Title
Analysis of the left-handed β-helical fold toward understanding of amyloid fibril structure

Author
Choi, Jay Hoon

Publication Date
2008

Peer reviewed|Thesis/dissertation

ANALYSIS OF THE LEFT-HANDED β-HELICAL FOLD TOWARD UNDERSTANDING OF AMYLOID FIBRIL STRUCTURE

by

Jay Hoon Choi

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological & Medical Informatics

in the

GRADUATE DIVISION

of the

Acknowledgements

First, I thank my advisor, Fred Cohen for providing years of support during my tumultuous time in graduate school. It is uncommon to find an advisor who could provide scientific advice as well as mental support. The first several years of my graduate school career was an extremely difficult time for me, as I dealt with medical injuries, accidents, and personal problems, all which had a profound effect on my research interest and performance. During this time, he gave me endless support at times by giving me complete freedom to explore what I really wanted to do, and at other times with specific instructions on what to do next day. I give him my deepest appreciation for believing in me and supporting me for the past six years. I would also like to thank my academic advisor, Patricia Babbitt, who granted me admission to UCSF, and provided me with support and encouragement when I needed it throughout my graduate career. I am truly grateful for her support, and I wish her a fast recovery from her recent surgery. I also thank Robert Fletterick and Charles Craik, who were very helpful in shaping my research project and who both served my oral and thesis committees.

The many members of the Cohen and Prusiner labs were terrific colleagues for the past five years and I appreciate all of their help. I would especially like to thank Sarit Helman who deserves deep appreciation for providing me with care, help and encouragement during my six years at the Cohen Lab. I would also like to extend great appreciation to Cedric Govaerts, who collectively taught me everything I know about protein structures and gave me countless like lessons.

A special thank you goes to Barnaby May, a great mentor whose advice and hard work have been enormously helpful, especially in the last three years. From building the experimental lab to planning the projects, designing and developing the assay to publishing scientific research, he basically thought me everything I needed to learn. Without his care and help, I would not be here.

I thank my parents, my sister, and my cousin, Eugene for always believing in me and encouraging me. Thanks also go to my life time friends back in Korea, Sangchul Lee, Wonsuk Kim, Jiwhan Kim, Kyoungwoo Seo, and Jeoungkyu Namgoong for so many great times in Seoul and elsewhere.

Finally, thank you my dear Jina, for years of daily companionship and nurturing a constant belief in me, that I could achieve anything I want. I love you with all my heart.

Analysis of the left-handed β-helical fold toward understanding of amyloid fibril structure

Jay Hoon Choi

The left-handed parallel β-helix (LβH) is a structurally repetitive, highly regular, symmetrical fold, formed by coiling of elongated β-sheets into helical 'rungs'. This canonical fold has recently received interest as a possible solution to the fibril structure of amyloid. However, sequence and structural requirements, and stability and folding of the LβH are not fully understood. To analyze this protein fold and investigate the possibility of its involvement in amyloid formations, combined approaches using computational and experimental methods were applied. (1) Sequence characteristics of the LβH were determined by analyzing known structures. A genome-wide sequence analysis was undertaken to determine the prevalence of this unique protein fold across the genomes. Molecular dynamics (MD) simulations were used to demonstrate the stabilizing effect of successive rungs and the hydrophobic core of the LβH. (2) An *in-vivo* folding assay using a known LβH protein was developed to investigate the residue tolerance and effect of mutations at structurally critical regions of the β-helical structure. Tyrosine fluorescence spectroscopy was also used to study the thermodynamic stability of the LβH protein. It showed that the β-helix could tolerate non-hydrophobic residues at interior positions, and proline residues in the β-helical domain were shown to be important, but not critical, in folding of the β-helix. A recombinant protein consisting of the LβH with a prion fragment was designed, constructed and tested. The results suggest that the amyloid prone PrP fragment may fold into a β-helix. (3) Modeling of possible monomeric

subunits of the insulin amyloid fibril was conducted using β-solenoid folds, namely the β-helix and β-roll. While both the β-helix and β-roll models agreed with currently available biophysical data from misfolded insulin, MD simulations showed that the β-roll model was relatively more stable. Insulin monomeric subunit models were incorporated into plausible models of insulin fibrils. Simulated X-ray fiber diffraction patterns of these models demonstrated that the model fibers constructed from a β-solenoid subunit provided the reasonable fit to available experimental diffraction patterns.

