UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Structural Studies of Proteins Involved in Diseases of Protein Deposition and a Protein Involved in Nitrogenase Assembly

Permalink

https://escholarship.org/uc/item/4rq0q454

Author

Phillips, Aaron Hale

Publication Date

2010

Peer reviewed|Thesis/dissertation

Structural Studies of Proteins Involved in Diseases of Protein Deposition and

a Protein Involved in Nitrogenase Assembly

by

Aaron Hale Phillips

A thesis submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David E. Wemmer

Professor Alex Pines

Professor Susan Marqusee

Spring 2010

Abstract

Structural Studies of Proteins Involved in Diseases of Protein Deposition and

a Protein Involved in Nitrogenase Assembly

by

Aaron Hale Phillips

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor David E. Wemmer, Chair

The work presented in this thesis describes structural investigations of proteins implicated in diseases of protein aggregation and a protein involved in the maturation of the nitrogenase enzyme complex. Each experimental chapter is full self-contained.

Protein aggregation into plaques has been identified as a central event in many human diseases such as Alzheimer's disease, type II diabetes, and the spongiform encephalopathies. Generally, the protein plaques display a common morphology consisting of long unbranched proteinaceous fibrils commonly referred to as amyloid. The intrinsically insoluble nature of proteins contained within amyloid fibrils has greatly impeded high resolution structural studies utilizing single crystal X-ray diffraction and high-resolution liquid NMR spectroscopy. Amide hydrogen exchange is often used to probe hydrogen bonding in proteins and does not strictly require soluble protein for study. In the first chapter of the thesis, I present an amide hydrogen exchange study conducted on an amyloid forming fragment of the protein implicated in the spongiform encephalopathies, PrP(89-143). These experiments identify the strongly hydrogen bonded fibrillar core.

The next section of the thesis contains work conducted on Aβ, the amyloid forming protein implicated in Alzheimer's disease. Due to increasing lines of evidence that the Aβ fibrils themselves do not cause neurodegeneration, much effort has been directed into studying the monomeric and pre-fibrillar oligomeric states of Aβ. This work is greatly impeded by both Aβ's strong propensity to aggregate as well as the inherent heterogeneity of the molecules in solution. In order to circumvent these issues the experiments presented herein are interpreted with respect to models of the structural ensemble obtained via molecular dynamics calculations, thereby yielding a model of the structural ensemble that is validated to an extent unavailable to either technique alone.

The final section of this thesis describes work conducted on the γ-subunit of nitrogenase, the enzyme complex responsible for the entry of nitrogen into the biosphere. The γ-subunit's

poposed role in the cell is to deliver the cofactor present at the active site of nitrogenase that is required for activity. I present biochemical experiments that demonstrate that the N-terminal domain of the γ-subunit is responsible for mediating the interaction between the γ-subunit and apo-nitrogenase, the solution structure of N-terminal domain of the γ-subunit, and NMR experiments that characterize the cofactor binding site on the C-terminal domain of the γ-subunit.

Acknowledgements

Table of Contents

Chapter 1

Preface

The work presented in this thesis describes structural investigations of proteins implicated in diseases of protein aggregation, and a protein involved in the maturation of the nitrogenase enzyme complex. Each of the three experimental chapters is fully self-contained.

Protein aggregation into plaques has been identified as a central event in many human diseases. These include Alzheimer's disease, type II diabetes, the spongiform encephalopathies, Parkinson's disease and Huntington's disease and can generally be categorized as diseases of protein deposition. With the exception of the spongiform encephalopathies, all of the plaques associated with these disorders are composed of long unbranched proteinaceous fibrils commonly referred to as amyloid (although the fibrils involved in Parkinson's and Huntington's disease are present within the intracellular space and therefore do not meet the most stringent definition of amyloid) [1, 2]. Upon treatment with proteases, plaques obtained from patients with spongiform encephalopathy readily form amyloid rods [3]. In addition to their common morphology the fibrils associated with all of these diseases all seem to have a common 'cross-β' architecture composed of β-strands running perpendicular to the axis of the fibril as characterized by X-ray fiber diffraction [4]. In recent years it has been shown that many proteins not known to form amyloid *in vivo* can form amyloid fibrils under non-native conditions, thus it has been proposed that the ability to polymerize into amyloid may be a general feature of polypeptides [5, 6]. Interestingly, the X-ray diffraction pattern of cooked egg whites shows the same characteristic cross-β diffraction pattern [7]. Due to both their prevalence in human disease and the intriguing notion that the ability to form amyloid fibrils may be a general aspect of polypeptide chains, understanding the structure of proteins within amyloid fibrils has been an intense area of research over the past few decades. The intrinsically insoluble nature of proteins contained within amyloid fibrils has greatly impeded high resolution structural studies utilizing single crystal X-ray diffraction and high-resolution liquid NMR spectroscopy, the common workhorses of structural biology. Amide hydrogen exchange is used to probe hydrogen bonding in proteins and has been used to probe protein conformation in several fibril forming peptides [8-10]. In chapter 2 of this thesis I present amide hydrogen exchange experiments on amyloid fibrils of a fragment of the protein implicated in the spongiform encephalopathies, PrP(89-143) [11]. The experiments presented here identify the location of the fibrillar core.

Chapter 3 of this thesis contains work conducted on Aβ, the amyloid fibril forming protein implicated in Alzheimer's disease. Due to an increasing number of lines of evidence that the plaque burden in the brains of patients affected with Alzheimer's disease does not correlate with the severity of disease symptoms, much effort has been directed into studying the monomeric and pre-fibrillar oligomeric states of Aβ [12-14]. The study of monomeric and oligomeric Aβ is impeded by the peptide's strong propensity to aggregate as well as the inherent heterogeneity of the molecules in solution, both with respect to peptide conformation and the specific stoichiometry of the oligomeric state. My coworkers and I have addressed this issue by using molecular dynamics calculations to generate models for conformational ensembles of Aβ and fragments, then using these ensembles to predict NMR measurables which are subsequently

compared to their experimental values thereby validating the model of the structural ensemble to an extent unavailable to either technique alone.

Chapter 4 describes studies on the γ-subunit of nitrogenase, the enzyme that, prior to the introduction of industrial fertilizers, was the central node for the entry of nitrogen into the biosphere. The nitrogenase enzyme complex consists of two proteins: dinitrogenase, which is responsible for the reduction of N_2 and dinitrogenase reductase, which provides electrons to dinitrogenase [15]. Dinitrogenase is an $\alpha_2\beta_2$ tetramer that contains four metalloclusters: two P-clusters and two copies of FeMo-co, present at the active site of the enzyme. FeMo-co contains iron, sulfur, molybdenum and *R*-homocitrate and is synthesized by the cell in a tightly regulated process [16]. Inactive dinitrogenase obtained from mutant strains of diazotrophic bacteria that cannot synthesize FeMo-co purifies as an $\alpha_2\beta_2\gamma_2$ hexamer. Upon the addition of FeMo-co *in vitro* the γ-subunit (NafY) dissociates from dinitrogenase leaving fully active dinitrogenase [17]. NafY is known to bind one molecule of FeMo-co tightly and has been proposed to function as a FeMo-co chaperone/insertase [18, 19]. In chapter 4 I present work that further refines the cellular function of NafY, including determining the solution structure of the N-terminal domain and characterizing the FeMo-co binding site on the C-terminal domain.

References

1.   Westermark, P., *Aspects on human amyloid forms and their fibril polypeptides.* FEBS Journal, 2005. **272**: p. 5942-5949.
2.   Eisenberg, D., et al., *The Structural Biology of Protein Aggregation Diseases: Fundamental Questions and Some Answers.* Accounts of Chemical Research, 2006. **39**: p. 568-575.
3.   Prusiner, S., et al., *Scrapie prions aggregate to form amyloid-like birefringent rods.* Cell, 1983. **35**: p. 349-358.
4.   Eanes, E.D. and G.G. Glenner, *X-ray diffraction studies on amyloid filaments.* Journal of Histochemistry and Cytochemistry, 1968. **16**: p. 673-677.
5.   Fändrich, M., M.A. Fletcher, and C.M. Dobson, *Amyloid fibrils from muscle myoglobin.* Nature, 2001. **410**: p. 165-166.
6.   Guijarro, J.I., et al., *Amyloid fibril formation by an SH3 domain.* Proceeding of the National Academy of Sciences, 1998. **95**: p. 4224-4228.
7.   Astbury, W.T., S. Dickinson, and K. Bailey, *The X-ray interpretation of denaturation and the structure of the seed globulins.* Biochemical Journal, 1935. **29**: p. 2351-2360.
8.   Ippel, J.H., et al., *Probing solvent accessibility of amyloid fibrils by solution NMR spectroscopy.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8648-8653.
9.   Hoshino, M., et al., *Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange.* Nature Structural Biology, 2002. **9**(5): p. 332-336.
10.   Del Mar, C., et al., *Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15477-15482.
11.   Prusiner, S.B., *Novel Proteinaceous Infectious Particles Cause Scrapie.* Science, 1982. **216**: p. 136-144.

12. Kayed, R., et al., *Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis.* Science, 2003. **300**(486-489).

13. Mucke, L., et al., *Highlevel neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation.* The Journal of Neuroscience, 2000. **20**: p. 4050–4058.

14. Walsh, D.M. and D.J. Selkoe, *Aß Oligomers – a decade of discovery.* Journal of Neurochemistry, 2007. **101**: p. 1172-1184.

15. Bulen, W.A. and J.R. Lecomte, *The Nitrogenase System From Azotobacter: Two-Enzyme Requirement for N2 Reduction, ATP-Dependent H2 Evolution, and ATP Hydrolysis.* Proceedings of the National Academy of Sciences, 1966. **56**: p. 979-986.

16. Rubio, L.M. and P.W. Ludden, *Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase.* Annual Reviews of Microbiology, 2008. **62**: p. 93-111.

17. Paustian, T.D., V.K. Shah, and G.P. Roberts, *Apodinitrogenase: Purification, Association with a 20-Kilodalton Protein, and Activation by the Iron-Molybdenum Cofactor in the Absence of Dinitrogenase Reductase.* Biochemistry, 1990. **1990**: p. 3515-3522.

18. Dyer, D.H., et al., *The Three-dimensional Structure of the Core Domain of NafY from Azotobacter vinelandii determined at 1.8-Å Resolution.* The Journal of Biological Chemistry, 2003. **278**(34): p. 32150-32156.

19. Rubio, L.M., S.W. Singer, and P.W. Ludden, *Purification and Characterization of NafY (Apodinitrogenase Gamma Subunit) from Azotobacter vinelandii.* The Journal of Biological Chemistry, 2004. **279**: p. 19739-19746.

Chapter 2

Probing the Conformation of a Prion Protein Fibril with Hydrogen Exchange

2.1 Summary

A fragment of the prion protein, PrP(89-143, P101L), bearing a mutation implicated in familial prion disease, forms fibrils that have been shown to induce prion disease when injected intracerebrally into transgenic mice expressing full-length PrP containing the P101L mutation [1]. In this study amide hydrogen exchange measurements were used to probe the organization of the peptide in its fibrillar form. The extent of hydrogen exchange at each time point was determined first by tandem proteolysis, liquid chromatography and mass spectrometry (HXMS) and then by exchange-quenched NMR. While single amide resolution is afforded by NMR measurements, HXMS is well suited to the study of natural prions because it does not require labeling with NMR active isotopes. Thus, natural prions obtained from infected animals can be compared to model systems such as PrP(89-143, P101L) studied here. This work reveals two segments of sequence that display a high level of protection from exchange, residues 102-109 and 117-136. There is in addition a region that displays exchange behavior consistent with the presence of a conformationally heterogeneous turn. The data are discussed with respect to several structural models proposed for infectious PrP aggregates.

2.2 Introduction

Prion diseases lead to neurodegenerative processes that result in spongiform degeneration and astrocytic gliosis in the central nervous system and are known to exist in genetic, sporadic and transmissible forms [2]. The agent that is implicated in initiating disease is the protein PrP, which can exist in two isoforms: the benign, native form PrP^C (PrP Cellular) and the infectious and neurotoxic form PrP^{Sc} (PrP Scrapie) [3]. The idea that a protein alone could transmit disease pathology was initially greeted with skepticism but is now generally accepted through the accumulation of a large body of evidence, culminating in the generation of disease symptoms in mice by inoculation with PrP aggregates formed from recombinant protein purified from *E. coli* [4]. PrP is found in all mammalian and avian genomes and, after maturation, consists of approximately two hundred residues with one disulfide bridge, two glycosylation sites, and is attached to the membrane via a C-terminal glycosylphosphatidylinositol (GPI) anchor. Solution structures of PrP^C, without post-translational modifications, from several organisms have been solved, showing it to be largely α-helical with an unstructured N-terminus of approximately one hundred residues [5-10]. Experiments have shown that PrP binds copper and may be involved in copper homeostasis [11-13]; however, PrP's precise biochemical functions remain unclear. For a detailed discussion see Aguzzi's review [14]. In contrast to the mostly α-helical, soluble PrP^C, PrP^{Sc} is β-rich, with a protease resistant core, and is present as an aggregate [15]. When PrP^{Sc} is proteolytically cleaved at its N-terminus, it readily associates to form fibrils or rods that retain

4

high levels of infectivity [16]. Both electron microscopy and molecular mechanics calculations have been used to generate model structures for infectious PrP [17-20]; however, due to their propensity to aggregate and intrinsic insoluble nature, it has been difficult to obtain any high resolution structural information on infectious PrP with which a single model can be validated.

Determining rates at which amide hydrogens exchange with solvent hydrogens provides a useful probe for hydrogen bonding in proteins. Exchange requires contact between the amide and solvent, and catalysis by an acid or base (which dominates at neutral pH). Amides involved in hydrogen bonds thus cannot exchange without first fluctuating to an 'open' state. Simple kinetic theories show that in the EX2 limit and at neutral pH rates of exchange are determined by the equilibrium constant for opening and the intrinsic rate of exchange of the amide (determined by the identity of the amino acid and its sequence context) [21]. Intrinsic exchange rates have been tabulated [22], so that observed exchange rates can be interpreted in terms of opening equilibria. At pH 7 and 25 °C, the intrinsic fully exposed amide hydrogen exchange rate is on the order of $10s^{-1}$. Recently, numerous groups have employed hydrogen exchange based experiments to study amyloid fibrils formed from a number of different proteins [23-32]. Interestingly, in each case it was shown that only subsets of residues in the protein were protected from exchange with solvent, with rather different patterns of protected residues for different proteins. The remaining residues exhibited rapid exchange behavior characteristic of exposed amide groups.

Work from Prusiner's lab to reconstitute prion infectivity from a synthetic protein identified a 55-residue PrP fragment, PrP(89-143, P101L), that readily associated into amyloid fibrils. Intracerebral injection of the aggregated fibrillar peptide initiated disease in mice that express full length PrP bearing the sensitizing mutation, identified in a familial form of prion disease, while the non-aggregated peptide did not initiate disease [1]. The amide hydrogen exchange behavior in the fibrillar form of PrP(89-143, P101L) was obtained first at medium sequence resolution with tandem proteolysis, liquid chromatography and mass spectrometry (HXMS) and then at single amide resolution with NMR spectroscopy. The HXMS technique has been described in detail in the literature previously [33]. It is clear that there are regions of the peptide which are highly protected from exchange and others that are not. The strongly protected regions identify the residues in the hydrogen bonded core of the amyloid fibril.

2.3 Materials and Methods

2.3.1 Amyloid Fibril Formation

Amyloid fibrils of PrP(89-143, P101L) were formed by dissolving 10 mg of peptide into 500 µL of 20 mM sodium acetate, 100 mM NaCl, pH 5 and adding 500 µL of acetonitrile. Samples were taken at several time points, and EM images were collected. Small fibrils were seen even a short time after preparing the sample. At longer times primarily longer fibrils with more clearly defined edges were seen. Fibrils used in hydrogen exchange experiments were collected after three weeks of incubation in the acetonitrile solution.

2.3.2 Electron Microscopy

Samples were loaded onto 1000 mesh copper grids, coated with formvar and glow discharged prior to use. The samples were negatively stained with 3% aqueous uranyl-acetate and images were collected on a FEI Tecnai 12 operating at 120 kV.

2.3.3 Hydrogen Exchange Mass Spectrometry

PrP(89-143, P101L) peptide used in MS analysis was synthesized using standard Fmoc chemistry and subsequent cleavage with trifluoracetic acid. PrP(89-143, P101L) fibrils were collected by centrifugation at 1300g in an Eppendorf tube and washed twice with 10mM sodium phosphate, pH 7.5. Deuterium exchange was initiated by resuspending the fibrils in 100% D_2O buffered with 10mM sodium phosphate at a pH* of 7.5 (pH* is the reading of a pH meter, uncorrected for isotope and solvent effects) to a final peptide concentration of 50 mg/mL. 5 µL aliquots of the peptide suspension were removed after 1 hour, 6 hours, 21 hours, 1 week, 3 weeks and 6 weeks. 5 µL of the peptide slurry was removed before exposure to deuterium to serve as the zero hour time point. Exchange samples were prepared for MS analysis by dilution with 55 µL of ice-cold 10 mM sodium phosphate buffer and rapid addition of 90 µL of quench buffer (6.8 M guanidine hydrochloride (GuHCl), 16.6% glycerol and 0.8% formic acid), yielding a final pH* of 2.1, followed by 5 minutes of incubation over thawing ice to fully dissolve the fibrils. Each solution was then divided into 50 µL aliquots in individual autosampler vials and rapidly frozen over powdered dry ice. Samples were stored under powdered dry ice in -80 ℃ freezers.

The extent of deuteration was determined by quickly thawing the frozen samples over melting ice and passing the solution through two immobilized protease columns (pepsin and fungal protease XIII from Sigma, 100 µL/min, 0.05% TFA). Peptides were separated by reverse-phase high pressure liquid chromatography (HPLC) (Vydac C-18 300A, 50 µL/min with a 6.4-38.4% acetonitrile gradient over 30 min, 4 ℃). The effluent was directly injected into a mass spectrometer (ThermoFinnigan LCQ, capillary temperature of 200 °C). Initial identification of proteolytic fragments was determined by tandem MS followed by data analysis using the SEQUEST and DXMS software programs (ThermoFinnigan, San Jose, CA and Sierra Analytics, Modesto, CA; respectively) [34, 35]. Peptide deuteration levels were extracted from the centroid of individual mass-envelopes according to the method of Zhang and Smith [36] after accounting for end effects and proline residues [22]. Peptide deuteration levels were corrected for back exchange during proteolysis and chromatography based on their retention time in the chromatography step. Back exchange at the beginning of chromatography was found to be 15% and 20% at the end according to fully deuterated reference protein samples.

2.3.4 Cloning, Expression and Purification

A synthetic gene for PrP(89-143, P101L) with an N-terminal six residue histidine tag and TEV cleavage site was designed based on *E. coli* optimum codon usage and synthesized by Genscript (Piscataway, NJ) in a pUC19 plasmid. This gene was subcloned into the pIVEX 2.4c vector between the NdeI and BamHI sites.

E. coli cells [BL21 (DE3)] containing a plasmid, pACYC, that constitutively expresses LacI were transformed using the hexaminecobalt chloride method [37] and grown on MDG [38] agar plates (1.5%) containg ampicillin and kanamyacin at 37 ℃ for 12 hours. A single colony was then picked and inoculated into 2 mL of MDAG media [38] (100 µg/mL ampicillin, 50 µg/mL kanamyacin) and allowed to grow at 37 ℃ for 6 hours. To obtain ^{15}N labeled protein, one mL of this starter culture was used to inoculate 1L of N-5052 media [38] contained in a 2.8L Fernbach flask. Cells were allowed to grow for 48 hours at 37 ℃ with shaking at 240 rpm. Cells expressing ^{13}C, ^{15}N labeled protein were obtained according to the method of Marley et al. [39] with the substitution of MDAG media for LB.

The cells were recovered by centrifugation at 5,000 rpm for 10 minutes and resuspended in ca. 40 mL of 50 mM KPi, pH 8.0, and lysed by sonication at 4 ℃. The resulting suspension

was centrifuged for 30 min at 30,000 rpm. The supernatant was discarded and the inclusion body was resolubilized by applying three rounds of resuspension in 6 M GuHCl, 200 mM NaPi pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol followed by centrifugation for 30 minutes at 30,000 rpm. The supernatants were pooled together and loaded onto Ni-NTA resin (6 mL). The column was then washed with a linear gradient from 6 to 0 M GuHCl, 50 mM NaP_i pH 8.0, 300 mM NaCl and 15 mM imidazole. The protein was eluted from the column with 50 mM NaP_i pH 8.0, 300 mM NaCl and 500 mM imidazole then dialyzed extensively against 50 mM Tris pH 8.0, 1 mM DTT, 1 mM EDTA at 4 ℃. One milligram of TEV protease was added to the dialysis bag, and the cleavage reaction was monitored by HPLC. After complete proteolysis to remove the His-tag, the solution was dialyzed against 50 mM NaP_i pH 8.0, 300 mM NaCl and 15 mM imidazole then passed over Ni-NTA resin to remove both the His-tag fragment and TEV and desalted using a C18 sep-pak cartridge (Waters). The solution was frozen and lyophilized. Dry peptide was stored in a desiccator. On average, 15 mg of purified ^{15}N labeled PrP (89-143, P101L) was obtained per 1 L growth. 0.5 mg of purified ^{13}C, ^{15}N labeled PrP (89-143, P101L) was obtained from 0.5L of M9 media.

2.3.5 NMR experiments (assignments)

      All data were collected on a Bruker Avance DRX-500 or Avance DRX-800 spectrometers equipped with a cryogenic probe, with samples kept at 25 ℃. A single ^{15}N, ^{13}C labeled sample of PrP(89-143, P101L) consisting of 0.5 mg of protein dissolved in 500 μL of DMSO (5% H2O, 0.03% TFA) was used for backbone assignments. Assignment of the [1]H[15]N HSQC spectrum was accomplished via interresidue connection of the Cα, Cβ, and N chemical shifts using the HNCA, CBCA(CO)NH, and HNCACB experiments [40-42]. Data were processed with NMRPipe [43]. For the 3D experiments, data were linear predicted to 128 pts in the ^{15}N dimension. Analysis and integration of data were performed with NMRView or CARA [44, 45].

2.3.6 Hydrogen Exchange NMR Spectroscopy

      Fibrillization of ^{15}N labeled PrP(89-143, P101L) was performed as outlined above. Approximately 0.5 mg of fibrillized ^{15}N labeled PrP(89-143, P101L) was aliquotted into a separate Eppendorf tube for each exchange time point. Fibrils were resuspended in 10mM KPi pD 7.5 and incubated at room temperature for varying periods of time. At the end of each time point, the fibrils were pelleted by centrifugation and washed twice with D_2O at 4 ℃. After removing the supernatant, the fibrils were frozen over dry ice and lyophilized. Lyophilized fibrils were dissolved in DMSO containing 5% D_2O and 0.03% TFA-D, pH* 5. A series of [1]H[15]N HSQCs was collected over 8 hours to correct for exchange with solvent under the quenching NMR sample conditions [25]. The dead time from the addition of quenching/denaturing solvent to the fibrils to the acquisition of the first time point of the 2D experiment was approximately 10 minutes (allowing for mixing to aid fibril dissolution and optimizing parameters for the NMR experiment). Each [1]H[15]N HSQC signal was the average of 4 transients, with 128 complex points acquired in the t_1 dimension, for a total acquisition time of 20 minutes per 2D data set.

      The intrinsic exchange rate in DMSO was determined by first integrating the volumes for each peak in each [1]H[15]N HSQC spectrum acquired over the 8 hours. These individual volumes vs. time were then plotted and fitted to single exponentials. The calculated rates from the exponential fits were then used to account for exchange during the dead time of the experiment.

The intrinsic exchange rates are shown in table 2.1. A 1-1-echo one-dimensional ^{1}H NMR spectrum was acquired for each sample and the methyl region of the spectrum integrated to normalize sample concentrations, which did not deviate by more than 3% between samples. Hydrogen exchange incubation times were 0 hours, 1 hour, 6 hours, 21 hours, 1 week and 6 weeks.

2.4 Results

2.4.1 Hydrogen Exchange Mass Spectrometry of PrP(89-143, P101L) fibrils

Synthetic PrP(89-143, P101L) fibrils that had been grown in H_2O were incubated in D_2O (buffered with sodium phosphate at pD 7.5) and aliquots removed after incubation for 1 hour, 6 hours, 21 hours, 1 week, 3 weeks, and 6 weeks. In order to determine the deuterium content of the fibrils using electrospay ionization mass spectrometry the fibrils were redissolved by the addition of an aqueous buffer containing guanidine hydrochloride at low pH to quench hydrogen exchange, limiting back exchange during the mass analysis. To identify sites retaining deuterium in the primary sequence, exchange quenched peptide solutions were fragmented using two protease columns containing immobilized pepsin and fungal protease XIII. Both proteases retain activity under quenching conditions. The peptide fragments were then separated by reverse phase HPLC with the eluent running directly into the mass spectrometer. The ability to resolve slowly exchanging sites depends on the degree of fragmentation obtained, which in optimal circumstances can discriminate single amides. Peptides were identified in a reference experiment in which non-deuterated PrP(89-143, P101L) was proteolyzed in the same manner and eluting fragments were then analyzed by tandem MS [46]. Peptides in the deuterated samples were then identified by matching both mass and retention time.

Figure 2.1 shows the extent of amide exchange of PrP(89-143, P101L) in fibrils after incubation in D_2O for six weeks observed using mass spectrometry. 34 'primary' peptide fragments were obtained with good signal to noise ratios in repeated experiments at each time point studied. Exchange in 23 additional 'secondary' peptide segments could be analyzed via deuterium-level subtraction of N-terminally aligned peptides of different length; N-terminal alignment of peptides is required to correct for end effects. Although PrP(89-143, P101L) was not proteolyzed sufficiently before mass spectrometry to afford single site resolution, a consensus exchange behavior can be approximated by calculating an average of the exchange behavior in which the contribution from each observed peptide is weighted according to the information content of the peptide. The longer the peptide, the less information for localization of exchange it contains. The averaging is performed according to

$$ HX = \sum_{i=1}^{n} \left(\frac{hx_i}{length_i} \right) \div \sum_{i=1}^{n} \frac{1}{length_i} \qquad \text{(Equation 2.1)} $$

where HX is the averaged percent protection, hx_i is the percent protection of each peptide i, n represents the number of peptides in which a given residue is observed, and $length_i$ is the length of each peptide i. Although the data do not afford single amide resolution, it is clear that there is a strongly protected region between residues 119 and 130, increased protection at the C-terminus, and two regions from 99 to 107 and 133 to 139 that show significantly lower protection from exchange with solvent.
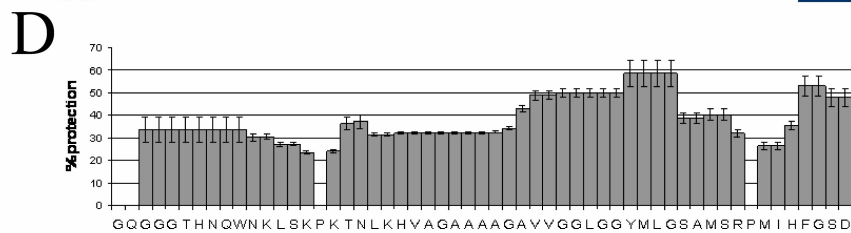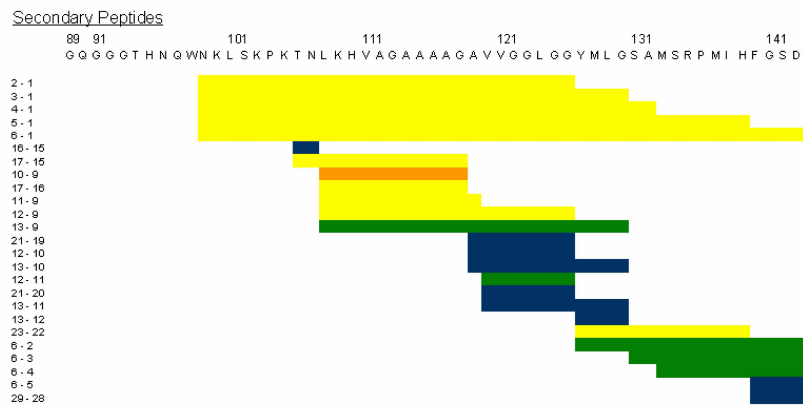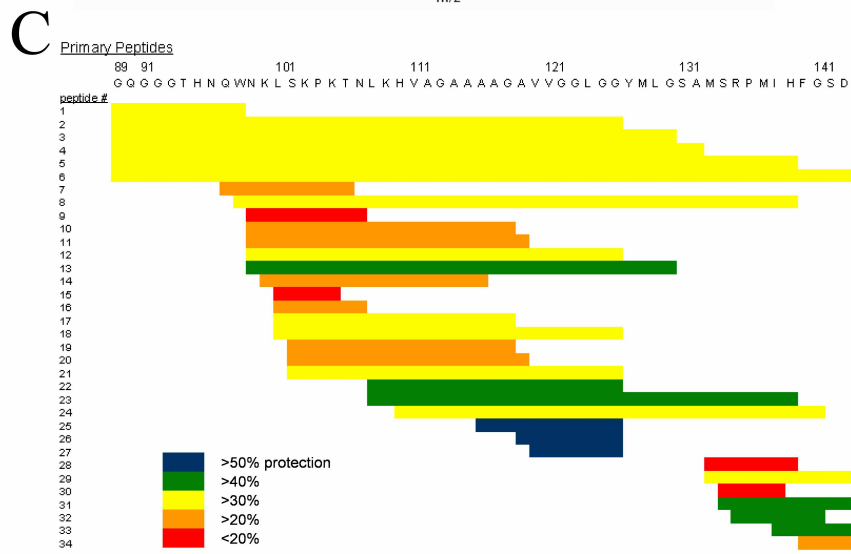
A

B

C Primary Peptides

>50% protection
>40%
>30%
>20%
<20%

Secondary Peptides

D

9
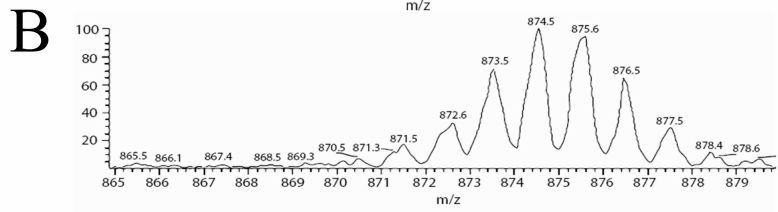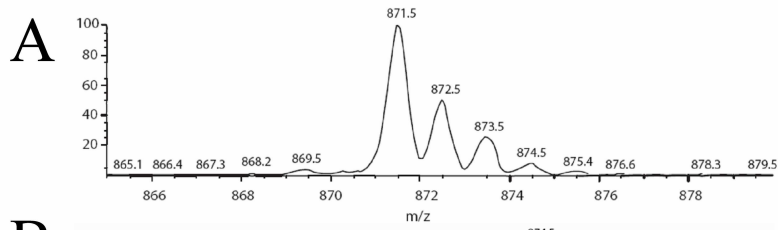
Figure 2.1 Exchange behavior of PrP(89-143, P101L) fibrils observed by mass spectrometry. A) Shows the mass envelope of the 1+ charge state of the proteolytic fragment of PrP(89-143, P101L) corresponding to residues 133-139. B) The mass envelope of the same peptide obtained from PrP fibrils that had been incubated in D_2O for one week. C) Exchange behavior of proteolytic fragments of PrP(89-143, P101L) after six weeks of incubation in D_2O. The extent of exchange is depicted as a percent protection of exchangeable backbone amides. Each 'primary' peptide is numbered. The identity of the two primary peptides used to obtain each 'secondary' peptide is indicated to the left of the 'secondary' peptides. D) Consensus exchange behavior calculated according to equation 1.

2.4.2 Recombinant Expression and Fibrillization of PrP(89-143, P101L)

Amide hydrogen exchange at individual sites can be detected with NMR, but requires isotopically labeled protein. In order to obtain uniformly isotopically labeled PrP(89-143, P101L) for heteronuclear NMR experiments an *E. coli* expression system was developed in which purification of the peptide was expedited by the presence of an N-terminal 6-His tag. Fibrils were grown with the same procedure used to fibrillize synthetic peptide that initiated prion disease in sensitized transgenic mice [1].

2.4.3 Sequence Specific Detection of Hydrogen Exchange in PrP(89-143, P101L) fibrils

As in the mass spectrometry exchange experiments, detecting the extent of hydrogen exchange in amyloid fibrils with NMR requires the denaturation of fibrils under exchange quenching conditions prior to data collection. Acidic dimethyl sulfoxide (DMSO) is an excellent quench solvent to dissolve amyloid fibrils [24, 25, 27, 47, 48]. The minimum rate for amide hydrogen exchange in DMSO occurs at pH* 5 and is typically on the order of 10^{-5} s^{-1} [49]. All non-proline residues have been assigned for the 55-mer with the exception of the N-terminal glycine that exchanges too quickly to be observed in the NMR experiment and four ambiguous glycine residues that occur in glycine-glycine pairs in the sequence. These residues have very similar exchange behavior; therefore, it is not important to assign them individually.

Dissolution of the fibrils and optimization of the NMR experiment lead to a 'dead time' of approximately ten minutes for each sample before NMR data collection. Since the DMSO contains a small amount of D_2O, multiple $^1H^{15}N$ HSQC spectra were collected over a period of six to eight hours after dissolving for each exchange time point to determine the intrinsic rate of exchange in DMSO for each amide. This allows one to discriminate between the exchange that is measured due to the exchange-in step of the fibrils and the exchange-in that occurs during the dead time for handling the sample. A 0-hour (fibrils treated with protonated buffer only) time point was utilized to confirm that the intrinsic rate of deuterium exchange-in under NMR conditions was slow for all residues. The measured rate of exchange for all residues was on the order of 10^{-5} s^{-1} (data not shown) which corresponds to a half-life of hours, so that little exchange occurs during the measurement time of the experiment.

