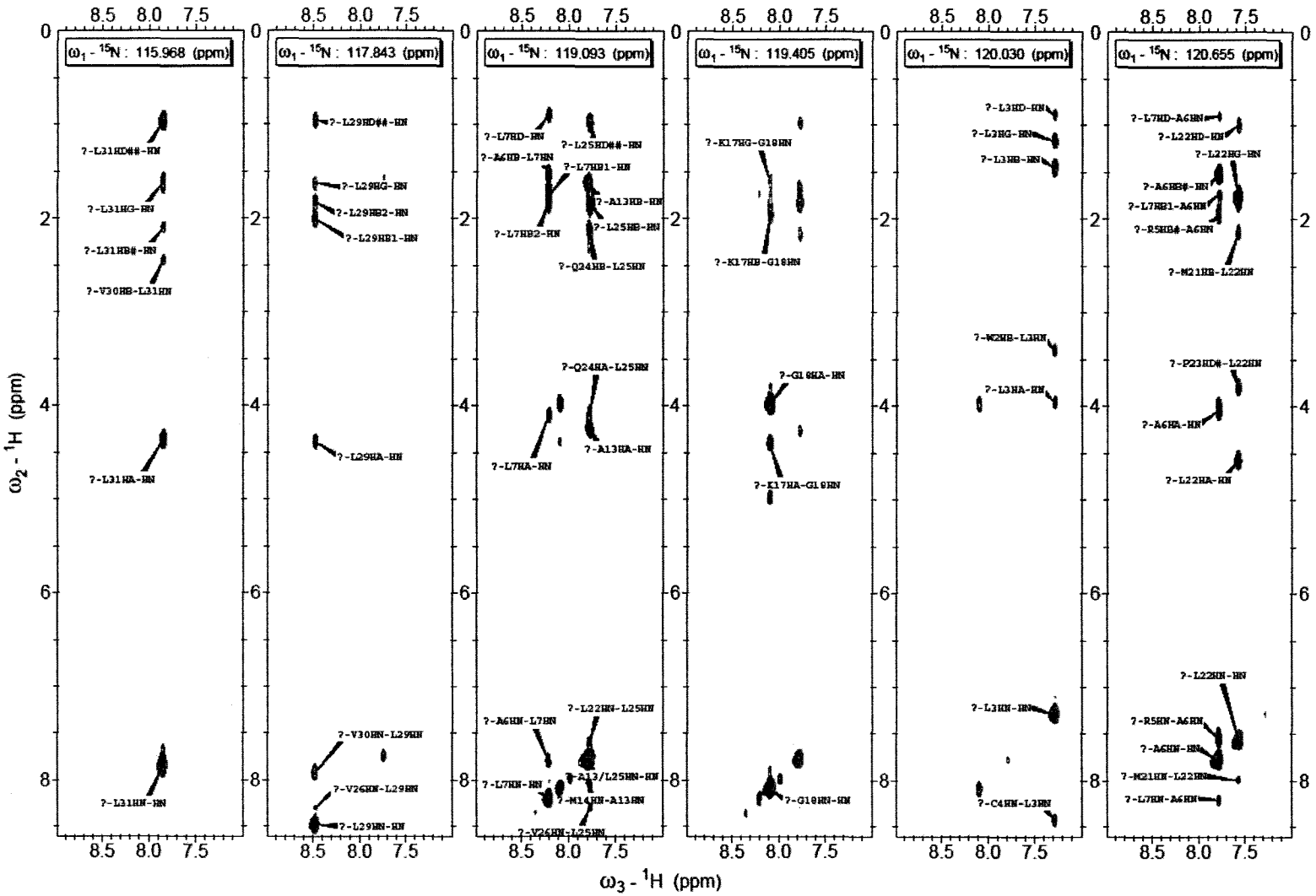


Figure 4.10 : Assigned slices of 3D ^{15}N-edited NOESY of Mini-B. NOEs only from the proton-pairs that were dipolar-coupled to the nine ^{15}N nuclei were retained in the experiment. Peaks for G18 were folded to ~ 119.5 ppm to reduce the spectral width.

Table 4.1 : Chemical shifts of the magnetic (spin 1/2) nuclei of Mini-B

AA	^{15}N	$^{1}H_N$	$^{1}H_A$	$^{1}H_B$	$^{1}H_G$	$^{1}H_D$	$^{1}H_E$	$^{1}H_{others}$
C1	---	7.36	4.46	3.61 3.58	Not found	---	---	---
W2	---	8.34	4.46	3.37 3.44	---	7.36	9.58 7.54	7.51, 7.15, 7.27
L3	120.2	7.29	3.97	1.51 1.42	1.17	0.91 0.85	---	---
C4	---	8.44	4.32	3.58 2.81	Not found	---	---	---
R5	---	7.53	3.99	2.01 1.94	1.67	3.18 3.14	Not found	Not found
A6	120.8	7.78	4.07	1.52	---	---	---	---
L7	119.0	8.21	4.13	1.77 1.85	1.69	0.90	---	---
I8	---	8.17	3.74	1.92	1.20 0.93	0.85	---	---
K9	---	7.93	4.09	1.98 2.04	1.63	1.76	3.04	7.36
R10	---	8.60	3.74	2.02	1.84	3.16	Not found	Not found
I11	---	7.88	4.11	2.17 1.85	1.76 1.13	0.91	---	---
Q12	---	8.60	3.97	2.34 2.09	2.82	---	6.51 6.32	---
A13	119.0	7.80	4.22	1.62	---	---	---	---
M14	---	8.02	4.42	2.34 2.23	2.69 2.82	---	1.82	---
I15	---	7.73	4.22	1.97	1.35 1.03	0.93	---	---
P16	---	---	4.47	1.98 2.35	2.05 2.16	3.67 4.01	---	---
K17	---	8.08	4.40	1.99	1.63	1.72	3.10	Not found
G18	126.4	8.09	3.98	---	---	---	---	---
G19	---	8.13	3.93 3.99	---	---	---	---	---
R20	---	7.83	4.36	1.96 1.89	1.92 1.77	3.28 3.21	7.21	Not found
M21	---	7.99	4.62	2.14	2.57 2.64	---	---	---
L22	120.7	7.57	4.55	1.83	1.77	1.02 0.99	---	---
P23	---	---	4.24	2.16 2.37	1.84 2.02	3.80	---	---
Q24	---	8.08	4.07	2.17	2.49	---	7.01 6.37	---
L25	119.3	7.77	4.29	1.84 1.87	1.71	0.99 0.94	---	---
V26	---	8.29	3.63	2.13	1.08 0.99	---	---	---
C27	---	8.31	4.47	3.27	Not found	---	---	---
R28	---	8.08	4.40	1.99 1.88	1.63 1.56	3.09	Not found	Not found
L29	117.9	8.48	4.40	1.99 1.82	1.63	0.95 0.92	---	---
V30	---	7.94	4.36	2.46	1.02 0.99	---	---	---
L31	116.2	7.84	4.37	2.10	1.59	0.95	---	---
R32	---	7.78	4.51	2.02 1.90	1.67 1.71	3.23	7.15	Not found
C33	---	8.30	4.63	3.27 3.19	Not found	---	---	---
S34	---	7.91	4.30	3.92	Not found	---	---	---

4.6 Structure Calculations

Three-dimensional structures of Mini-B were calculated by dynamic simulated annealing. As explained in Section 4.4, after completing the spectral assignments of the three NOESYs using Sparky 3.110 [77], the spectrum with 200 ms mixing time was selected for generating the restraints to be used in the structure calculation. All the assigned NOEs were classified into strong, medium and weak categories depending on the peak intensities (heights). Proton-pairs were assigned a distance range according to which of the categories they fell in. A second list of restraints was generated from the hydrogen bonds between residues that were known to be within α-helical regions of the peptide. This was based on the preliminary data about the conformational mapping of Mini-B obtained by Alonso *et al.* using Fourier Transform Infrared (FTIR) Spectroscopy (PDB ID 1SSZ) [81] and confirmed using local NOE patterns. A third list of restraints was generated from the allowed values of backbone dihedral angles corresponding to the known secondary structure. Table 4.2 summarizes the experimental restraints used in structure calculation.