Table of Contents

Introduction 1

Chapter 1 Analysis of the sequence and structural features of the left-handed β-helical fold 19

Chapter 2 Stability analysis of the left-handed β-helical fold using molecular dynamics simulations 49

Chapter 3 *In-vivo* folding study of the left-handed β-helical fold . . . 77

Chapter 4 Design and engineering of a recombinant left-handed β-helical protein with inclusion of a prion fragment. 122

Chapter 5 Molecular modeling of the misfolded insulin monomeric subunit 133

Chapter 6 Molecular modeling and fiber diffraction analysis of the insulin amyloid fibril 166

List of Tables

Table 1 Sequence statistics of LβH turn regions 29

Table 2 Structural statistics from the MD simulation of LβH 54

Table 3 Amino acid residue tolerance of LpxA LβH domain from the *in-vivo LpxA* folding assay. 103

Table 5 Structural statistics from the MD simulation of the β-roll and β-helix insulin models 147

List of Figures

Figure 1.1 Side and cross-section view of LβH structures 28

Figure 1.2 Genome-wide search for LβH domains using sequence patterns and structural constraints 34

Figure 1.3 The distribution of LβH loop regions and LβH domain . . 36

Figure 2.1 Molecular dynamics simulation of LβH models with increasing numbers of rungs 53

Figure 2.2 Molecular dynamics simulation of elongating LβH models 55

Figure 2.3 Molecular dynamics simulations of LβH models with varying cavity sizes 58

Figure 2.4 Stability analysis of type-I LβH vs. type-II LβH 62

Figure 2.5 Structural variations of the LβH fold 65

Figure 3.1 Monomeric and trimeric forms of the LβH protein, UDP-*N*-acetylglucosamine acyltransferase (LpxA) 87

Figure 3.2 LpxA *in-vivo* folding assay for wild type LpxA and arginine LpxA mutant proteins 91

Figure 3.3 *In-vivo* folding assay for LpxA mutant proteins with mutations in the hydrophobic core 95

Figure 3.4 *In-vivo* folding assay and spectroscopic characterization of wild type LpxA and LpxA proline mutant proteins 100

Figure 4.1 Threading the PrP sequence onto the LpxA protein and *in-vivo* folding assay of the LpxA-PrP recombinant protein 126

Figure 5.1 Insulin monomer structure and β-solenoid–based monomeric subunit models for the insulin amyloid fibril . 142

Figure 5.2 Molecular dynamics simulation of the β-roll and β-helix insulin monomeric models with increasing temperature in acidic conditions. 146

Figure 5.3 Molecular dynamics simulation of single chains from the β-roll and β-helix monomeric models at 345 K under acidic conditions. 149

Figure 6.1 Quaternary structure arrangements to simulate insulin amyloid fibrils 169

Figure 6.2 Experimental and simulated diffraction patterns of an insulin amyloid fibril and a six-protofilament β-roll–based insulin fibril model. 174

Figure 6.3 Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of β-roll–based insulin fibril models. 176

Figure 6.4 Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of β-helix–based insulin fibril models. 177

Figure 6.5 Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of multifold-symmetric β-helix–based insulin fibril models. 178

Introduction:

Protein misfolding, amyloid, and β-helices

Protein misfolding

Proteins are essential parts of living organisms and participate in virtually every process within living cells. In the last several decades, we have learned a great deal about the various structural forms that the protein molecules adopt in their native states and their functional roles in cellular systems. We are also beginning to have a basic understanding of the complex protein folding mechanism at a molecular level and how biological systems have evolved to optimize the protein structure and to ensure the correct folding of proteins. In recent years, this knowledge has led to the development of numerous computational methods that predict the structures and functions of proteins from their genomic information, which has become largely available from the various genome projects in the past decade. Due to the importance of proteins and their roles in cellular systems, failure to synthesize, or misfolding of proteins, may cause the malfunctioning of cellular mechanisms and serious harm to the host organism. Thus, biological systems have evolved elaborate protection procedures to prevent this failure to synthesize or misfolding of proteins by detection and degradation of proteins when misfolding occurs.