Given the small number of time points in the experimental data set and the limited range of time sampled, the slow exchange rates of the backbone cannot be quantified with certainty. Therefore, the exchange data were analyzed using a percent intensity (I_P) given by

$$I_P = I_t/I_0 \text{ x } 100\% \qquad \text{(Equation 2)}$$

where I_t is the integrated intensity (corrected for back exchange and normalized by concentration) for a given amide cross peak at time t and I_0 is the integrated intensity (corrected

for back exchange and normalized by concentration) for the 0-hour time point. The calculated I_P values for each residue for each experimental time point are shown in Table 2.2. Four types of exchange behavior were observed and are summarized in figure 2.2. The amide hydrogens of residues 90 and 143, the terminal amides, exchange completely before the first time point. Residues 95-101 and residues 137-141 exchange more slowly, with a half-life of approximately a week, but are completely replaced after six weeks. Residues 110-116 exchange slowly as well, with similar extents of exchange after a week as residues 95-101, but retain I_P ~50 even after six weeks, which implies a multi-exponential behavior in the rates. Finally, residues 102-109 and 117-135 remain highly protected even after six weeks ($I_P > 90$).
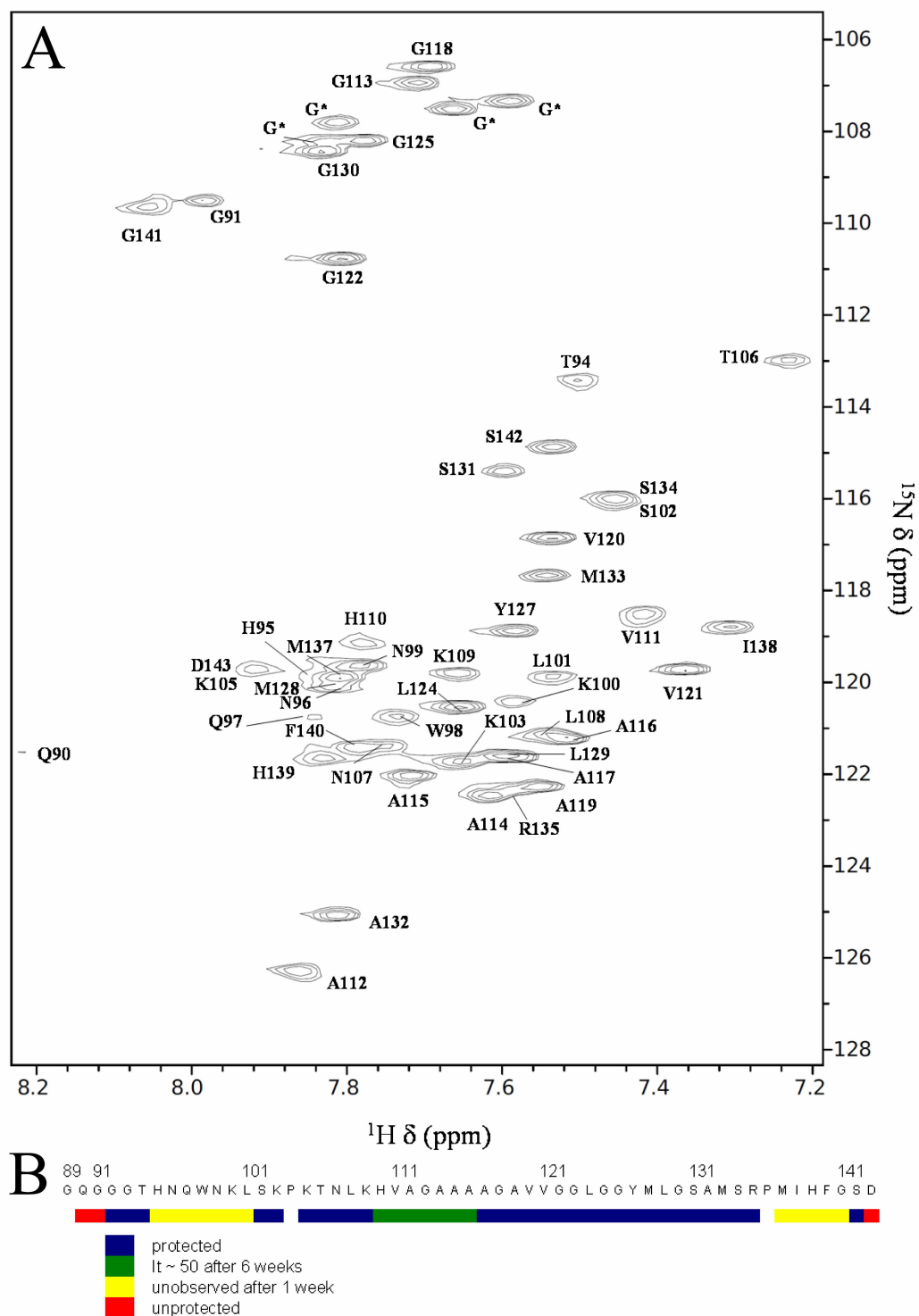
Figure 2.2 Exchange behavior of PrP(89-143, P101L) fibrils observed with NMR. A) Shows the assigned ^{1}H^{15}N HSQC spectrum of PrP(89-143, P101L) under the conditions of a typical 0-hour time point, dissolving fibrils that had not been exposed to deuterium in DMSO containing 5% D$_2$O and 0.03% TFA-D, pH* 5. The four ambiguous glycine residues are denoted with an asterisk. B) A summary of the exchange behavior of PrP(89-143, P101L) fibrils.

Table 2.1. Intrinsic Exchange Rates (1/s)			
Residue	DMSO pH* 5	H$_2$O pH 3.0	H$_2$O pH 7.5
G89		-	
Q90	1.1E-03	4.9E-02	3.1E+02
G91	8.0E-05	2.9E-03	6.4E+01
G92*	2.4E-04	3.1E-03	6.0E+01
G93*	2.4E-04	3.1E-03	6.0E+01
T94	3.0E-05	1.7E-03	2.7E+01
H95	7.0E-05	7.6E-03	3.1E+01
N96	9.6E-05	1.5E-02	1.0E+02
Q97	1.6E-04	2.5E-03	5.2E+01
W98	6.0E-05	1.3E-03	1.3E+01
N99	9.7E-05	2.5E-03	5.2E+01
K100	2.2E-04	2.2E-03	4.1E+01
L101	4.0E-05	1.1E-03	7.5E+00
S102	2.9E-04	1.9E-03	3.1E+01
K103	4.0E-05	2.1E-03	3.9E+01
P104		-	
K105	4.2E-04	1.2E-03	1.1E+01
T106	3.0E-05	1.6E-03	2.4E+01
N107	9.0E-06	4.2E-03	1.1E+02
L108	1.0E-03	1.2E-03	1.2E+01
K109	8.0E-06	1.2E-03	1.2E+01
H110	2.0E-05	6.5E-03	2.6E+01
V111	7.6E-05	1.7E-03	6.6E+00
A112	4.0E-05	1.5E-03	1.6E+01
G113	3.0E-06	2.3E-03	4.0E+01
A114	3.0E-05	2.5E-03	3.2E+01
A115	1.6E-04	1.9E-03	2.2E+01
A116	3.3E-04	1.9E-03	2.2E+01
A117	3.0E-05	1.9E-03	2.2E+01
G118	5.0E-05	2.3E-03	4.0E+01
A119	6.0E-05	2.5E-03	3.2E+01
V120	1.0E-05	9.8E-04	4.3E+00
V121	4.0E-05	9.0E-04	3.1E+00
G122	1.3E-04	1.8E-03	2.9E+01
G123*	2.4E-04	3.1E-03	6.0E+01
L124	1.2E-04	1.2E-03	8.4E+00
G125	1.4E-04	1.7E-03	2.5E+01
G126*	2.4E-04	3.1E-03	6.0E+01
Y127	2.8E-04	1.6E-03	1.7E+01
M128	1.5E-04	1.6E-03	2.4E+01
L129	1.4E-04	1.1E-03	7.3E+00
G130	9.0E-05	1.7E-03	2.5E+01
S131	5.0E-05	3.4E-03	7.5E+01
A132	9.7E-05	2.3E-03	4.3E+01
M133	3.8E-04	1.5E-03	2.1E+01
S134	3.0E-05	2.9E-03	6.5E+01
R135	2.2E-04	2.5E-03	5.2E+01
P136		-	
M137	2.0E-05	1.2E-03	1.2E+01
L138	3.2E-05	9.6E-04	5.2E+00
H139	3.0E-06	3.3E-03	1.2E+01
F140	3.0E-06	3.5E-03	1.9E+01
G141	5.0E-05	2.3E-03	4.6E+01
S142	6.0E-05	3.4E-03	7.5E+01
D143	1.0E-03	2.3E-03	3.4E+01

Table 2.1 Experimental amide exchange rates for all the residues of DMSO-denatured PrP(89-143, P101L) and, for comparison, the calculated exchange rates in water at pH 3.0 where exchange rate is at the minimum [50] and at pH 7.5. It is evident that the quenching in DMSO is much more efficient than in water.

13

Table 2.2. I_p as a function of time						
Residue	0-hour	1-hour	6-hour	21-hour	1 week	6 weeks
G89				-		
Q90	100	0	0	0	0	0
G91	100	103	0	0	0	0
G92*	100	99	96	89	89	80
G93*	100	99	96	89	89	80
T94	100	87	93	85	87	75
H95	100	86	80	75	38	0
N96	100	89	89	82	68	3
Q97	100	93	82	80	59	0
W98	100	95	94	80	52	0
N99	100	95	96	79	54	0
K100	100	100	99	83	55	0
L101	100	108	99	90	61	6
S102	100	100	99	93	90	80
K103	100	92	97	80	86	85
P104				-		
K105	100	100	100	87	82	79
T106	100	99	99	99	96	88
N107	100	97	99	92	80	84
L108	100	92	96	93	93	90
K109	100	100	102	99	90	79
H110	100	93	99	82	50	52
V111	100	82	90	90	44	38
A112	100	91	88	93	49	43
G113	100	100	92	87	58	53
A114	100	95	101	90	47	41
A115	100	100	95	97	50	30
A116	100	99	104	95	60	58
A117	100	101	99	88	85	80
G118	100	99	90	93	89	87
A119	100	100	103	95	90	90
V120	100	96	101	85	90	88
V121	100	92	92	86	90	82
G122	100	103	92	90	90	83
G123*	100	99	96	89	89	80
L124	100	100	85	95	90	95
G125	100	103	90	89	90	83
G126*	100	99	96	89	89	80
Y127	100	100	93	81	77	72
M128	100	89	85	80	80	65
L129	100	92	90	86	81	69
G130	100	100	99	94	89	81
S131	100	93	93	89	80	76
A132	100	99	100	90	90	88
M133	100	88	93	82	91	86
S134	100	99	94	92	84	82
R135	100	99	88	87	75	81
P136				-		
M137	100	97	94	90	56	0
L138	100	100	100	89	60	2
H139	100	97	93	95	49	0
F140	100	98	92	88	59	4
G141	100	103	100	85	43	0
S142	100	92	92	83	84	78
D143	100	0	0	0	0	0

Table 2.2 A listing of measured I_P (equation 2) values for each residue of PrP(89-143, P101L) amyloid fibrils from the hydrogen exchange experiment. The exchange behavior of the four ambiguous glycine residues (marked by an asterisk) is reported as an average, the exchange behavior of each was similar.

2.5 Discussion

Previous solid-state NMR studies of selectively isotopically labeled PrP(89-143, P101L) fibrils from the Wemmer laboratory have demonstrated that residues 112 to 124 adopt an extended β-sheet conformation with evidence for conformational heterogeneity between residues 112-114. Furthermore, multiple quantum NMR experiments indicated that PrP(89-143, P101L) does not adopt an in-register parallel β-sheet [51]. To extend the current knowledge of the conformation of fibrillar PrP(89-143, P101L) the extent of backbone hydrogen exchange was measured at various time points up to 6 weeks of exchange. In the most favorable cases, following hydrogen exchange with tandem proteolysis and LC-MS (HXMS) can lead to single amide resolution. Moreover, HXMS is ideally suited to study prions as it does not require isotopically labeled samples; thus it is one of the few techniques suited to studying the conformation of PrP^{Sc} obtained directly from infected animals. The exchange behavior of natural PrP^{Sc} could then be compared to that of different constructs of recombinant PrP, which are much more tractable to analysis.

In the MS study of exchange in fibrillar PrP(89-143, P101L) the limited number of peptide fragments observed with sufficient signal to noise for analysis precluded single amide resolution of exchange behavior. However, the data from mass spectrometry clearly demonstrate that there are regions of it that are strongly protected from exchange with solvent, while other regions are much less protected from exchange. This is expected as the 55 residue peptide is too long to exist as a single, stably hydrogen-bonded, extended strand within the fibril as the fibril diameter is known to be approximately 8 nm. The exchange profile reveals a highly protected β-strand region with residues 119-130, flanked by two segments which are within a less protected region, encompassing residues 99 to 107 and 133 to 139. Fragmentation was poor near the N-terminus preventing analysis of that region.

To generate a higher resolution map of protection more detailed exchange data were required. After assigning the resonances of PrP(89-143, P101L) amide exchange profiles were collected by following exchange with NMR. Many of the residues show the slow exchange behavior expected from amides involved in stable hydrogen bonds. Two regions, encompassing residues 102-109 and 117-135, remained protected from exchange after 6 weeks and likely constitute the beta sheet core of the fibrils. The high degree of protection is not surprising in that some residues in β$_2$-microglobulin amyloid have been shown to have exchange half-lives of weeks [32] and some residues in Aβ amyloid have been shown to have half-lives of months [26, 27]. Another set of residues, close to the N-terminus and at the C-terminus, were completely exchanged after six weeks of incubation in D$_2$O. While residues 95-101 and 137-141 are moderately protected from exchange with half-lives that are typically afforded by secondary structure conformations in soluble proteins, but they exchange more rapidly than the β-sheet core residues of other reported amyloid fibrils. The lower protection in these two regions is consistent with the mass spectrometry measurements.

Interestingly, residues 110-116 show exchange behavior characteristic of multiple exponentials. This group of residues undergoes exchange to ~50% after 1 week. However, after six weeks, no further exchange has occurred. This behavior can be explained by the presence of

two (or more) conformations for these residues present in the sample with different exchange rates. One component has a half-life of approximately one week and the second component has a much longer half-life. The mass spectrometry data in this region are also consistent with intermediate protection, but the data lack sufficient sequence resolution to observe a bimodal deuterium distribution for this region of the peptide that would confirm the presence of two populations of conformers, one more accessible to exchange with solvent than the other. Furthermore, the residues 110-116 encompass the region of the peptide that displayed conformational heterogeneity in past solid-state NMR experiments [51].

Currently, there are several different models for the conformation of PrP in prions [52]. One is a left handed β-helix proposed by Govaerts and coworkers from 2D electron crystallography data on proteolyzed native PrP^{Sc} [17]. Another is a 'spiral' model identified in a molecular dynamics study of PrP fibrillization that was also shown to be consistent with the electron crystallography data [20]. Neither the β-helix or 'spiral' model requires the fibril to be formed by an in-register parallel β-sheet as seen in Aβ fibrils [27], consistent with previous observations from solid state NMR [51]. Additionally, solid-state NMR work on fibrils of another prion variant, human PrP(Y145Stop), has demonstrated that the N-terminus of the protein, up to residue 112, remains highly mobile in the fibril and that residues 112-115, 118-122 and 130-139 form β-strands [53]; the corresponding regions in the mouse protein are residues 111-114, 117-121 and 129-138. Lastly, a low-resolution model has been generated from hydrogen exchange experiments on prion fibrils that reproduced prion disease; protection from exchange is only observed in the C-terminal region of the protein, residues ca. 169-221 [4, 54]. This in disagreement with the previous models and will not be considered in this discussion as the region of interest in this work, residues 89-143, is completely outside of the protected region deduced from that work.

In the β-helical model, three 5-residue strands are joined by a single residue that has torsion angles characteristic of a left-handed helix to form a triangular 18-residue 'rung', these rungs are then stacked in a stably hydrogen bonded network, including a hydrogen bond at each residue, to form the β-helix. The exchange data shows that the region from 102-136 is resistant to exchange with some degree of conformational heterogeneity from 110-116. This span of the peptide is essentially long enough to support two rungs of a β helix; but, PrP(89-143, P101L) is rather unlikely to adopt this conformation as in nature the minimal number of 'rungs' found experimentally to date is four (seen in the structure of the C-terminal domain of *N*-acetylglucosamine 1-phosphate uridyltransferase from *E. coli,* PDB ID code 1FXJ). However, in the fibril, monomers could be packing against each other to extend the helix. More difficult to reconcile than the length issue is the fact that the previous solid state NMR experiments on PrP(89-143, P101L) did not find any evidence of turn conformations between residues 112-118 and between residues 122-124 [51]. If the core of the fibril is formed by a β-helix, some of these residues should be clearly involved in turns.

The core of the 'spiral' model of PrP^{Sc} consists of four short β strands in which three short strands (E1, E2, E3; residues 115-118, 128-131, 159-163, respectively) form an anti-parallel intra-molecular sheet where the extra strand (E4, residues 134-139) adds to the sheet on an adjacent molecule through a parallel E1:E4 interface. E3 is not essential for the formation of the inter-molecular interface; however, the position of the protected residues in PrP(89-143, P101L) fibrils are not consistent with the location of E1, E2 and E4 in the model although the spiral architecture could explain the protection factors and the observation that PrP(89-143, P101L) fibrils do not form an in-register β-sheet.

16

2.6 Conclusion

This study has utilized amide hydrogen exchange to study the conformation of a model prion, PrP(89-143, P101L), in its fibrillar state. Two regions of the protein between residues 102-109 and 117-135 have high protection from exchange with solvent even after six weeks and are likely to constitute the core of the fibril. The high degree of protection is comparable to that observed in the core of Aβ fibrils [26, 27]. Two other regions, encompassing residues 95-101 and 137-141, exchange completely with solvent after 1 week, an exchange time expected for residues involved in hydrogen bonds but not in the amyloid core. Interestingly, the region of the protein between residues 110-116 exhibits intermediate exchange behavior that can be explained by the existence of two populations of conformers in the fibril which exchange with solvent at different rates. This interpretation of the intermediate protection factors is corroborated by the previous finding that this region possesses some conformational heterogeneity [51]. The core of the amyloid fibril consists of two highly protected β-strands (β1 and β2, residues 102-109 and 117-136, respectively) which are joined by a hydrogen bonded, yet conformationally heterogeneous turn. The overall fibril architecture must be different than that seen in fibrils of Aβ as the PrP fibrils have been shown to not be in register. Assuming slightly different locations of strands E1, E2, and E4 from the 'spiral' model, the core of the fibril could form from an intramolecular sheet between β1 and the first section of β2 where the intermolecular interface is between the second section of β2 and β1 on the adjacent molecule.

2.7 Acknowledgements

2.8 Abbreviations Used

HXMS	tandem proteolysis, liquid chromatography and mass spectrometry
PrP	Prion Protein
PrP^C	PrP Cellular
PrP^{Sc}	PrP Scrapie
GuHCl	guanidine hydrochloride
DMSO	dimethyl sulfoxide
HSQC	Heteronuclear Single Quantum Correlation

2.9 References

1. Kaneko, K., et al., *A synthetic peptide initiates Gerstmann-Straussler-Scheinker (GSS) disease in transgenic mice.* Journal of Molecular Biology, 2000. **295**(4): p. 997-1007.
2. DeArmond, S.J. and S.B. Prusiner, *Transgenetics and neuropathology of prion diseases*, in *Prions Prions Prions*. 1996. p. 125-146.

3.   Prusiner, S.B., *Development of the Prion Concept*, in *Prion Biology and Diseases*, S.B. Prusiner, Editor. 2004, Cold Spring Harbor Laboratory Press: Woodbury, NY. p. 89-141.
4.   Legname, G., et al., *Synthetic Mammalian Prions.* Science, 2004. **305**: p. 673-676.
5.   Lysek, D.A., et al., *Prion protein NMR structures of cats, dogs, pigs, and sheep.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(3): p. 640-645.
6.   Calzolai, L., et al., *Prion protein NMR structures of chickens, turtles, and frogs.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(3): p. 651-655.
7.   Gossert, A.D., et al., *Prion protein NMR structures of elk and of mouse/elk hybrids.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(3): p. 646-650.
8.   James, T.L., et al., *Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(19): p. 10086-10091.
9.   Hornemann, S., C. Schorn, and K. Wuthrich, *NMR structure of the bovine prion protein isolated from healthy calf brains.* Embo Reports, 2004. **5**(12): p. 1159-1164.
10.   Riek, R., et al., *NMR structure of the mouse prion protein domain PrP(121-231).* Nature, 1996. **382**(6587): p. 180-182.
11.   Brown, D.R., *Prion protein expression modulates neuronal copper content.* Journal of Neurochemistry, 2003. **87**(2): p. 377-385.
12.   Garnett, A.P. and J.H. Viles, *Copper binding to the octarepeats of the prion protein - Affinity, specificity, folding, and cooperativity: Insights from circular dichroism.* Journal of Biological Chemistry, 2003. **278**(9): p. 6795-6802.
13.   Sakudo, A., et al., *Prion protein suppresses perturbation of cellular copper homeostasis under oxidative conditions.* Biochemical and Biophysical Research Communications, 2004. **313**(4): p. 850-855.
14.   Aguzzi, A., F. Baumann, and J. Bremer, *The Prion's Elusive Reason for Being.* Annual Reviews of Neuroscience, 2008. **31**: p. 439-477.
15.   Pan, K.-M., et al., *Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**: p. 10962-10966.
16.   Prusiner, S.B., et al., *Scrapie prions aggregate to form amyloid-like birefringent rods.* Cell, 1983. **35**: p. 349-358.
17.   Govaerts, C., et al., *Evidence for assembly of prions with left-handed ß-helices into trimers.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**: p. 8342-8347.
18.   Wille, H., et al., *Structural studies of the scrapie prion protein by electron crystallography.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**: p. 3563-3568.
19.   DeMarco, M.L., et al., *Structural Properties of Prion Protein Protofibrils and Fibrils: An Experimental Assessment of Atomic Models.* Biochemistry, 2006. **2006**(45): p. 15573-15582.

20. DeMarco, M.L. and V. Dagget, *From conversion to aggregation: Protofibril formation of the prion protein.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(8): p. 2293-2298.

21. Englander, S.W., et al., *Hydrogen exchange: The modern legacy of Linderstrom-Lang.* Protein Science, 1997. **6**(5): p. 1101-1109.

22. Bai, Y., et al., *Primary Structure Effects on Peptide Group Hydrogen Exchange.* Proteins-Structure Function and Genetics, 1993. **17**: p. 75-86.

23. Del Mar, C., et al., *Structure and properties of alpha-synuclein and other amyloids determined at the amino acid level.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15477-15482.

24. Hoshino, M., et al., *Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange.* Nature Structural Biology, 2002. **9**(5): p. 332-336.

25. Ippel, J.H., et al., *Probing solvent accessibility of amyloid fibrils by solution NMR spectroscopy.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(13): p. 8648-8653.

26. Kheterpal, I., et al., *A beta amyloid fibrils possess a core structure highly resistant to hydrogen exchange.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(25): p. 13597-13601.

27. Luhrs, T., et al., *3D structure of Alzheimer's amyloid-beta(1-42) fibrils.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(48): p. 17342-17347.

28. Nazabal, A., et al., *Probing the structure of the infectious amyloid form of the prion-forming domain of HET-s using high resolution hydrogen/deuterium exchange monitored by mass spectrometry.* Journal of Biological Chemistry, 2005. **280**(14): p. 13220-13228.

29. Olofsson, A., et al., *Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy.* Journal of Biological Chemistry, 2004. **279**(7): p. 5699-5707.

30. Olofsson, A., A.E. Sauer-Eriksson, and A. Ohman, *The solvent protection of Alzheimer amyloid-beta-(1-42) fibrils as determined by solution NMR spectroscopy.* Journal of Biological Chemistry, 2006. **281**(1): p. 477-483.

31. Wang, S.S.S., et al., *Hydrogen exchange-mass spectrometry analysis of beta-Amyloid peptide structure.* Biochemistry, 2003. **42**(31): p. 9507-9514.

32. Yamaguchi, K.I., et al., *Core and heterogeneity of beta(2)-microglobulin amyloid fibrils as revealed by H/D exchange.* Journal of Molecular Biology, 2004. **338**(3): p. 559-571.

33. Virgil Woods, J. and Y. Hamuro, *High Resolution, High-Throughout Amide Deuterium Exchange-Mass Spectrometry (DXMS) Determination of Protein Binding Site Structure and Dynamic: Utility in Pharmaceutical Design.* Journal of Cellular Biochemistry Supplement, 2001. **37**: p. 89-98.

34. Pantazatos, D., et al., *Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**: p. 751-756.

35. Englander, J.J., et al., *Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**: p. 7057-7062.

36. Zhang, Y.Z. and D.L. Smith, *Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation.* Protein Science, 1993. **2**: p. 522-531.

37.   Hanahan, D., *Studies on Transformation of Escherichia-Coli with Plasmids.* Journal of Molecular Biology, 1983. **166**(4): p. 557-580.
38.   Studier, F.W., *Protein production by auto-induction in high-density shaking cultures.* Protein Expression and Purification, 2005. **41**(1): p. 207-234.
39.   Marley, J., M. Lu, and C. Bracken, *A method for efficient isotopic labeling of recombinant proteins.* Journal of Biomolecular Nmr, 2001. **20**: p. 71-75.
40.   Wittekind, M. and L. Mueller, *Hncacb, a High-Sensitivity 3d Nmr Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins.* Journal of Magnetic Resonance Series B, 1993. **101**(2): p. 201-205.
41.   Grzesiek, S. and A. Bax, *Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance Nmr.* Journal of the American Chemical Society, 1992. **114**(16): p. 6291-6293.
42.   Kay, L.E., et al., *3-Dimensional Triple-Resonance Nmr-Spectroscopy of Isotopically Enriched Proteins.* Journal of Magnetic Resonance, 1990. **89**(3): p. 496-514.
43.   Delaglio, F., et al., *Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes.* Journal of Biomolecular Nmr, 1995. **6**(3): p. 277-293.
44.   Johnson, B.A. and R.A. Blevins, *Nmr View - a Computer-Program for the Visualization and Analysis of Nmr Data.* Journal of Biomolecular Nmr, 1994. **4**(5): p. 603-614.
45.   Keller, R., *The Computer Aided Resonance Assignment Tutorial.* 2004, Cantina: Verlag.
46.   Eng, J.K., A.L. McCormack, and J.R. Yates, *An Approach to Correlate Tandem Mass-Spectral Data of Peptides with Amino-Acid-Sequences in a Protein Database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.
47.   Hirota-Nakaoka, N., et al., *Dissolution of beta(2)-microglobulin amyloid fibrils by dimethylsulfoxide.* Journal of Biochemistry, 2003. **134**(1): p. 159-164.
48.   Ritter, C., et al., *Correlation of structural elements and infectivity of the HET-s prion.* Nature, 2005. **435**(7043): p. 844-848.
49.   Zhang, Y.Z., Y. Paterson, and H. Roder, *Rapid Amide Proton-Exchange Rates in Peptides and Proteins Measured by Solvent Quenching and 2-Dimensional Nmr.* Protein Science, 1995. **4**(4): p. 804-814.
50.   Bai, Y.W., et al., *Primary Structure Effects on Peptide Group Hydrogen-Exchange.* Proteins-Structure Function and Genetics, 1993. **17**(1): p. 75-86.
51.   Lim, K.H., et al., *Solid-state NMR structural studies of the fibril form of a mutant mouse prion peptide PrP89–143(P101L).* Solid State Nuclear Magnetic Resonance, 2006. **29**: p. 183-190.
52.   Cobb, N.J. and W.K. Surewicz, *Prion Diseases and Their Biochemical Mechanisms.* Biochemistry, 2009. **2009**(48): p. 2574–2585.
53.   Helmus, J.J., et al., *Molecular conformation and dynamics of the Y145Stop variant of human prion protein in amyloid fibrils.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**: p. 6284–6289.
54.   Lu, X., P.L. Wintrode, and W.K. Surewicz, *Beta-Sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **105**(5): p. 1510-1515.

Chapter 3

Insights into the Conformational Ensemble of Monomeric Aβ from a Dual Molecular Dynamics and NMR Study

3.1 Summary

This chapter describes structural studies on amyloid-β peptides that utilize a comparison between experimentally determined NMR observables (chemical shifts, ^{3}J-couplings, relaxation parameters and Overhauser cross peaks) with those predicted from simulated ensembles obtained via molecular dynamics calculations. New generation TIP4P-Ew water and Amber ff99SB protein force fields provide a good prediction of the ^{13}C relaxation parameters and Overhauser cross peaks for Aβ$_{21-30}$, the peptide fragment first studied to validate this technique. Inspection of the simulated ensemble shows that the interactions that lead to the observation of medium-range Overhauser cross peaks cannot occur in a single molecule simultaneously, demonstrating that in the study of flexible peptides the calculation of a single structure from Overhauser distance restraints, as routinely performed for well-folded proteins, is essentially meaningless. A more robust view of peptide conformation is provided by the comparison of NMR observables with those calculated from simulated ensembles. In the case of Aβ$_{21-30}$, this type of analysis reveals a structural ensemble that contains a majority population (~60%) of unstructured conformers, lacking any secondary structure or persistent hydrogen-bonding networks. The remaining minority population contains conformers with a β-turn centered at Val24 and Gly25, as well as evidence of the Asp23 to Lys28 salt bridge important to the fibril structure. The work with Aβ$_{21-30}$ provides a framework for a similar study on Aβ$_{1-40}$ and Aβ$_{1-42}$, of which the NMR data has already been collected and is discussed herein.

3.2 Introduction

3.2.1 General Introduction to Amyloid-β

The major constituent of amyloid fibril plaques found in the brains of patients with Alzheimer's disease is the amyloid-β peptide (Aβ) [1]. Aβ is formed by the successive proteolytic action of β- and γ-secretase on the amyloid precursor protein (APP), an integral membrane protein concentrated at the synapses of neurons. While there is some evidence that APP is involved in the formation of functional synapses and neural plasticity [2, 3], the precise role of APP in the cell remains controversial. The cleavage of APP yields Aβ peptides of varying lengths from 39-43 residues, the most common forms of which are the 40- and 42-mer, Aβ$_{1-40}$ and Aβ$_{1-42}$, respectively [4]. Although much work has been undertaken to study the conformation of Aβ in fibrils, several recent lines of evidence, including the generation of AD symptoms in mice in which no amyloid plaques are present, suggest that it may be the pre-

fibrillar monomer and oligomeric states of the Aβ peptides, rather than fibrillar Aβ, that give rise to the symptoms of the Alzheimer's disease [5-7].

Over the past decade significant strides have been made in determining the atomic level structure of Aβ fibrils using solid-state NMR [8-10]; however, characterizing the monomeric and early oligomeric species of Aβ remains a significant challenge. This is due both to the difficulty of obtaining homogenous samples of monomers or oligomers for study because of their strong propensity to aggregate and the complimentary challenge of interpreting bulk structural data collected on heterogeneous distributions of conformers. Recent NMR studies on the monomeric form of $Aβ_{1-40}$ and $Aβ_{1-42}$ have found evidence of some regular structure, albeit in the presence of high concentrations of non-aqueous solvents such as hexafluoroisopropanol, trifluroethanol, and SDS micelles [11-14]. Other, more recent studies focused on aqueous environments have predicted averaged quantities such as scalar coupling constants and spin relaxation constants for calculated ensembles of the peptide for comparison to NMR experiments [15-17]. These data measure only local rather than tertiary structure and hence provide an incomplete description of the structural ensemble. Nuclear Overhauser cross peaks, arising from protons in close proximity, are better suited to probe tertiary structure in proteins; hence, they are heavily relied upon in NMR structure determination. In addition to their sensitivity to longer-range structure, nuclear Overhauser cross peaks are averaged over both population and the inverse of the sixth power of internuclear distance. This second averaging provides a powerful tool to study a heterogeneous distribution of interconverting conformers, as expected to be relevant for the Aβ peptides studied here, and provides significant insight into the structure of the monomeric form of Aβ.

In this chapter of the thesis I present studies combining molecular dynamics calculations (MD) and NMR spectroscopy conducted on a fragment of Aβ that has been extensively studied. This work paves the way for similar studies with the full-length forms of the peptide, $Aβ_{1-40}$ and $Aβ_{1-42}$, which are underway. The next section contains an introduction to the shorter construct, which served as a tractable model system to demonstrate that our combined MD and NMR strategy provides insight into the peptide's conformational distribution unavailable to either technique alone, followed by an introduction to similar work with the longer peptides. Lastly, I introduce the NMR observables that are used to probe structure in tandem with MD calculations

3.2.2 Introduction to $Aβ_{21-30}$

In an attempt to probe for any significant structure in the monomeric state, Lazo *et al*. subjected $Aβ_{1-40}$ and $Aβ_{1-42}$ peptides to digestion by multiple proteases [18]. Despite containing a large number of potential proteolytic sites throughout the sequence, the peptide fragment spanning residues 21 to 30 (AEDVGSNKGA) was a significant product for each enzyme reaction, indicating relative resistance to cleavage despite containing several possible cleavage sites for the different proteases employed. Furthermore, little digestion was seen when the synthetic $Aβ_{21-30}$ fragment itself was subjected to the same protease conditions. These results were interpreted to mean that some structure in the 21 to 30 region protects $Aβ_{1-40}$ and $Aβ_{1-42}$ from protease degradation and that this structure is retained in the $Aβ_{21-30}$ fragment.