Table 4.2 : Experimental restraints for Mini-B structure calculation

Intra-residue	559
Sequential	127
Medium-range	74
Long-range	00
Unambiguous	760
Ambiguous	37
Total NOE distance restraints	797
Dihedral-angle restraints	48
Hydrogen-bond restraints	32

Distance restraints corresponding to the sequential and medium-range NOEs, backbone dihedral angle restraints and hydrogen-bond restraints constituted the required additional input files for Mini-B structure calculation using the simulated annealing

algorithm within CNS 1.1 [78]. Although most of the NOEs (559 out of 797) were intra-residue, those were not added into the CNS input as they contribute little structural information but their inclusion tends to over-constrain the structure calculations.

A total of nine rounds of structure calculations for Mini-B were performed in CNS. A total of 100 lowest energy structures were determined in each round from the first to the eighth and 200 in the final round. The first round began without the ambiguous NOEs and the resonances were also not stereospecifically assigned. The calculated structures were viewed by MOLMOL 2k.2 [79]. Some stereospecific assignments followed from the observation and some corrections in the distances assigned to restraints were made as well. Then the second round of calculation was performed and the structures were re-viewed. This enabled the making of more stereospecific assignments and helped again to correct more distance restraints. This process continued up to the fourth round. The ambiguous NOEs were then added and the fifth round of calculation was carried out. Additional NOE information obtained from the 3D ^{15}N-edited NOESY were incorporated in the sixth round. Three additional rounds of structures were calculated and, each time, subtle adjustments to the distances assigned to the restraints were made. After the final round, the lowest-energy 15 structures were retained for analysis. In this set, all the experimentally derived restraints were fulfilled, except for one minor NOE violation. The coordinates of these 15 lowest energy structures of Mini-B have been deposited in the Protein Data Bank (PDB) [PDB ID 2A2H].

An important set of secondary structure indicators that came from the selected distance restraints is summarized in Figure 4.11. The height of the bars is proportional to the spatial proximity of the protons in pairs i.e., higher bars correspond to pairs of protons that are relatively close together. The resonance frequencies themselves also provided indications for the 3D structure. The chemical shift index (CSI) for the H_α resonances, which was obtained by subtracting the chemical shifts of H_α for residues in random coil configuration [82] from that of Mini-B, is shown at the bottom of Figure 4.11. CSI values

close to zero are indicative of a random coil configuration while significantly positive values of a helical (α-type) and significantly negative values of an extended (β-strand) conformations. These analyses confirmed that Mini-B consists of two helices, the first one runs through amino acids 2 to 14 while the second one stretches from amino acids 23 to 33 in the chain.

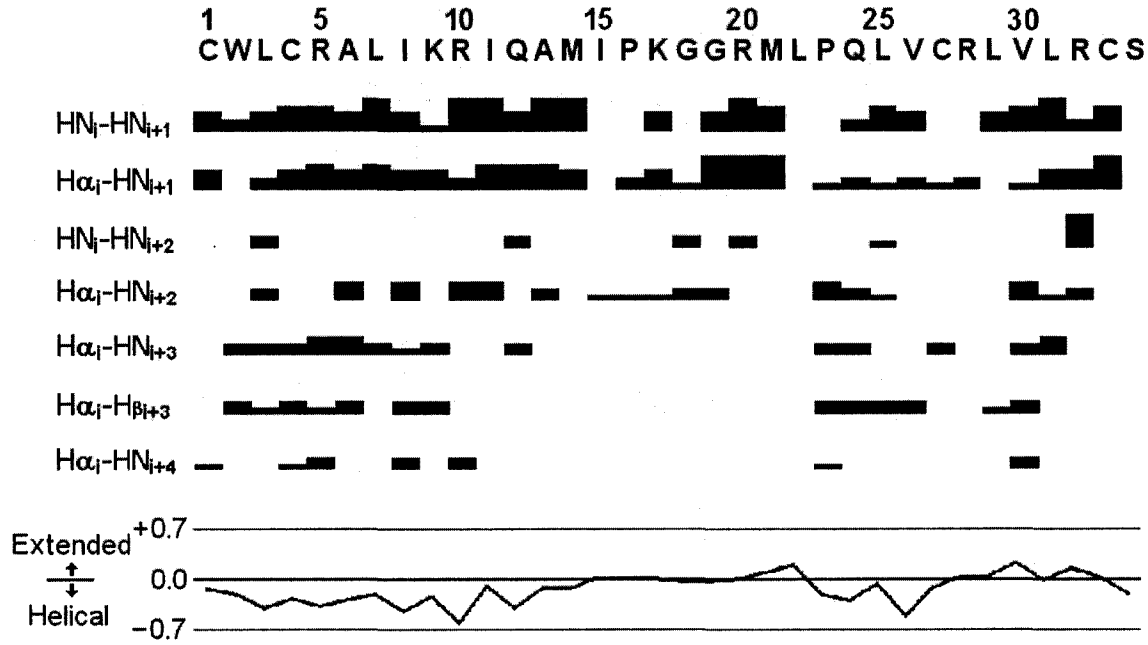


Figure 4.11 : Structure indicators for Mini-B in HFIP. Selected distance restraints are shown at the top. The height of the bars correspond to the intensity of the NOEs. Strong NOEs (i.e., higher bars) for HN_i-HN_{i+1} and $H\alpha_i$-HN_{i+1} proton-pairs are indicative of α-helical conformation. Prevalence of NOEs from the other proton-pairs shown here are also indicative of α-helical conformation. The chemical shift index (CSI) for the H_α resonances are shown at the bottom. This curve is obtained by subtracting the random coil H_α chemical shifts [82] from that of Mini-B. CSI values considerably lower than zero indicates helical conformation and that considerably higher than zero indicates extended conformation. These indicators suggest that Mini-B consists of two terminal α-helical regions connected by an unstructured loop at the middle.

The Ramachandran plot (a graph generated for the allowed values of dihedral angles ϕ and ψ) for the lowest energy structure of Mini-B is shown in Figure 4.12. All the residues found to form the helices in the structure are seen within the allowed regions for α-helix in the plot with most of them in the most favored region.