Despite these inherent cellular level controls to prevent protein misfolding events, a number of human diseases are still associated with protein misfolding and the subsequent aggregation of proteins (Dobson 1999; 2001; 2003b). It includes a wide range of diseases from cystic fibrosis and familial emphysema to non-neuropathic systemic and localized diseases such as Type II diabetes, and neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and the prion diseases (Dobson 2003a; Prusiner 2004; Ross and Poirier 2004; Chiti and Dobson 2006). These diseases are often inherited and sporadic, but they can also be infectious. Each disease is associated with a particular protein or aggregates of these proteins, which are thought to be the direct or indirect cause of the pathological conditions associated with the disease (Dobson 1999). In recent years, most research has been focused on a group of diseases that involve protein misfolding or the destabilization of the normal soluble structure of proteins followed by subsequent conversions to insoluble fibrillar aggregations, which accumulate in a variety of organs including the liver, spleen and brain (Ross and Poirier 2004). These fibrillar forms of ordered aggregates are known as amyloid. Although proteins that are involved in these amyloid diseases are structurally and evolutionarily unrelated in their native states, the misfolded fibrillar form of proteins found

in the disease states share common structural characteristics with similar overall morphology (Sunde et al. 1997; Dobson 2001).

In the last decades, many of the studies on amyloid diseases have been focused on the structural understanding of amyloid fibrils and the mechanism of the conversion from native proteins into fibrillar aggregates. In many cases, the structural conversion and formation of the amyloid fibrils of the disease-associated proteins have been reproduced *in-vitro* (Kelly 1998; Dobson 2004; Luheshi et al. 2008). Recent studies have revealed that proteins, under conditions that destabilize the native state, undergo a process of unfolding or misfolding which forms intermediate aggregates prior to the formations of well-defined amyloid fibrils. It is also becoming increasingly evident that these intermediate aggregates or oligomeric fibril precursors are the toxic species and the primary origin of pathological conditions (Caughey and Lansbury 2003; Arrasate et al. 2004). However, it remains unclear whether the amyloid formation is part of the cellular mechanism which removes toxic intermediate aggregates, or the amyloid formation is itself responsible for the toxicity.

Structural study of amyloid

Amyloid fibrils are generally unbranched, protease-resistant filaments with dominant β-sheet structures organized in a cross-β fashion, in which the β-strands run perpendicular to the fibril axis (Serpell and Smith 2000; Jimenez et al. 2002; Murali and Jayakumar 2005; Serpell et al. 2007).

Electron microscopy (EM), cryo-EM, and atomic force microscopy (AFM) have revealed that the protofilament units are compact and the β-strand repeats with typical diameters ranging 15 to 40 Å (Jimenez et al. 1999; Jimenez et al. 2002; Khurana et al. 2003). Although amyloid fibrils share common structural properties, their overall morphology can vary depending on the precursor protein, conditions of fibrillization, and the number and arrangement of protofilaments (Goldsbury et al. 1997; Jimenez et al. 1999; Bouchard et al. 2000; Jimenez et al. 2002).

Over the last few decades a great deal of effort has been invested in attempting to gain a structural understanding of the mechanisms of protein misfolding and subsequent amyloid fibril formation. For example, recent X-ray crystallography studies of microcrystals formed from the short peptides of amyloid-related proteins have revealed a distinctive packing in the β-structure core (Nelson et al. 2005; Sawaya et al. 2007; Wiltzius et al. 2008).

EM and solid-state nuclear magnetic resonance studies of A$\beta_{1\text{-}40}$ and the fungal HET-s prion protein showed that the misfolded amyloid conformation may adopt an architecture that is structurally unrelated to the native conformations, but similar to β-structures previously observed in nature, namely β-helix or β-solenoid folds (Jenkins and Pickersgill 2001; Hennetin et al. 2006; Kajava and Steven 2006; Sachse et al. 2008; Wasmer et al. 2008).

β-helix: why is it interesting in structural study of amyloid?