One of the primary benefits of studying this small hydrophilic peptide fragment is that it is readily soluble in its monomeric form at concentrations in which the longer and more hydrophobic $Aβ_{1-40}$ and $Aβ_{1-42}$ sequences aggregate quickly. In addition, this sequence encompasses many of the amino acids that are thought to be important for understanding the

Alzheimer's disease state. The $A\beta_{21-30}$ sequence consists mainly of a charged region that, in the context of the fibril structure of the full-length $A\beta_{1-40}$, comprises a turn connecting the two flanking hydrophobic β-strand regions that is necessary for generating the observed cross-β structure down the fibril axis [10]. Additionally, as observed by solid state NMR, in the fibrillar form there is a buried salt bridge between Asp23 and Lys28; these residues are also encompassed by this peptide [8, 19, 20]. It has been shown that mutating either residue dramatically affects fibril formation [19, 21]. It is also noteworthy that many of the familial Alzheimer's disease (FAD) mutants of the APP protein are located between residues 21 to 23, each of which leads to dramatically different *in vitro* fibril formation properties and *in vivo* clinical outcomes [4, 21-27].

Although absolute proteolysis is not a perfect metric to probe protein structure, a number of experimental and computational studies have attempted to determine if the presence of some stable residual structure in the $A\beta_{21-30}$ monomer can account for its protease resistance. Teplow and co-workers studied the wild-type (WT) peptide as well as five FAD mutants of the $A\beta_{21-30}$ peptide using rotating frame nuclear Overhauser effect spectroscopy (ROESY) NMR experiments [18, 28]. They proposed that WT $A\beta_{21-30}$ folds into a single conformer corresponding to a unique bend structure, identified through long range (i,i+8) cross peaks for Glu22 H_α to Ala30 HN, and (i,i+6) sidechain-sidechain cross peaks for Glu22 to Lys28 in the ROESY spectrum of WT $A\beta_{21-30}$ which were absent in the FAD mutants. Replica exchange molecular dynamics simulations by Baumketner et al. using the OPLS and TIP3P all-atom models for peptide and water, found that 40% of the peptide ensemble is folded into two distinct bend structures stabilized primarily by Asp23 sidechain interactions with the Ser26 sidechain and backbone [29]. Borreguero and coworkers studied the peptide by a coarse-grained model in which they find collapsed structures stabilized by hydrophobic interactions between Val24 and Lys28, as well as electrostatic interactions between Asp23 to Lys28 [30]. They also simulated five ~100 ns molecular dynamics trajectories using CHARMM-27 and TIP3P for the peptide and water model, without using any accelerated sampling technique, each with a different combination of density, starting structure and salt concentration. Though the authors acknowledge that their simulations are far too short to converge to the equilibrium ensemble, they report contacts between hydrophobic regions of Val24 and Lys28 that are more stable on the nanosecond timescale than charged interactions between Lys28 and Glu22 or Asp23 [31]. Finally Mousseau, Derreumaux and coworkers used an activation-relaxation sampling technique combined with the OPEP coarse-grained model of $A\beta_{21-30}$ and found several clusters of structures, all sharing a turn formed by stabilizing interactions between Val24 and Lys28 of many different sidechain contact combinations [32]. In summary, simulations found little evidence of the long-range interactions observed in the NMR experiment but provided little consensus on the solution structure of the $A\beta_{21-30}$ peptide.

3.2.3 Introduction to $A\beta_{1-40}$ and $A\beta_{1-42}$

As mentioned above, Aβ is found in several different isoforms, the two most common of which are $A\beta_{1-40}$ and $A\beta_{1-42}$. Although the longer isoform differs from the shorter only by the presence of an additional two C-terminal residues, Ile41 and Ala42, $A\beta_{1-40}$ and $A\beta_{1-42}$ show vastly different behavior. $A\beta_{1-42}$ aggregates much more quickly than $A\beta_{1-40}$ [33]. $A\beta_{1-42}$ is the major species present in intracerebral amyloid deposits seen in AD patients [34] whereas $A\beta_{1-40}$ is the major species in blood plasma [35]. $A\beta_{1-42}$ has been shown to be significantly more toxic

in animal models of AD and to neurons in cell culture than $A\beta_{1-40}$ [7, 36-38]. Interestingly, the ratio of $A\beta_{1-42}$ to $A\beta_{1-40}$ correlates with the presence of AD [39, 40].

Solution NMR experiments that have been conducted on monomeric $A\beta_{1-42}$ and $A\beta_{1-40}$ in aqueous environments failed to identify significant differences between the average structures of the two forms of $A\beta$, although they did establish that the oxidation state of Met35 strongly affects the aggregation propensity [41, 42] . However, differences relating to protein dynamics between the two forms of $A\beta$ have been reported. A recent ^{15}N relaxation study established that the C-terminus of $A\beta_{1-42}$ is more rigid than $A\beta_{1-40}$ and may be involved in aggregation [15, 16]. Similar ^{13}C methyl-group relaxation experiments by the same group also were consistent with additional rigidity in the C-terminus of $A\beta_{1-42}$ [16]. By using the combined MD and NMR approach employed here it should be possible to observe structural motifs in subpopulations of peptides as well as probe dynamics in the ensemble.

3.2.4 NMR Measurables Employed

The NMR measurables utilized in this work are chemical shifts, three bond J-couplings, ^{13}C relaxation times, and nuclear Overhauser effect cross peaks. This section of the thesis will introduce how they probe protein structure.

The chemical shift of a given protein spin (HN, N, Hα, Hβ, Cα, and Cβ) is affected by the electronic environment around the spin; this electronic environment is determined by the secondary and tertiary structure in which the spin is involved. Backbone chemical shifts can be calculated for a protein of known structure by quantum mechanical calculations, approaches utilizing empirical relationships derived between chemical shift data and classical physics, or by hybrid approaches. The large collection of NMR assignments of proteins of known structure is sufficient in size that if the protein structure is known, then chemical shifts can be reliably predicted by hybrid techniques [43-46].

The three bond J-coupling between HN and Hα, $^3J_{HN-H\alpha}$, is related to the φ dihedral angle through the well-known Karplus relationship [47, 48] (equation 3.1),

$$J(\phi) = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C \qquad \text{(equation 3.1)}$$

where *A, B,* and *C* are empirically determined constants. Along with ψ, the φ dihedral angle is a measure of secondary structure in proteins. While there are three bond J-couplings which depend on ψ, ($^3J_{C\beta i-Ni+1}$, $^3J_{H\alpha i-Ni+1}$ and $^3J_{Ni-Ni+1}$), they are not proton-only J-couplings and thus require at least one labeled heteronucleus to measure and were not used in this study. When measured on a distribution of conformers interconverting quickly with respect to the NMR time scale, as is the case here, J-couplings and chemical shifts are simple population averages. If the interconversion between conformers was in slow exchange with respect to the NMR time scale, multiple sets of NMR resonances would be observed.

NMR relaxation parameters are widely used in the study of protein dynamics [49]. In this chapter the focus is on the [13]Cα spin-lattice and spin-spin relaxation times (T_1 and T_2, respectively) for use in the comparison of time-scales between MD simulations and the NMR experiments. The relaxation of a ^{13}C that is directly bonded to a single proton is dominated by the dipolar interaction between the bound proton and the chemical shift anisotropy (CSA) of the ^{13}C spin. The relaxation times depend on the spectral density function, the Fourier transform of the correlation function for the directly bonded [1]H-^{13}C pair given by

$$J(\omega) = \int_{-\infty}^{\infty} C(\tau) e^{i\omega\tau} d\tau \qquad \text{(equation 3.2)}$$

The spectral density function (equation 3.2) controls the components of the spin relaxation rates, defined as the inverse of the T_1 and T_2 relaxation times at the relevant ^{1}H and ^{13}C Larmor frequencies.

$$R_1 = \frac{1}{T_1} = R_1^{\,DD} + R_1^{\,CSA} \qquad \text{(equation 3.3)}$$

$$R_2 = \frac{1}{T_2} = R_2^{\,DD} + R_2^{\,CSA} + R_a \qquad \text{(equation 3.4)}$$

where DD and CSA denote the dipolar and CSA components of the spin relaxation rate constants, respectively. R_a is the sum of the relaxation rate constants for pseudo-first order processes such as chemical exchange and diffusion; these are ignored in this work. For directly bonded ^{1}H-^{13}C pairs containing a single proton, the dipolar relaxation rates given by

$$R_1^{\,DD} = (1/20)K^2\left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)\right] \qquad \text{(equation 3.5)}$$

$$R_2^{\,DD} = (1/40)K^2\left[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 3J(\omega_H) + 6J(\omega_H + \omega_C)\right] \quad \text{(equation 3.6)}$$

are the main contributors to the overall relaxation. The constant factor, K, is defined as

$$K = \frac{\mu_0}{4\pi\, r_{eff}^{\,3}} \hbar \gamma_a \gamma_b \qquad \text{(equation 3.7)}$$

where μ_0 is the permeability of free space, $\hbar$ is Planck's constant, and γ_a and γ_b are the gyromagnetic ratios for the nuclei of interest which for the ^{13}C relaxation experiment are carbon and hydrogen, and r_{eff},

$$r_{eff} = \left\langle \frac{1}{r^6(t)} \right\rangle^{-\frac{1}{6}} \qquad \text{(equation 3.8)}$$

is the appropriately averaged internuclear distance between atoms. The CSA contributions to the spin-lattice and spin-spin relaxation rates from an axially symmetric chemical shift tensor are given by

$$R_1^{\,CSA} = (2/5)\Delta\delta^2 \omega_C^{\,2} J(\omega_C) \qquad \text{(equation 3.9)}$$

$$R_2^{\,CSA} = (2/5)\left(\Delta\delta^2 \omega_C^{\,2}/6\right)\left[4J(0) + 3J(\omega_C)\right] \qquad \text{(equation 3.10)}$$

where $\Delta\delta$ is the difference between the parallel and perpendicular components on the chemical shift tensor. Although there is an additional averaging over distance implied by equation 3.8, it is worthwhile to note that for directly bonded spin pairs, the distance is approximately constant. (However, the Amber ff99SB force field does include bond vibration and libration motions, which cause the fastest decay of correlation function for directly bonded spin pairs.)

25

The final and most information-rich NMR observable utilized in this work is the nuclear Overhauser effect, which depends on the distance between protons in space. The nuclear Overhauser effect is most often observed by two-dimensional Nuclear Overhauser Effect spectroscopy (NOESY) in which magnetization is transferred between resonances via cross relaxation through a transient dipolar interaction. The intensity of a given cross peak is related to the distance between the protons. Typically cross peaks between protons within 6 Å can be observed. In practice, the observation of Overhauser cross peaks can be encumbered if molecular tumbling occurs on the same time scale as the Larmor frequency, thereby interfering with the Overhauser cross-relaxation mechanism. At magnetic fields strengths currently employed in high resolution NMR, molecules with a molecular weight around 1 kDa, as is the case for Aβ$_{21-30}$, are affected by this issue. This is avoided by transferring magnetization while the spins are made to precess about an applied external field during the mixing time, thereby changing the effective Larmor frequency and allowing for magnetization to transfer via a dipolar mediated cross-relaxation mechanism. This technique is referred to as rotating frame Nuclear Overhauser spectroscopy (ROESY) [50, 51].

Overhauser cross peaks are typically used as the most important structural restraint in biomolecular structure determination with NMR where many loose distance restraints are incorporated into a pseudo-energy function used to restrain a structure calculation. This approach to interpret Overhauser cross peaks is relatively straightforward when the molecule of interest is a single stable conformer. When the molecule of interest is quite flexible and exists as an ensemble of interconverting conformers, this approach is intractable. For example, a cross peak between two protons that spend 100% of the time 6 Å apart will have the same intensity a cross peak between two protons that spend 9% of the time 4 Å apart. There is no way to extract this information from the spectra alone. However, if we have the knowledge of the peptide's structure and dynamics from MD simulations providing a correlation function for each spin pair, the intensity of Overhauser cross peaks can be calculated from its Fourier transform, the spectral density function (equation 3.2), according to the method of van Gunsteren and coworkers [52, 53]. The normalized correlation function is

$$C(\tau) = \left\langle \frac{1}{r^6(t)} \right\rangle^{-1} \left\langle \frac{P_2(\cos\chi_{t,t+\tau})}{r^3(t)r^3(t+\tau)} \right\rangle \qquad \text{(equation 3.11)}$$

where P$_2$ is the 2nd order Legendre polynomial, χ is the angle between the interspin vector in the laboratory reference frame at times t and t + τ, r(t) is the instantaneous pair distance, and the angle brackets denote a thermal average. Rotating frame nuclear Overhauser cross peaks are then predicted by calculating the intensity for each spin pair,

$$I(t_{\text{mix}}) = Xe^{-\Lambda t_{\text{mix}}} X^{-1} I(0) \qquad \text{(equation 3.12)}$$

where X and Λ are the eigenvectors and eigenvalues of the full relaxation matrix, R, composed of the diagonal elements

$$\rho_{ii} = \sum_{j=1,\neq i}^{n} \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(2\omega_0) + \frac{9}{4} J_{ij}(\omega_0) + \frac{5}{4} J_{ij}(0) \right] \qquad \text{(equation 3.13)}$$

and off-diagonal elements

$$\sigma_{ij} = \frac{1}{10} K^2 \left[\frac{3}{2} J_{ij}(\omega_0) + J_{ij}(0) \right]$$ (equation 3.14)

which sample the spectral density function at different frequencies. The direct dipolar relaxation rate is described by ρ, σ is the cross-relaxation rate for all proton pairs as described by van Gunsteren and coworkers, and K is defined as above with $\gamma_a\gamma_b$ equal to γ_H^2. The expressions for calculating the relaxation rates in NOESY spectra differ slightly with the diagonal elements given by

$$\rho_{ii} = \sum_{j=1, \neq i}^{n} \frac{1}{10} K^2 \left[3J_{ij}(2\omega_0) + \frac{3}{2} J_{ij}(\omega_0) + \frac{1}{2} J_{ij}(0) \right]$$ (equation 3.15)

and the off-diagonal elements given by

$$\sigma_{ij} = \frac{1}{10} K^2 \left[3J_{ij}(\omega_0) - \frac{1}{2} J_{ij}(0) \right]$$ (equation 3.16)

After cross peak intensities are predicted, it is possible to compare them to those obtained experimentally, thereby validating a structural ensemble calculated with MD.

3.3 Materials and Methods

3.3.1 NMR spectroscopy of $A\beta_{21-30}$

The $A\beta_{21-30}$ peptide (AEDVGSNKGA) was synthesized and purified to 98% purity by reverse-phase HPLC (Anaspec, San Jose CA). NMR samples contained 10 mM $A\beta_{21-30}$ and 25 mM ammonium d_4-acetate in 90% H_2O, 10% ^2H_2O or 100% ^2H_2O. Upon dissolution, the pH was adjusted to 6.0 with 1M NaOH. NMR data were collected at 283 K on Bruker Avance 500 or 600 MHz spectrometers and Bruker Avance II 800 or 900 MHz spectrometers. All data were processed with NMRPipe [54] and analyzed with NMRView [55] or CARA [56] . All spectra were recorded at 10°C to facilitate comparison with previous NMR studies on this peptide.

Chemical shift assignments were obtained with 2D ^1H-^1H TOCSY [57] and ROESY [50] experiments. All resonances have been assigned. Distance restraints were obtained from two 2D ^1H-^1H ROESY experiments in both 90% ^1H_2O, 10% ^2H_2O and 100% ^2H_2O with a composite pulse spin lock applied during the 300 ms mixing time to avoid TOCSY artifacts (see the H/HA correlations in the ROESY spectrum depicted in figure 3.4 for an example) [58]. Additional ^1H_2O ROESY spectra were also collected with 200 ms and 400 ms mixing times but were not used in generating distance restraints. A total of 4096 and 1400 points (States-TPPI) were collected in t_2 and t_1, respectively. Spectral widths in both dimensions were 7184 Hz and 8503 Hz on the 800 MHz and 900 MHz spectrometers, respectively.

To build a single structural model fitting all of the data simultaneously using a standard NMR structure determination approach, ROESY cross peaks were classified as strong, medium, weak, and very weak based on peak intensity and converted to 2.9, 3.3, 5.0, and 6.0 Å upper distance restraints, respectively. The set of 155 manually assigned distant restraints was used to calculate 1000 structures with the program CYANA [59]. The 50 lowest energy structures of the minimized ensemble were analyzed with the program Pymol (DeLano Scientific: San Carlos,

CA). Structural statistics and hydrogen bonds present in the structure ensemble, detected with the Amber [60] suite program *ptraj* are presented in the Results.

Spin-lattice (T_1) and spin-spin (T_2) relaxation times for natural abundance ^{13}C at the C_α position were measured for all non-glycine amino acids using the same 100% ^2H_2O sample described above. T_1 was measured at 500 MHz by inverse-detected inversion recovery with delay times 5, 25, 50, 150, 400, 600, 1000, 2000, and 2500 ms. T_2 was measured at 600MHz by an inverse-detected CPMG experiment with delay times 0, 20.48, 40.96, 61.44, 82.92, 122.88, 143.36, 163.84, 204.8 ms. Relaxation parameters were fit from the data as described previously [61].

Three bond scalar coupling constants, $^3J_{HN-H\alpha}$, were measured from the multiplet structure in a 2D double-quantum filtered COSY measured at 500 MHz with a spectral width of 6127 Hz in both dimensions. To ensure that the experimental lineshape was not adversely affected by limited digital resolution 4096 points were collected in both t_1 and t_2. Quadrature detection in t_1 was obtained according to the States-TPPI method and the digital resolution was matched to t_2 with linear prediction.

3.3.2 MD simulations of $A\beta_{21-30}$

In this study the zwitterionic $A\beta_{21-30}$ peptide is represented with the Amber ff99SB fixed charge empirical force field [62]. Amber ff99SB is a recent reparameterization of the backbone dihedral angles of proteins by Simmerling and coworkers to correct previous problems with secondary structure propensities of the original ff99 parameters. Amber ff99SB quantitatively captures the distribution of backbone ϕ/ψ angles compared with quantum mechanical calculations and validation on model peptide and protein systems. Two separate sets of simulations were run in which the peptide was solvated with TIP3P [63] and TIP4P-Ew [64] water models, respectively. TIP3P is an older water model that is popularly used in aqueous protein simulations, while TIP4P-Ew is a newer re-parameterized version of the standard TIP4P water model for use with Ewald summation techniques. The alternate TIP4P-Ew water model was chosen since it reproduces many salient thermodynamic and dynamic features of bulk water properties when compared with experiment, and also its excellent performance for temperature trends of these properties is especially relevant for this experimental study which is conducted at 10°C [64].

In this work the AMBER9 molecular dynamics simulation package [60] was used to generate the structural and dynamical ensembles of $A\beta_{21-30}$ peptide fragment in water. The system is prepared by solvating the $A\beta_{21-30}$ structure with 1578 TIP3P or 1579 TIP4P-Ew water molecules, respectively. A single sodium (Na^+) ion is included in the system to balance the peptide net charge. Each system is briefly equilibrated using Andersen thermostats [65] to bring the system up to 300K temperature, then equilibrated for 125ps at constant pressure with Berendsen (weak) coupling at 1 bar (default parameters) and 300K temperature. The average density of the last 100ps of the constant pressure simulation is then calculated and a snapshot containing position and velocity information with that density (within $0.001g/cm^3$) is selected as the starting structure. In all simulations the equations of motion are integrated with 1fs timesteps, the long-range electrostatic interactions are calculated using Particle Mesh Ewald method (PME) [66], and a cutoff of 9.0 Å is used for real space electrostatics and LJ interactions.

Replica exchange through the *sander* module of Amber9 was used to improve convergence at the lower temperatures, using 64 temperature replicas exponentially spaced between 270 K and 507 K with exchange attempts every 1 ps [67, 68]. We ran two independent

replica exchange simulations for a time between 45ns and 50ns per replica exchange simulation, of which the first 20ns of each replica is treated as equilibration. The second replica exchange simulation was started from configurations acquired after 20ns equilibration in the first replica exchange simulation, but using a new set of randomized velocities. Convergence to equilibrium of the two independent simulations was tested by whether they both reach the same linear average of the pair distances over their structural ensembles. The equilibrium populations of the two independent runs differ by no more than 7% in this quantity.

Microcanonical ensemble (NVE) trajectories were run to measure dynamical quantities since coupling to a thermal bath, especially by the Andersen or Langevin thermostats provided in the AMBER codes, can perturb system dynamics. For each peptide-water model combination we ran thirty NVE trajectories of 20ns in length. Starting structures for these trajectories were selected from the 284K replica of each water model, separated by 1ns of replica exchange simulation to ensure structural decorrelation. Since only coordinate information (not velocity) was saved for the structural ensemble, structures were equilibrated at 284K for 100ps prior to the 20ns constant energy runs. The *ptraj* module of AMBER was used to analyze the DSSP defined secondary structure, hydrogen-bonds and electrostatic/saltbridge interactions over the generated structural ensembles [69]. All possible donors and acceptors for the hydrogen bonds are specified in the analysis.

3.3.3 Prediction of NMR Observables for the calculated $A\beta_{21-30}$ ensemble

Using the SHIFTS [44-46] and SHIFTX [43] programs, the chemical shifts for all protons and C_α and C_β carbon atoms were predicted by averaging these quantities over the members of the Boltzmann-weighted simulation ensemble. Two types of parameter sets for *A*, *B*, and *C* in equation 3.1 were used. The first type corresponds to fits of the Karplus equation by using experimentally measured scalar coupling constants and known reference protein structures, a parameter set for $^3J_{HN-H\alpha}$ was derived by Vuister and Bax with *A* = 6.51, *B* = -1.76, and *C* = 1.60 [47], although other choices are possible [70-72]. In fact, it has been shown that variations between the different sets of Karplus parameters likely reflect differences in the magnitude of dihedral angle fluctuations about the reference equilibrium folded structure used to calculate the values of *A*, *B*, and *C* [70, 72]. As our ensemble includes dihedral angle fluctuations, using these Karplus parameters to calculate the ensemble averaged J-couplings results in a double counting of such motions. Thus, a second parameter set in which a harmonic approximation to this dynamical motion has been removed, where *A* = 9.5, *B* = -1.4, and *C* = 0.30, was also considered [70].

To calculate the spectral density functions for ^1H-^1H spin pairs to predict ROESY data and for ^{13}C-^1H pairs to predict the relaxation parameters (T_1 and T_2), the method of Peter *et al.* [52] is employed in evaluating the normalized correlation functions (equation 3.11). The correlation functions were calculated for each of the approximately 1800 ^1H-^1H spin pairs as well as the eight non-glycine directly bonded $^{13}C\alpha$-$^1H\alpha$ spin pairs. The evaluation of equation 3.11 was an average over thirty independent constant energy (NVE) trajectories at 284K of equation 3.11, for correlation times τ up to 5ns. Using many trajectories run in parallel, launched with different representative members of the Boltzmann distribution of peptide conformations, is preferable to a single long trajectory in the NVE ensemble, which will not exhibit a canonical structural ensemble. Averaging multiple trajectories, each with a slightly different total energy, also has the minor added benefit of sampling the energy distribution of the canonical ensemble despite using constant energy trajectories. It is important to note that equation 3.11 was

"unnormalized" by the $1/r^6$ average spin-spin distance in each trajectory in order to compute the average over trajectories.

   Each resulting average numerical correlation function for a given atom pair is then fit to a triple exponential form using a shell script invoking the *fit* routine in Gnuplot. A triple exponential form was selected since multiple relaxation modes with different timescales invariably exist for a peptide. In our case, a very fast (<1ps) mode due to vibration and libration exists, as well as several reorientational modes arising from anisotropic tumbling due to the non-spherical shape of the peptide. Using a quadruple exponential form did not show substantially different fits for a test group of data, and hence were not used. In some cases only two exponentials were required, although three exponential fits were typical. The fitted time correlation function is then Fourier transformed to define the spectral density functions (equation 3.2) according to equation 3.17.

$$\int_{-\infty}^{\infty} e^{-\tau/b} e^{i\omega\tau} d\tau = \frac{2b}{1+\omega^2 b^2} \qquad \text{(equation 3.17)}$$

It is important to note that the peptide tumbles and locally reorients rapidly enough for all relevant spin-spin vector time-correlation functions to decay to zero within the time of the 20ns dynamics simulations.

   The T_1 and T_2 relaxation times were calculated from the appropriate spectral density functions using equations 3.2-10. In equations 3.9 and 3.10, $\Delta\delta$ is taken to be 25ppm [61]. It is also interesting to note that the fitted full time correlation functions are explicitly calculated and analytically Fourier transformed, and hence do not need to fit our spectral density functions through a Lipari-Szabo model-free analysis [49] which has limited applicability when the system of interest lacks the separation between internal and external motion timescales, as is the case for $A\beta_{21-30}$. In order to compare the simulated T_1 and T_2 relaxation times to the experiment performed in ^2H_2O, the time correlation functions are scaled by a factor of 1.2 as a simple approximation for the larger viscosity of ^2H_2O compared to H_2O at 10°C.

   The ROESY cross peaks were calculated from the spectral density functions in the process described by equations 3.12-14. Unlike the extended atom model (no aliphatic hydrogens) of the previous studies MD studies on $A\beta_{21-30}$, here all hydrogen atoms have been simulated explicitly for each methyl group and hence we calculate all pair correlation functions, including neighboring methylene and methyl group protons. Water proton coordinates have been ignored as is the standard assumption in the NMR experiment.

   This coupled system of differential equations (3.12-3.14) was solved for the magnetization matrix at the mixing time used by the NMR experiments for both H_2O and ^2H_2O. The experimental conditions of heavy water solvation on the relaxation matrix was simulated by removing the exchangeable hydrogens including backbone amides (HN), hydrogens in basic NH_3^+ groups, and hydroxyl hydrogens (HO) from the spin-matrix, accomplished by setting all pair distances with these protons to 30Å. In order to generate peak predictions, the peak volume contributions (including positive contributions from cross peaks dominated by spin diffusion) are summed for degenerate spins (methyl groups) as well as those from spins within the same residue that are indistinguishable at the resolution of our NMR experiments. Note that this method explicitly accounts for peak intensity effects caused by methyl group rotation since spectral density functions are calculated for each proton in a methyl group and the individual intensities from indistinguishable peaks are summed to compare to experimental spectra. Amine peak volumes in the H_2O spectra were scaled by a factor of 0.9 to approximately account for the

presence of 10% deuterium exchanged amide protons from the 10% ^2H_2O used for NMR lock. Predicted cross peaks to basic amine and hydroxyl groups were filtered from the predictions since these cross peaks would be significantly broadened by exchange with solvent protons on the NMR timescale.

In order to directly compare simulations to experiment, the constant relating the arbitrary experimental cross peak intensity scale to the simulated intensities for diagonal magnetization of unity at mixing time of 0 ms must be determined. This is especially important because the experimental noise level must be determined to evaluate the ability of the predictions to separate observed and unobserved peaks. The constant relating the simulation and experimental scales is calculated by determining the slope of least-squares fit line (constrained to pass through the origin) [73] for the experimental intensities versus simulated cross peak volumes of all cross peaks for all distinguishable pairs separated by four or more bonds, including all long-range, medium-range, and sequential peaks, as well as distant intraresidue pairs. Cross peaks between pairs separated by fewer than four bonds are eliminated because they are subject to significant TOCSY intensity contributions, evident from line shape distortion (see figure 3.4) [58]. All remaining experimental cross peak intensities (peak height) are assumed to be proportional to the volume, which assumes no significant line shape distortion. This procedure was repeated for the two water model simulations and for the two experimental solvation conditions (H_2O and ^2H_2O). The derived constant, multiplied by the weakest experimentally assigned peak intensity for the appropriate H_2O and ^2H_2O experiment, then provides an estimate of the noise level in the simulation. In the Results section, the simulated and experimental intensities are now comparable as a multiplicative factor of the noise.

3.3.4 NMR spectroscopy of Aβ$_{1-40}$ and Aβ$_{1-42}$

Synthetic Aβ$_{1-40}$ and Aβ$_{1-42}$ peptides were purchased from Anaspec (San Jose, CA). Prior to use peptides were resuspended at a concentration of 1 mg/mL in degassed 2 mM NaOH and sonicated gently to dissolve. This solution was then flash frozen in liquid N_2 and lyophilized. This process is designed to remove any potential fibril seeds from the peptide preparation [74]. NMR samples were obtained by dissolving NaOH treated peptides in 20mM sodium phosphate buffer at pH 7.2 in either 90% H_2O/10% ^2H_2O or 100% ^2H_2O at a protein concentration of 600μM and then filtering with a 0.22 μm Spin-X filter (Corning Inc., Corning, NY).

Chemical shift assignments and Overhauser cross peaks for Aβ$_{1-40}$ and Aβ$_{1-42}$ were collected from 2D ^1H-^1H TOCSY [57] and 2D ^1H-^1H NOESY experiments [75] measured on a 900MHz Bruker Avance II spectrometer. TOCSY spectra were collected with 40, 60 and 80 ms mixing times. H_2O NOESY spectra were collected with mixing times of 100 and 200ms. ^2H_2O NOESY spectra were collected with mixing times of 200ms. Aggregation of Aβ$_{1-40}$ and Aβ$_{1-42}$ NMR samples was followed by one-dimensional ^1H NMR spectra. As expected, Aβ$_{1-42}$ aggregated more rapidly than Aβ$_{1-40}$. Multiple data sets were collected on a single Aβ$_{1-40}$ over 3 days of measuring time whereas Aβ$_{1-42}$ samples were only used for 24 hours. More than 90% of the proton resonances of Aβ$_{1-40}$ have been assigned. Assignment of Aβ$_{1-42}$ is ongoing.

3.4 Results

3.4.1 Aβ$_{21-30}$ Chemical Shifts

A robust chemical shift calculation must describe the anisotropic shielding of the applied magnetic field for the given atom, a quantity that depends sensitively on the local electronic

structure environment. Even for folded proteins with a dominant native conformer, each atom type can exist in many different local environments, making an accurate calculation of chemical shifts for a protein quite a challenge. This problem is exacerbated in the case of disordered proteins that have a far greater diversity of conformations and interconverting local environments, thus hybrid empirical chemical shift packages that perform well for globular proteins such as SHIFTX [44-46] and SHIFTS [43] may inadequately predict chemical shifts of disordered proteins.

In Figure 3.1 the Cα, Cβ, Hα and HN experimental chemical shifts are shown for A$\beta_{21\text{-}30}$ as compared to the calculated chemical shifts over our simulated ensemble for different force fields. To obtain the experimental chemical shift values the reference value of the chemical shift of a random coil at 25°C was subtracted from the average shift for each amino acid [76]; while the carbon shifts show a very weak dependence on temperature and the 25°C random coil reference shift is used, the amide proton random coil shift reference for each amino acid is adjusted to a value appropriate for 10°C [77]. Together the chemical shift data emphasize that the peptide is largely unstructured.

In order to take into account any anticipated limitations of the chemical shift calculation algorithms the average chemical shifts were calculated for the unstructured subset of each simulated ensemble, comprising ~60% of the TIP4P-Ew and ~40% of the TIP3P ensemble. These values were used as a random coil reference state that was subtracted from the total simulated ensemble with the hope of some cancellation of errors. Figure 3.1 shows that the resulting ensemble averaged chemical shifts calculated with SHIFTS and SHIFTX are less than ~1ppm for C$_\alpha$ and C$_\beta$ shifts from the calculated random coil shifts, are similar between the TIP4P-Ew and TIP3P simulations, and show at best qualitative agreement with experiment. While the calculated amide proton shift values deviate significantly from the NMR values, this may be due to the inability of the theory to capture all of the chemically relevant contributions to amide shifts, a notorious problem [46], as opposed to an inadequacy of the structural ensemble.
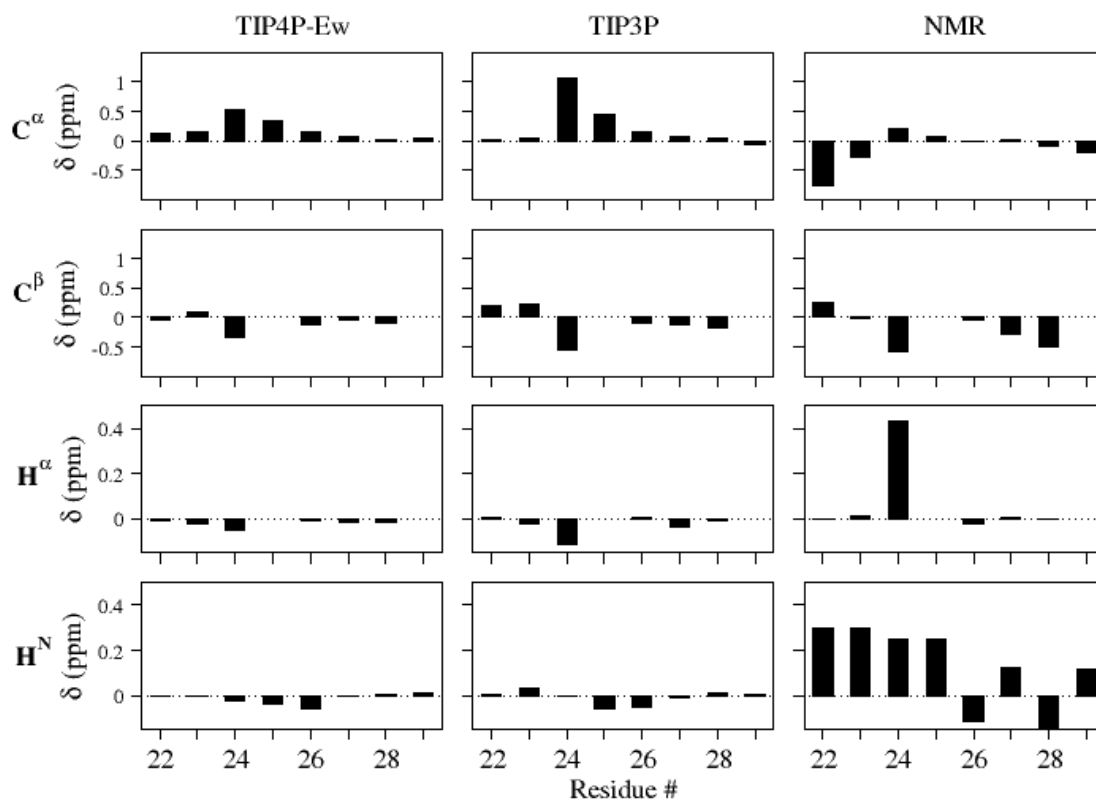
Figure 3.1 ^{13}C and ^{1}H chemical shifts from experiment and simulation for Aβ$_{21-30}$. Experimental NMR shifts are calculated as a difference from the peptide measurements and tabulated random coil values. For ^{13}C nuclei, which are insensitive for temperature changes, the tabulated 25°C random coil shifts were used, while for the amide proton shifts the random coil values of 10°C were used to account for the experimental temperature used in this study. The predicted chemical shifts are SHIFTS calculations averaged over the full ensemble after subtracting SHIFTS calculations averaged over the unstructured subpopulations (as defined by lack of DSSP secondary structure for all residues to represent the calculated random coil population).