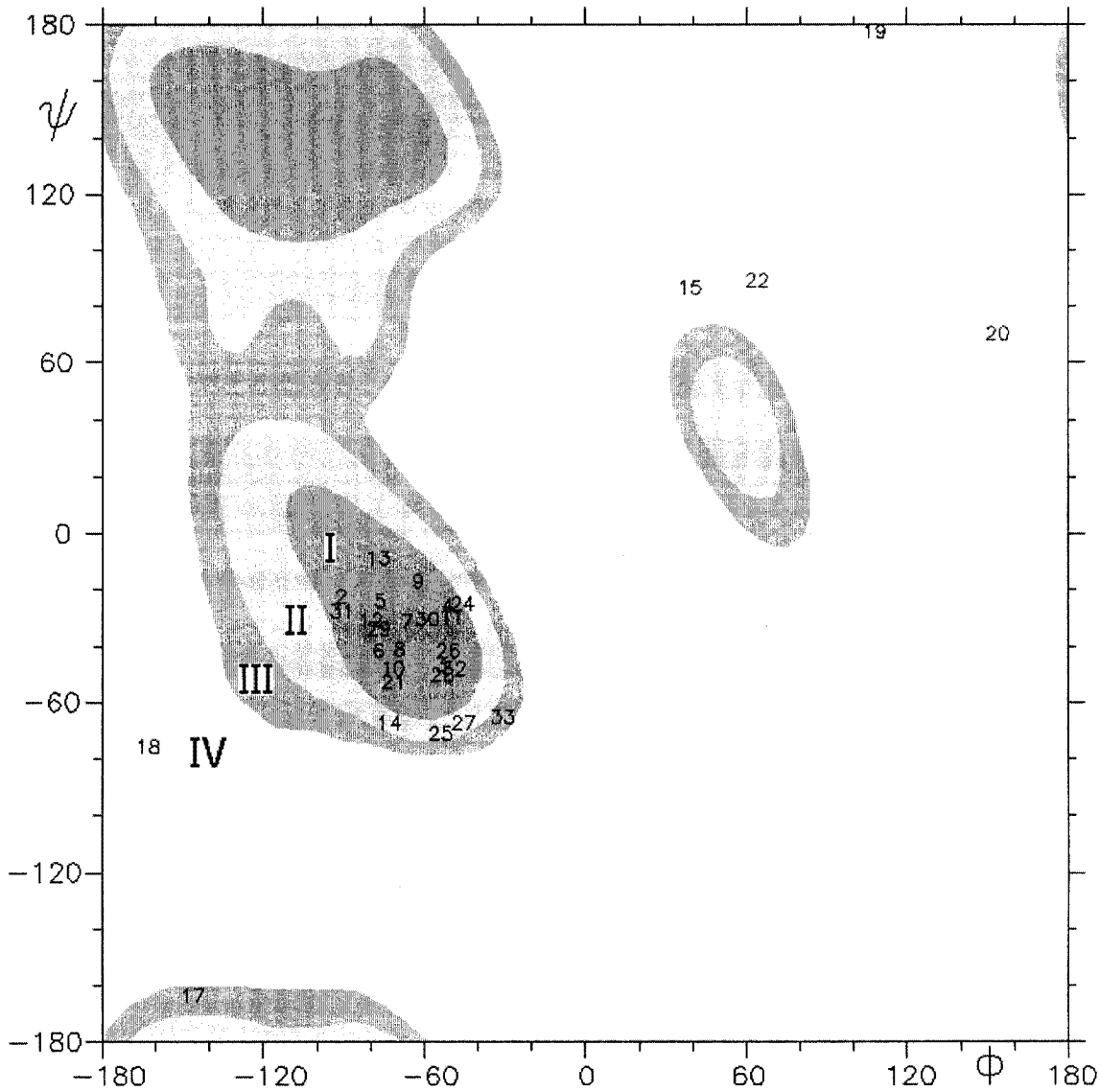


Figure 4.12 : Ramachandran plot for the lowest energy structure of Mini-B in HFIP. Code for the regions for α-helical structure : I – most favored, II – less favored, III – allowed and IV – not allowed. Residues are represented by the numbers representing their positions in the sequence. Most of the residues forming the two terminal helices according to the structure are seen within the most favored region. The rest of the helix-forming residues are also seen within the less favored and allowed regions. Residues in the middle portion are seen in the not allowed region. The terminal residues are missing as one of the two angles for them is not defined.

4.7 Structure of Mini-B in HFIP

The high resolution structure of Mini-B in aqueous solvent containing HFIP was found to consist of two well-defined α-helical terminal regions connected by an unstructured loop at the middle. Figure 4.13 shows a ball-stick structure of Mini-B that represents the lowest energy conformation from the 200 structures calculated in the last round. The atoms are shown as balls (the color code is explained in the figure caption) and the bonds between the atoms are shown as sticks.

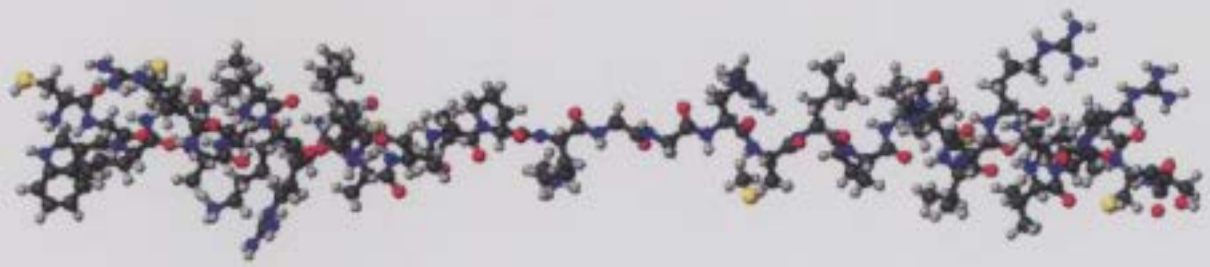


Figure 4.13 : The lowest energy ball-stick structure of Mini-B in HFIP. Color code for atoms : C – black, N – blue, O – red, S – yellow and H – gray. The terminal regions are seen to be α-helical.

Figure 4.14 shows the electrostatic surface charge plots for the lowest energy structure of Mini-B. The positively charged surface is shown in blue and the negatively charged surface in red. Mini-B carries a net charge of +7 (same as the full-length SP-B) at neutral pH and five of the seven positively charged residues are distributed at the terminal α-helical regions. This makes the helices electrostatically biased. Models aimed to explain the functional characteristics of the native SP-B depict these cationic properties of the two terminal regions crucial for the interactions between the protein and the anionic phospholipids in the lung surfactant. It is the combination of structure and charge for the α-helical N- and C-terminal domains that is the key to SP-B's function [76].

Figure 4.15 shows the relative positions of the helix-forming residues of Mini-B plotted on helical wheels. For the first helix, 8 of the 13 residues are nonpolar and mostly

positioned at the right part of the wheel. The 3 positively charged residues are localized at the top. For the second helix, 5 out of 11 residues are nonpolar and mostly localized at the left half of the wheel. The 2 positively charged residues are positioned at the bottom. Although Mini-B is dominated by nonpolar residues and the hydrophobic nature of the peptide is thus obvious, the helical domains possess amphiphilic properties as well because of the presence of a few polar and charged residues.

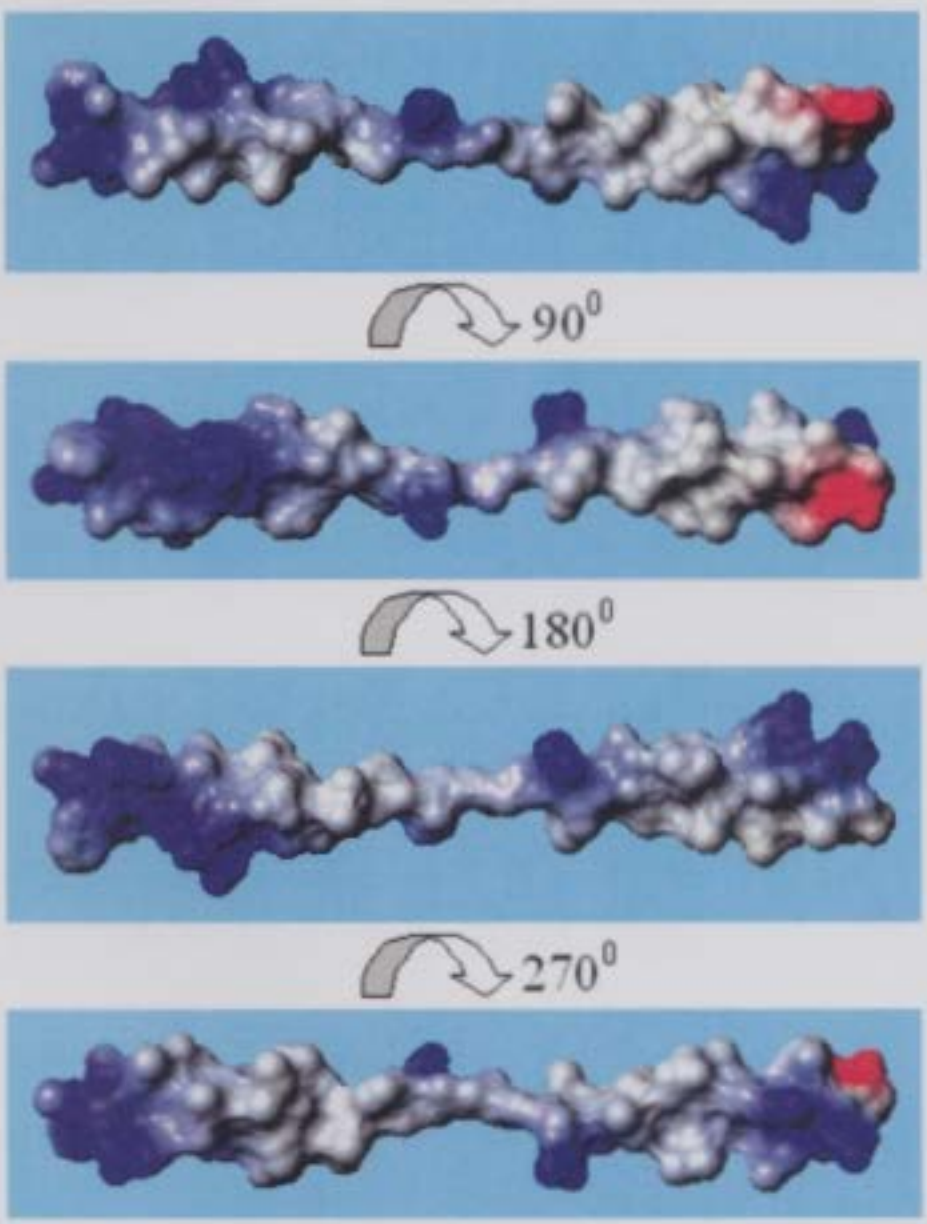


Figure 4.14 : Electrostatic surface charge plots for the lowest energy structure of Mini-B in HFIP. Color code : blue – positively charged surface and red – negatively charged surface. Mini-B contains seven positively charged residues which are mostly distributed in the terminal regions.