The limitations of traditional biophysical methods for studying misfolded proteins have led to the use of molecular modeling to probe plausible structural solutions. In the last decades, a number of different theoretical models of the structural subunit of the amyloid fibrils have been proposed, such as a trimeric β-helical model of the prion protein (Govaerts et al. 2004) and the β-arch model of Aβ peptide (Petkova et al. 2002). Increasing evidence from EM and solid-state NMR studies indicates that a parallel β-helix or β-solenoid-like structure may serve as a monomeric assembling subunit of amyloid fibrils, with subtle variations, providing the diversity of amyloid fibrils observed (Jimenez et al. 2002; Sachse et al. 2008; Wasmer et al. 2008).

The left-handed parallel β-helix (LβH) is a structurally repetitive, highly regular and symmetrical fold formed by the coiling of elongated β–sheets into helical 'rungs'. Recent modeling studies have proposed LβH folds as plausible solutions to the structures of misfolded proteins associated with the prion and Huntington's diseases (Govaerts et al. 2004; Stork et al.

2005; Yang et al. 2005; Langedijk et al. 2006; Merlino et al. 2006). These studies have converged on the β-helical architecture due to the structural features that these folds share with the unresolved structure of the longer chain amyloids. For example, biophysical studies have indicated that the longer chain amyloids are β-rich with extensive cross-β structure, a feature that could be accounted for by end-to-end polymerization of β-helical-like subunits. However, the accuracy of modeling studies employing LβH folds in the fields of amyloid research has been limited by the relative absence of information pertaining to the sequence and structural features of this relatively rare protein fold.

Overview

In this dissertation, a variety of different approaches are brought together by the single theme of understanding sequence and structural characteristics of the LβH to investigate its structural involvement in amyloid fibril formation.

Chapter 1 is focused on sequence and structural analysis of the LβH fold by inspecting available LβH structures and by conducting genome-wide sequence analysis. The LβH fold has recently received interest as a possible solution to the fibril structure of amyloid. As an aim to understand the structural requirements of the LβH fold, sequence characteristics of the LβH were determined by analyzing known structures and identifying positional preferences of specific residues types. In addition, the structure-based sequence preference of the LβH was defined, and a genome-wide sequence search was undertaken to determine the prevalence of this unique protein fold. This profile-based LβH search algorithm predicted a large fraction of LβH proteins from microbial origins. However, the relative number of predicted LβH proteins per species was approximately equal across the genomes from prokaryotes to eukaryotes.

Chapter 2 extends this work on analysis of the LβH to further study the structural stability of the LβH. Molecular dynamics (MD) simulations were used to demonstrate the stabilizing effect of successive rungs and the hydrophobic core of the LβH. It is also shown that a two-rung structure is the minimally stable LβH structure.

In Chapter 3, residue tolerance and the effect of mutations on the folding of β-helix are discussed. The structural requirements of the LβH fold were examined experimentally using *in-vivo* and *in-vitro* systems. An *in-vivo* folding assay using a known LβH protein, UDP-*N*-acetylglucosamine acyltransferase (LpxA), from *E. coli* was developed to investigate the tolerance for and the effect of mutations at structurally critical regions of the β-helical structure. Tyrosine fluorescence spectroscopy was also used to study the *in-vitro* folding of LpxA protein to determine the effect of proline mutations on the thermodynamic stability of the LβH protein. In addition to previously known sequence preferences at each canonical position of the β-helix, it was discovered that the β-helix could tolerate non-hydrophobic residues at interior positions, although this resulted in a decrease of the LpxA enzymatic activity. In addition, proline

residues in the β-helical domain are shown to be important, but not critical, in folding of the β-helix.

Chapter 4 extends the folding studies of the LβH and discusses the design of a recombinant protein consisting of LpxA with a prion fragment (LpxA-PrP). Using the information obtained about the residue tolerance of the β-helix, a new sequence threading of a prion fragment (residues 104-143) on LpxA was conducted. Based on the threading result, the recombinant LpxA-PrP was constructed and the activity of the recombinant LpxA was tested using our *in-vivo* folding assay. The results of these experiments suggest that the amyloid prone PrP fragment may fold into a β-helix.