3.4.2 Aβ$_{21-30}$ J-couplings

In Figure 3.2, the ^{3}J$_{HN-Hα}$ scalar coupling constants, measured from a high resolution COSY spectrum, are compared to the simulations using the dynamically uncorrected and harmonically corrected Karplus parameters (see Methods, section 3.3.3). It is apparent that the dynamically uncorrected and harmonically corrected Karplus parameter sets work equally well on this disordered system (Figure 3.2a). This is because disordered systems, with the greater likelihood that all dihedral angles sample a much larger range of φ values, more closely approach the uniform sampling limit of J = A/2+C. This limit is very similar between very different Karplus parameter sets and, for the sets examined here, are 4.9Hz and 5.05Hz for the corrected and uncorrected sets, respectively. In Figure 3.2b the simulations have well converged average scalar coupling values between trajectories, and the two different water models show only small differences, less than 1 Hz. The overall agreement between simulation and experiment is quite good, making clear that both are consistent with an ensemble that is largely random coil, consistent with the chemical shift data. This is especially clear when we compare the experimental ^{3}J$_{HN-Hα}$ values to that calculated from a single structure based on incorporating all of the ROESY restraints (shown in figure 3.6 and developed further below).
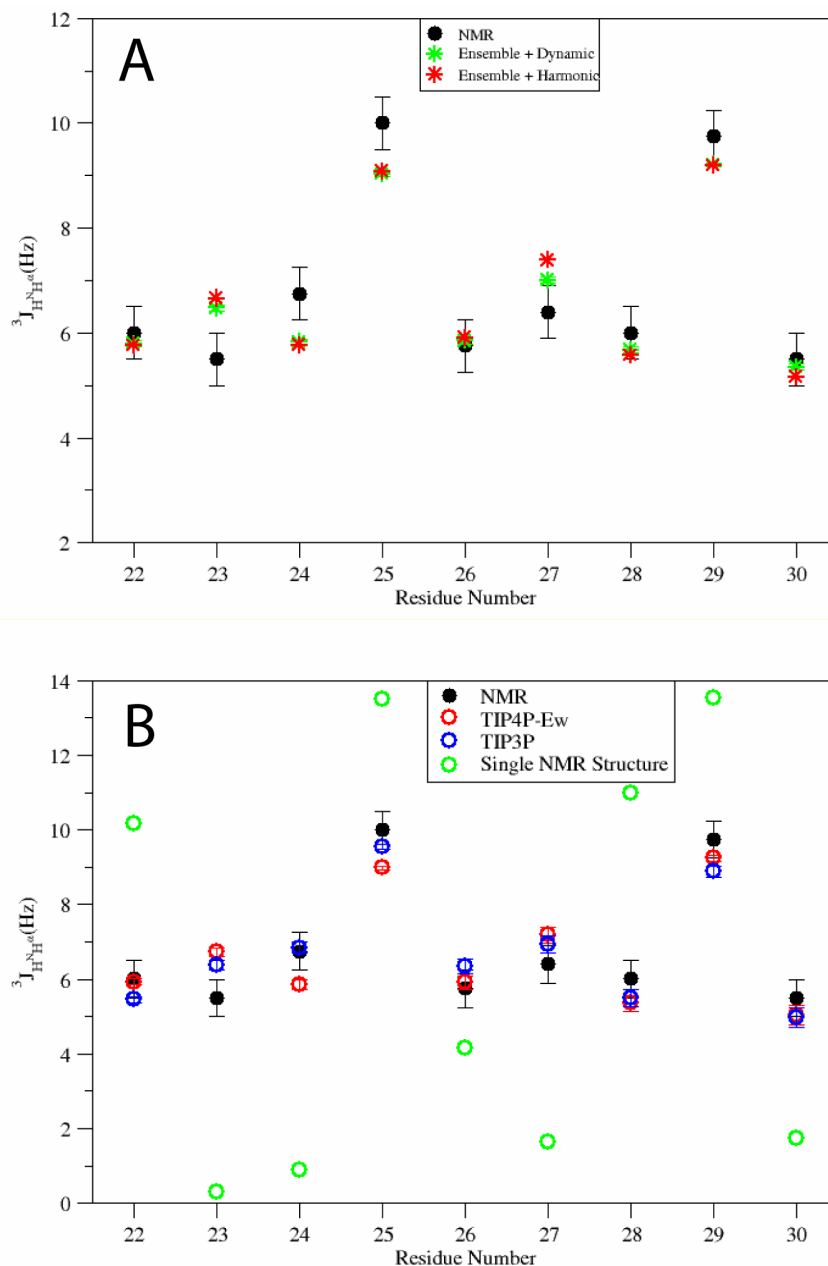
Figure 2. Comparison of experimental scalar coupling constant $^3J_{HN\text{-}H\alpha}$ and that calculated from the simulated ensembles. Error bars are experimental uncertainty for NMR values as well as simulated standard deviations calculated for trajectories split into three sections. The calculated coupling constants for the Hα protons of glycine are added to compare to experiment in which they are indistinguishable. (a) two parameterizations of the Karplus equation averaged over a single replica exchange ensemble using the TIP4P-Ew model. It is apparent that the dynamically uncorrected and harmonically corrected Karplus parameter sets work equally well on this disordered system. (b) average over the two independent replica exchange ensembles for different empirical force fields and compared to experimentally determined coupling constants. It is evident that an ensemble measurement gives far better agreement with the experimental $^3J_{HN\text{-}H\alpha}$ values than that calculated from a single structure based on incorporating all of the ROESY restraints (shown in Figure 6).

34

3.4.3 Aβ$_{21-30}$ T$_1$ and T$_2$

The experimental Cα spin-lattice and spin-spin relaxation times measured for the non-glycine positions of Aβ$_{21-30}$ are presented in Table 3.1. Since the chemical shifts for ^{13}Cα-^{1}Hα pairs for Glu22 and Lys28 overlap, T$_1$ and T$_2$ at these positions cannot be distinguished and the parameters are treated as an average of relaxation times at the two positions. As the combined data for these positions fit well to a single exponential form, T$_1$ and T$_2$ relaxation times are similar for these positions. The experimentally determined relaxation parameters vary by a maximum of 25% for non-terminal residues positions, indicating that the peptide does not contain significantly stiffer and/or slower moving regions on average.

T$_1$ and T$_2$ calculated from simulations of the peptide solvated with TIP4P-Ew model water show excellent agreement with the experimental values (Table 1). Non-terminal amino acid relaxation times are within ±10% for both T$_1$ and T$_2$. Consistent with experiment, the terminal amino acids show significantly longer relaxation times (indicating faster motion) than non-terminal ones, although the simulated relaxation times are larger than the experimental values. The discrepancy is greatest for the C-terminal alanines, which may indicate that the simulations predict less structure in this region than in the experiment, although viscosity differences at lower temperature in ^{2}H$_2$O may influence the dynamics. To test this hypothesis, a time scaled correlation function was used for all proton pairs to account for the larger solvent viscosity (see Methods). We find that the C-terminal T$_1$ and T$_2$ are now within 15% of the experimental values while the non-terminal relaxation parameters change by only a few percent (Table 3.1).

		A21	E22	D23	V24	S26	N27	K28	A30
T1	**Experiment**	415	298*	244	291	285	274	298*	475
at	**TIP4P-Ew**	492	292	268	272	274	273	287	530
500MHz	**Scaled**	457	276	253	260	262	261	275	485
	TIP3P	853	469	398	389	386	396	413	907
T2	**Experiment**	403	265*	269	230	235	241	265*	372
at	**TIP4P-Ew**	445	262	248	236	245	239	242	475
600MHz	**Scaled**	398	238	225	215	224	218	220	425
	TIP3P	860	474	403	386	386	393	402	858

Table 3.1. ^{13}C NMR Spin Relaxation times T$_1$ and T$_2$ for non-glycine C$_\alpha$ positions from experiment and TIP4P-Ew, TIP3P and time scaled TIP4P-Ew, in milliseconds. Glu22 and Lys28 resonances overlapped such that T$_1$ and T$_2$ could not be independently measured. The relaxation times calculated from the overlapped peaks are indicated with an asterisk.

T$_1$ and T$_2$ relaxation times calculated from the TIP3P simulations result are ~1.8 times larger than experimentally observed, for both terminal and non-terminal positions. This overestimate of relaxation time is a result of the faster dynamics of peptide motion in the TIP3P solvent. To demonstrate this difference the averaged vector autocorrelation function for the ^{13}C$_\alpha$ and H$_\alpha$ pair at the Val24 position for both the TIP3P and TIP4P-Ew simulation is presented in Figure 3.3. The TIP3P simulations result in time correlation functions with fitted decay parameters more than twice that of the TIP4P-Ew simulations. The faster peptide motion in TIP3P is likely a result of the unrealistically low viscosity (faster self-diffusion) properties of the TIP3P model, speeding up the peptide dynamics.
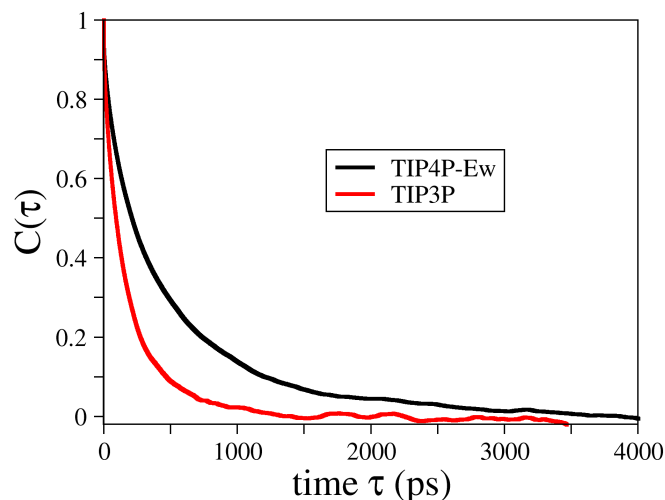
Figure 3.3. Normalized average vector time correlation function for Val24 C_α to H_α position for constant energy trajectories solvated with TIP4P-Ew and TIP3P. The TIP3P model (red) shows a dramatically faster decay for all vector time correlation functions relative to TIP4P-Ew, and we present this as an example.

3.4.4 Aβ$_{21-30}$ Experimental and Simulated ROESY Cross peaks

The 2D ROESY experiments in 90% H_2O:10% ^2H_2O and 100% ^2H_2O yielded a set of 155 assigned ROESY cross peaks. Although the majority of the cross peaks were for intra-residue (83) and sequential (44) pairs, 28 weak, medium range ROE interactions were also detected. These medium range ROE cross peaks comprise several i,i+2 and i,i+3 interactions and two extremely weak i,i+4 interactions; no longer range ROE cross peaks were observed and no strong patterns of α-helical or β-sheet contacts are evident. Lazo *et al*. report a long-range i,i+8 Glu22 Hα to Ala30 HN cross-peak [18], and more recent work from the same group assigns the peak to an overlap of the original Glu22 Hα to Ala30 HN interaction and i,i+2 Lys28 Hα to Ala30 HN in spectra collected at 500MHz. The observation of the long-range ROE is critical to their proposed NMR model, which is significantly collapsed. However, due to the higher resolution of the spectrum collected at 900 MHz, the cross peak is interpreted to be solely attributed to the i,i+2 contact between Lys28 Hα and Ala30 HN, as shown in Figure 3.4.

An additional set of long-range interactions between the Glu22 and Lys28 sidechains are reported by Lazo *et al*. In the higher field spectra, the small chemical shift difference between Hβ3 Lys28 and Hβ2 Glu22 are distinguishable (Figure 3.5). At lower field, these resonances would overlap very close to the limit of the resolution of the experiment and hence be difficult to distinguish. The 900MHz data contains no evidence of true Glu22 to Lys28 ROE interactions. It is likely that the previously reported long range cross peaks between these residues are due to misassignment of peaks too close to distinguish at lower field.

The new spectra show that a majority of the medium range cross peaks suggest turn or partial collapse structures for residues 22 through 27, although medium-range interactions are also seen from Ala21 to Asp23 and Val24, indicating the peptide backbone is not simply extended in the N-terminal region. Two additional pairs of interactions are found in the C-terminal region between the side chain and Hα of Lys28 with the backbone of Ala30, as well as the methyl group (Hβ) of Ala30 with both Hβ methylene atoms of Asn27.

36
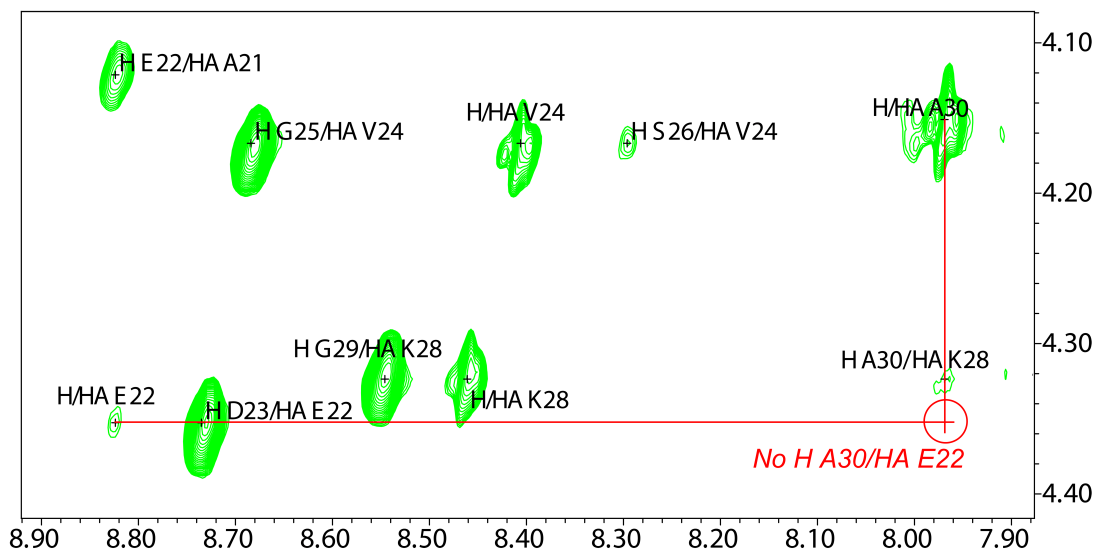
Figure 3.4. Fingerprint region of ROESY spectrum in H_2O of $A\beta_{21-30}$ demonstrating the cross peak interpreted by Lazo et al. and Grant et al. as Hα Glu22 with HN Ala30 in their 500MHz experiments is clearly resolved as only Hα Lys28 to HN Ala30 in this 900MHz experiment. It is also clear from the lineshape of intra-residue HN-Hα correlations that significant TOCSY artifacts do occur during the ROESY mixing time.
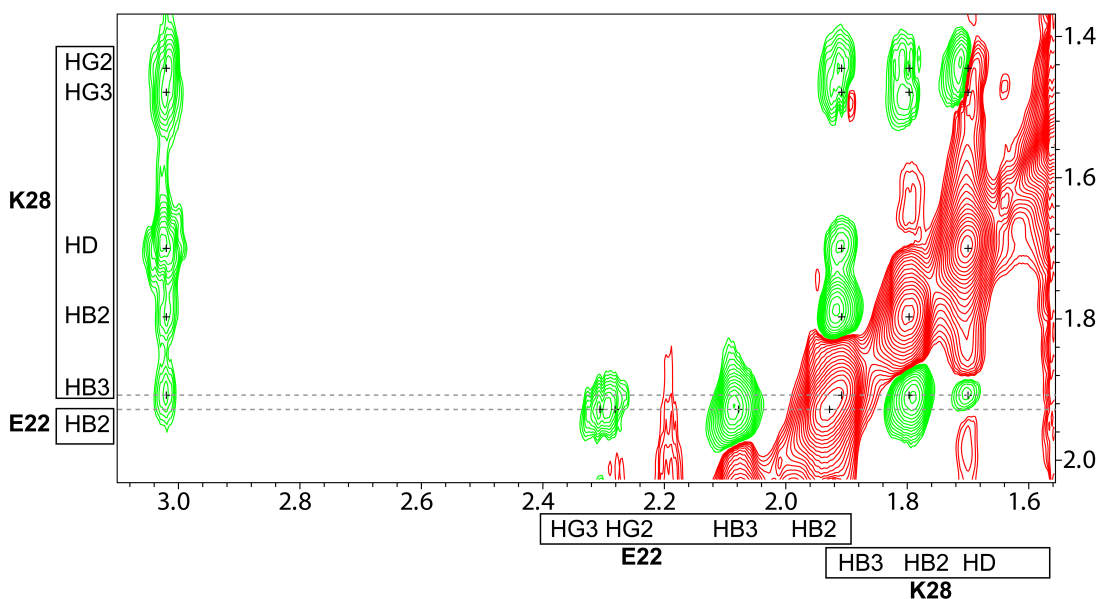
Figure 3.5 Hβ3 Lys28 has a nearly overlapping chemical shift with Hβ2 Glu22, potentially leading to cross peak misassignment in previous studies conducted at lower fields.

Predicted ROESY cross peaks from simulation were calculated for a 300 ms mixing time and compared to the cross peaks observed in the H_2O and ^2H_2O experimental spectra in Tables 3.2 and 3.3. Since medium and long range ROE interactions contain the most information about peptide structure, this discussion focuses only on these interactions and ignores intra residue and

sequential interactions in the comparisons of simulations to experiment. The top-ranked intensities for the ROE cross peaks predicted to be above the noise level are tabulated along with the experimentally measured value; a dash indicates no cross peak was seen experimentally, while "<1.0" denotes a potential experimental cross peak so weak that it cannot reliably be definitively assigned. Because we have two independent replica exchange calculations for each water model, we can explicitly evaluate convergence of the ROESY cross peaks. Although the r^1 (linear) pair distance averages vary by less than a few percent between the simulations and suggest good convergence, a small number of r^{-6} averaged pair distances vary by >30% and translate into noticeable changes in the corresponding peak volumes, and thus peak ranking. Therefore the few cross peaks where the combined ensemble cross peak predictions do not make evident the significantly different predictions from the independent simulations for both TIP4P-Ew and TIP3P are specifically highlighted below.

Tables 3.2 and 3.3 show that the TIP4P-Ew and TIP3P ensembles predict 12 and 14 of the experimentally assigned cross peaks, respectively, from the 900 MHz H_2O spectra, and 14 and 15, respectively, of the assigned cross peaks in the 800 MHz ^2H_2O experiment (note that there is redundancy of cross peaks between the two experiments so that there are only 28 distinct cross-peaks in total). The true positive cross peaks encompass i,i+2, i,i+3 and i,i+4 medium range cross peaks from across the entire peptide. Thus both TIP3P and TIP4P-Ew ensembles pick out the majority of the 28 experimentally observed medium-range ROE interactions from the 600 possible distinguishable medium and long-range interactions.

The predictions also show a number of "false positive" cross peaks *i.e.* cross peak predicted above the estimated noise level in the simulations that are not experimentally observed. In the TIP4P-Ew predictions, all of the false positive cross peaks, excepting one from the H_2O spectrum and four additional from the ^2H_2O spectrum, are i,i+2 or i,i+3 assignments that involve residues and regions of the chain that have other observed ROE interactions. This suggests that the simulations are bringing together the correct residues of the peptide and that these false positives are due to small differences in the detailed structural distances. In fact the r^{-6} distance averaging to estimate peak volumes is extremely sensitive to distance, such that changes of $2^{1/6}$ (~1.12) in distance can translate to a doubling of the calculated peak volume. In other words, false positive cross peaks predicted to be weak but still above the noise by our simulations may be just below the background noise in the experiment if the simulated distances are closer than in the experiment by only a factor of 1.12.

Therefore, it is reasonable to focus on false positives predicted to be more than a factor of 2 above the noise. In this case the differences in quality of the water models are revealed in that the TIP3P simulations have far too many false positives as strongly predicted cross peaks, suggesting a structural ensemble that is far more collapsed than observed experimentally. The TIP3P model shows far more false positives above the estimated noise, and poor ranking for what are the strongest experimentally observed peaks. By contrast TIP4P-Ew has fewer false positives and gets the intensity rankings right for the most prominent experimental peaks. In fact, the TIP4P-Ew simulations of the cross peak of strongest intensity between the Hβ methyl group of Ala21 and the Hγ methyl groups of Val24 led to the assignment of this peak in the experimental spectra which was initially hidden beneath an experimental artifact. Therefore, the remainder of the Results section focuses on the TIP4P-Ew simulations.

The more significant discrepancies between the experiments and simulations involve "false negatives"- experimentally observed cross peaks that are not predicted to be observable by simulation. Since the noise in the experiment is not uniform over the spectra, if we dip just below

38

the estimate for noise by a factor of 2, the TIP4P-Ew simulation predictions show an additional four H_2O and two ^2H_2O experimentally assigned cross peaks. Based on the highly sensitive cross peak intensity discussed above, these could be classified as true positives as well. However, five additional missing cross peaks are a full factor of 10 below the noise level, and hence are genuine false negatives. One such set of false negatives are the interactions between β methylene protons of Asp 23 and Ser 26, which involve four cross peaks seen experimentally. Since one of the possible four cross peaks of methylene pair interactions for these two residues is predicted among the top 10 of the ^2H_2O cross peaks predicted by the TIP4P-Ew simulations, it is apparent that the simulation is bringing together the correct areas of the side-chains, but not the correct detailed geometry in this region.

The experimental cross peaks, however, may imply more equivalent distances than are actually present in the underlying structural ensemble. Magnetization selectively ROE transferred through space to a single hydrogen in a methylene pair would subsequently be more evenly redistributed between the pair due to TOCSY type (through bond) transfer. TOCSY transfer is created by the rotating frame pulse during ROESY mixing, is difficult to remove completely, and is an effect that is not accounted for by the predicted ROESY spectra. These same arguments apply to the β methylene protons of Asn 27 and Ala 30, which involve four cross peaks seen experimentally, while the simulations find one of the cross peaks above the noise and two cross peaks just below the noise. The missing cross peak between protons is too spatially distant in the simulated ensemble suggesting either that the simulations have imperfect local geometry or that selectively ROE transferred magnetization has been redistributed through TOCSY type transfer among methylene protons.

Table 3.2. Proton cross-peak predictions for TIP3P and TIP4P-Ew for 900MHz in H_2O. Experimental intensities (Iexp) are normalized to the intensity of the weakest assigned peak. Simulation intensities (Isim) are normalized to experimental intensity as described in Methods. Experimental intensities are labeled as a dash if the peak is absent or "<1.0" if some evidence of a peak is present but is too weak to be assigned with confidence. Simulated intensities are marked with a # if the H_2O prediction is found in the ^2H_2O experiment.

TIP3P					
Isim	**Iexp**	**Proton 1**		**Proton 2**	
5.4	<1.0	HA	VAL24	HB3	ASN27
4.4	#1.8	HB3	ASP23	HB2	SER26
4.0	–	HA	ASP23	H	GLY25
3.0	2.6	HG	VAL24	H	SER26
2.8	15.7	HB	ALA21	HG	VAL24
2.7	<1.0	HG3	GLU22	H	VAL24
2.7	–	HB3	ASP23	H	SER26
2.7	–	HA	GLU22	HA	ASN27
2.6	2.9	HB	ALA21	H	ASP23
2.6	–	HA	VAL24	H	ASN27
2.6	1.3	HB3	ASP23	H	GLY25
2.4	4.9	HG3	GLU22	HG	VAL24
2.3	#1.4	HA	VAL24	HB2	ASN27
2.0	2.6	HA	VAL24	H	SER26
2.0	1.2	HA	GLY25	H	ASN27
1.6	–	HA	ASN27	H	GLY29
1.4	#3.8	HB3	GLU22	HG	VAL24
1.4	7.2	HG2	GLU22	HG	VAL24
1.4	2.8	HG	VAL24	HB3	ASN27
1.4	–	1HD2	ASN27	H	GLY29
1.4	–	HB3	GLU22	H	VAL24
1.3	–	HA	ASP23	HA	LYS28
1.3	–	HB3	ASP23	HE2	LYS28
1.2	–	HG	VAL24	H	ASN27
1.2	2.6	HG	VAL24	HB2	ASN27
1.2	–	H	VAL24	H	SER26
1.2	#2.4	HB2	GLU22	HG	VAL24
1.2	–	1HD2	ASN27	HA	GLY29
1.2	–	H	ASP23	HB2	SER26
1.2	–	HG2	GLU22	HA	ASN27
1.1	–	HG3	LYS28	H	ALA30
1.1	–	HB3	ASN27	H	GLY29
1.1	–	HB	ALA21	HA	ASN27
1.0	–	HA	GLU22	H	LYS28
1.0	–	H	ASP23	H	SER26
1.0	–	H	GLY25	H	ASN27
1.0	–	HB	ALA21	HB2	SER26
1.0	–	HA	ASP23	H	SER26
1.0	–	HG	VAL24	HA	GLY29
False Negatives:					
0.9	1.1	HA	SER26	H	LYS28
0.6	3.8	HA	GLY25	HD2	LYS28
0.6	3.8	HB3	ASN27	HB	ALA30
0.5	1.7	HA	LYS28	H	ALA30
0.3	2.5	HB3	LYS28	H	ALA30
0.3	1.1	HB2	ASN27	H	GLY29
0.0	3.3	HB2	ASN27	HB	ALA30

TIP4P-Ew

Isim	Iexp	Proton 1		Proton 2	
4.1	15.7	HB	ALA21	HG	VAL24
2.8	2.6	HG	VAL24	H	SER26
2.6	2.9	HB	ALA21	H	ASP23
1.9	4.9	HG3	GLU22	HG	VAL24
1.9	1.3	HB3	ASP23	H	GLY25
1.7	<1.0	HG3	GLU22	H	VAL24
1.6	–	2HD2	ASN27	H	GLY29
1.3	–	HB3	GLU22	H	VAL24
1.3	#2.4	HB2	GLU22	HG	VAL24
1.3	–	HA	GLU22	HA	ASN27
1.3	1.2	HA	GLY25	H	ASN27
1.3	#1.4	HA	VAL24	HB2	ASN27
1.2	<1.0	HA	VAL24	HB3	ASN27
1.2	#1.8	HB3	ASP23	HB2	SER26
1.2	2.6	HA	VAL24	H	SER26
1.1	–	HG	VAL24	H	ASN27
1.1	–	2HD2	ASN27	HA	GLY29
1.1	#3.8	HB3	GLU22	HG	VAL24
1.1	7.2	HG2	GLU22	HG	VAL24
1.0	–	HB	VAL24	H	SER26
1.0	–	HB3	ASP23	H	SER26
False Negatives:					
0.9	1.1	HB2	ASN27	H	GLY29
0.7	2.8	HG	VAL24	HB3	ASN27
0.6	2.6	HG	VAL24	HB2	ASN27
0.5	3.3	HB2	ASN27	HB	ALA30
0.4	1.7	HA	LYS28	H	ALA30
0.4	3.8	HA	GLY25	HD2	LYS28
0.2	2.5	HB3	LYS28	H	ALA30
0.0	1.1	HA	SER26	H	LYS28
0.0	3.8	HB3	ASN27	HB	ALA30

Table 3.3. Proton cross-peak predictions at 800MHz in ^2H_2O. See Table 3.2 for additional details.

TIP3P					
Isim	Iexp	Proton 1		Proton 2	
11.7	<1.0	HA	VAL24	HB3	ASN27
10.0	1.8	HB3	ASP23	HB2	SER26
6.1	–	HA	GLU22	HA	ASN27
5.8	7.4	HB	ALA21	HG	VAL24
5.0	5.1	HG3	GLU22	HG	VAL24
5.0	1.4	HA	VAL24	HB2	ASN27
3.0	3.8	HB3	GLU22	HG	VAL24
3.0	1.8	HG	VAL24	HB3	ASN27
3.0	4.3	HG2	GLU22	HG	VAL24
2.9	–	HA	ASP23	HA	LYS28
2.7	–	HB3	ASP23	HE2	LYS28
2.5	–	HG2	GLU22	HA	ASN27
2.5	1.6	HG	VAL24	HB2	ASN27
2.5	2.4	HB2	GLU22	HG	VAL24
2.4	–	HB	ALA21	HA	ASN27
2.3	–	HA	VAL24	HA	ASN27
2.2	–	HB	ALA21	HB2	SER26
2.1	–	HG	VAL24	HA	GLY29
2.0	–	HA	GLY25	HA	GLY29
2.0	3.1	HG	VAL24	HE2	LYS28
1.9	–	HB3	ASP23	HG3	LYS28
1.8	–	HA	GLU22	HA	GLY29
1.8	1.3	HG	VAL24	HB2	SER26
1.8	–	HG3	GLU22	HB2	SER26
1.8	–	HG3	LYS28	HB	ALA30
1.6	–	HG2	GLU22	HB2	ASN27
1.6	–	HG	VAL24	HD2	LYS28
1.6	0.8	HA	GLU22	HG	VAL24
1.6	–	HG3	GLU22	HA	ASN27
1.5	–	HG	VAL24	HB3	SER26
1.5	–	HB	ALA21	HB	ALA30
1.5	–	HB2	SER26	HE2	LYS28
1.5	–	HA	GLY25	HE2	LYS28
1.4	–	HB2	GLU22	HA	GLY29
1.3	–	HA	ASP23	HB2	SER26
1.3	1.8	HA	GLY25	HD2	LYS28
1.3	–	HB3	ASP23	HB2	LYS28
1.3	–	HB	VAL24	HB3	ASN27
1.2	–	HB	ALA21	HB2	ASN27
1.2	–	HA	SER26	HE2	LYS28
1.2	–	HG	VAL24	HA	ASN27
1.2	–	HA	GLU22	HA	LYS28
1.2	3.9	HB3	ASN27	HB	ALA30
1.2	–	HB	ALA21	HB3	ASP23
1.2	2.2	HB3	ASP23	HB3	SER26
1.2	–	HB3	ASP23	HD2	LYS28
1.2	–	HB2	SER26	HG3	LYS28
1.1	–	HB3	ASP23	HA	GLY25
1.1	–	HE2	LYS28	HB	ALA30
1.1	–	HB	ALA21	HB2	LYS28
1.1	–	HB3	GLU22	HA	ASN27
1.1	–	HB3	ASP23	HG2	LYS28
1.1	–	HG	VAL24	HA	SER26
1.0	–	HA	VAL24	HE2	LYS28
1.0	–	HB	ALA21	HA	ASP23
1.0	–	HG3	GLU22	HB2	ASN27
1.0	–	HB	ALA21	HA	ALA30
1.0	–	HB3	GLU22	HB2	ASN27

```
 1.0      -    | HB    ALA21  | HA    VAL24
 1.0      -    | HG2   GLU22  | HA    GLY29
False Negatives:
 0.9     1.0   | HG    VAL24  | HA    LYS28
 0.3     1.0   | HB3   SER26  | HG3   LYS28
 0.1     2.5   | HB2   ASN27  | HB    ALA30
 0.0     0.8   | HB2   ASP23  | HB3   SER26
-0.7     1.7   | HB2   ASP23  | HB2   SER26
```

		TIP4P-Ew			
Isim	**Iexp**	**Proton 1**		**Proton 2**	
8.0	7.4	HB	ALA21	HG	VAL24
3.9	5.1	HG3	GLU22	HG	VAL24
3.1	–	HA	GLU22	HA	ASN27
3.0	1.4	HA	VAL24	HB2	ASN27
2.8	1.8	HB3	ASP23	HB2	SER26
2.7	<1.0	HA	VAL24	HB3	ASN27
2.7	2.4	HB2	GLU22	HG	VAL24
2.3	3.8	HB3	GLU22	HG	VAL24
2.2	4.3	HG2	GLU22	HG	VAL24
2.2	0.8	HA	GLU22	HG	VAL24
1.8	–	HB	VAL24	HB2	ASN27
1.6	1.3	HG	VAL24	HB2	SER26
1.5	1.8	HG	VAL24	HB3	ASN27
1.5	3.1	HG	VAL24	HE2	LYS28
1.5	–	HB3	ASP23	HA	LYS28
1.4	–	HG	VAL24	HA	SER26
1.4	1.6	HG	VAL24	HB2	ASN27
1.4	–	HA	VAL24	HB2	LYS28
1.4	–	HG2	GLU22	HA	ASN27
1.3	–	HB	ALA21	HB2	SER26
1.3	–	HA	SER26	HE2	LYS28
1.3	–	HB	ALA21	HB3	ASP23
1.3	–	HB	ALA21	HA	ASP23
1.2	–	HG	VAL24	HA	ASN27
1.2	–	HA	GLY25	HE2	LYS28
1.1	0.9	HG	VAL24	HA	LYS28
1.1	–	HB2	SER26	HE2	LYS28
1.0	–	HA	ASP23	HB2	SER26
1.0	–	HB	ALA21	HA	SER26
1.0	2.5	HB2	ASN27	HB	ALA30
1.0	–	HG	VAL24	HB3	SER26

```
False Negatives:
 0.9     1.8   | HA    GLY25  | HD2   LYS28
 0.4     1.0   | HB3   SER26  | HG3   LYS28
 0.1     2.2   | HB3   ASP23  | HB3   SER26
-0.1     0.8   | HB2   ASP23  | HB3   SER26
-0.1     3.9   | HB3   ASN27  | HB    ALA30
-0.3     1.7   | HB2   ASP23  | HB2   SER26
```

3.4.5 Aβ21-30 Structure from Simulation and Experiment

Determining a single structure from multiple weak, medium range ROE experimental cross peaks for disordered peptides can be misleading since it is also possible that ROE cross peaks arise from multiple distinct populations or perhaps a fully heterogeneous structural ensemble. Typical structure calculations on structured proteins assume that weak NOE or ROE cross peaks correspond to large (~4.5-6.0Å) upper distance restraints on a single well-defined structural population, and hence all the restraints should be applied simultaneously. Though the

MD simulations show that the $A\beta_{21-30}$ peptide ensemble involves significant disorder and hence it is inappropriate to use the standard structure determination methods [78-80], I calculated a single best-fit structure for purposes of comparison.