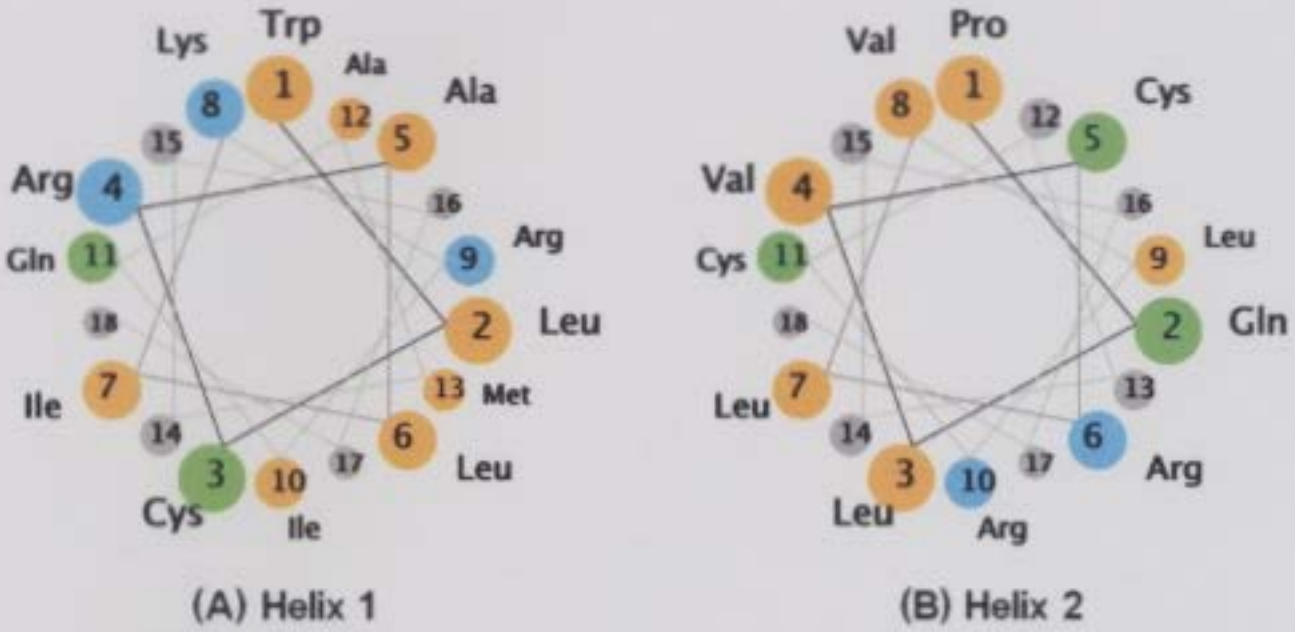


Figure 4.15 : Relative positions of the residues in the Mini-B helices plotted on helical wheels. Color code : – nonpolar, – polar but uncharged and – positively charged. The gray-colored numbers do not represent any residues.

An ensemble of the 15 lowest energy structures of Mini-B that was deposited to the PDB is shown in Figure 4.16. The backbone (gray) and the heavy atoms in the side-chains (colored) are shown while the hydrogens are omitted for clarity in Figure 4.16 (A). It is noted that the middle segment that joins the two terminal helices is unstructured. Thus the relative orientation of the two helices is not defined and the angle between the helices shown here is arbitrary. The backbone ribbon representation of these 15 lowest energy structures of Mini-B is shown in Figure 4.16 (B). Again, the two termini of the peptide are clearly seen to consist of two α-helices connected by a randomly oriented middle region.

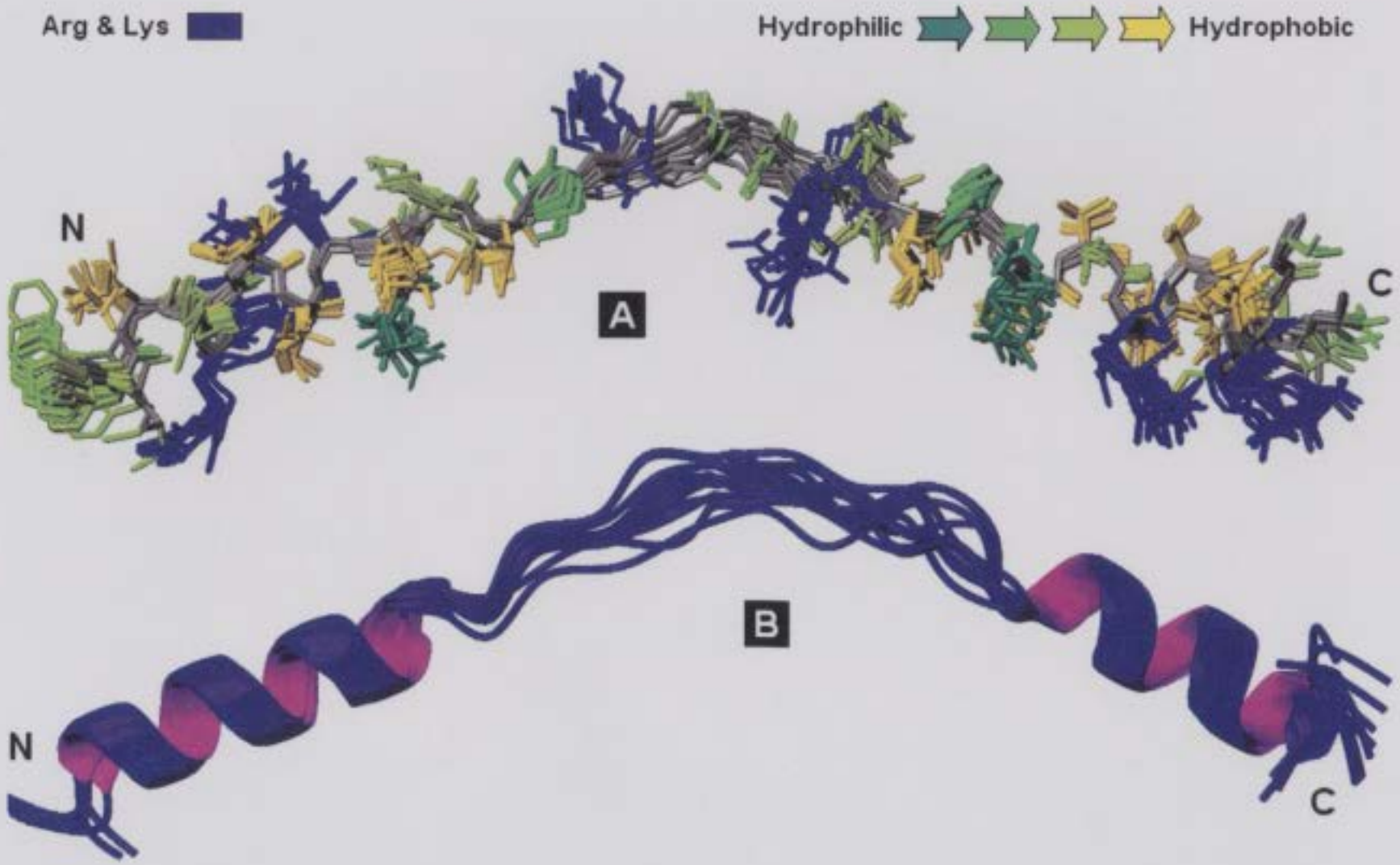


Figure 4.16 : An ensemble of 15 lowest energy structures of Mini-B in HFIP. (A) The backbone (gray) and the side-chain heavy atoms (colored) are seen while the side-chain hydrogens are omitted for clarity. (B) The backbone ribbon structures. The peptide is clearly seen to consist of two terminal α-helices which are connected by an unstructured loop. The angle between the helices shown here is arbitrary as the relative orientation of the two helices is not defined.

4.8 Discussion

It is a fundamental axiom of biology that the function of a protein depends on its 3D structure [44]. Understanding function through structure is thus a primary goal of structural biology. However, the structure-function relationships of SP-B have not yet been established because the 3D structure of the native protein is still undetermined. Currently, researchers have attempted to characterize the peptide fragments of SP-B. The major emphasis of studying these peptides is to identify minimal regions of the protein retaining the interfacial properties thought to be important for exogenous surfactant activity [41, 42, 83 - 87]. Mini-B is a fragment of human SP-B, exhibiting significant biological activity of the full-length protein [76]. The first detailed structural characterization of this peptide has been obtained in the present research using solution NMR. This high-resolution structure of Mini-B will help to identify the structural basis that underlies SP-B's mechanisms of function.