In Chapter 5, molecular modeling of the misfolded insulin monomeric subunit using β-helix and β-roll folds is discussed. Insulin, a small hormone protein with 51 residues in two disulfide-linked polypeptide chains, adopts a predominantly α-helical conformation in its native state. It readily undergoes protein misfolding and aggregates into amyloid fibrils under a variety of conditions. Insulin is a unique model system in which to study protein fibrillization, since its three disulfide bridges are retained in the fibrillar state and thus limit the conformational space available to the protein

during misfolding and fibrillization. Taking into account this unique conformational restriction, modeling of possible monomeric subunits of the insulin amyloid fibril using β-solenoid folds, namely the β-helix and β-roll, was conducted. Both the β-helix and β-roll models agreed with currently available biophysical data from misfolded insulin. However, molecular dynamics simulations showed that the β-roll model was relatively more stable suggesting this may provide the structural stability expected from a fiber subunit.

Chapter 6 extends the modeling studies of the β-solenoids and discusses how to incorporate these motifs into plausible models of insulin fibrils using available biophysical data. Simulated X-ray fiber diffraction patterns of these models demonstrate that the model fibers constructed from a β-roll subunit provide the most reasonable fit to available experimental diffraction patterns. This study suggests that the β-roll or similar β-solenoid-type subunit may provide a solution to the structure of misfolded insulin.

References

Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431:** 805-810.

Bouchard, M., Zurdo, J., Nettleton, E.J., Dobson, C.M., and Robinson, C.V. 2000. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* **9:** 1960-1967.

Caughey, B., and Lansbury, P.T. 2003. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26:** 267-298.

Chiti, F., and Dobson, C.M. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75:** 333-366.

Dobson, C.M. 1999. Protein misfolding, evolution and disease. *Trends Biochem Sci* **24:** 329-332.

Dobson, C.M. 2001. The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356:** 133-145.

Dobson, C.M. 2003a. Protein folding and disease: a view from the first Horizon Symposium. *Nat Rev Drug Discov* **2:** 154-160.

Dobson, C.M. 2003b. Protein folding and misfolding. *Nature* **426:** 884-890.

Dobson, C.M. 2004. Experimental investigation of protein folding and misfolding. *Methods* **34:** 4-14.

Goldsbury, C.S., Cooper, G.J., Goldie, K.N., Muller, S.A., Saafi, E.L., Gruijters, W.T., Misur, M.P., Engel, A., Aebi, U., and Kistler, J. 1997. Polymorphic fibrillar assembly of human amylin. *J Struct Biol* **119:** 17-27.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. 2004. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101:** 8342-8347.

Hennetin, J., Jullian, B., Steven, A.C., and Kajava, A.V. 2006. Standard conformations of beta-arches in beta-solenoid proteins. *J Mol Biol* **358:** 1094-1105.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

Jimenez, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M., and Saibil, H.R. 1999. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *Embo J* **18:** 815-821.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. 2002. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99:** 9196-9201.

Kajava, A.V., and Steven, A.C. 2006. Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv Protein Chem* **73:** 55-96.

Kelly, J.W. 1998. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* **8:** 101-106.

Khurana, R., Ionescu-Zanetti, C., Pope, M., Li, J., Nielson, L., Ramirez-Alvarado, M., Regan, L., Fink, A.L., and Carter, S.A. 2003. A general model for amyloid fibril assembly based on morphological studies using atomic force microscopy. *Biophys J* **85:** 1135-1144.

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Luheshi, L.M., Crowther, D.C., and Dobson, C.M. 2008. Protein misfolding and disease: from the test tube to the organism. *Curr Opin Chem Biol* **12:** 25-31.

Merlino, A., Esposito, L., and Vitagliano, L. 2006. Polyglutamine repeats and beta-helix structure: molecular dynamics study. *Proteins* **63:** 918-927.

Murali, J., and Jayakumar, R. 2005. Spectroscopic studies on native and protofibrillar insulin. *J Struct Biol* **150:** 180-189.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435:** 773-778.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. 2002. A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **99:** 16742-16747.

Prusiner, S.B. 2004. *Prion biology and diseases*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., pp. xiii, 1050 p.

Ross, C.A., and Poirier, M.A. 2004. Protein aggregation and neurodegenerative disease. *Nat Med* **10 Suppl:** S10-17.