The set of restraints was used to calculate 1000 structures of $A\beta_{21-30}$ and the 50 lowest energy structures were aligned. For the entire peptide, the superposition of the final 50 structures has an RMSD of 0.81 ± 0.42 Å for the backbone atoms and 1.15 ± 0.61 Å for all heavy atoms. Sixteen of the twenty structures are within 1.0Å RMSD for all heavy atoms, forming the dominant cluster whose lowest energy structure has three major bends, pinching together Asp23 to the Ser26, Gly25 to Lys28 and Asn27 to Ala30 (Figure 3.6). These bends are created by 12 unique (i, i+3) and (i,i+4) ROE interactions between residues Asp23-Ser26, Val24-Asn27, Val24-Lys28, Gly25-Lys28, and Asn27-Ala30. Many of the cross peaks for these interactions are very weak and detectable in the ROESY spectrum only when the peptide is dissolved in 100% ^2H_2O, and there is only one cross peak between Gly25 HAs and Lys28 HDs detected. As a result, one of the three bends in the minimized structure backbone is stabilized by one backbone hydrogen bond, and a second by a single Lys28 NH_3^+ interaction, but most of the pair distances for which ROE interactions are observed are not restrained by any favorable intermolecular interactions. This lack of stabilizing interactions was also evident for the minimized structures for the $A\beta_{21-30}$ peptide model proposed previously by Lazo *et al*.
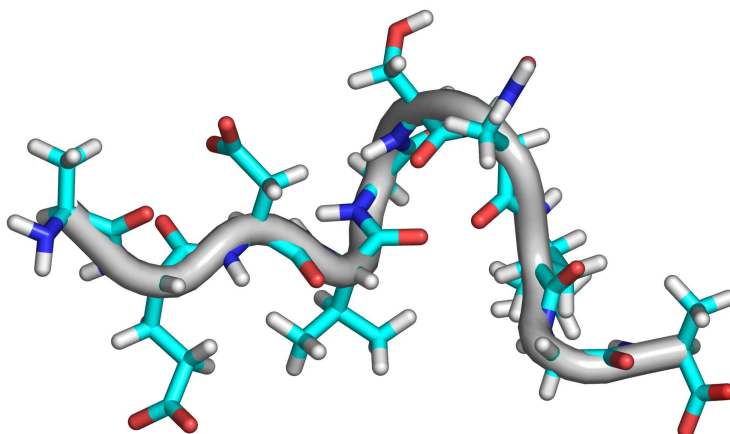

Figure 3.6. Representative structure from restraint energy minimized ensemble, calculated by simultaneously applying all observed ROE interactions.

Given the good quality of the TIP4P-Ew simulated experimental observables presented above, the experimental picture can be refined by analyzing the underlying simulated ensemble for structural populations. Standard clustering by RMSD is not informative due to the lack of order in much of the ensemble. The two first principal components in a PCA analysis yielded only a single large population, giving little information about the underlying structure. Since hydrophobic collapse is unlikely to be the dominant structuring force in a peptide that has only a single large aliphatic residue and no aromatic residues, hydrogen bond interactions are hypothesized to stabilize the structure that gives rise to the ROE interactions. Thus the most useful tool for partitioning the structures is the patterning of the hydrogen bonding and electrostatic interactions that may persist in sub-populations. It is important to emphasize that 60-65% of the TIP4P-Ew population is unstructured by these metrics, however the remaining 35-

40% of the population does explain the presence of the medium-range restraints observed in the experimental ROESY cross peaks with a large degree of success.

In the TIP4P-Ew simulations, the most populated hydrogen-bond is between the sidechain carboxyl oxygens of Asp23 and the backbone amide of G25 (Figure 7A). This interaction is found in 17% of the ensemble, and stabilizes the backbone dihedral angles near to those of a type I β-turn for residues Asp23 to Ser26. A true type I β-turn, which is found in 5% of the population (Figure 3.7B), is defined by backbone hydrogen bonding between the backbone carbonyl oxygen of residue i (Asp23) with the backbone amide hydrogen of residue $i+3$ (Ser26), resulting in the amide hydrogen of $i+2$ (Gly25) pointing toward the sidechain of i, this is most populated hydrogen bond found in the simulated ensemble.

The peptide maintains a conformation near a type I β-turn if interactions between the carboxyl oxygens of Asp23 to the sidechain hydroxyl of Ser26 are found, consistent with the observed ROEs, which brings together the sidechain hydrogens of Asp23 and Ser26. If structures with at least one of these Asp23 to Ser26 interactions are also considered, the type I β-turn population increases from 5% to 14% of the ensemble, much higher than any other turn region in the peptide. While the TIP3P structural ensemble shows this turn in approximately 35% of the ensemble, its overrepresentation in the structured population most likely contributes to the poor agreement of ROE cross peak volumes when compared to experiment.

Despite the prevalence of structures with type I β-turn structure, the β-turn does not nucleate a β-sheet, which would be characterized by backbone contacts between Glu22 and Asn27. Instead, in the structures observed, the backbone amide of Asn27 hydrogen bonds to the backbone carbonyl oxygen of Asp23, precluding the formation of β-sheet structure. Furthermore, this interaction brings the Val24 methyl hydrogens near the Asn27 sidechain hydrogens, accounting for those observed ROE interactions. Other smaller groups of covarying hydrogen bonds are observed, including simultaneous interactions between the backbone carbonyl of Val24 with the backbone amide of both Asn27 and Lys28, bringing in proximity the Val24 sidechain with Asn27 sidechain as observed in the ROESY spectra.

Salt bridge formation between Asp23 and Lys28, observed in the solid state NMR structure of the Aβ_{1-40} fibril, is found in 7% of the ensemble (Figure 3.8), while the competing salt bridge between Glu22 and Lys28 is found 1.5% of the time. Together these salt bridge structures are observed with comparable frequency to the turn populations, but the salt-bridge contacts do not stabilize either hydrogen bonding structure or close proximity of other protons in the intervening region. In principle a close contact involving the basic lysine amine hydrogens may be observed in a ROESY spectrum, but salt bridges, unless stable enough to prevent proton exchange at neutral pH of the basic amine on the NMR timescale as can occur in well-folded proteins, do not typically bring together NMR visible protons.
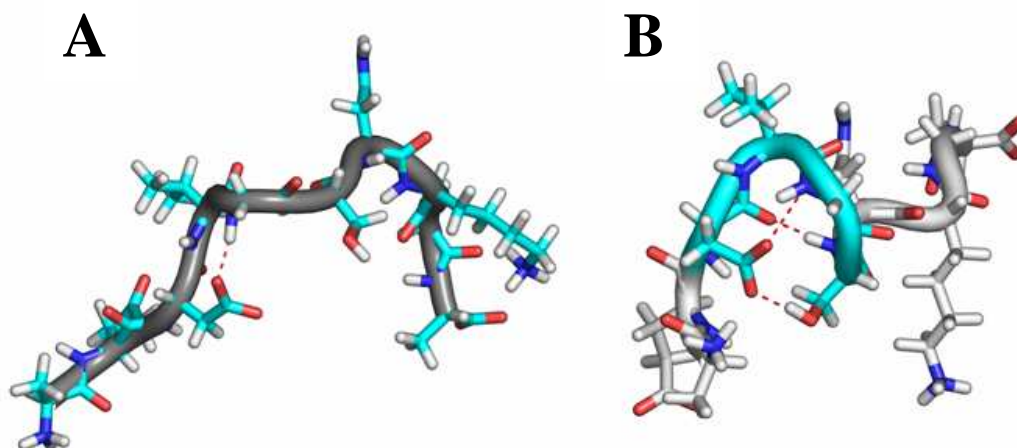
Figure 3.7 Representative structures showing (A) the most populated N-terminal hydrogen bond (dotted red) and (B) the hydrogen bonds and electrostatic interactions (dotted red) stabilizing the type I β-turn (cyan) centered at Val24 and Gly25.
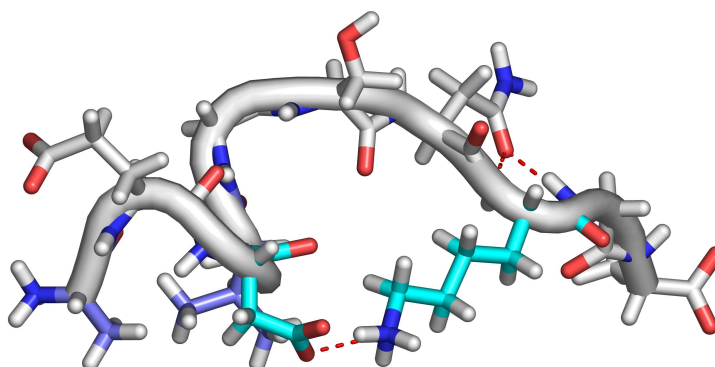

Figure 3.8 Asp23 to Lys28 salt-bridge (shown by red dotted line between aqua amino acids) found in 7% of the TIP4P-Ew ensemble, and Ala21 HB to Val24 HG (both in light blue) van der Waals contacts which give rise to the strongest observed ROE interaction

3.4.6 Aβ$_{1-40}$ and Aβ$_{1-42}$ NMR Assignment and Preliminary Structural Analysis

      Work with Aβ peptides has long been plagued by the difficulty of obtaining preparations of the peptide and even more so by batch to batch differences in biophysical properties such as aggregation kinetics [81-83]. In recent years, these problems have been largely circumvented when working with Aβ peptides prepared via solid-phase peptide synthesis by resuspending the HPLC purified peptide in weakly basic solution, sonicating gently to dissolve and then freezing and lyophilizing the solution [74]. This procedure is designed to prevent a freshly prepared

solution of Aβ from passing through its isoelectric point, approximately pH 5, due to residual trifluoroacetic acid, a byproduct of both SPPS and HPLC purification.

Approximately 90% of the proton resonance assignments of Aβ$_{1-40}$, containing assignments for each residue, were obtained from 2D NOESY and TOCSY spectra and a complete set of NOESY cross peaks assigned. The quality of the experimental data is demonstrated by figures 3.9 and 3.10, which depict the fingerprint region of a TOCSY spectrum collected in H$_2$O and the amide-methyl correlations from a NOESY spectrum collected in H$_2$O, respectively. In total, ~3700 cross peaks have been assigned, with ~1000 medium range cross peaks (where the two residues are between 2 and 4 residues apart in sequence) and 162 long range NOE's. These contacts are summarized schematically in figure 3.11 and will be used in an analysis similar to that described above in section 3.4.4 upon the completion of the more computationally intense molecular dynamics calculations of the longer Aβ$_{1-40}$ peptide.
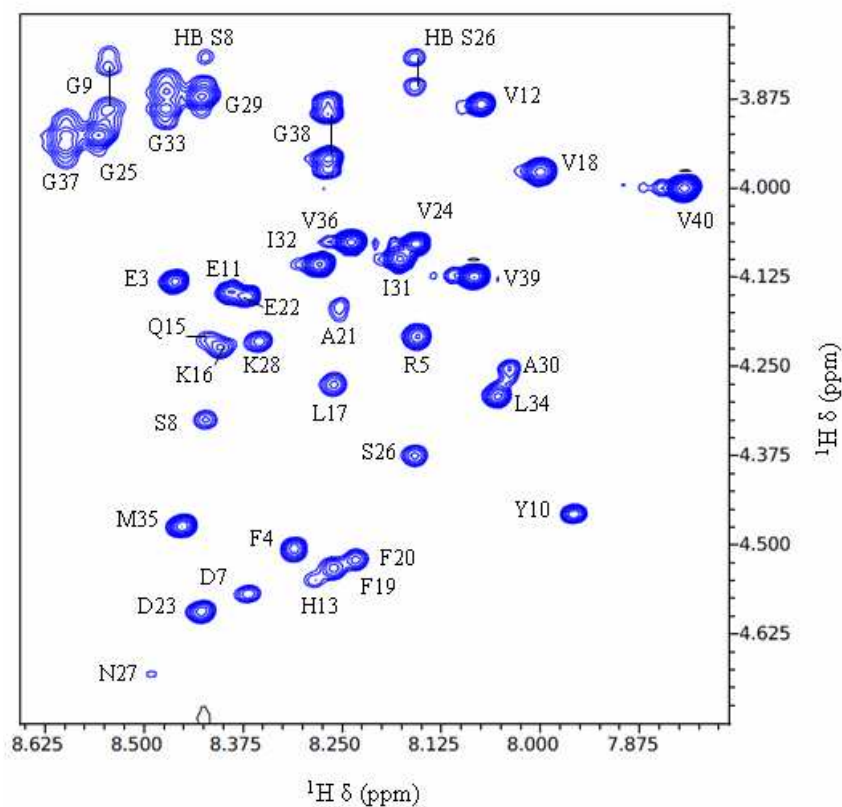

Figure 3.9. The fingerprint region of a ^{1}H-^{1}H TOCSY spectrum collected at 900MHz in H$_2$O. The majority of correlations are between amide and alpha protons. Correlations between amide and beta protons are labeled HB. All residues have been assigned.
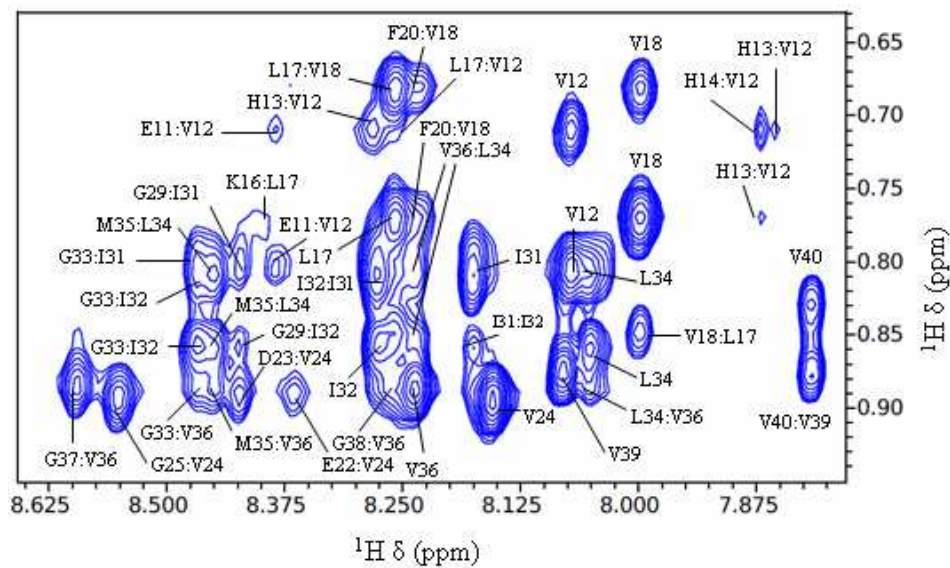
Figure 3.10. The amide-methyl region from a ^1H-^1H NOESY spectrum collected at 900MHz in H_2O. Intraresidue correlations are denoted by a single label. The first residue listed in each label contains the amide, the second contains the methyl group.
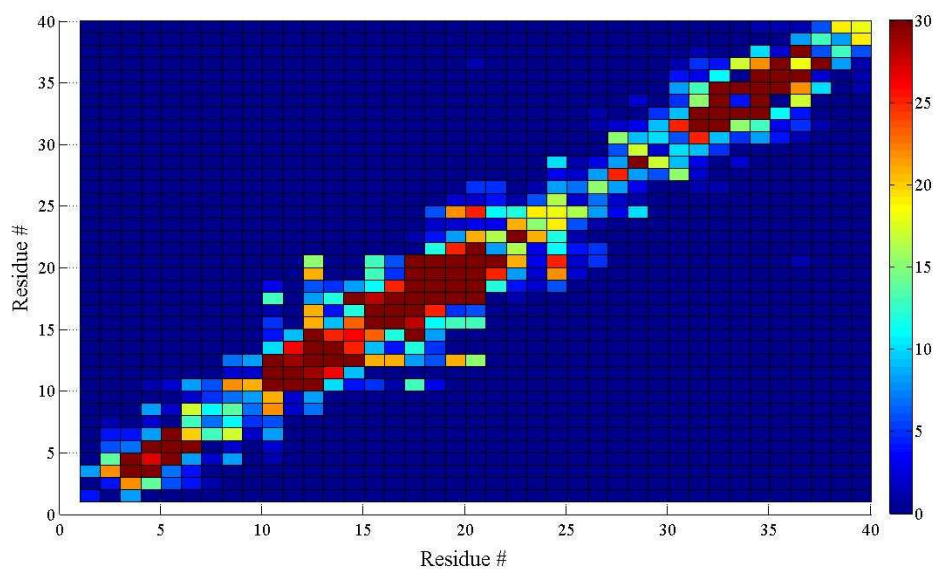

Figure 3.11. Schematic NOE contact map of $A\beta_{1-40}$. The color represents the number of NOE's between a given set of residues. The cut off point of 30 contacts was chose as it represents the maximum number of (i, i+2) contacts observed.

3.5 Discussion

According to the high field NMR experiments, $A\beta_{21\text{-}30}$ shows no long-range and only weak medium-range ROE interactions, demonstrating none of the features of a protein with a single native state. It is therefore evident that the presence of a singly populated collapsed structure incorporating a unique bend due to a i,i+8 Glu22 H_α to Ala30 HN cross peak and i,i+6 Glu22 sidechain to Lys28 sidechain cross peak reported by Lazo *et al.* and Grant *et al.* is incorrect on two levels. The first is a problem of misassignment in their lower resolution ROESY spectra in which they propose a i,i+8 interaction that is instead revealed to be a weak i,i+2 interaction and a i,i+6 interaction where intra-residue peaks are overlapping, both of which were distinguishable by the higher resolution 800 and 900 MHz spectra used here.

More significantly, peptides and disordered protein systems should not conform to a single dominant structure, and should only be described by appropriate ensembles. The poor quality of a single structure becomes evident when simultaneously applying all ROE interactions as distance restraints to give a minimized structure with surprisingly few favorable inter-residue interactions. Since there are only a few restraints that are all "weak" and hence provide only a loose upper bound on the distance, all of the restraints are satisfied by pair distances near this bound. This loose bounding results in a dominant structure with no consensus stabilizing contacts, hydrogen bonds, regular secondary structures or reverse turns.

In this work it is demonstrated that there is a good match between the TIP4P-Ew/ff99SB simulated and experimentally observed structure and dynamics, as measured by ROE cross peaks and ^{13}C relaxation, indicating that these simulations faithfully approximate the ensemble of structures investigated by the experiments, allowing them to be used to describe the full structural ensemble diversity. The structural ensemble of the $A\beta_{21\text{-}30}$ peptide involves a majority (~60%) of unstructured population according to lack of any DSSP secondary structure assignment or hydrogen-bonding patterns. However the remaining minority population involves ~14% population of β-turn structure centered at Val24 and G25 bringing together Asp23 and S26. The simulations also indicate that the Asp23 to Lys28 salt bridge, important to the fibril structure, is formed in ~7% of the ensemble. Finally a separate set of structures populated only by a few percent brings together the Val24 and Asn27 regions. It is clear the $A\beta_{21–30}$ system is highly disordered, and that the ~5-15% of distinct structural populations measured here have been overrepresented in the previous experimental work on this system due to both the misinterpretation of a medium-range Overhauser crosspeak as a long-range contact as well as the more fundamental error of utilizing NOE's observed within a small, flexible peptide to restrain a structure calculation that attempts to satisfy all Overhauser contacts simultaneously.

The preliminary NMR work described here with $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ demonstrates that as the homonuclear correlation spectra of the peptides are of sufficient quality to assign the peptide resonances it will be possible, albeit more computationally challenging, to study the conformational ensemble of the full-length peptide monomers in a similar manner to that which was used to study $A\beta_{21\text{-}30}$. Experimental and simulation data reported in the literature suggest that the structural populations increase upon lengthening of the Aβ peptide to larger fragments [84-87], this is corroborated by the greater number of Overhauser contacts observed in the spectra of $A\beta_{1\text{-}40}$, including ~150 long range contacts (those over four residues apart in sequence), indicating that there are significantly more structural motifs in this peptide than observed in $A\beta_{21\text{-}30}$. No long-range contacts were identified in the Overhauser spectra of $A\beta_{21\text{-}30}$. The Overhauser contacts observed in the spectra of $A\beta_{1\text{-}40}$, depicted schematically in figure 3.11,

indicate that there is an increased interaction within the region encompassing residues 12-22 relative to the remainder of the peptide. The data do not reveal contacts between this 'sticky' region in the center of the peptide and the C-terminus as described in the MD work by Sgourakis, *et al.* [17]. However, a full characterization of the structural ensemble that monomeric $A\beta_{1-40}$ adopts in solution will require completed simulations of the peptide for a comparison with the NMR data. These calculations are underway.

3.6 Conclusion

In the typical, historical application of biomolecular NMR, that of structure determination, the biomolecule of interest most often occupies a single, well-defined conformation. It is becoming increasingly clear that proteins that are intrinsically disordered or contain disordered regions are involved in important biological processes [53, 88, 89]. While protein dynamics involving the interconversion between stable states or the fluctuation about a stable state have been well studied by NMR [90], despite significant progress in the field [91-93], the interpretation of NMR observables for disordered protein systems combining multiple structural constraints for a system with significant disorder often leads to an inadequate description of the ensemble diversity [94]. Molecular dynamics simulations of disordered systems have the opposite challenge; the ensemble is directly observable but the accuracy is difficult to compare to experiments and is often biased by the empirical force fields employed, which were often developed with respect to well-folded proteins. While simulations of folded proteins in their native state have been shown to quantitatively reproduce NMR relaxation parameters [62, 95], many empirical force fields have had difficulty reproducing the behavior of simple disordered systems such as trialanine [96].

In the work described here, the more recent generation protein and water force fields employed allow for the prediction of NMR experimental observables to judge the accuracy of the MD simulations. The Amber ff99SB/TIP4P-Ew simulations do not predict an overly collapsed structure with regions where there are no experimentally observed ROEs, while false negatives are weak but often just buried in the noise, reflecting the difficulty of converging $1/r^6$ distance averages. Additionally, the TIP4P-Ew water model correctly predicts the observed ^{13}C relaxation times. Despite the difficulty of converging the $1/r^6$ average, the double averaging of predicted Overhauser cross peaks provide a much more sensitive measure of underlying structure in a mostly unstructured ensemble. By contrast, chemical shifts and scalar coupling observables are simple population averages, making prediction of these quantities from simulations much easier to converge, but far less structurally informative. In general, the comparison of molecular simulations with a variety of NMR observables is well suited to the study of proteins with a significant degree of disorder and Overhauser cross peaks are the most sensitive and informative observable to compare in systems such as Aβ.

3.7 Acknowledgments

3.8 Abbreviations used

AD   Alzheimer's Disease
APP   Amyloid Precursor Protein
FAD   Familial Alzheimer's Disease
MD   Molecular Dynamics
CSA   Chemical Shift Anisotropy
WT   Wild-Type

3.9 References

1.   Goedert, M. and M.G. Spillantini, *A century of Alzheimer's disease.* Science, 2006. **314**: p. 777-781.
2.   Turnera, P.R., et al., *Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory.* Progress in Neurobiology, 2003. **70**: p. 1-32.
3.   Priller, C., et al., *Synapse Formation and Function Is Modulated by the Amyloid Precursor Protein.* The Journal of Neuroscience, 2006. **26**: p. 7212-7221.
4.   Bugiani, O., et al., *An Italian type of HCHWA.* Neurobiology of Aging, 1998. **19**: p. S238.
5.   Kayed, R., et al., *Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis.* Science, 2003. **300**(486-489).
6.   Walsh, D.M. and D.J. Selkoe, *Aß Oligomers – a decade of discovery.* Journal of Neurochemistry, 2007. **101**: p. 1172-1184.
7.   Mucke, L., et al., *Highlevel neuronal expression of abeta 1-42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation.* The Journal of Neuroscience, 2000. **20**: p. 4050–4058.
8.   Petkova, A.T., et al., *Structural investigations of Alzheimer's beta-amyloid fibrils by solid state NMR.* Biophysical Journal, 2004. **86**(506A).
9.   Balbach, J.J., et al., *Supramolecular structure in full-length Alzheimer's beta-amyloid fibrils: Evidence for a parallel beta-sheet organization from solid-state nuclear magnetic resonance.* Biophysical Journal, 2002. **83**: p. 1205-1216.
10.   Petkova, A.T., et al., *A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**: p. 16742-16747.
11.   Sticht, H., et al., *Structure of amyloid A4-(1-40)-peptide of Alzheimer's disease.* European Journal of Biochemistry, 1995. **233**: p. 293-298.
12.   Shao, H., et al., *Solution structures of micelle-bound amyloid beta-(1-40) and beta-(1-42) peptides of Alzheimer's disease.* Journal of Molecular Biology, 1999. **285**: p. 755-773.
13.   Coles, M., et al., *Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is?* Biochemistry, 1998. **37**: p. 11064-11077.
14.   Crescenzi, O., et al., *Solution structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain.* European Journal of Biochemistry, 2002. **269**: p. 5642-5648.

15. Yan, Y. and C. Wang, *Abeta42 is more rigid than Abeta40 at the C terminus: implications for Abeta aggregation and toxicity.* Journal of Molecular Biology, 2006. **364**: p. 853-862.

16. Yan, Y., et al., *Methyl dynamics of the amyloid-beta peptides Abeta40 and Abeta42.* Biochemical and Biophysical Research Communications, 2007. **362**: p. 410-414.

17. Sgourakis, N.G., et al., *The Alzheimer's peptides Abeta40 and 42 adopt distinct conformations in water: a combined MD / NMR study.* Journal of Molecular Biology, 2007. **368**: p. 1448-1457.

18. Lazo, N.D., et al., *On the nucleation of amyloid beta-protein monomer folding.* Protein Science, 2005. **14**: p. 1581-1596.

19. Sciarretta, K.L., et al., *A beta 40-Lactam(D23/K28) models a conformation highly favorable for nucleation of amyloid.* Biochemistry, 2005. **44**: p. 6003-6014.

20. Petkova, A.T., W.M. Yau, and R. Tycko, *Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils.* Biochemistry, 2006. **45**: p. 498-512.

21. Grabowski, T.J., et al., *Novel amyloid precursor protein mutation in an Iowa family with dementia and severe cerebral amyloid angiopathy.* Annals of Neurology, 2001. **49**: p. 697-705.

22. Hendriks, L., et al., *Presenile-dementia and cerebral-hemorrhage linked to a mutation at codon-492 of the beta-amyloid precursor protein gene.* Nature Genetics, 1992. **1**: p. 218-221.

23. Kamino, K., et al., *Linkage and mutational analysis of familial Alzheimer-Disease kindreds for the APP gene region.* American Journal of Human Genetics, 1992. **51**: p. 998-1014.

24. Walsh, D.M., et al., *In vitro studies of amyloid beta-protein fibril assembly and toxicity provide clues to the aetiology of Flemish variant (Ala(692) -> Gly) Alzheimer's disease.* Biochemical Journal, 2001. **355**: p. 869-877.

25. Van Nostrand, W.E., et al., *Pathogenic effects of cerebral amyloid angiopathy mutations in the amyloid beta-protein precursor*, in *Alzheimer's Disease: Vascular Etiology and Pathology*. 2002, New York Academy of Sciences: New York. p. 258-265.

26. Van Nostrand, W.E., et al., *Localization of a fibrillar amyloid beta-protein binding domain on its precursor.* The Journal of Biological Chemistry, 2002. **277**: p. 36392-36398.

27. Levy, E., et al., *Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type.* Science, 1990. **248**: p. 1124-1126.

28. Grant, M.A., et al., *Familial Alzheimer's disease mutations alter the stability of the amyloid beta-protein monomer folding nucleus.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**: p. 16522-16527.

29. Baumketner, A., et al., *Structure of the 21-30 fragment of amyloid beta-protein.* Protein Science, 2006. **15**(1239-1247).

30. Borreguero, J.M., et al., *Folding events in the 21-30 region of amyloid beta-protein (Abeta) studied in silico.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **2005**: p. 6015-6020.

31. Cruz, L., et al., *Solvent and mutation effects on the nucleation of amyloid beta-protein folding.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(18258-18263).

32.   Chen, W., N. Mousseau, and P. Derreumaux, *The conformations of the amyloid-beta (21-30) fragment can be described by three families in solution.* The Journal of Chemical Physics, 2006. **125**: p. 084911.

33.   Jarrett, J.T., E.P. Berger, and P.T. Lansbury, Jr, *The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease.* Biochemistry, 1993. **32**: p. 4693-4697.

34.   Selkoe, D.J., *The molecular pathology of Alzheimer's disease.* Neuron, 1991. **6**: p. 487-498.

35.   Wang, R., et al., *The profile of soluble amyloid beta protein in cultured cell media. Detection and quantification of amyloid beta protein and variants by immunoprecipitation-mass spectrometry.* The Journal of Biological Chemistry, 1996. **271**: p. 31894–31902.

36.   El-Agnaf, O.M., et al., *Oligomerization and toxicity of betaamyloid-42 implicated in Alzheimer's disease.* Biochemical and Biophysical Research Communications, 2000. **273**: p. 1003-1007.

37.   McGowan, E., et al., *Abeta42 is essential for parenchymal and vascular amyloid deposition in mice.* Neuron, 2005. **47**: p. 191-199.

38.   Zhang, Y., et al., *Selective cytotoxicity of intracellular amyloid beta peptide1-42 through p53 and Bax in cultured primary human neurons.* The Journal of Cell Biology, 2002. **156**: p. 519-529.

39.   Jonghe, C.D., et al., *Pathogenic APP mutations near the gamma-secretase cleavage site differentially affect Abeta secretion and APP C-terminal fragment stability.* Human Molecular Genetics, 2001. **10**: p. 1665-1671.

40.   Suzuki, N., et al., *An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants.* 1994.

41.   Hou, L., et al., *Solution NMR studies of the A beta(1-40) and A beta(1-42) peptides establish that the Met35 oxidation state affects the mechanism of amyloid formation.* Journal of the American Chemical Society, 2004. **126**: p. 1992-2005.

42.   Riek, R., et al., *NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, A beta(1-40)(ox) and A beta(1-42)(ox).* European Journal of Biochemistry, 2001. **268**: p. 5930-5936.

43.   Neal, S., et al., *Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts.* Journal of Biomolecular NMR, 2003. **26**(215-240).

44.   Osapay, K. and D.A. Case, *Analysis of proton chemical shifts in regular secondary structure of proteins.* Journal of Biomolecular NMR, 1994. **4**: p. 215-230.

45.   Xu, X.P. and D.A. Case, *Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database.* Journal of Biomolecular NMR, 2001. **21**: p. 321-333.

46.   Moon, S. and D.A. Case, *A new model for chemical shifts of amide hydrogens in proteins.* Journal of Biomolecular NMR, 2007. **38**: p. 139-150.

47.   Vuister, G.W. and A. Bax, *Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins.* Journal of the American Chemical Society, 1993. **115**: p. 7772-7777.

48.   Karplus, M., *Contact Electron-Spin Coupling of Nuclear Magnetic Moments.* Journal of Chemical Physics, 1959. **30**(1): p. 11-15.

49. Lipari, G. and A. Szabo, *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity.* Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.

50. Bax, A. and D.G. Davis, *Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy.* Journal of Magnetic Resonance, 1985. **65**(2): p. 355-360.

51. Bothnerby, A.A., et al., *Structure Determination of a Tetrasaccharide - Transient Nuclear Overhauser Effects in the Rotating Frame.* Journal of the American Chemical Society, 1984. **106**(3): p. 811-813.

52. Peter, C., X. Daura, and W.F. van Gunsteren, *Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations.* Journal of Biomolecular Nmr, 2001. **20**(4): p. 297-310.

53. Daura, X., et al., *Unfolded state of peptides*, in *Unfolded Proteins*. 2002, ACADEMIC PRESS INC: San Diego. p. 341-360.

54. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**: p. 277-293.

55. Johnson, B.A. and R.A. Blevins, *NMR VIEW - A computer-program for the visualization and analysis of NMR data.* Journal of Biomolecular NMR, 1994. **4**: p. 603-614.

56. Keller, R., *Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment*. 2004/2005, Swiss Federal Institute of Technology Zurich: Zurich. p. 159.

57. Braunschweiler, L. and R.R. Ernst, *Coherence Transfer by Isotropic Mixing - Application to Proton Correlation Spectroscopy.* Journal of Magnetic Resonance, 1983. **53**(3): p. 521-528.

58. Hwang, T.L. and A.J. Shaka, *Cross Relaxation without Tocsy - Transverse Rotating-Frame Overhauser Effect Spectroscopy.* Journal of the American Chemical Society, 1992. **114**(8): p. 3157-3159.

59. Guntert, P., *Automated NMR structure calculation with CYANA.* Methods in Molecular Biology, 2004. **278**: p. 353-378.

60. Case, D.A., et al., *The Amber Biomolecular Simulation Programs.* Journal of Computational Chemistry, 2005. **26**: p. 1668-1688.

61. Palmer, A.G., M. Rance, and P.E. Wright, *Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-13 Heteronuclear NMR Spectroscopy.* Journal of the American Chemical Society, 1991. **113**: p. 4371-4380.

62. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**: p. 712-725.

63. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water.* Journal of Chemical Physics, 1983. **79**: p. 926-935.

64. Horn, H.W., et al., *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew.* Journal of Chemical Physics, 2004. **120**: p. 9665-9678.

65. Andersen, H.C., *Molecular-Dynamics Simulations at Constant Pressure and-or Temperature.* Journal of Chemical Physics, 1980. **72**: p. 2384-2393.

66. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* Journal of Chemical Physics, 1993. **98**: p. 10089-10092.

67. Geyer, C.J. and E.A. Thompson, *Annealing Markov-Chain Monte-Carlo with Applications to Ancestral Inference.* Journal of the American Statistical Society, 1995. **90**: p. 909-920.

68. Hukushima, K. and K. Nemoto, *Exchange Monte Carlo method and application to spin glass simulations.* Journal of the Physical Society of Japan, 1996. **65**: p. 1604-1608.

69. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**: p. 2577-2637.

70. Bruschweiler, R. and D.A. Case, *Adding Harmonic Motion to the Karplus Relation for Spin-Spin Coupling.* Journal of the American Chemical Society, 1994. **116**(24): p. 11199-11200.

71. Habeck, M., W. Rieping, and M. Nilges, *Bayesian estimation of Karplus parameters and torsion angles from three-bond scalar couplings constants.* Journal of Magnetic Resonance, 2005. **177**(1): p. 160-165.

72. Lindorff-Larsen, K., R.B. Best, and M. Vendruscolo, *Interpreting dynamically-averaged scalar couplings in proteins.* Journal of Biomolecular Nmr, 2005. **32**(4): p. 273-280.

73. Feenstra, K.A., et al., *A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides.* Journal of Biomolecular NMR, 2002. **23**: p. 181-194.

74. Teplow, D.B., *Preparation of Amyloid Beta-Protein for Structural and Functional Studies.* Methods in Enzymology, 2006. **413**: p. 20-33.

75. Lippens, G., C. Dhalluin, and J.M. Wieruszeski, *Use of a Water Flip-Back Pulse in the Homonuclear Noesy Experiment.* Journal of Biomolecular Nmr, 1995. **5**(3): p. 327-331.

76. Wishart, D.S., et al., *H-1, C-13 and N-15 Chemical-Shift Referencing in Biomolecular NMR.* Journal of Biomolecular NMR, 1995. **6**: p. 135-140.

77. Merutka, G., H.J. Dyson, and P.E. Wright, *Random Coil H-1 Chemical-Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series Ggxgg.* Journal of Biomolecular NMR, 1995. **5**: p. 14-24.

78. Torda, A.E., R.M. Scheek, and W.F. van Gunsteren, *Time-dependent distance restraints in molecular-dynamics simulations.* Chemical Physics Letters, 1989. **157**: p. 289-294.

79. Torda, A.E., R.M. Scheek, and W.F. van Gunsteren, *Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat.* Journal of Molecular Biology, 1990. **214**: p. 223-235.

80. Nanzer, A.P., W.F. van Gunsteren, and A.E. Torda, *Parameterization of time-averaged distance restraints in MD simulations.* Journal of Biomolecular NMR, 1995. **6**: p. 313-320.

81. Soto, C., et al., *Fibrillogenesis of synthetic amyloid-Beta peptides is dependent on their initial secondary structure.* Neuroscience Letters, 1995. **200**: p. 105-108.

82. Simmons, L.K., et al., *Secondary structure of amyloid Beta peptide correlates with neurotoxic activity in vitro.* Molecular Pharmacology, 1994. **45**: p. 373-379.

83. Howlett, D.R., et al., *Aggregation state and neurotoxic properties of Alzheimer Beta-amyloid peptide.* Neurodegeneration, 1995. **4**: p. 23-32.

84. Tarus, B., J.E. Straub, and D. Thirumalai, *Dynamics of Asp23-Lys28 salt-bridge formation in A beta(10-35) monomers.* Journal of the American Chemical Society, 2006. **128**(50): p. 16159-16168.

85. Massi, F., et al., *Simulation study of the structure and dynamics of the Alzheimer's amyloid peptide congener in solution.* Biophysical Journal, 2001. **80**(1): p. 31-44.

86. Kirschner, D.A., et al., *Synthetic Peptide Homologous to Beta-Protein from Alzheimer-Disease Forms Amyloid-Like Fibrils In vitro.* Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(19): p. 6953-6957.

87. Barrow, C.J., et al., *Solution Conformations and Aggregational Properties of Synthetic Amyloid Beta-Peptides of Alzheimers-Disease - Analysis of Circular-Dichroism Spectra.* Journal of Molecular Biology, 1992. **225**(4): p. 1075-1093.

88. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm.* Journal of Molecular Biology, 1999. **293**(2): p. 321-331.

89. Xie, Q., et al., *The Sequence Attribute Method for Determining Relationships Between Sequence and Protein Disorder.* Genome Informatics, 1998. **9**: p. 201-213.

90. Brüschweiler, R., *New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins.* Current Opinion in Structural Biology, 2003. **13**: p. 175–183.

91. Marsh, J.A., et al., *Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation.* Protein Science, 2006. **15**: p. 2795-2804.

92. Choy, W.Y. and J.D. Forman-Kay, *Calculation of ensembles of structures representing the unfolded state of an SH3 domain.* Journal of Molecular Biology, 2007. **308**: p. 1011-1032.

93. Bezsonova, I., et al., *Oxygen as a paramagnetic probe of clustering and solvent exposure in folded and unfolded states of an SH3 domain.* Journal of the American Chemical Society, 2007. **129**: p. 1826-1835.

94. Marsh, J.A., et al., *Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure.* Journal of Molecular Biology, 2007. **367**: p. 1494-1520.

95. Case, D.A., *Molecular dynamics and NMR spin relaxation in proteins.* Accounts of Chemical Research, 2002. **35**: p. 325-331.

96. Mu, Y.G., D.S. Kosov, and G.Z. Stock, *Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments.* Journal of Physical Chemistry B, 2003. **107**: p. 5064-5073.

Chapter 4

Towards the Mechanism of Cofactor Insertion to Nitrogenase: The Solution Structure of the SAM domain of NafY and NMR characterization of the Cofactor binding site

4.1 Summary

Catalysis at the active site of the molybdenum nitrogenase requires a complicated iron-sulfur cluster containing an additional molybdenum atom coordinated by R-homocitrate that is able to store electrons for the reduction of diatomic nitrogen to ammonia. This cofactor, FeMo-co, is synthesized by a tightly regulated biochemical pathway involving the activities of at least ten proteins. When this pathway is impaired the cell is no longer able to synthesize FeMo-co and accumulates an inactive, $\alpha_2\beta_2\gamma_2$ hexameric form of dinitrogenase that lacks FeMo-co. Upon FeMo-co insertion *in vitro* the γ-subunit dissociates from the reconstituted dinitrogenase. In *Azotobacter vinelandii* this γ-protein is NafY. It has been proposed that NafY participates in the final steps of dinitrogenase maturation by acting as both a chaperone that stabilizes the FeMo-co deficient apo-dinitrogenase and as a FeMo-co carrier protein. NafY contains two domains: an N-terminal domain whose specific function had not been determined (n-NafY) and a C-terminal domain of known structure (core-NafY) that is able to bind FeMo-co. In this chapter I focus on the biochemical characterization of n-NafY, the structure determination of n-NafY by NMR, as well as insights into the process of FeMo-co binding by NMR. The structure of n-NafY reveals that it belongs to the sterile alpha motif (SAM) family of protein interaction domains and in this work it is shown that the N-terminal domain of NafY mediates the interaction with apo-dinitrogenase. As FeMo-co is paramagnetic I have been able to use the broadening of NMR resonances in close proximity to the bound cofactor to localize the binding site to a helical lobe of core-NafY.

4.2 Introduction

Nitrogenase is the enzyme in diazotrophs that catalyzes the reduction of diatomic nitrogen to ammonia and is composed of two oxygen sensitive metalloproteins, dinitrogenase reductase and dinitrogenase. Dinitrogenase reductase (NifH) provides electrons to dinitrogenase (NifDK), where the reduction of diatomic nitrogen occurs [1-4]. Holo-dinitrogenase contains two metal centers termed P-clusters as well as two copies of one of the most complicated metal clusters known, FeMo-co, which is able to store electrons sequentially delivered to dinitrogenase by NifH [2, 5]. FeMo-co is a $Mo\text{-}Fe_7\text{-}S_9\text{-}X$ metal cluster coordinated at its molybdenum atom by homocitrate where X denotes a central atom in the iron-sulfur core which is thought to be either a carbon, nitrogen or oxygen atom [6]. The structure of FeMo-co is shown below in figure 4.1. The biosynthesis of FeMo-co is performed by a tightly regulated biochemical pathway involving the activities of at least ten proteins. The current model of this process in *A. vinelandii* is
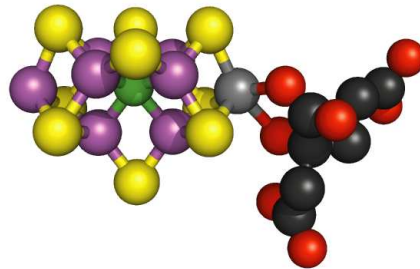
summarized below in figure 4.2 adapted from [7].

Figure 4.1. The structure of FeMo-co. Sulfur atoms are shown in yellow, Iron in lavender, Molybdenum in gray, Oxygen in red and carbon in black. There is insufficient experimental data to distinguish the identity of central atom in green between Carbon, Nitrogen or Oxygen.
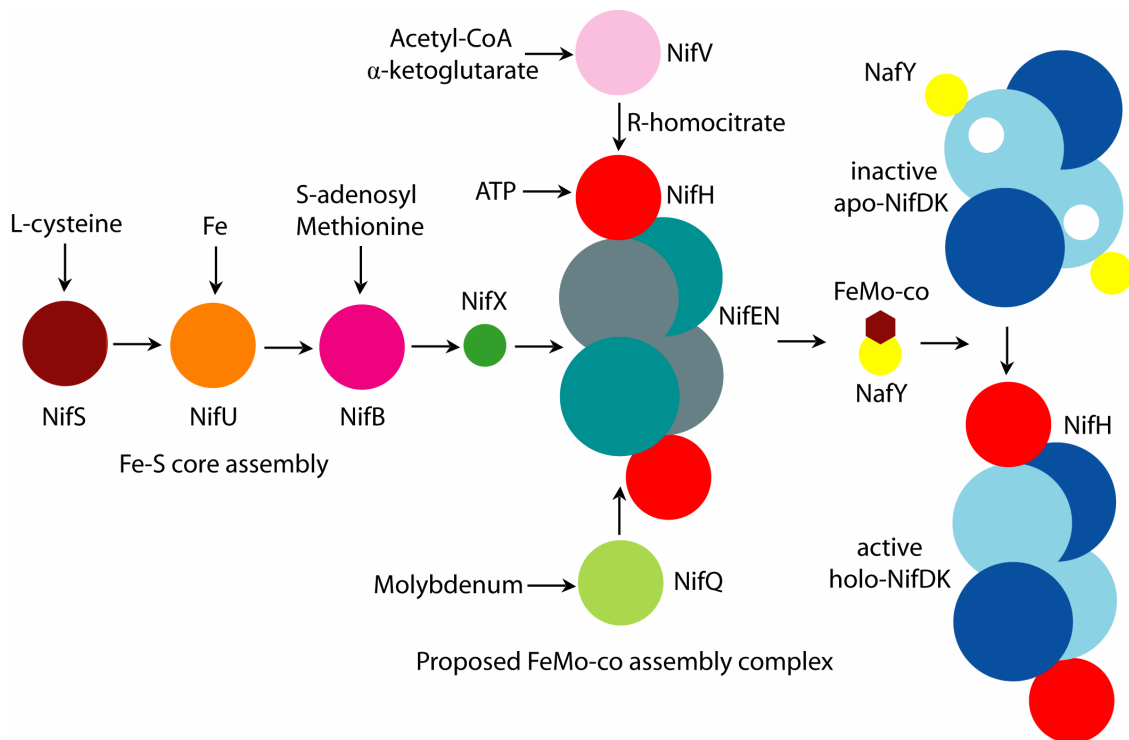

Figure 4.2. Overview of the biosynthesis of FeMo-co. *In vitro*, the synthesis of FeMo-co requires only NifB, NifEN, and NifH when supplied with Fe^{2+}, S^{2-}, MoO_4^{2-}, R-homocitrate, S-adenosyl methionine, and ATP. *In vivo*, the proteins NifS and NifU provide sulfur and iron to NifB, which requires S-adenosyl methionine for the synthesis of NifB-co, an essential precursor to FeMo-co. NifX delivers NifB-co is to the complex of NifEN and NifH where molybdenum and R-homocitrate are added to the cluster after delivery by NifV and NifQ, respectively. NafY is bound to apo-NifDK but has also been proposed as the carrier of FeMo-co from NifEN and to apoNifDK. Upon FeMo-co insertion, NafY disassociates from NifDK yielding the active protein.

As outlined in figure 4.2, the *nif*B gene product is required for the synthesis of an essential precursor to FeMo-co called NifB-co. Strains of both *Azotobacter vinelandii* (UW45) and *Kiebsiella pneumoniae* (UN1655) that contain a mutant, inactive form of the *nif*B gene accumulate an apo-dinitrogenase which contains its P-clusters yet lacks FeMo-co. This apo form of NifDK co-purifies with another approximately 20 kDa protein, termed the γ subunit, which dissociates from dinitrogenase upon the addition of FeMo-co [8-11]. The *K. pneumoniae* γ-subunit has been shown to be the product of *nifY* [11] whereas the *A. vinelandii* γ-subunit has been shown to be the protein NafY which is not encoded by a previously sequenced *nif* (nitrogen fixation) gene [12-14].

While NafY is not required for the biosynthesis of an active *A. Vinelandii* dinitrogenase *in vitro* [15], a mutant strain of *A. vinelandii* which does not express active NafY (UW45) exhibits a 2-fold reduction in growth rate under stressing nitrogen fixing conditions, in which the growth medium is not supplemented with 10mM sodium molybdate. A strain which lacks both active NafY and NifX (UW146) exhibits a 5-fold reduction in growth rate under stress. Furthermore, when activated *in vitro* by the addition of FeMo-co, cellular extracts obtained from UW146 *A. vinelandii* develop only half the nitrogenase activity of extracts from the UW45 strain [12]. These observations suggest that NafY plays a role in stabilizing the apo-form of dinitrogenase.

Recent work has shown that NafY possesses an independently folding C-terminal domain (core-NafY). Core-NafY's structure was determined by X-ray crystallography and it has been shown that, unlike full-length NafY, core-NafY does not bind tightly enough to apo-dinitrogenase to co-electrophorese. Core-NafY has also been shown to bind one molecule of FeMo-co with high affinity through contacts including a surface exposed histidine residue (H121) [16, 17]. These observations have led to the current hypothesis that NafY acts as FeMo-co chaperone insertase, binding FeMo-co and maintaining apo-dinitrogenase in a conformation that is optimal for cofactor insertion.

In this work the solution structure of the N-terminal NafY domain (n-NafY) is presented and its biochemical function is further characterized, demonstrating that it binds tightly to apo-dinitrogenase and inhibits the formation of an active dinitrogenase by blocking the insertion of FeMo-co. I have also utilized NMR spectroscopy to refine the region of core-NafY is involved in coordinating FeMo-co. These results enhance the understanding of NafY's role in dinitrogenase maturation.

4.3 Materials and Methods

4.3.1 Bacterial strains and growth conditions

A. vinelandii strains DJ (wild type) [18], UW146 [19], and DJ1143 [20] have been previously described. *A. vinelandii* cells were cultivated in a 150-L fermentor (New Brunswick Scientific IF-150) in 120-L batches of modified Burk's medium with limiting ammonium (80 µg/ml N). Alternatively, *A. vinelandii* cells from a nitrogen-replete medium (400 µg/ml N) grown in 1 L baffled flasks were washed with and resuspended into 250 ml of nitrogen-depleted modified Burk's medium. Growth in the presence of molybdate, *nif* derepression, and cell breakage have been described [21]. *Escherichia coli* DH5α and BL21 (pREP-4) strains were grown in Luria-Bertani medium at 37˚C with shaking (250 rpm). For growth of *E. coli* on plates, medium solidified with 1.5% agar was used. Antibiotics were used at standard concentrations [22].

4.3.2 Expression of NafY and NafY domains in E. coli

Overproduction of NafY and NafY domains in *E. coli* was accomplished by fusing a glutathione-S-transferase (GST) domain to the respective protein sequences, in vector pRHB153 (Hernandez *et al.*, 2007). A Tobacco Etch Virus (TEV) protease cleavage site was included in between the tag and the protein of interest. The *A. vinelandii nafY* gene was PCR amplified with oligonucleotides 247 (5'-GCCCAGGAGGCCGCATATGGTAACCCCCGTGAAC -3') and 248 (5'-ACGGCCCTGGCGCTCGAGCGCCACGGCGATCAGC -3'), and ligated into the *Nde*I-*Xho*I sites of plasmid pRHB153 to generate plasmid pRHB159. The *A. vinelandii nafY* N-terminal domain (n-NafY) was amplified with oligonucleotides 498 (5'-GGAGGCCGCATATG GTAACCCCCGTGAACATGAG -3') and 499 (5'-CCGGCACGCGCTCGAGCTTTCACACC GGCTGG -3'), and ligated into the *Nde*I-*Xho*I sites of plasmid pRHB153 to generate plasmid pRHB261. The *A. vinelandii nafY* core domain was amplified with oligonucleotides 500 (5'-CC AGCCGGTGAAGCATATGCGCGTGCCGGAAGG -3') and 502 (5'- CCTCGTCGTCGGACT CGAGTCAGAAGCGCACG -3'), and ligated into the *Nde*I-*Xho*I sites of plasmid pRHB153 to generate plasmid pRHB263. *E. coli* BL21 (pREP-4) cells were transformed with plasmid pRHB159, pRHB261 or pRHB263, respectively, and over expression of the GST-tagged fusion protein was induced by addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) to the culture medium. Isotopic labeling was performed according to Marley *et al.* [23]. Triply labeled NafY-ΔC was prepared by growth in a modified, more strongly buffered M9 medium containing 13.0 g/L KH_2PO_4, 10.0 g/l K_2HPO_4, 9.0 g/L Na_2HPO_4 and 2.4 g/L K_2SO_4.

4.3.3 Purification of NafY-ΔC, n-NafY, and core-NafY

E. coli cells induced for the overexpression of the corresponding polypeptide were resuspended in buffer A (10 mM sodium phosphate, 1.8 mM potassium phosphate buffer, pH 7.3, 140 mM NaCl, 2.7 mM KCl) and disrupted at 12,000 psi in a French press. Cell debris was removed by two centrifugation steps at 25,000 x *g* for 40 min. The cell-free extracts were loaded onto a 25-ml glutathione (GSH)-Sepharose 4 Fast Flow affinity column (Amersham Biosciences). The column was washed with 250 ml of buffer A supplemented with 1% Triton X-100, and the GST-tagged fusion protein was eluted in 75 ml of buffer B (50 mM Tris-HCl pH 8.0, 10 mM reduced glutathione (GSH)). The GST tag was cleaved by addition of 10 μg of pure TEV protease per mg of GST-tagged protein, followed by 2h of incubation at 30 ℃. The protein mixture was then subjected to a series of chromatographic steps: gel-filtration on Sephadex G-25 (to remove glutathione), affinity to GSH-sepharose (to remove the GST tag), and Ni-affinity chromatography (to remove the TEV protease). A typical purification procedure yielded 10 mg of protein from 6g of *E. coli* cell paste. NafY-ΔC or the NafY domains were estimated to be >98% pure based on SDS-PAGE analysis. The resulting protein was concentrated by ultrafiltration using an Amicon cell equipped with YM10 membrane or an Ultrafree-0.5 device equipped with a PBCC Biomax 5 kDa membrane (Millipore) to ca. 1 mM for NMR analysis.

4.3.4 Purification of other components

His-tagged apo-dinitrogenase was purified from *A. vinelandii* strain DJ1143 as previously described [24]. NifH was purified as described [25] with minor modifications. Methods for the purification of FeMo-co [26] and NifB-co [27] have been previously described, the cofactors were kindly provided by Dr. Luis Rubio.

4.3.5 Complex formation between FeMo-co and NafY or core-NafY

In a typical procedure, the NafY protein or core-NafY was diluted in 50 mM MOPS pH 7.4, 2 mM dithionite (DTH) (buffer C) in an Amicon cell equipped with an YM10 membrane. Aliquots of purified FeMo-co were added in sequential steps to the solution containing the protein, maintaining an N-methylformamide (NMF) final concentration below 3% in order to prevent protein denaturation. After concentration of the mixture in the Amicon cell, the sample was re-diluted in buffer C and a new aliquot of FeMo-co was added. Sequential steps of dilution and concentration were repeated until a 1:1.5 NafY:FeMo-co ratio was achieved, in order to ensure full occupancy of the protein component. The final sample was subjected to gel-filtration on Sephadex G-25 to remove the residual NMF.

4.3.6 SDS-PAGE, anaerobic native-gel electrophoresis, and immunoblot analyses

The procedure for SDS-PAGE has been described [28]. Immunoblot analysis was performed as described in [29]. Purified preparations of NafY described in this work were used to raise anti-NafY antibodies at Capralogics Inc, (Hardwick, MA). For anaerobic native-gel electrophoresis, proteins were separated on gels with 7-20% acrylamide and 0-20% sucrose gradients as described [30]. Native gels were then stained for proteins with Coomassie R-250 by standard procedures, or stained for iron as described in [31].

4.3.7 In vitro FeMo-co insertion and nitrogenase activity assays

The *in vitro* FeMo-co insertion reactions were performed in 3-ml serum vials sealed with rubber stoppers. The vials were repeatedly evacuated and flushed with argon gas, and finally rinsed with 0.3 ml of anaerobic 25 mM Tris, pH 7.5, 1 mM DTH (buffer D). The complete reactions contained 25 mM Tris pH 7.5, 1 mM DTH, 9 mg/ml bovine serum albumin, 50 μg apo-dinitrogenase, 0.215 nmol of FeMo-co and the corresponding amount of purified N-terminal domain of NafY. The reactions (total volume of 325 μl) were incubated at 30°C for 30 min to allow for the FeMo-co insertion. The resulting activation of apo-dinitrogenase present in the reaction mixture was analyzed by the acetylene reduction assay after addition of 0.4 mg of NifH and 0.8 ml of ATP-regenerating mixture [25].

4.3.8 EPR spectroscopy

EPR analyses of FeMo-co bound to NafY in the presence/absence of apo-dinitrogenase were performed. The NafY/FeMo-co complex was generated as described above. EPR spectra were recorded on a Bruker ECS-106 spectrometer equipped with an ER 4116 dual-mode X-band cavity and an Oxford Instruments ESR-900 helium flow cryostat. X-band perpendicular mode EPR spectra were recorded using a microwave frequency of 9.69 GHz, at 11 ± 0.2 K, a modulation frequency of 100 kHz, a modulation amplitude of 1.0 mT (10 Gauss), a sweep rate of 12 mT/s, and a microwave power of 10 mW, with each spectrum being the sum of eight scans. The program IGOR Pro (WaveMetrics, Lake Oswego, OR) was used for handling of all EPR data. These measurements were done by the Britt Lab, UC Davis.

4.3.9 Miscellaneous assays

Protein concentrations in the samples were determined by the bicinchoninic acid method (BCA reagent, Pierce) using bovine serum albumin (BSA) as a standard [32]. The procedure for iron determination was performed as described [33], with minor modifications. UV-visible spectroscopy was carried out in a Shimadzu (Kyoto) UV1601V spectrophotometer.

4.3.10 NMR spectroscopy and structure determination

NMR samples of n-NafY and core-NafY were obtained by concentrating purified protein in buffer (pH 7.3) containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2 mM EDTA, and 0.02% (w/v) NaN_3 in either 100% ^2H_2O or 10% (v/v) ^2H_2O to a protein concentration of ca. 1 mM. NMR samples of NafY-ΔC were kept in 25mM imidazole pH 7.3 with 150mM NaCl. All n-NafY NMR data were collected on a 98-residue polypeptide containing the first 96 residues (1-96) of NafY plus two N-terminal residues introduced during cloning. All core-NafY NMR data sets were collected on a 135-residue polypeptide containing 132 residues (R100 through F231 out of 243 residues) of NafY plus three N-terminal residues introduced during the cloning procedure. NMR experiments were recorded at 298 K on either a Bruker Avance 800, a Bruker Avance DMX 600 equipped with a cryoprobe or a Bruker Avance II 900 also equipped with a cryoprobe. NMR data were processed using NMRPipe [34] or rNMRtk (http://webmac.rowland.org/rnmrtk/) and spectra were analyzed with NMRDraw and CARA [35].

The backbone resonance assignments of n-NafY were obtained with 3D HNCA [36], CBCAcoNH [37], ^{15}N-resolved TOCSY [38], and HBHAcoNH spectra [39]. Sidechain resonance assignments were obtained with 3D CcoNH and HccoNH spectra [40] recorded with non-linear sampling [41] and 3D HcCH-COSY and HcCH-TOCSY spectra [42]. 71% of all protons were assigned; excluding the N-terminus and the loop between residues 34-43, 85% of protons have assignments. NOE crosspeaks were obtained from both ^{15}N-resolved NOESY and ^{13}C-resolved NOESY spectra collected with mixing times of 100 ms. Torsion angle restraints were obtained from three-bond J couplings measured from an HNHA spectrum [43] and from chemical shift derived TALOS restraints [44]. Residual dipolar couplings (rdcs) were measured from a set of IPAP ^{1}H ^{15}N-HSQC [45] spectra recorded with and without ca. 17 mg/mL of pf1 phage present in the sample.

Approximately 700 NOE restraints were manually assigned and used to generate structures through simulated annealing with CYANA [46], which were then used as initial structures in an automated NOE assignment routine implemented in CYANA. Automatically assigned NOE's were then checked manually before being accepted. Rdc restraints were included in the structure calculation using Xplor-NIH [47]. The anisotropy and rhombicity of the anisotropy tensor were estimated by the histogram method [48] and then optimized to best fit the structural data. 200 structures were calculated, statistics are reported in Table I for the 20 lowest energy structures. Structure figures were constructed with MOLMOL (http://hugin.ethz.ch/wuthrich/software/molmol/) or PYMOL (www.pymol.org). Electrostatic plots were calculated with APBS [49] in 150mM salt at 300K.

Resonance assignments of core-NafY were obtained with 3D ^{15}N-resolved NOESY, HNCA, CBCAcoNH and HNCACB spectra [50]. A ^{1}H-^{15}N HSQC spectrum of apo-core-NafY and 2 spectra of core-NafY bound to FeMoCo were collected and the peak heights measured. The level of attenuation of each peak was calculated as a ratio between the two heights. Resonance assignments of full-length NafY-ΔC were obtained through TROSY-HNCA spectrum [51] and a non-linear sampled TROSY-HNCACB spectrum [52] collected on a deuterated sample of ^{2}H,^{13}C,^{15}N triply labeled NafY-ΔC. T1 and T2 relaxation times were measured on ^{15}N labeled samples by inversion recovery [53] and CPMG [54] experiments, respectively. The collection of NafY-ΔC rdc's required the addition of 500 mM NaCl to the NMR sample to screen electrostatic interactions between phage and the core domain of NafY.

4.4 Results

4.4.1 Domains of NafY

Full-length NafY has been previously shown to bind both apo-dinitrogenase and FeMo-co and to consist of two distinct domains. The three dimensional structure of core-NafY has been determined by X-ray crystallography [16, 17]. ^{1}H-^{15}N HSQC spectra of the N-terminal domain (n-NafY), core-NafY, and a full-length construct of NafY with a short C-terminal truncation (NafY-ΔC) are shown in figure 4.3. Approximately 90% of amide resonances have been assigned for both the individual domains as well as the full-length NafY-ΔC, which required three-dimensional spectra to be acquired with TROSY techniques on deuterated protein for robust coherence transfer. Both core-NafY and n-NafY give well-resolved spectra while the spectrum of NafY-ΔC is broadened more significantly than would be expected from solely the difference in the correlation time between the individual domains and the 26 kDa full length protein, possibly suggesting the presence some inter-domain interaction fluctuating on the microsecond to millisecond time scale although the behavior can also be explained by a significant anisotropy in the tumbling of NafY-ΔC. The data discussed below confirm that the broadening observed in spectra of NafY-ΔC is due to anisotropic tumbling rather than conformational exchange.

No significant chemical shift changes were observed in either domain when n-NafY and core-NafY were mixed 1:1 in solution (figure 4.4A vs. figure 4.4B), consistent with the conclusion that the two domains fold independently and do not interact to form a defined quaternary structure. The spectrum of NafY-ΔC, figure 4.4C, was assigned and the assignments compared with those from the individual domains, n-NafY and core-NafY. Small shift differences were detected, which are plotted as a function of residue in figure 4.5D. There are relatively few residues affected, with shift changes that are modest in magnitude. The residues with the largest shifts are indicated on the structure in figure 4.5B. These do not define a contiguous surface, suggesting that only transient interactions occur. ^{1}H-^{15}N residual dipolar coupling measurements on NafY-ΔC ordered in a phage medium showed that the two domains have very different order tensors, and hence are not well ordered with respect to one another. ^{15}N T1/T2 ratios for NafY-ΔC clearly indicated different effective rotational correlations times for the two domains, again supporting flexible linkage of n-NafY and core-NafY. However, the derived rotational correlation times for the individual domains are substantially longer than expected if there was no interaction at all between them. This behavior is analogous to the WW and prolyl-isomerase domains of Pin1, also disordered but with some weak interaction dependent on the domains being tethered [55]. Taken together with the observation that the region of the protein between the two domains is sensitive to proteolytic cleavage [16], these data show that the two domains of NafY do fold independently and are likely to function with some degree of independence.
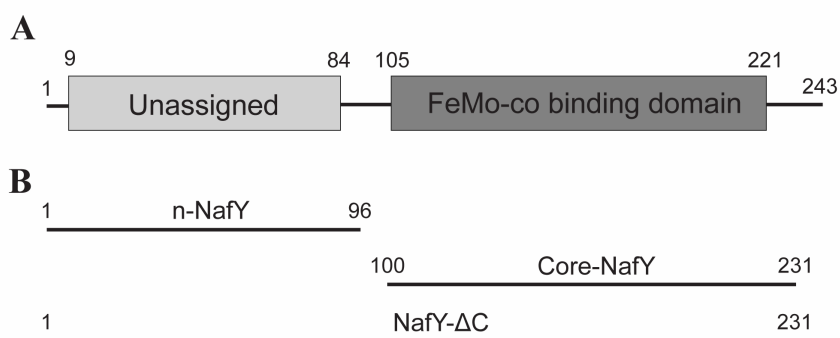
Figure 4.3. NafY domain architecture. (A) Diagram illustrating the domain organization of NafY. (B) Truncated NafY variants used in this study. Amino acid residues comprising constructs studied are indicated.
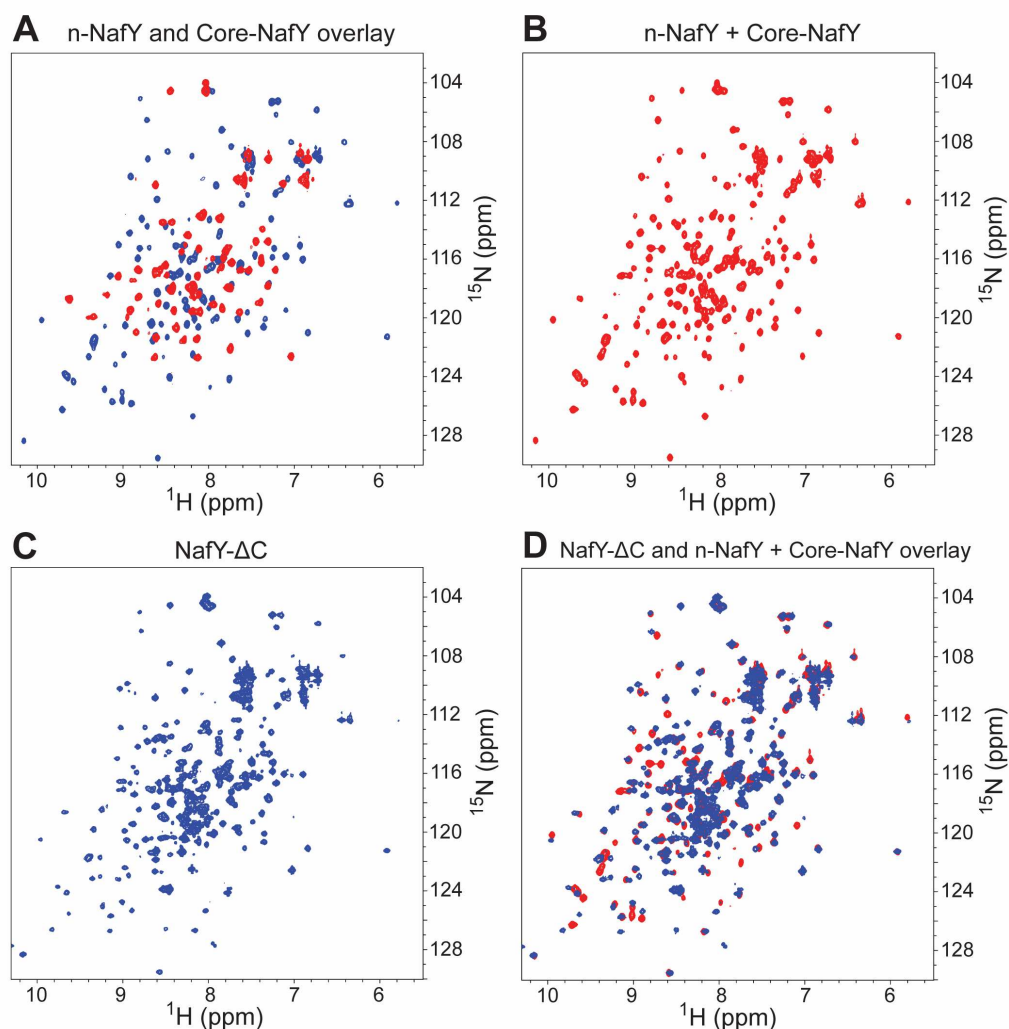

Figure 4.4. ^{1}H-^{15}N HSQC spectra of NafY domains and domain combinations. (A) Overlay of n-NafY (red) and core-NafY (blue). (B) Spectrum of an equimolar mixture of n-NafY and core-NafY. (C) NafY-ΔC spectrum. (D) Overlaid spectra of NafY-ΔC (blue) and the equimolar mixture of n-NafY and core-NafY (red).
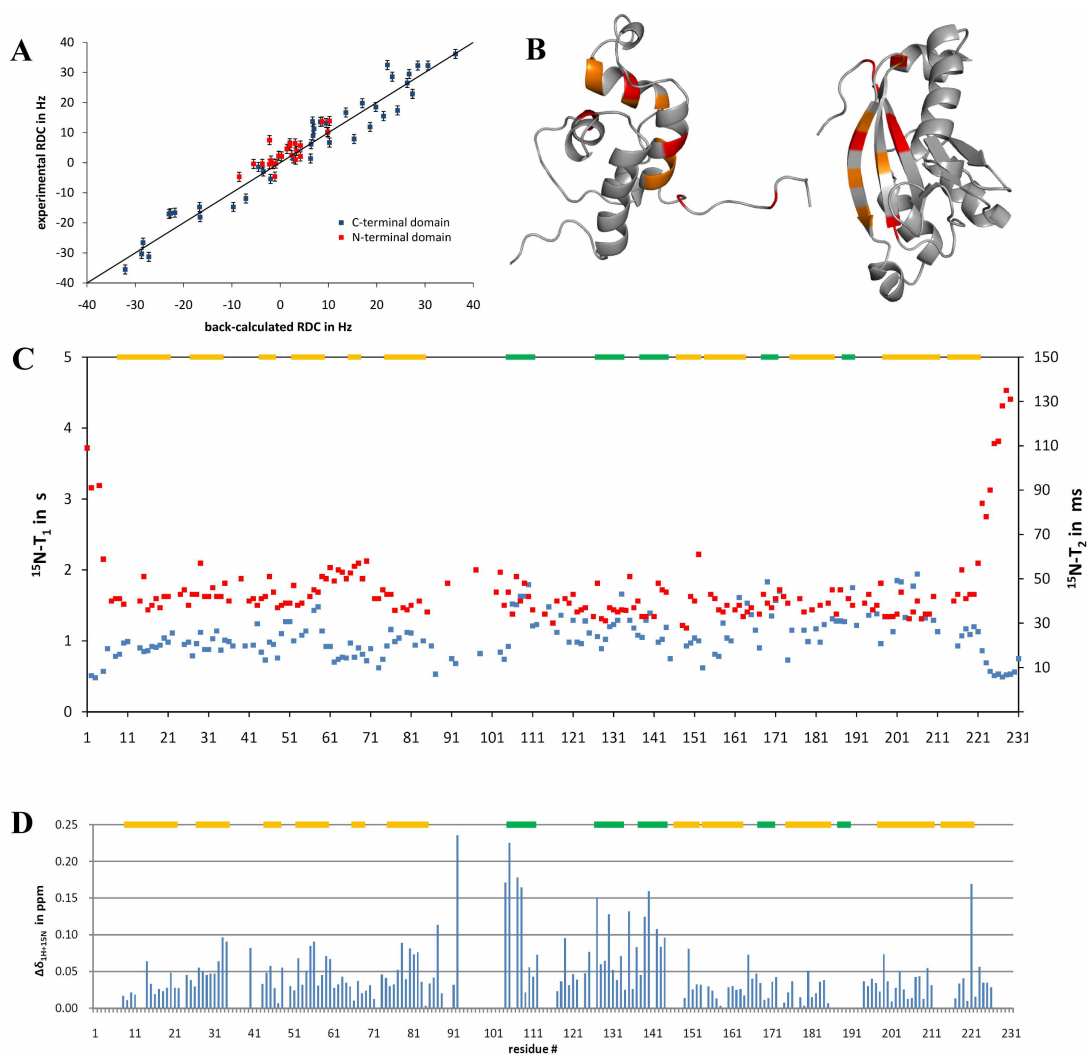
Figure 4.5. Domain Interactions in full-length NafY-ΔC. (A) Experimental and back-calculated RDC values. (B) Cartoon diagram showing the residues in n-NafY and core-NafY that exhibit chemical shift differences when compared to full-length NafY. Residues more than two standard deviations outside of the mean are show in red, those one deviation outside are shown in orange. (C) T1 plotted in blue, T2 plotted in red. (D) Chemical shift changes between individual domains and NafY-ΔC.