The sequence of SP-B is homologous to some other proteins, all of which belong to the Saposin protein superfamily characterized by a high content of α-helical secondary structure [88]. All of them also possess six cysteine residues, the relative positions of which are highly conserved to form three intramolecular disulfide bridges [89]. However, SP-B contains one extra cysteine and the native form of SP-B is a homodimer stabilized by a disulfide bridge through the fifth of the seven cysteines [75]. The members from the Saposin superfamily whose structures have been determined so far are : Saposin B (PDB entry 1N69 [37]), Saposin C (PDB entry 1SN6 [38]), NK-Lysin (PDB entry 1NKL [39]) and Amoebapore (PDB entry 1OF9 [40]). The structures of these proteins including Mini-B are shown in Figure 4.17. All the Saposin proteins contain four (or more) α-helices connected by loops. Table 4.3 lists the amino acid sequences of these proteins [88]. The helix forming residues are shown in red color and the disulfide bridge forming cysteines with gray shade in Table 4.3. The Mini-B residues of SP-B are shown underneath, with the two α-helices obtained from the structure determined in the present research in blue.

NMR structural studies of membrane or lipid-associated proteins/peptides are often attempted in organic solvent rather than in micelles. However, the organic solvent usually induces the secondary structure only, not the overall 3D conformation. The fluorinated alcohol (TFE) is a commonly used organic solvent to induce the secondary structure. HFIP, although less commonly used in NMR studies, appears to be an even stronger secondary structure inducer. Presently, no widespread consensus has been reached on how these solvent systems promote the secondary structures. However, strengthening of both intra-peptide hydrogen bonds and local hydrophobic interactions is likely to be important in TFE or HFIP stabilization of secondary structures [90, 91].

Table 4.3 : Amino acid sequence of SP-B and other Saposin family members [88]

Protein	Amino acid sequence
SP-B	FPIPLPYCWLCRALIKRIQAMIPKGA-----LAVAVAQVCRVVPLVA-GGICQCLAERYSVILLDTLLGRML-PQLVCRLVLRCS-M
NKL	G-YFCESCRKIIQKLEDMVGPQPN-EDTVTQAASQVCDKL-K-ILRGLCKKIMRSFLRRISWDILTGK-KPQAICVDIKICK-E
SapA	SLP-CDICKDVVTAAGDMLKDNAT-EEEILVYLEKTCDWLPKPNMSASCKEIVDSYLPVILDIIKGEMSRPGEVCSALNLCESLQ
SapB	GDVCQDCIQMVTDIQTAVRTNSTFVQALVEHVKEECDRLG-PGMADICKNYISQYSEIAIQMMMH--MQPKEICALVGFCD--E
SapC	SDVYCEVCEFLVKEVTKLIDNNKT-EKEILDAFDKMCSKLPKS-LSEECQEVVDTYGSSILSILLEEVS-PELVCSMLHLCSGLVPR
SapD	DGGFCEVCKKLVGYLDRNLEKNST-KQEILAALEKGCSFLPDP-YQKQCDQFVAEYEPVLIEILVEVMD-PSFVCLKIGACPSAH
PFP	GEILCNLCTGLINYLENLLTTK-G-ADKVKDYISSLCNKA-SGFIATLCTKVLDFGIDKLI-QLIEDKVDANALCAKIHAC
Mini-B	-------CWLCRALIKRIQAMIPKG----------GRML-PQLVCRLVLRCS-M

As the main part of present research, the structure of Mini-B was determined in organic solvent HFIP. Mini-B was found to contain two α-helices at the terminal regions connected by an unstructured loop. The amino acid distribution along the sequence of Mini-B (and also SP-B) supports the formation of amphiphilic helical segments which are hydrophobic at one side and hydrophilic at the other. This set of motifs is common in other proteins of the Saposin superfamily and is implicated in the interaction of these proteins with lipids [75]. From the helical wheel diagrams (Figure 4.15), it is seen that the nonpolar residues of Mini-B are indeed mostly localized at one side of the helices making them amphiphilic. The amphiphilicity of the helical segments of Mini-B supports a

relatively superficial disposition of SP-B in lipid bilayers and monolayers in the lung. The helical segments of the protein may interact primarily with the polar headgroups of the phospholipids and slightly with their acyl chains in the hydrophobic regions of the surfactant structures.

Mini-B contains a total of seven positively charged amino acids while the net charge of SP-B is also + 7. Five of the seven cationic residues are localized in the helical regions of Mini-B (Figure 4.15). This can define a preferential electrostatic interaction of SP-B terminal helices with anionic phospholipids, such as PG, in both monolayers and bilayers. The SP-B – PG interaction has been proposed to promote enrichment of the interfacial film with the most surface active phospholipids during compression [16, 72]. Selective interaction of SP-B with negatively charged phospholipids could also be important for the formation of nonbilayer intermediates required in bilayer-monolayer transition [81].

Mini-B retained significant surface activity of the native SP-B in treating surfactant deficient rats [Waring *et al.*, unpublished]. This indicates that Mini-B preserves crucial characteristic properties of the full-length protein. SP-B likely binds anionic/zwitterionic phospholipids of lung surfactant to ensure rapid adsorption of the material and formation of specific structures that reduce the surface tension at the air-water interface in alveoli. This ability to bind other molecules is the most fundamental property which underlies all the biochemical functions of proteins [44]. Binding property arises from the complementarity of shape and charge distribution between the other molecule and its binding site on the protein and from the arrangement of donors and acceptors for hydrogen bonds. Retention of the activity of the full-length protein to a significant extent indicates that Mini-B likely contains the binding sites of the native SP-B. The two helices of Mini-B likely correspond to the two terminal helices of SP-B which are thought to be the active regions of the full-length protein [76].

The conformational features identified from this study provide insight into the structural basis of SP-B that determines its function. The present research would also provide a primary source of information for any future NMR structural studies of Mini-B in lipid micelles mimicking the physiological environment.