4.4.2 The N-terminal domain of NafY exhibits an apo-dinitrogenase binding activity

It has been previously shown by anaerobic native gel electrophoresis that full length NafY binds apo-dinitrogenase, but core-NafY does not bind sufficiently strongly to co-electrophorese with apo-dinitrogenase [16]. In order to address the function of n-NafY in the absence of core-NafY pull-down experiments were performed in which both NafY-ΔC and n-NafY were expressed as fusions to glutathione-S-transferase (GST). After binding each protein to a GSH-Sepharose resin through its GST-moiety, the resin was exposed to an *A. vinelandii*

65

cell-free extract containing apo-dinitrogenase. The extract of apo-nitrogenase used in this experiment was obtained from the *A. vinelandii* strain UW146 (*nifB⁻ ΔnafY*) which accumulates the apo-form of dinitrogenase, due to the deletion of the *nifB* gene from the chromosomal DNA of the bacteria [12]. When the extract of UW146 was applied to a chromatography resin containing the fusion protein GST-NafY-ΔC, apo-dinitrogenase was specifically recovered from the cell-free extract. Similar results were observed when the fusion protein GST-n-NafY was used, demonstrating the ability of the amino terminal domain of NafY to bind apo-dinitrogease without the presence of core-NafY (figure 4.6). No binding was observed when either NafY-ΔC or n-NafY was not bound to the resin, indicating the absence of any nonspecific binding of apo-dinitrogenase to the GSH-Sepharose resin. The efficiency of apo-dinitrogenase binding to n-NafY was lower than with the NafY- ΔC construct, suggesting that apo-dinitrogenase binding occurs with some interaction with both domains despite any strong association between them as discussed in section 4.4.1.
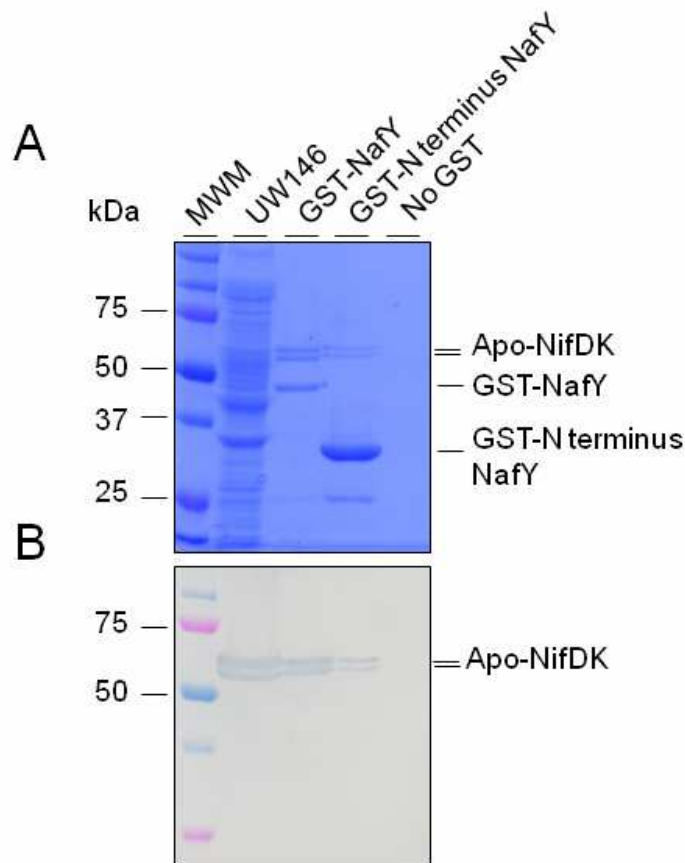

Figure 4.6. Apo-NifDK pull-down assay (A) SDS-PAGE analysis of the interaction between n-NafY and apo-NifDK. An UW146 (*nifB⁻ ΔnafY*) cell-free extract containing apo-NifDK (lane 1), was applied to a GSH-Sepharose resin carrying bound GST-NafY (lane 2), bound GST-n-NafY (lane 3), or no protein (lane 4). After thorough washing, proteins bound to the resin were eluted by addition of reduced glutathione and analyzed by SDS-PAGE. Molecular weight markers are indicated to the left. (*B*) Immunoblot analysis of SDS-gel from panel *A* developed with antibodies to NifDK. The position of apo-NifDK is indicated to the right.

4.4.3 The presence of n-NafY inhibits nitrogenase activity by blocking FeMo-co insertion

The effect of n-NafY on FeMo-co insertion was studied by measuring the nitrogenase activity of different preparations of the reconstituted holo-NifDK obtained under different conditions. Two versions of apo-dinitrogenase were obtained from different strains of *A. vinelandii*, Δ*nifB* (DJ1143) [56] and Δ*nifB* Δ*nafY* (UW269). Both variants were purified as tetramers free from bound NafY to facilitate comparison of results obtained from FeMo-co insertion experiments. Extensive washing with imidazole-containing buffer (which weakens the NafY-NifDK interaction) was required to completely remove NafY from the Δ*nifB* apo-NifDK.

First, the specific activity of the protein was measured in both cases by changing the FeMo-co/protein ratio (figure 4.7A). The maximum activation of the nitrogenase was achieved with a 2:1 FeMo-co:apo-dinitrogenase molar ratio, as expected. The total value of the specific activity is about two-fold higher when the apo-dinitrogenase is purified from a *nafY* $^+$ background, as previously reported [12].

Next, similar reactions were set up in order to determine whether the presence of n-NafY had any positive or negative effect on the insertion of FeMo-co into apo-dinitrogenase. The reaction mixtures, containing a sub-saturating amount of FeMo-co:apo-NifDK in a 1:1 molar ratio (apo-NifDK contains two FeMo-co binding sites), were titrated with increasing amounts of n-NafY (Figure 4.7B). Remarkably, the nitrogenase specific activity, measured by the production of ethylene from acetylene reduction, exhibited a 50% decrease when a 2:1 stoichiometry of monomeric n-NafY relative to tetrameric ($\alpha_2\beta_2$) apo-dinitrogenase was present in the reaction. The degree of inhibition continued to increase when larger amounts of n-NafY were added in the mixture, reaching a complete inhibition at a 50:1 ratio of n-NafY to tetrameric apo-dinitrogenase.

These *in vitro* FeMo-co insertion and nitrogenase activity experiments were scaled up for spectroscopic analysis in order to better understand the inhibitory effect of n-NafY on nitrogenase activity. Figure 4.7C shows the electron paramagnetic resonance (EPR) signatures of apo-dinitrogenase reconstituted with FeMo-co under different experimental conditions. The reaction mixtures contained a 1:1 FeMo-co:apo-dinitrogenase molar ratio, in the presence or absence of a 50:1 n-NafY:apo-dinitrogenase molar ratio. The electron paramagnetic resonance (EPR) signatures of the apo and holo forms of dinitrogenase have been thoroughly reported and analyzed [56], as well has the EPR signal for isolated or NafY-bound FeMo-co [17]. The blue trace in figure 4.7C shows the EPR spectra of purified apo-dinitrogenase, as expected, with the P-clusters in the DTH-reduced state. The green trace is the spectrum of apo-dinitrogenase reconstituted by the addition of FeMo-co. The holo-form of dinitrogenase was generated in this case as shown by the presence of the typical signal for the DTH-reduced iron-sulfur clusters contained in dinitrogenase in the EPR spectrum. The orange trace is the spectrum of apo-dinitrogenase reconstituted with FeMo-co in the presence of n-NafY. Interestingly, the EPR signature of this sample lacks the features present in holo- dinitrogenase and more closely resembles that of the apo-dinitrogenase. This result strongly suggests that the inhibitory role of the n-NafY on nitrogenase activity is caused by hindering the insertion of FeMo-co into the apo-dinitrogenase.
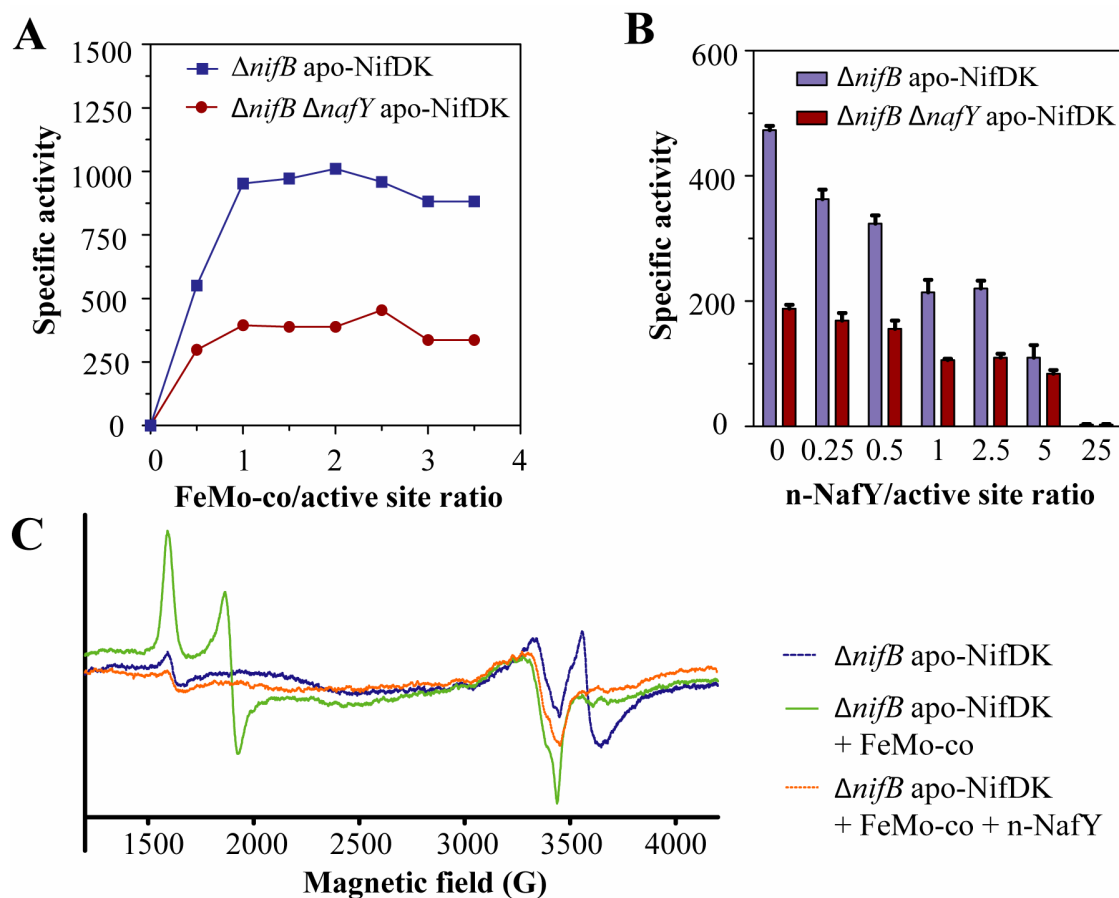
Figure 4.7. The effect of n-NafY on apo-NifDK reconstitution. (A) Activation by FeMo-co of Δ*nifB* apo-NifDK and Δ*nifB* Δ*nafY* apo-NifDK. Specific activity is given in nmol C_2H_4 produced per min per mg of apo-NifDK added to the reaction mixture. Data are the average of two independent determinations. (B) Effect of n-NafY on the activation of Δ*nifB* apo-NifDK and Δ*nifB* Δ*nafY* apo-NifDK by sub-optimal amounts of FeMo-co. Specific activity is given in nmol C_2H_4 produced per min per mg of apo-NifDK added to the reaction mixture. Data are the average of two to four independent determinations. (C) CW-EPR spectra from apo-NifDK (blue line) and apo-NifDK samples incubated with FeMo-co (green line) or n-NafY plus FeMo-co (orange line).

4.4.4 The Solution Structure of n-NafY

To better address the function of n-NafY I solved its three-dimensional structure by NMR spectroscopy. A stereo image of a superposition of the best 20 structures calculated by simulated annealing in Xplor-NIH [47] is shown in Figure 4.8A. Figure 4.9 shows the assigned ^{1}H-^{15}N HSQC spectrum of n-NafY. Twelve amide resonances including those from the first eight residues, with two extra N-terminal residues introduced during the cloning, and most of the residue 35-42 loop region remain unassigned. The E39 and G40 resonances were assigned by their distinctive chemical shifts. The E39 and G40 amides as well as several of the unassigned resonances are significantly broadened and residues 34-43 are poorly restrained by NOE crosspeaks. These observations are consistent with this region of the protein undergoing conformational exchange on the microsecond to millisecond timescale. Statistics of the structure

calculation are summarized in table 4.1. The structural coordinates have been deposited in the protein data bank with accession code 2KIC. The NMR assignments have been deposited at the Biological Magnetic Resonance Bank with accession code 16269.

The experiments discussed above have established that n-NafY binds to apo-dinitrogenase and prevents FeMo-co insertion. The 3-dimensional structure of n-NafY suggests a possible mechanism for this. Figure 4.10 shows the electrostatic surface of n-NafY and highlights a large patch of negatively charged surface flanked by the flexible loop from residues 35 through 42.   The crystal structure of apo-dinitrogenase from *A. vinelandii* [57] has been solved and contains a region of positive charge on its surface which has been proposed to act as a funnel for FeMo-co insertion as the surface charge in the same region of holo-dinitrogenase is negative after FeMo-co insertion [58]. N-NafY, like FeMo-co, is negatively charged and is likely to interact with the positive surface of apo-dinitrogenase. However, it is also worth noting that in the crystal structure of apo-dinitrogenase there is an unstructured loop of approximately 30 residues with a net charge of ca. -7 not seen in the density which may play a role along with the flexible loop in n-NafY in the interaction between apo-dinitrogenase and NafY and subsequent insertion of FeMo-co.

The function of structural homologues of n-NafY further supports the idea that the insertion of FeMo-co to apo-dinitrogenase is mediated by the binding of n-NafY to apo-dinitrogenase.   Secondary structure matching (SSM) [59] of the well-defined region of n-NafY (residues 8-88) has shown that despite any significant sequence homology n-NafY possesses structural homology to the sterile alpha motif (SAM), a domain involved in mediating protein-protein interactions [60, 61]. SAM domains are most often found in eukaryotes but there are a small number identified in bacteria. The fold of typical SAM domains consists of a five-helix bundle, with a helical hairpin formed by the first two helices and the three additional helices pack on one side of the hairpin [62]. SAM domains have been generalized to include proteins with four helices (generally two helical hairpins), with somewhat different packing than the typical five-helix version. n-NafY also has some differences relative to a typical SAM domains: first, $\alpha 1$ and $\alpha 2$ are at nearly a right angle, rather than being packed close to antiparallel in a hairpin; second, $\alpha 3$ has been reduced to a single helical turn; third, there is an extra short helical turn, $\alpha 4'$ inserted between $\alpha 4$ and $\alpha 5$; and fourth, the loop linking $\alpha 2$ and $\alpha 3$ is poorly ordered. Figure 4.8B shows an overlay of the n-NafY structure nearest to the mean and the SAM domain with closest structural similarity, that of the *Sacharomyces cerevisiae* protein Ste11 [63] (PDB code 1OW5). In spite of the differences, the basic fold of n-NafY is relatively close to typical SAM domains.

Eukaryotic genomes of known sequence contain similar numbers of SAM domains and SH2 domains, another common protein-protein interaction motif. However, unlike SH2 domains which bind only to phosphorylated tyrosine residues, SAM domains have been shown to bind many different types of proteins and have also been implicated in protein-RNA [64] and protein-lipid [65] interactions. In addition to the possessing diverse binding partners, the mechanism of the interactions between SAM domains and their binding partners vary widely with well studied cases of binding mediated though electrostatics [64] or through surface exposed hydrophobics [63]. The only hydrophobic side chains of n-NafY that are significantly exposed to solvent are those of M68, M73 and W84 which neither form a contiguous surface nor are conserved among its homologs. Therefore it is likely that the interaction between apo-dinitrogenase and NafY is driven primarily by electrostatics as discussed previously.   Results of the SSM analysis are presented in table 4.2.
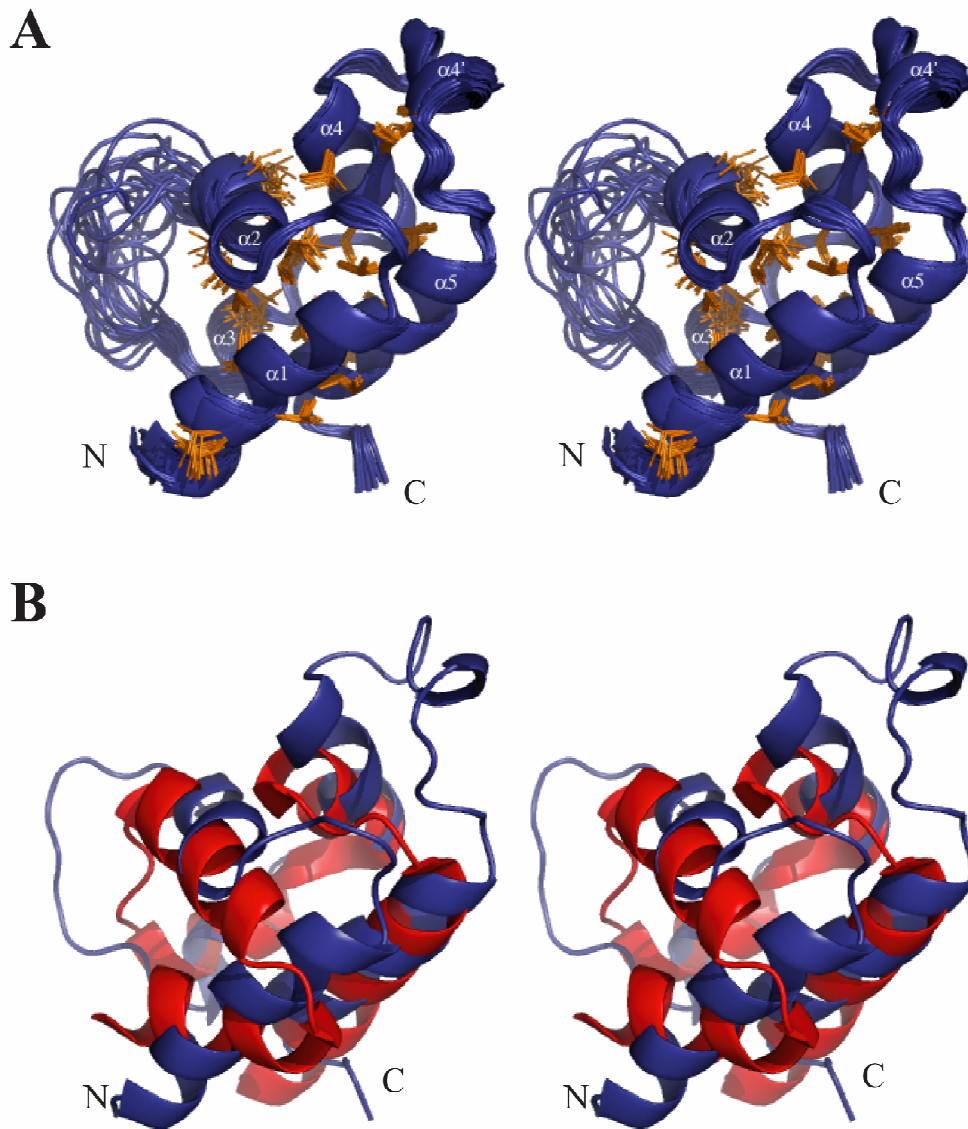
Figue 4.8. (A) A stereo, cartoon representation of the superposition of the 20 lowest energy structures of n-NafY. The side chains of well-ordered hydrophobic residues are shown in orange (B) An overlay of the closest structural homolog of n-NafY, the SAM domain of Ste11 from *S. cerevisiae*, PDB code 1OW5. The Ste11 SAM domain is depicted in red. The lowest energy structure of n-NafY is shown in blue.
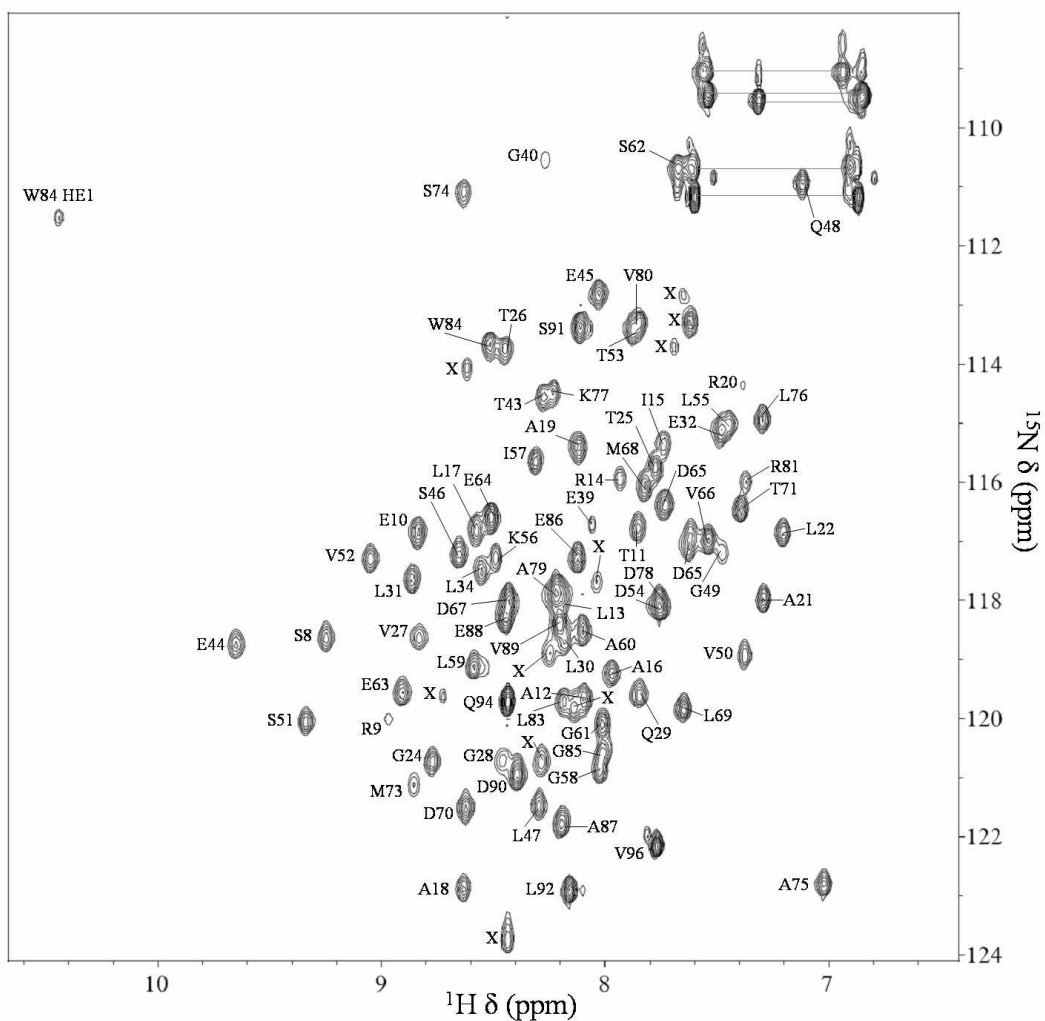
Figure 4.9. ^{1}H-^{15}N HSQC spectrum of n-NafY from *A. Vinelandii* collected at 800 MHz with residue assignments indicated. Ten amide peaks remain unassigned due to an absence of any connectivity in both HNCA and 3D-^{15}N NOESY spectra and are marked with an 'X'. The pairs of the side chain peaks from the glutamine (4) and asparagine (1) residues are shown with horizontal lines connecting them. Residues 1-6, 35-38, 42, as well as the extra N-terminal residue introduced during cloning remain unassigned.
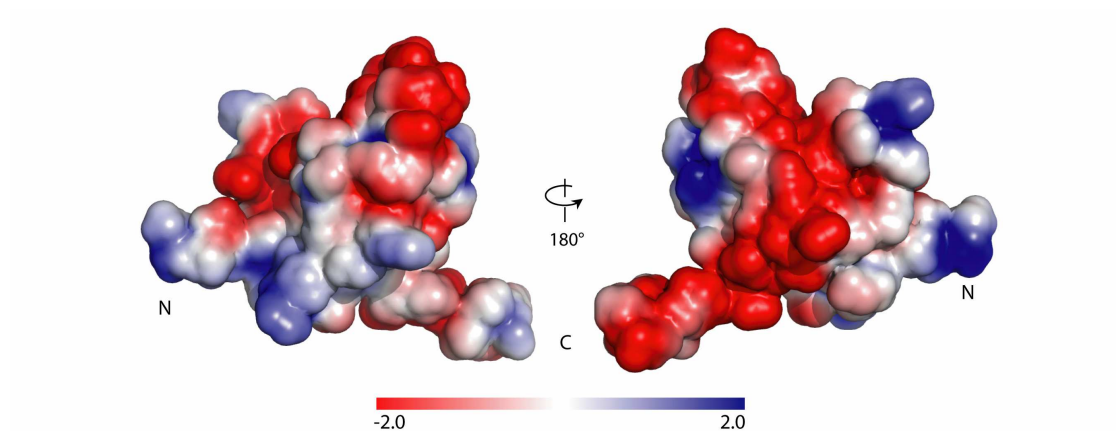
Figure 4.10. Electorstatic surface of the lowest energy structure of n-NafY. The scale is in units of kT/e$^-$. The left view is in the same orientation as figure 4.8.

Table 4.1. Structural statistics for the N-terminal domain of NafY from *A. vinelandii*

Restraints Used	
NOE	1084
Intraresidue	276
Sequential	319
Medium Range	316
Long Range	173
Dihedrals	141
$^{3}J_{HNHA}$ (φ)	54
TALOS (φ/ψ)	87
HN RDC	42

RMSD from restraints	
NOE (Å)	0.0271 ± 0.0012
Dihedrals (º)	0.813 ± 0.158
HN RDC's (Hz)	0.659 ± 0.056
HN RDC R_{dip}[a]	11.0%

RMSD from idealized geometry	
Bond (Å)	0.0038 ± 0.0001
Angle (º)	0.5925 ± 0.0059
Improper (º)	0.3944 ± 0.0076

Violations per structure	
NOE (> 0.3 Å)	0
Dihedrals (> 5º)	$0.400 \pm .0503$
HN RDC (> 2 Hz)	0

Coordinate RMSDs to the mean[b]	
Backbone (Å)	0.45 ± 0.09
Heavy Atoms (Å)	0.96 ± 0.08

Structure Quality Factors[c]

	Mean score for all models	Standard Deviation	Z-score[d]
Procheck G-factor[e] (phi / psi only)	-0.66	N/A	-2.28
Procheck G-factor[e] (all dihedral angles)	-0.70	N/A	-4.14
Verify3D	0.21	0.0224	-4.01
ProsaII (-ve)	0.52	0.0656	-0.54
MolProbity clashscore	56.91	6.6726	-8.24

Ramachandran Plot Summary for ordered residues[e]	
Most favored regions (%)	83.3
Additionally allowed regions (%)	16.4
Generously allowed regions (%)	0.3
Disallowed regions (%)	0.0

a) R_{dip} is defined as the ratio of the r.m.s. difference between observed and calculated RDC values to the r.m.s. difference if the bond vectors were distributed randomly [66] and was calculated using the iDC toolkit [67].

b) Calculated using the well-structured region consisting of residues 9-33 and 44-85

c) Determined using Protein Structure Validation Suite version 1.3 [68]. The suite includes Procheck [69], Verify3D [69], ProsaII [70] and MolProbity [71].

d) With respect to mean and standard deviation for for a set of 252 X-ray structures < 500 residues, of resolution <= 1.80 Å, R-factor <= 0.25 and R-free <= 0.28; a positive value indicates a 'better' score

e) Residues with sum of phi and psi order parameters > 1.8

Table 4.2. Structural homologs of n-NafY identified through SSM

#	Scoring			Rmsd	N$_{algn}$	N$_g$	%$_{seq}$	Query %$_{sse}$	Target (PDB entry)			
	Q	P	Z						Match	%$_{sse}$	N$_{res}$	Name
1	0.27	0.3	2.6	3.09	57	4	12	67	1ow5:A	80	60	NMR structure of the *Saccharomyces cerevisiae* SAM (sterile alpha motif) domain
2	0.26	0.4	2.7	3.40	57	3	12	67	1dxs:A	80	57	Crystal structure of the C-terminal SAM domain of human p73 alpha splice variant
3	0.23	0.0	1.7	4.02	70	3	9	67	1v38:A	80	78	Solution structure of the SAM domain of mouse SAMSN1
4	0.22	0.0	1.4	3.86	62	7	8	67	1cok:A	80	68	Structure of the C-terminal domain of p73
5	0.21	0.1	2.8	2.98	50	4	10	67	2qkq:A	80	61	Structure of the SAM domain of human ephrin type-B receptor 4
6	0.21	0.0	1.6	3.69	56	4	11	67	2qkq:B	80	61	Structure of the SAM domain of human ephrin type-B receptor 4
7	0.21	0.1	2.2	2.94	53	4	13	83	1wcn:A	100	70	NMR structure of the carboxy terminal domains of *Escherichia coli* NusA
8	0.21	0.0	1.2	3.91	63	4	11	67	1b0x:A	80	72	The crystal structure of an eph-receptor SAM domain reveals a mechanism for modular dimerization
9	0.21	0.0	2.0	3.64	58	3	9	67	1pk1:A	67	68	Hetero SAM domain structure of PH (Polyhomeotic-proximal chromatin protein) and SCM (Sex comb on midleg CG9495-PA)
10	0.21	0.0	1.5	4.09	62	3	10	67	3bq7:D	80	68	SAM domain of diacylglycerol kinase delta1 (E35G)

The various scores output by the SSM algorithm are defined as the following:

- RMSD is the room mean square deviation between aligned α-carbons in angstroms.
- N$_g$ is the number of gaps between aligned secondary structure elements between the query and target.
- %$_{seq}$ is the percent of aligned residues that are identical between the query and the target
- %$_{sse}$ describes what fraction of secondary structure elements the align
- N$_{algn}$ is the number of residues in a given alignment
- $Q = (N_{algn})^2 / [(1+(RMSD/R_0)^2)*N_{query}*N_{res}$ where $R_0 = 3.0$Å. Identical structures have a Q-score of 1.0.
- P = -log(P-value) where the P-value measures the probability of achieving the same or better quality of match at random.
- The Z-score is a measure of the statistical significance of a match in terms of Gaussian statistics. The higher the Z-score, the higher the statistical significance of the match.