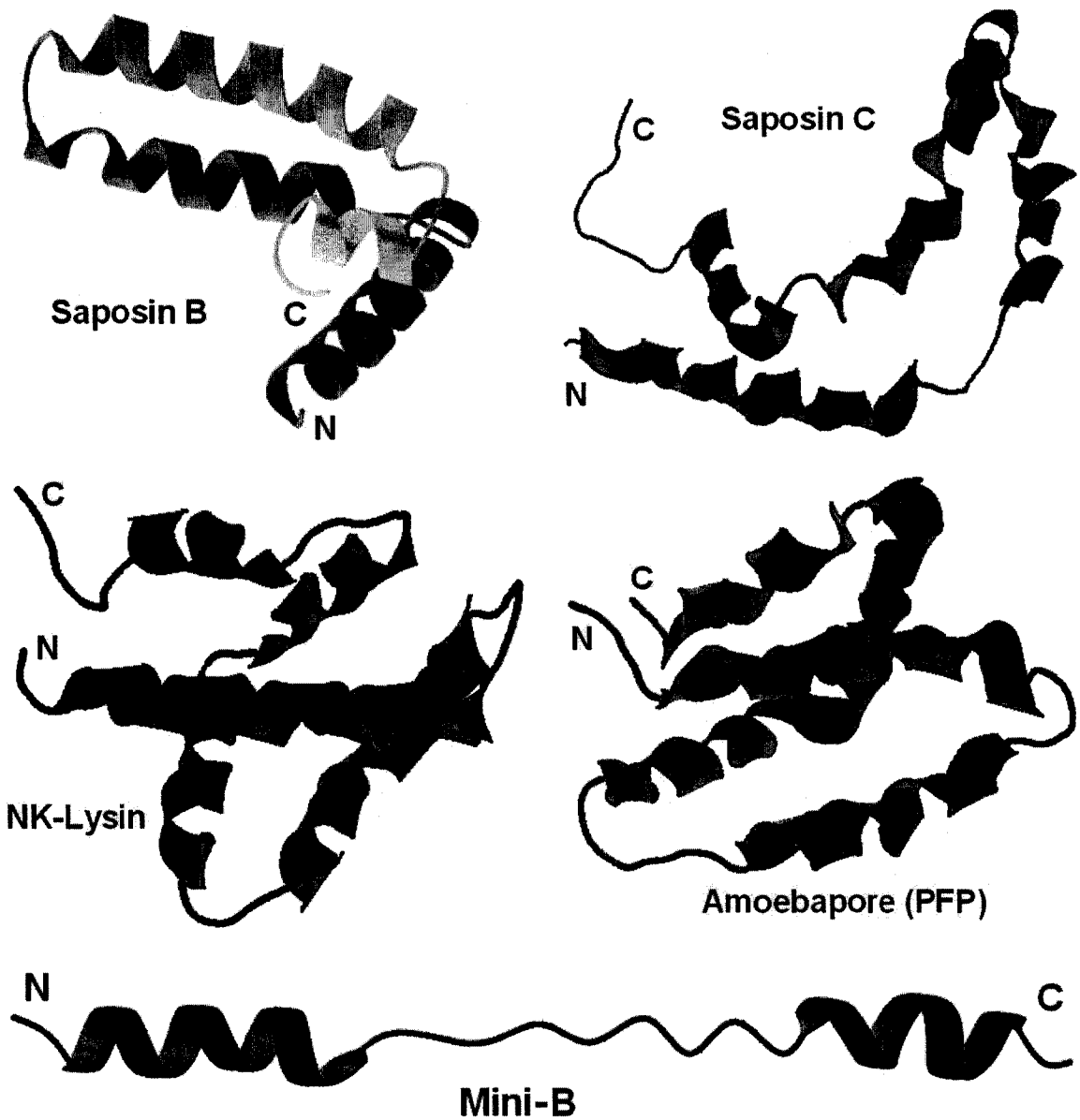


Figure 4.17 : Structures of the proteins belonging to the Saposin superfamily [diagrams are taken from Protein Data Bank]. The structure of Mini-B that was determined in the present research is shown at the bottom. The relative orientation of the two helices in Mini-B was not determined since HFIP induced only the secondary structure and not the overall 3D structure of the peptide.

Chapter 5

Conclusion

5.1 Summary and Remarks

Lung surfactant is a mixture of lipids and proteins that lines the air-water interface in alveoli. This material reduces the surface tension generated at the interface to extremely low values and thus prevents lung collapse during expiration and eases the work of breathing during inspiration. The protein SP-B is an essential component without which the surfactant cannot work. It is found as a homodimer in the mammalian lung. The 3D structure of this protein is not yet known and hence the structural basis for its crucial role is still unresolved. In the present research, the structure and lipid interactions of SP-B were probed through studies of smaller fragments of the full-length protein. Use of these protein-fragments had two main advantages; firstly, they are biologically active and hence definitely contain the functional regions of full-length SP-B and secondly, their small sizes allow them to be produced synthetically and facilitates solution NMR studies which are hampered by large protein size.

The structures of SP-BCTERM were determined earlier by Booth *et al.* in organic solvent HFIP and lipid micelles formed from detergent SDS [42]. The structures in both environments showed that the peptide is mostly α-helical. In the first phase of present research, the conformational and lipid-binding features of SP-BCTERM were studied in four other lipid micelle systems. These lipids, namely DPC, LPG, LPA and DHPC, are micelle-forming analogs of the physiological phospholipids PC and PG with which the native SP-B preferentially interacts. Thus the responses of SP-BCTERM in these lipid micelle systems are indicative of the interactions between SP-B and the physiological phospholipids.

SP-BCTERM caused an aggregation of micelles formed from each of the four lipids, as indicated by the reduced intensity of the NMR signals and confirmed (at least for DPC) by DLS. The NMR spectra also indicated that the peptide took on multiple conformations in DPC and LPG micelles, though CD analyses indicated that these multiple

conformations remained highly helical. The aggregation was most pronounced for LPA micelles. In DHPC, although aggregation was still present, SP-BCTERM took on a well-defined structure. These studies, together with the previous studies, showed that not all micelle types result in aggregation induced by SP-BCTERM in a similar manner. In particular, the peptide did not cause any noticeable aggregation of SGS micelles.

The findings from these investigations demonstrate the ability of SP-BCTERM to firmly bind certain lipid molecules analogous to the physiological phospholipids. Thereby, this biologically active fragment of SP-B confers on the full-length protein the ability to make large-scale lipid rearrangements and maintain the structures at alveolar air-water interface at various stages of lung surfactant operation.

In the second and main phase of present research, the high-resolution structure of Mini-B was determined in structure-promoting fluorinated organic solvent HFIP by solution NMR. Mini-B, which is an N-terminal – C-terminal construct of the native SP-B, was found to consist of two α-helices connected by an unstructured loop. The two helices correspond to the terminal helical regions of the full-length protein. These helices are amphiphilic in nature because of the localization of the polar and nonpolar residues on opposite sides. The two helices are cationic and contain five of the seven positively charged amino acids. Both the amphiphilic and cationic properties of Mini-B provide important clues on the lipid-protein interactions that are necessary to explain SP-B's functional mechanisms.

The HFIP-induced structure of Mini-B is the first high-resolution structure determined for this peptide. This structure will provide a platform for any future investigations on the conformational features of Mini-B in lipid mimetic environments. Since the peptide demonstrated significant biological activity, it likely consists of the most important regions of the full-length protein. Hence, its structure provides vital information on the structural features of the native SP-B and the mechanisms underlying the function.

SP-B is essential for life but the molecular basis for its activity is not yet understood. The current research will enhance the ongoing endeavor in defining the structural features that lead to SP-B's function. Knowledge of the molecular mechanisms involved in the function, in turn, would be of great utility in developing better artificial lung surfactants.

5.2 Future Work

The structures of SP-BCTERM were determined earlier in organic solvent HFIP and micelles composed from the detergent SDS. The first phase of this research explored the binding ability of this cationic peptide by inducing aggregation of anionic phospholipids LPG and LPA and zwitterionic phospholipid DPC. The peptide, however, took on a well-defined homogeneous conformation in another zwitterionic environment offered by the phospholipid DHPC. Thus, the DHPC micelle system appears to be a potential lipid mimetic for future structural studies of SP-B peptides.

The structure of Mini-B was determined in HFIP in the second and main phase of present research. Mini-B was synthesized without the disulfide bonds between the cysteines. Presently, work is ongoing to determine the structure of Mini-B with the disulfide bonds between the cysteines in the micelle system formed from SDS. Once this is done, structures of Mini-B in HFIP and SDS can be compared.

As the next step, the conformational features of Mini-B can be studied in lipid micelles that mimic the physiological environment of SP-B better than SDS, in particular DHPC. An exploration of the specific lipid binding ability of this peptide can be studied in DPC and LPG micelles as well. The concentrations of the phospholipids used to study SP-BCTERM were 150 mM in the present research. For the next study, noting the number of lipid molecules required to form a single micelle, the concentrations of peptide and

lipid can be chosen such that one peptide molecule be available per lipid micelle. In the case of aggregation, DLS experiments can be performed to determine the sizes of the aggregated complexes and to compare those with the sizes of the micelles without the peptide.

The final target of the ongoing research is determining the structure of the full-length SP-B. Hence after structural characterization of Mini-B, efforts can be taken to study a larger peptide encompassing more regions of SP-B. One such peptide, named Maxi-B, is in the process of being synthesized. Structural studies of this peptide, once it is available, can begin in HFIP environment to provide the first detailed structural characterization. Later on, more physiologically relevant structures can be determined in lipid micelle mimetics.

A clear understanding of SP-B's structural characteristics will help unveil the structural basis for its functional properties. This knowledge will contribute directly into the ongoing endeavor to obtain better surfactant replacement treatments.

Bibliography

[1] Possmayer F; Physicochemical aspects of pulmonary surfactant; In: Polin RA, Fox WW, eds.; Fetal and Neonatal Physiology; WB Saunders Company 1997; 1259-1275.

[2] Schürch S, Bachofen H, Goerke J and Green F; Surface properties of rat pulmonary surfactant studied with the captive bubble method: adsorption, hysteresis, stability; Biochim Biophys Acta 1992; 1103 : 127-136.

[3] von Neergaard K; Neue Auffassungen über einen Grundbegriff der Atemmechanik. Die Retraktionskraft der Lunge, abhangig von der Oberflachenspannung in den Alveolem; Z Gesamte Exp Med 1929; 66 : 373-394.

[4] Pattle R; Properties, function and origin of the alveolar lining layer; Nature 1955; 175 : 1125-1126.

[5] Clements J; Surface tensions of lung extracts; Proc Soc Exp Biol Med 1957; 95 : 170-172.

[6] Wright JR and Clements JA; Lung surfactant turnovers and factors that affect turnover; In: Massaro D, ed. Lung Cell Biology; New York: Marcel Dekker Inc.; 1989 : 655-699.

[7] Schmitz G and Muller G; Structure and function of lamellar bodies, lipid-protein complexes involved in storage and secretion of cellular lipids; Journal of Lipid Research 1991; 32 : 1539-1570.