4.4.5 Insights into FeMo-co binding to the Core-NafY

Although the structure of core-NafY in the absence of FeMo-co is known [16] the structural details of how core-NafY is able to bind FeMo-co are not well understood. In dinitrogenase FeMo-co is coordinated by one histidine and one cysteine residue [72]. In a previous study on core-NafY several mutations of histidine and cysteine residues were made in order to elucidate which residues were involved in FeMo-co binding. That work identified histidine 121 as clearly being involved in FeMo-co binding. The histidine residue is located within an HFG motif that is strictly conserved throughout the family of similar *A. vinelandii* proteins (NifB, NifX, and NafY) known NifB-co or FeMo-co [17]. In order to yield further insight into how core-NafY binds FeMo-co I assigned 108 amide resonances out of the possible 125 non-proline residues in a core domain construct which terminated at residue 231 and then collected ^{1}H-^{15}N HSQC spectra of core-NafY in complex with FeMo-co. The resonances of spins that are in close proximity to the paramagnetic cofactor are attenuated due to the enhanced

74

relaxation caused by the presence of unpaired electron spins [73] thereby allowing for the localization of the FeMo-co binding site. The overlaid spectra and a depiction of which residues are affected by the paramagnetic cofactor are shown in figure 4.11. Histidine 121 is strongly attenuated by FeMo-co and the data also identify two cysteine residues that are strongly attenuated, cysteine 125 and cysteine 196. In the mutational study by Rubio et al. a C125A mutant retained FeMo-co binding; however no mutant was constructed of cysteine 196 because it is not conserved over the family of homologous *A. vinelandii* FeMo-co binding proteins.

Interestingly, the core domain displays modest structural homology to a family of vitamin B12-binding proteins, most significantly to MutS from *Clostridium tetanomorphum* [74, 75], GlmS from *Clostridium cochlearium* [76], and MtaC from *Moorella thermoacetica* [77]. The backbone RMSD in each case is approximately 3.5 Å. MutS and GlmS are B12-binding domains of glutamate mutase whereas MtaC is the B12-binding component of a methyltranfserase. The topologies of these domains are similar to core-NafY albeit with some insertions. Interestingly, in each case the proteins bind vitamin B12 through a surface exposed histidine residue located in close proximity to the analogous residue in core-NafY, histidine 121. Upon B12 binding, the flexible loop immediately following the metal-coordinating histidine assumes a stable helical structure which is sampled in the unbound protein structure [74, 75]. The analogous loop in core-NafY (residues 120-126) is shorter than the loop which fluctuates to a helix observed in the B12 binders. Moreover, the Cα and Cβ chemical shift index in this region of core-NafY indicate that the helical conformation is not detectibly sampled by these residues [78]. However, some helical NOEs are observed in NOESY spectra of core-NafY for this region. The overlay of the core-domain with MtaC is shown in figure 4.12.
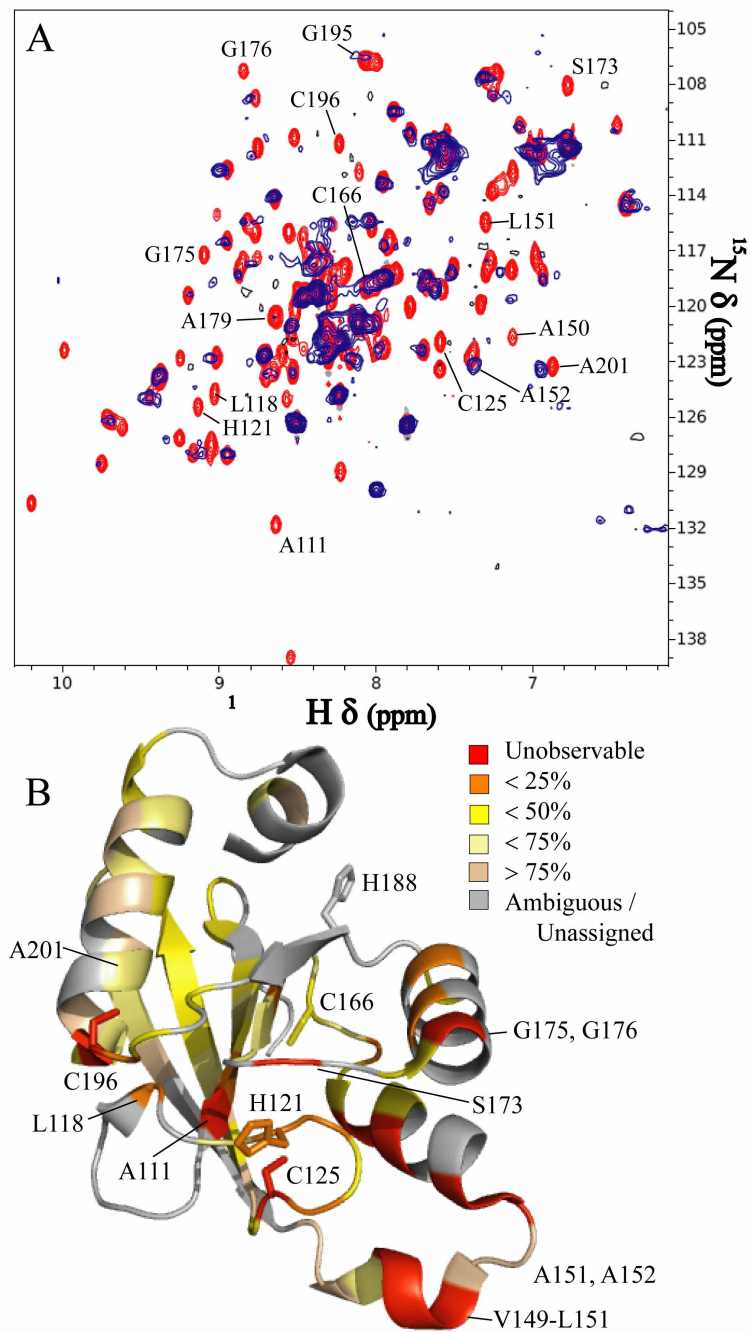
Figure 4.11. Localization of the FeMo-co binding site in core-NafY. (A) Overlaid ^{1}H-^{15}N HSQC spectra of apo-core-NafY (red) and core-NafY complexed with FeMo-co (blue). Several residues are labeled for illustration. (B) Crystal structure of core-NafY color coded by the level of attenuation upon FeMo-co binding. The increments of the color coding are half standard deviations about the mean denoted by arbitrarily setting the mean to 50%. Functionally significant histidine and cysteine side chains are labeled.
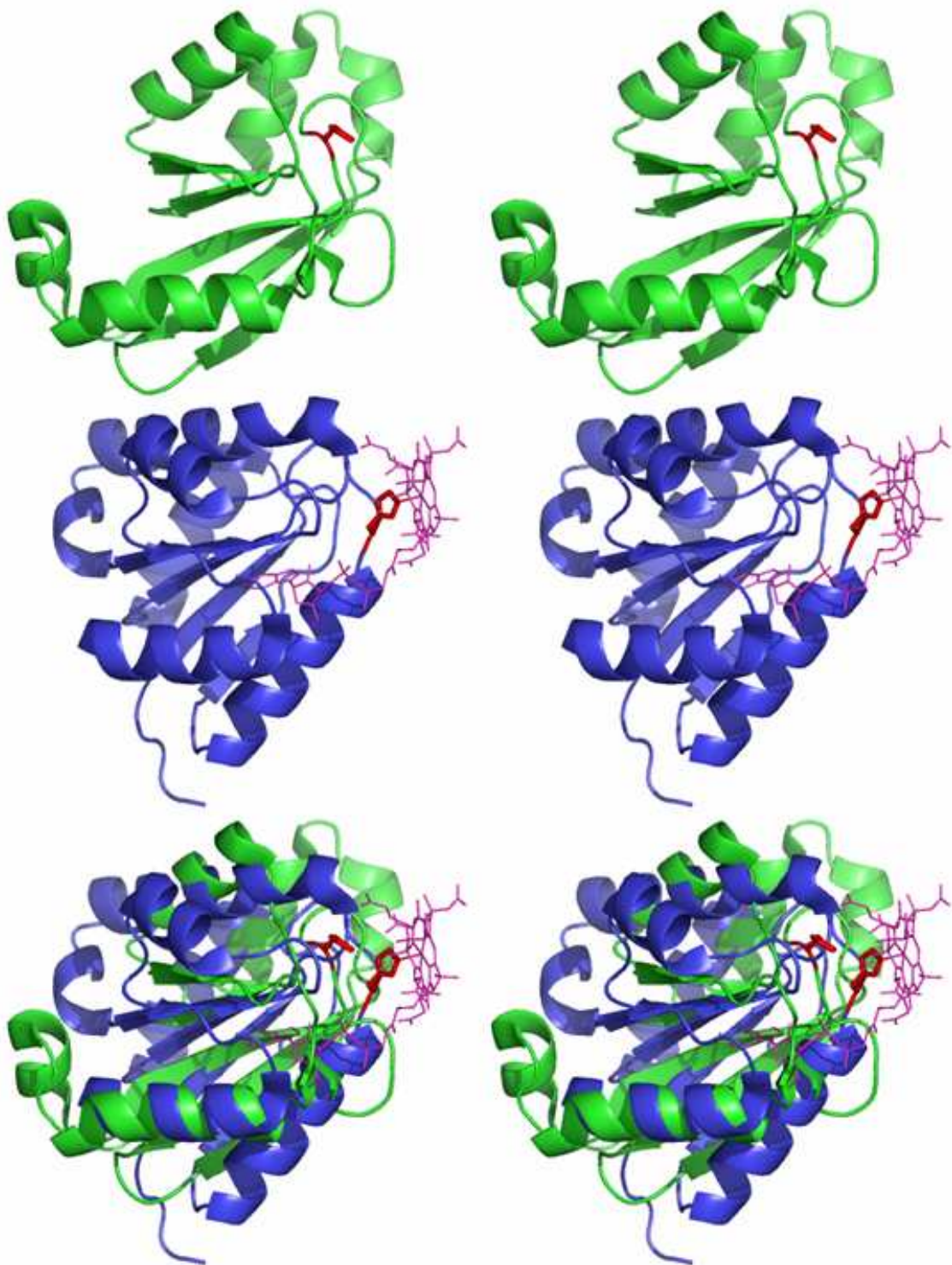
Figure 4.12. Stereo representations of core-NafY (green) with corrinoid bound MtaC (blue, PDB code 1Y80) as well as the overlay of the two structures. The histidine residues implicated in binding are shown in red. Vitamin B12 is in pink.

4.5 Discussion

This work has utilized biochemistry and EPR and NMR spectroscopies to further understand the role of NafY in the maturation of the *A. vinelandii* nitrogenase. Previously proposed roles for NafY are the stabilization of apo-dinitrogenase in a conformation amenable to FeMo-co insertion and also as a chaperone insertase that delivers FeMo-co to apo-dinitrogenase [10, 12]. The structure of core-NafY is known and it is clear that the core domain is sufficient to bind FeMo-co but does not efficiently bind apo-dinitrogenase [16]. This chapter describes the structure and function of the N-terminal domain of NafY and provides further information on where within the core domain FeMo-co binds.

After the observation that full-length NafY co-electrophoreses with apo-dinitrogenase while the core domain does not [16], a pull-down assay was developed in which both immobilized n-NafY and full-length NafY are able to extract apo-dinitrogenase from *A. vinelandii* cell extracts. This assay demonstrated that while the N-terminal domain is sufficient to bind apo-dinitrogenase, full-length NafY binds more efficiently than the N-terminal domain alone. It follows that while the core domain does not bind apo-dinitrogenase on its own there is some cooperativity in binding which allows the full-length protein to bind more tightly than the first domain alone.

Elucidating the role of NafY in the cell has been impeded by two somewhat contentious pieces of biochemical data. The first being that apo-dinitrogenase can be reconstituted with FeMo-co in the absence of NafY to an active holo-form that retains 80% of normal activity [56]. However, a mutant strain of *A. vinelandii* which lacks an active NafY exhibits a 2-fold reduction in growth rate under stressing nitrogen fixing conditions, where the growth medium is not supplanted with 10mM sodium molybdate; whereas a strain which lack both active NafY and NifX exhibits a 5-fold reduction in growth rate under stress. In order to specifically address the role of the N-terminal domain of NafY on nitrogenase maturation, the effects of various reconstitution conditions on dinitrogenase activity were measured. These experiments confirmed the previously reported observation that apo-dinitrogenase purified from a Δ*nifB* Δ*nafY* mutant exhibits half the activity of apo-dinitrogenase purified from a Δ*nifB* mutant upon reconstitution with FeMo-co [12]. This observation supports the chaperone role of NafY providing stability to apo-dinitrogenase. This assay was followed by determining the effect of the addition of n-NafY to apo-dinitrogenase during FeMo-co insertion. As it was shown that the N-terminal domain is sufficient to bind to apo-dinitrogenase and is hypothesized to stabilize the apo-form, a promotion of nitrogenase activity upon the addition of n-NafY prior to FeMo-co was expected. Furthermore, if NafY stabilizes the apo-protein, the presence of the N-terminal domain may be sufficient to restore the activity of the apo-dinitrogenase obtained from a Δ*nifB* Δ*nafY* strain to that of protein obtained from a Δ*nifB* strain. Surprisingly, this experiment showed that the presence of n-NafY inhibited nitrogenase activity with protein obtained from both strains.

By repeating the same FeMo-co insertion assays with sufficient quantities for EPR analysis it was shown that n-NafY inhibits nitrogenase activity by preventing the insertion of FeMo-co into apo-dinitrogenase. This is in contrast to the behavior of full-length NafY which dissociates from hexameric apo-dinitrogenase ($\alpha_2\beta_2\gamma_2$) upon the addition of FeMo-co, leaving behind $\alpha_2\beta_2$ holo-dinitrogenase [10]. From these observations it is clear that the dissociation of NafY from dinitrogenase requires the presence of the core domain and that n-NafY is likely binding in close proximity to the proposed FeMo-co insertion funnel seen in the structure of apo-dinitrogenase thereby blocking FeMo-co insertion [57].

In order to further understand how n-NafY binds to apo-dinitrogenase, I determined its solution structure using NMR spectroscopy. The domain adopts an all α-helical fold containing 5 helices. The two longer helices, α1 and α5, are separated by 3 short helices and a structured helical turn, designated α4'. There is a flexible loop between helices α2 and α3 which the NMR data show to be involved in some motion on the μs-ms timescale as the resonances in this region are exchange broadened. Interestingly, the three-dimensional structure reveals unpredicted structural homology to the sterile alpha-motif domain, a protein domain associated with the binding of various macromolecular ligands, the generalized structure of which consists of a five helix bundle with a helical hairpin formed by the first two helices and the three additional helices packed on one side of the hairpin. These general features are reproduced in the structure of the N-terminal domain of NafY with the extra insertion of the helical turn, α4'. This unpredicted homology to a protein interaction domain is consistent with the N-terminal domain's role in mediating the apo-NifDK:NafY interaction. It is also interesting to note that SAM domains are widely used in eukaryotes, but the occurrence of SAM domains in bacteria is much rarer. Although SMART and Pfam databases identify more than 2500 SAM domains, only 11 of them are from prokaryotic proteins and, at present, only the AR1 domain of the NusA protein from *Escherichia coli* [79] and a 4-helix SAM domain, part of TipαN34 from *Heliobacter pylori* [80], have been structurally characterized.

In order to further the knowledge of what part of the core domain is involved in FeMo-co binding I assigned the NMR resonances of the apo form and then observed the paramagnetic relaxation enhancement that occurs when core-NafY is in complex with FeMo-co. The amide resonance of histidine 121, implicated by mutagenesis to bind FeMo-co, is strongly attenuated once the cofactor is bound. Two other cysteine residues are affected by FeMo-co binding, residues 125 and 196. These data are consistent with the mode of FeMo-co binding proposed previously [17] and show clearly that the first three helices α1, α2 and α3 and the loop between β1 and β2 are in close proximity to the cofactor. Interestingly, core-NafY displays some structural homology to proteins responsible for the delivery of vitamin B12. These proteins share a common topology with the core domain, albeit with insertions, and all bind B12 through a surface exposed histidine located in the same region of the protein as in core-NafY. As FeMo-co and vitamin B12 are similar in that they are highly labile and reactive and must be delivered to their respective destinations via carrier proteins, it is not surprising that these processes may be evolutionarily related.

4.6 Conclusion

This work has established that the N-terminal domain of NafY is sufficient to bind apo-dinitrogenase by using a pull-down assay in which apo-dinitrogenase is selectively removed from *A. vinelandii* extracts by exposure to immobilized n-NafY. Nitrogenase activity assays show that the presence of n-Nafy inhibits nitrogenase activity. The mechanism of this inhibition was elucidated using EPR spectroscopy which shows that the presence of n-NafY blocks FeMo-co insertion into apo-dinitrogenase. The solution structure of n-NafY was then determined, revealing unpredicted structural homology to the sterile alpha motif, a protein domain involved in binding a diverse set of macromolecular ligands. Utilizing the enhanced NMR relaxation that occurs in close proximity to a paramagnetic center, I have also localized the binding of FeMo-co to a region of the core domain that consists of helices α1, α2, and α3 and the loop between sheets β1 and β2 containing FeMo-co known ligand, histidine 121. Taken together with the previously

published data on the full length protein and the core domain [16, 17], the data presented here are consistent with a model of NafY functioning as a chaperone insertase that both stabilizes the apo form of dinitrogease and delivers FeMo-co to it.

4.7 Acknowledgements

4.8 Abbreviations Used

SAM	Sterile alpha motif
GST	Glutathione-S-transferase
TEV	Tobacco Etch Virus
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
GSH	Reduced glutathione
PAGE	Poly-acrylamide gel electrophoresis
DTH	Dithionite
NMF	N-methylformamide
EPR	Electron paramagnetic resonance
BSA	Bovine serum albumin

4.9 References

1. Bulen, W.A. and J.R. Lecomte, *The Nitrogenase System From Azotobacter: Two-Enzyme Requirement for N2 Reduction, ATP-Dependent H2 Evolution, and ATP Hydrolysis.* Proceedings of the National Academy of Sciences, 1966. **56**: p. 979-986.
2. Hageman, R.V. and R.H. Burris, *Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle.* Proceedings of the National Academy of Sciences, 1978. **75**: p. 2699-2702.
3. Kirn, J. and D.C. Rees, *Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from Azotobacter vinelandii.* Nature, 1992. **360**: p. 553-560.
4. Shah, V.K. and W.J. Brill, *Nitrogenase. IV. Simple method of purification to homogeneity of nitrogenase components from Azotobacter vinelandii.* Biochimica et Biophysica Acta, 1973. **305**: p. 445-454.
5. Igarashi, R.Y. and L.C. Seefeldt, *Nitrogen Fixation: The Mechanism of the Mo-Dependent Nitrogenase.* Critical Reviews in Biochemistry and Molecular Biology, 2003. **38**: p. 351-384.
6. Chan, M., J. Kim, and D. Rees, *The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures.* Science, 1993. **260**: p. 792-794.

7. Rubio, L.M. and P.W. Ludden, *Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase.* Annual Reviews of Microbiology, 2008. **62**: p. 93-111.

8. Paustian, T.D., V.K. Shah, and G.P. Roberts, *Apodinitrogenase: Purification, Association with a 20-Kilodalton Protein, and Activation by the Iron-Molybdenum Cofactor in the Absence of Dinitrogenase Reductase.* Biochemistry, 1990. **1990**: p. 3515-3522.

9. Hawkes, T.R. and B.E. Smith, *Purification and characterization of the inactive MoFe protein (NifB-Kp1) of the nitrogenase from nifB mutants of Klebsiellapneumoniae.* Biochemical Journal, 1983. **209**: p. 43-50.

10. Homer, M.J., D.R. Dean, and G.P. Roberts, *Characterization of the Gamma Protein and Its Involvement in the Metallocluster Assembly and Maturation of Dinitrogenase from Azotobacter vinelandii.* The Journal of Biological Chemistry, 1995. **270**: p. 24745-24752.

11. Homer, M.J., et al., *The nifY Product of Klebsiella pneumoniae Is Associated with Apodinitrogenase and Dissociates upon Activation with the Iron-Molybdenum Cofactor.* Journal of Bacteriology, 1993. **175**: p. 4907-4910.

12. Rubio, L.M., et al., *Cloning and Mutational Analysis of the Gamma Gene from Azotobacter vinelandii Defines a New Family of Proteins Capable of Metallocluster Binding and Protein Stabilization.* The Journal of Biological Chemistry, 2002. **277**: p. 14299-14305.

13. Jacobson, M.R., et al., *Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii.* Journal of Bacteriology, 1989. **171**(2): p. 1017-1027.

14. Joerger, R.D. and P.E. Bishop, *Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii.* Journal of Bacteriology, 1988. **170**: p. 1475-1487.

15. Curatti, L., et al., *In vitro synthesis of the iron–molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins.* Proceedings of the National Academy of Sciences, 2007. **104**: p. 17626-17631.

16. Dyer, D.H., et al., *The Three-dimensional Structure of the Core Domain of NafY from Azotobacter vinelandii determined at 1.8-Å Resolution.* The Journal of Biological Chemistry, 2003. **278**(34): p. 32150-32156.

17. Rubio, L.M., S.W. Singer, and P.W. Ludden, *Purification and Characterization of NafY (Apodinitrogenase Gamma Subunit) from Azotobacter vinelandii.* The Journal of Biological Chemistry, 2004. **279**: p. 19739-19746.

18. Goodwin, P.J., et al., *The Azotobacter vinelandii NifEN Complex Contains Two Identical [4Fe-4S] Clusters.* Biochemistry, 1998. **37**: p. 10420-10428.

19. Nagatani, H.H., V.K. Shah, and W.J. Brill, *Activation of inactive nitrogenase by acid-treated component I.* J. Bacteriol., 1974. **120**(2): p. 697-701.

20. Jacobson, M.R., et al., *Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii.* J. Bacteriol., 1989. **171**(2): p. 1017-1027.

21. Shah, V.K., L.C. Davis, and W.J. Brill, *Nitrogenase. I. Repression and derepression of the iron-molybdenum and iron proteins of nitrogenase in Azotobacter vinelandii.* Biochim. Biophys. Acta, 1972. **256**(2): p. 498-511.

22. Ausubel, F.N., et al., eds. *Current Protocols in Molecular Biology.* 2000, Greene Publishing and Wiley Interscience: New York.

23. Marley, J., M. Lu, and C. Bracken, *A method for efficient isotopic labeling of recombinant proteins.* Journal of Biomolecular NMR, 2001. **20**: p. 71-75.

24. Christiansen, J., et al., *Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a nifB-deletion mutant of Azotobacter vinelandii.* Biochemistry,

1998. **37**(36): p. 12611-12623.

25.   Shah, V.K. and W.J. Brill, *Nitrogenase. IV. Simple method of purification to homogeneity of nitrogenase components from Azotobacter vinelandii.* Biochim. Biophys. Acta, 1973. **305**: p. 445-454.

26.   Shah, V.K. and W.J. Brill, *Isolation of an iron-molybdenum cofactor from nitrogenase.* Proc. Natl. Acad. Sci. USA, 1977. **74**(8): p. 3249-3253.

27.   Shah, V.K., et al., *In vitro synthesis of the iron-molybdenum cofactor of nitrogenase. Purification and characterization of NifB cofactor, the product of NIFB protein.* J. Biol. Chem., 1994. **269**(2): p. 1154-1158.

28.   Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(259): p. 680-685.

29.   Brandner, J.P., et al., *Expression of the Rhodobacter sphaeroides cytochrome c2 structural gene.* J. Bacteriol., 1989. **171**(1): p. 360-368.

30.   Allen, R.M., et al., *Dinitrogenase reductase- and MgATP-dependent maturation of apodinitrogenase from Azotobacter vinelandii.* J. Biol. Chem., 1993. **268**: p. 23670-23674.

31.   Kuo, C.F. and I. Fridovich, *A stain for iron-containing proteins sensitive to nanogram levels of iron.* Anal. Biochem., 1988. **170**(1): p. 183-185.

32.   Smith, P.K., et al., *Measurement of protein using bicinchoninic acid [published erratum appears in Anal Biochem 1987 May 15;163(1):279].* Anal. Biochem., 1985. **150**(1): p. 76-85.

33.   Lyon, E.J., et al., *UV-A/blue-light inactivation of the 'metal-free' hydrogenase (Hmd) from methanogenic archaea.* Eur. J. Biochem., 2004. **271**(1): p. 195-204.

34.   Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system basedon UNIX pipes.* Journal of Biomolecular NMR, 1995. **6**(3): p. 277-293.

35.   Keller, R., *Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment.* 2004/2005, ETH Zürich.

36.   Grzesiek, S. and A. Bax, *An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins.* Journal of Magnetic Resonance, 1992. **99**: p. 201-207.

37.   Grzesiek, S. and A. Bax, *Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR.* Journal of the American Chemical Society, 1992. **114**: p. 6291-6293.

38.   Driscoll, P.C., et al., *Complete Resonance Assignment for the Polypeptide Backbone of Interleukin 16 Using Three-Dimensional Heteronuclear NMR Spectroscopy.* Biochemistry, 1990. **29**: p. 3542-3556.

39.   Grzesiek, S. and A. Bax, *Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins.* Journal of Biomolecular NMR, 1993. **3**: p. 185-204.

40.   Grzesiek, S., J. Anglister, and A. Bax, *Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization.* Journal of Magnetic Resonance, Series B, 1993. **101**: p. 114-119.

41.   Sun, Z.-Y.J., et al., *High-resolution aliphatic side-chain assignments in 3D HCcoNH experiments with joint H–C evolution and non-uniform sampling.* Journal of Biomolecular NMR, 2005. **32**: p. 55-60.

42.   Kay, L.E., et al., *A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-*

Chain 1H and 13C Correlations in H2O Samples of Proteins. Journal of Magnetic Resonance, Series B, 1993. **101**: p. 333-337.

43. Vuister, G.W. and A. Bax, *Measurement of four-bond HN-HA J-couplings in staphylococcal nuclease.* Journal of Biomolecular NMR, 1994. **4**: p. 193-200.

44. Cornilescu, G., F. Delaglio, and A. Bax, *Protein backbone angle restraints from searching a database for chemical shift and sequence homology.* Journal of Biomolecular NMR, 1999. **13**: p. 289-302.

45. Ottinger, M., F. Delaglio, and A. Bax, *Measurement of J and Dipolar Couplings from Simplified Two-Dimensional NMR Spectra.* Journal of Magnetic Resonance, 1998. **131**(2): p. 373-378.

46. Günter, P., *Automated NMR Structure Calculation With CYANA.* Methods in Molecular Biology, 2004. **278**: p. 353-378.

47. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package.* Journal of Magnetic Resonance, 2003. **160**: p. 65-73.

48. Clore, G.M., A.M. Gronenborn, and A. Bax, *A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information.* Journal of Magnetic Resonance, 1998. **133**: p. 216-221.

49. Baker, N., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proceedings of the National Academy of Sciences, 2001. **98**: p. 10037-10041.

50. Wittenkind, M. and L. Mueller, *HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins.* Journal of Magnetic Resonance, Series B, 1993. **101**(121-227).

51. Eletsky, A., A. Kienhöfer, and K. Pervushin, *TROSY NMR with partially deuterated proteins.* Journal of Biomolecular NMR, 2001. **20**: p. 177-180.

52. Rovnyak, D., et al., *Accelerated acquisition of high resolution triple-resonance spectra using nonuniform sampling and maximum entropy reconstruction.* Journal of Magnetic Resonance, 2004. **170**: p. 15-21.

53. Hahn, E.L., *An accurate nuclear magnetic resonance method for measuring spinlattice relaxation times.* Physical Review, 1949. **76**: p. 145.

54. Kay, L.E., et al., *Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins.* Journal of Magnetic Resonance, 1992. **97**: p. 359-375.

55. Bernadoa, P., et al., *Interpretation of NMR relaxation properties of Pin1, a two-domain protein, based on Brownian dynamic simulations.* Journal of Biomolecular NMR, 2004. **29**: p. 21-35.

56. Christiansen, J., et al., *Catalytic and Biophysical Properties of a Nitrogenase Apo-MoFe Protein Produced by a nifB-Deletion Mutant of Azotobacter Vinelandii.* Biochemistry, 1998. **1998**(37): p. 12611-12623.

57. Schmid, B., et al., *Structure of a Cofactor-Deficient Nitrogenase MoFe Protein.* Science, 2002. **296**: p. 352-356.

58. Peters, J.W., et al., *Redox-Dependent Structural Changes in the Nitrogenase P-Cluster.* Biochemistry, 1997. **36**(6): p. 1181-1187.

59. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallographica, 2004. **D60**: p.

2256-2268.

60. Qiao, F. and J.U. Bowie, *The Many Faces of SAM*. Science's STKE, 2005. **2005**(286): p. p. re7.

61. Kim, C.A. and J.U. Bowie, *SAM domains: uniform structure, diversity of function.* TRENDS in Biochemical Sciences, 2003. **28**(12): p. 625-628.

62. Ponting, C.P., *SAM: A novel motif in yeast sterile and Drosophila polyhomeotic proteins.* Protein Science, 1995. **4**: p. 1928-1930.

63. Kwan, J.J., et al., *The Solution Structure of the S. cerevisiae Ste11 MAPKKK SAM Domain and its Partnership with Ste50.* Journal of Molecular Biology, 2004. **342**: p. 681-693.

64. Green, J.B., et al., *RNA Recognition via the SAM Domain of Smaug.* Molecular Cell, 2003. **11**: p. 1537-1548.

65. Barrera, F., et al., *Binding of the C-terminal sterile alpha motif (SAM) domain of human p73 to lipid membranes.* The Journal of Biological Chemistry, 2003. **278**: p. 46878-46885.

66. Clore, G. and D. Garret, *R-factor, Free R, and Complete Cross-Validation for Dipolar Coupling Refinement of NMR Structures.* Journal of the American Chemical Society, 1999. **1999**(39): p. 9008-9012.

67. Wei, Y. and M. Werner, *iDC: A comprehensive toolkit for the analysis of residual dipolar couplings for macromolecular structure determination.* Journal of Biomolecular NMR, 2006. **35**(1): p. 17-25.

68. Bhattacharya, A., R. Tejero, and G. Montelione, *Evaluating protein structures determined by structural genomics consortia.* Proteins: Structure, Function, and Bioinformatics, 2007. **66**(4): p. 761-1007.

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

70. Sippl, M., *Recognition of errors in three-dimensional structures of proteins.* Proteins: Structure, Function, and Genetics, 1993. **17**(4): p. 355-362.

71. Lovell, S., et al., *Structure validation by Calpha geometry: phis,psi and Cbeta deviation.* Proteins: Structure, Function, and Genetics, 2003. **50**(3): p. 437-450.

72. Einsle, O., et al., *Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor.* Science, 2002. **297**: p. 1696-1700.

73. Bertini, I., C. Luchinat, and M. Piccioli, *Paramagnetic Probes in Metalloproteins.* Methods in Enzymology, 2001. **339**.

74. Tollinger, M., et al., *How a protein prepares for B12 binding: structure and dynamics of the B12-binding subunit of glutamate mutase from Clostridium tetanomorphum.* Structure, 1998. **6**(1021-1033).

75. Hoffmann, B., et al., *A Protein Pre-Organized to Trap the Nucleotide Moiety of Coenzyme B12 : Refined Solution Structure of the B12-Binding Subunit of Gutamate Mutase from Clostridium tetanomorphum.* Chembiochem, 2001. **2**: p. 643-655.

76. Hoffmann, B., et al., *Structure and dynamics of the B12-binding subunit of glutamate mutase from Clostridium cochlearium.* European Journal of Biochemistry, 1999. **263**: p. 178-188.

77. Das, A., et al., *Characterization of a Corrinoid Protein Involved in the C1 Metabolism of Strict Anaerobic Bacterium Moorella thermoacetica.* Proteins: Structure, Function, and

Bioinformatics, 2007. **67**: p. 167–176.

78.   Wishart, D. and B. Sykes, *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.* Journal of Biomolecular NMR, 1994. **4**: p. 171-180.

79.   Bonin, I., et al., *Structural basis for the interaction of Escherichia coli NusA with protein N of phage lambda.* Proceedings of the National Academy of Sciences, 2004. **101**: p. 13762-13767.

80.   Tosi, T., et al., *Structures of the tumor necrosis factor alpha inducing protein Tipa: A novel virulence factor from Helicobacter pylori.* FEBS Letters, 2009. **583**: p. 1581-1585.

The End