[8] Goerke J; Pulmonary surfactant: functions and molecular composition; Biochim Biophys Acta 1998; 1408 : 79-89.

[9] Massaro D, Clerch L and Massaro GD; Surfactant secretion: evidence that cholinergic stimulation of secretion is indirect; Am J Physiol 1982; 243 : C39-C45.

[10] Wirtz H and Schmidt M; Ventilation and secretion of pulmonary surfactant; Clin Invest 1992; 70 : 3-13.

[11] Chander A and Fisher AB; Regulation of lung surfactant secretion; Am J Physiol 1990; 258 : L241-L253.

[12] Lumb RH; Phospholipid transfer proteins in mammalian lung; Am J Physiol 1989; 257 : L190-L194.

[13] Orgeig S, Daniels CB and Sullivan LC; The development of the pulmonary surfactant system; In: Harding R, Pinkerton K, Plopper C, eds; The lung: Development, Aging and the Environment, London: Academic Press, 2004; Chapter 10, 150-167.

[14] Johansson J and Curstedt T; Molecular structures and interactions of pulmonary surfactant components. Eur J Biochem 1997; 244 : 675-693.

[15] Schürch S, Qanbar R, Bachofen H and Possmayer F; The surface-associated surfactant reservoir in the alveolar lining; Biol Neonate 1995; 67 (suppl) : 61-76.

[16] Pérez-Gil J and Keough KMW; Interfacial properties of surfactant proteins; Biochim Biophys Acta 1998; 1408 : 203-217.

[17] Lang CJ, Daniels CB and Orgeig S; New insights into the thermal dynamics of the Surfactant system from warm and cold animals; Lung Surfactant Function and Disorder; Lung Biology in Health and Disease, Volume 201; 17-57.

[18] Pérez-Gil J; Molecular Interactions in pulmonary surfactant films; Biol Neonate 2002; 81 (suppl 1) : 6-15.

[19] Takamoto DY, Lipp MM, von Nahmen A, Lee KY, Waring AJ and Zasadzinski JA; Interaction of lung surfactant proteins with anionic phospholipids; J Biophys 2001; 81 : 153-169.

[20] Schwartz RM, Luby AM, Scanlon JW and Kellogg RJ; Effect of surfactant on morbidity, mortality, and resource use in newborn infants weighing 500 to 1500 g; N Engl J Med 1994; 330 : 1476-1480.

[21] Günther A, Ruppert C, Schmidt R, Markart P, Grimminger F, Walmrath D and Seegeer W; Surfactant alteration and replacement in acute respiratory distreess syndrome; Respir Res 2001; 2 : 353-364.

[22] Lewis JF and Veldhuizen R; The role of exogenous surfactant in the treatment of acute lung injury; Annu Rev Physiol 2003; 65 : 613-642.

[23] Veldhuizen R, Nag K, Orgeig S and Possmayer F; The role of lipids in pulmonary surfactant; Biochim Biophys Acta 1998; 140 8: 90-108.

[24] Mohwald H; Phospholipid monolayers: in Lipowski R, Sackmann E (eds.); Structure and dynamics of membranes: from cells to vesicles; Amsterdam, Elsevier 1995; 161-211.

[25] Notter RH, Finkelstein JN and Taubold RD; Comparative adsorption of natural lung surfactant, extracted phospholipids, and artificial phospholipid mixtures to the air-water interface; Chem Phys Lipids 1983; 33 : 67-80.

[26] Korfhagen TR, Bruno MD, Ross GF, Huclsman KM, Ikegami M, Jobe AH, Wert SE, Stripp BR, Morris RE, Glasser SW, Bachurski CJ, Iwamoto HS and Whitsett JA; Altered surfactant function and structure in SP-A gene targeted mice; Proc Natl Acad Sci USA 1996; 93 : 9594-9599.

[27] McCormack FX; Structure, processing and properties of surfactant protein A; Biochimica et Biophysica Acta 1998; 1408 : 109-131.

[28] Nogee LM, Garnier G, Dietz HC, Singer L, Murphy AM, deMello DE and Colten HR; A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds; J Clin Invest 1994; 93 : 1860-1863.

[29] Clark JC, Wert SE, Bachurski CJ, Stahlman MT, Stripp BR, Weaver TE and Whitsett JA; Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice; Proc Natl Acad Sci USA 1995; 92 : 7794-7798.

[30] Robertson B, Kobayashi T, Ganzuka M, Grossmann G, Li WZ and Suzuki Y; Experimental neonatal respiratory failure induced by a monoclonal antibody to the hydrophobic surfactant-associated protein SP-B; Pediatr Res 1991; 30 : 239-243.

[31] Glasser SW, Burhans MS, Korfhagen TR, Na CL, Sly PD, Ross GF, Ikegami M and Whitsett JA; Altered stability of pulmonary surfactant in SP-C-deficient mice; Proc Natl Acad Sci USA 2001; 98 : 6366-6371.

[32] Nogee LM, Dunbar AEIII, Wert SE, Askin F, Hamvas A and Whitsett JA; A mutation in the surfactant protein C gene associated with familial interstitial lung disease; N Engl J Med 2001; 344 : 573-579.

[33] Oosterlaken-Dijksterhuis MA, Haagsman HP, van Golde LMG and Demel RA; Characterization of lipid insertion into monomolecular layers mediated by lung surfactant proteins SP-B and SP-C. Biochemistry 1991; 30 : 10965-10971.

[34] Hawgood S, Benson BJ, Schilling J, Damm D, Clements JA and White RT; Proc Natl Acad Sci USA 1987; 84 : 66-70.

[35] Weaver TE and Conkright JJ; Functions of surfactant proteins B and C; Annu Rev Physiol 2001; 63 : 555-578.

[36] Hawgood S, Derrick M and Francis P; Structure and properties of surfactant protein B; Biochim Biophys Acta 1998; 1408 : 150-160.

[37] Ahn VE, Faull KF, Whitelegee JP, Fluharty AL and Prive GG; Crystal structure of Saposin B reveals a dimeric shell for lipid binding; Proc Natl Acad Sci USA 2003; 100 : 38-43.

[38] Hawkins CA, Alba E and Tjandra N; Solution structure of human saposin C in a detergent environment; J Mol Biol 2005; 346 : 1381-1392.

[39] Liepinsh E, Andersson M, Ruysschaert JM and Otting G; Saposin fold revealed by the NMR structure of NK-Lysin; Nat Struct Biol 1997; 4 : 793-795.

[40] Hecht O, Nuland NV, Schleinkofer K, Dingley AJ, Bruhn H, Leippe M and Grotzinger J; Solution structure of the pore-forming protein of Entamoeba Histolytica; J Biol Chem 2004; 279 : 17834-17841.

[41] Kurtz JW and Lee KYC; NMR structure of lung surfactant peptide SP-B_{11-25}; Biochemistry 2002; 41 : 9627-9636.

[42] Booth V, Waring AJ, Walther FJ and Keough KMW; NMR structures of the C-terminal segment of surfactant protein B in detergent micelles and hexafluoro-2-propanol; Biochemistry 2004; 43 (48) : 15187-15194.

[43] Kuchel PW and Ralston GB; Schaum's Outlines Biochemistry Second Edition; Tata McGraw-Hill Publishing Company Limited, New Delhi; 2003.

[44] Petsko GA and Ringe D; Protein Structure and Function; Primers in Biology, New Science Press Ltd; 2004.

[45] Cavanagh J, Fairbrother WJ, Palmer III AG and Skelton NJ; Protein NMR Spectroscopy Principles and Practice; Academic Press, Inc.; 1996.

[46] Branden C and Tooze J; Introduction to protein Structure; Garland Publishing, Inc.; 1991.

[47] James TL; Chapter 1 - Fundamentals of NMR; Department of Pharmaceutical Chemistry, University of California, San Francisco; http://www.biophysics.org/education/james.pdf.

[48] Booth VK; Structure and Function of Transcription Elongation Factor TFIIS and Methanobacterium thermoautotrophicum Protein 1615; PhD Thesis 2000; Department of Medical Biophysics, University of Toronto.

[49] Hornak JP; The Basics of NMR; http://www.cis.rit.edu/htbooks/nmr/bnmr.htm

[50] Wishart DS and Sykes BD; The 13C chemical shift index. A simple method for the identification of protein secondary structure using 13C chemical shift data; J Biol NMR 1993; 4 : 171-175.

[51] Wüthrich K; NMR of Proteins and Nucleic Acids; John Wiley & Sons; 1986.

[52] Renner C; Bioorganic NMR; http://www.biochem.mpg.de/moroder/nmr/main_engl.html.

[53] Cornilescu G, Delaglio F and Bax A; Protein backbone angle restraints from searching a database for chemical shift and sequence homology; J Biomol NMR 1999; 13 : 289-302.

[54] Fesik SW and Zuiderweg ERP; Quart Rev Biophys 1990; 23 : 97-131.

[55] Clore GM and Gronenborn AM; Prog NMR Spectrosc 1991; 23 : 43-92.

[56] Griesinger C, Sørensen OW and Ernst RR; J Magn Reson 1989; 84 : 14-63.

[57] Brünger AT; X-PLOR Version 3.1 : A system for X-ray Crystallography and NMR; Yale University, New Haven 1992.

[58] Schirra HJ; Structure Determination of Proteins with NMR Spectroscopy; http://www.cryst.bbk.ac.uk/PPS2/projects/schirra/html/home.htm

[59] Anmin T, Ziegler A, Steinbauer B and Seelig J; Thermodynamics of Sodium Dodecyl Sulfate Partitioning into Lipid Membranes; Biophysical Journal 2002; 83 (3) : 1547-1556.

[60] Bonard JM, Stora T, Salvetat JP, Maier F, Stocki T, Duschl C, Forro L, De Heer WA and Chatelain A; Purification and size selection of carbon nanotubes; Advanced Materials 1997; 9 (10) : 827-831.

[61] Marrink SJ, Tieleman DP and Mark AE; Molecular Dynamics Simulation of the Kinetics of Spontaneous Micelle Formation; Journal of Physical Chemistry 2000; 104 : 12165-12173.

[62] Clayton D, Brereton IM, Kroon PA and Smith R; NMR Studies of the Low Density Lipoprotein Receptor Binding Peptide of Apolipoprotein E bound to Dodecylphosphocholine Micelles; Protein Science 1999; 8 : 1797-1805.

[63] Chou JJ, Baber JL and Bax A; Characterization of Phospholipid Mixed Micelles by Translational Diffusion; Journal of Biomolecular NMR 2004; 29 (3) : 299-308.

[64] Perugini MA, Schuck P and Howlett GJ; Self Association of Human Apolipoprotein E3 and E4 in the Presence and Absence of Phospholipid; Journal of Biological Chemistry 2000; 275 (47) : 36758-36765.

[65] Li Z, Mintzer E and Bittman R; The Critical Micelle Conctrations of Lysophosphatidic Acid and Spingosylphosphorylcholine; Chemistry and Physics of Lipids 2004; 130 : 197-201.

[66] Delaglio F; NMRPipe 2.2; NIH Laboratory of Chemical Physics, NIDDK.

[67] Pérez-Gil J, Cruz A and Casals C; Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in aqueous organic solvents and lipids; Biochim Biophys Acta 1993; 1168 : 261-270.

[68] Morrow MR, Pérez-Gil J, Simatos G, Boland C, Stewart J, Absolom D, Sarin V and Keough KM; Pulmonary surfactant-associated protein SP-B has little effect on acyl chains in dipalmitoylphosphatidylcholine dispersions; Biochemistry 1993; 32 : 4397-4402.

[69] Pérez-Gil J, Casals C and Marsh D; Interactions of hydrophobic lung surfactant proteins SP-B and SP-C with dipalmitoylphosphatidylcholine and dipalmitoylphosphatidylglycerol bilayers studied by electron spin resonance spectroscopy; Biochemistry 1995; 34 : 3964-3971.

[70] Rodriguez-Capote K, Nag K, Schurch S and Possmayer F. Surfactant protein interactions with neutral and acidic phospholipid films; Am J Physiol Lung Cell Mol Physiol 2001; 281 : L231-L242.

[71] Schram V and Hall SB; Thermodynamic effects of the hydrophobic surfactant proteins on the early adsorption of pulmonary surfactant; Biophys J 2001; 81 : 1536-1546.

[72] Nag K, Munro JG, Inchley K, Schurch S, Peterson NO and Possmayer F; SP-B refining of pulmonary surfactant phospholipid films; Am J Physiol 1999; 277 : L1179-L1189.

[73] Suzuki Y, Fujita Y and Kogishi K; Reconstitution of tubular myelin from synthetic lipids and proteins associated with pig pulmonary surfactant; Am Rev Respir Dis 1989; 140 : 75-81.

[74] Williams MC, Hawgood S and Hamilton RL; Changes in lipid structure produced by surfactant proteins SP-A, SP-B, and SP-C; Am J Respir Cell Mol Biol 1991; 5 : 41-50.

[75] Zaltash S, Palmblad M, Curstedt T, Johansson J and Persson B; Pulmonary surfactant protein B: a structural model and a functional analogue; Biochim Biophys Acta 2000; 1466 : 179-186.

[76] Waring AJ, Walther FJ, Gordon LM, Hernandez-Juviel J, Hong T, Sherman MA, Alonso C, Alig T, Braun A, Bacon D and Zasadzinski JA; The role of charged amphipathic helices in the structure and function of surfactant protein B; J Pept Res; online publication date 7 Oct 2005.

[77] Goddard TD and Kneller DG; Sparky 3; University of California, San Francisco, USA.

[78] Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T and Warren GL; Crystallography & NMR System (CNS) CNSsolve, Version: 1.1; Yale University.

[79] Koradi R; MOLMOL 2K.2; MOLMOL Manual; Institut für Molekularbiologie und Biophysik; ETH-Hönggerberg, Zürich.

[80] Billeter M, Braun W and Wüthrich; Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra : Computation of sterically allowed proton-proton distances and statistical analysis of proton-proton distances in single crystal protein conformations; J Mol Biol 1982; 155 : 321-346.

[81] Alonso C, Alig T, Bacon D, Hong T, Hernandez-Juviel J, Gordon LM, Walther FJ, Sherman MA, Waring AJ and Zasadzinski JA; Conformational mapping of Mini-b: an N-terminal/C-terminal construct of Surfactant Protein B using 13c-enhanced Fourier Transform Infrared (FTIR) Spectroscopy; To be published; PDB ID : 1SSZ.

[82] Merutka G, Dyson HJ and Wright PE; "Random coil" ^{1}H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG; J Biomol NMR 1995; 5 : 14-24.

[83] Lipp MM, Lee KY, Waring A and Zasadzinski JA; Fluorescence, polarized fluorescence, and Brewster angle microscopy of palmitic acid and lung surfactant protein B monolayers; Biophys J 1997; 72 : 2783-2804.

[84] Longo MM, Waring A and Zasadzinski JA; Lipid bilayer surface association of lung surfactant protein SP-B, amphipathic segment detected by flow immunofluorescence;
Biophys J 1992; 63 : 760-773.

[85] Waring, A, Taeusch W, Bruni R, Amirkhanian J, Fan B, Stevens R and Young J; Synthetic amphipathic sequences of surfactant protein-B mimic several physicochemical and *in vivo* properties of native pulmonary surfactant proteins; Peptide Res 1989; 2 : 308-313.

[86] Baatz JE, SarinV, Absolom DR, Baxter C and Whitsett JA; Effects of surfactant-associated protein SP-B synthetic analogs on the structure and surface activity of model membrane bilayers; Chem Phys Lipids 1991; 60 : 163-178.

[87] Cochrane CG and Revak SD; Pulmonary Surfactant Protein B (SP-B): Structure-Function Relationships; Science 1991; 254 : 566-568.

[88] Pérez-Gil J, Cruz A and Plasencia I; Structure-function relationships of hydrophobic proteins SP-B and SP-C in pulmonary surfactant; Lung Surfactant Function and Disorder; Lung Biology in Health and Disease, Volume 201; 125-142.

[89] Andersson M, Curstedt T, Jornvall H and Johansson J; An amphipathic helical motif common to tumourolytic polypeptide NK-lysin and pulmonary surfactant polypeptide SP-B; FEBS Lett 1995; 362 : 328-332.

[90] Chiti F, Taddei N, Webstar P, Hamada D, Fiaschi T, Ramponi G and Dobson CM; Acceleration of the folding of acylphopsphatase by stabilization of local secondary structure; Nat Struct Biol 1999; 6 : 380-387.

[91] Gast K, Zirwer D, Muller-Frohne M and Damaschun G; Trifluoroethanol-induced conformational transitions of proteins: insights gained from the differences between α-lactalbumin and ribonuclease A; Protein Sci 1999; 8 : 625-634.