Sachse, C., Fandrich, M., and Grigorieff, N. 2008. Paired beta-sheet structure of an Abeta(1-40) amyloid fibril revealed by electron microscopy. *Proc Natl Acad Sci U S A* **105:** 7462-7466.

Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., et al. 2007. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447:** 453-457.

Serpell, L.C., Benson, M., Liepnieks, J.J., and Fraser, P.E. 2007. Structural analyses of fibrinogen amyloid fibrils. *Amyloid* **14:** 199-203.

Serpell, L.C., and Smith, J.M. 2000. Direct visualisation of the beta-sheet structure of synthetic Alzheimer's amyloid. *J Mol Biol* **299:** 225-231.

Stork, M., Giese, A., Kretzschmar, H.A., and Tavan, P. 2005. Molecular dynamics simulations indicate a possible role of parallel beta-helices in seeded aggregation of poly-Gln. *Biophys J* **88:** 2442-2451.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C. 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273:** 729-739.

Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319:** 1523-1526.

Wiltzius, J.J., Sievers, S.A., Sawaya, M.R., Cascio, D., Popov, D., Riekel, C., and Eisenberg, D. 2008. Atomic structure of the cross-{beta} spine of Islet Amyloid Polypeptide (Amylin). *Protein Sci*.

Yang, S., Levine, H., Onuchic, J.N., and Cox, D.L. 2005. Structure of infectious prions: stabilization by domain swapping. *FASEB J* **19:** 1778-1782.

Chapter 1

Analysis of the sequence and structural features of the left-handed β-helical fold

Introduction

The parallel β-helix is a repetitive fold where the repeating unit is a β-helical coil formed by segments of β-strand (Jenkins and Pickersgill 2001; Iengar et al. 2006; Kajava and Steven 2006). With few exceptions, both the right-handed β-helix (RβH) and LβH share common structural features. Each rung of the canonical β-helix consists of two to three β-strands interrupted by turn or loop regions (Simkovsky and King 2006). The β-helical rungs are aligned to form a cross-β structure such that elongated β-strands connected by hydrogen bonds lie parallel to the helical axis (Govaerts et al. 2004; Simkovsky and King 2006). Structural repetition of coils creates a cylindrical hydrophobic core. The hydrophobic core of the β-helical proteins is characterized by buried stacks of similar side chains (Jenkins and Pickersgill 2001). While RβH is generally characterized by β-strands connected by variable length of turns and loops, LβH is more rigid and repetitive than the RβH variant (Iengar et al. 2006).

Since the first crystal structure of a LβH protein was determined: UDP-*N*-acetylglucosamine acyltransferase, LpxA, from *E. coli* (Raetz and Roderick 1995), the structures of nine different proteins and their homologs have been reported to contain the LβH fold. To date, all known LβH

structures are bacterial in origin and share a similar transferase activity (Jenkins and Pickersgill 2001; Kajava and Steven 2006). One exception is the antifreeze protein from the spruce budworm (Graether et al. 2000; Jenkins and Pickersgill 2001; Leinala et al. 2002a; Leinala et al. 2002b). All known bacterial LβH folds (referred to as a type-I LβH) have six-residues per strand (18-residue per rung), described as an imperfect repeating hexapeptide motif, [LIV]-[GAED]-X_2-[STAV]-X. The smaller LβH structure of spruce budworm antifreeze protein (referred to as a type-II LβH) consists of five-residues per strand (15-residue per rung). Type-I LβH and type-II LβH folds share a similar basic architecture and structural pattern. Each rung (or coil) of the canonical β-helix consists of three flat and untwisted parallel β-strands connected by either a one- or two-residue turn or a long external loop region (Jenkins and Pickersgill 2001; Parisi and Echave 2004; Iengar et al. 2006; Kajava and Steven 2006). In this study, loops are defined as a stretch of sequence containing more than two residues whose backbone alignment deviates from the normal LβH turn. Loop regions are a unique feature of the type-I LβH fold and also occur uniformly across all known type-I LβH structures. The largest LβH domain identified to date is from the bifunctional *N*-acetylglucosamine 1-phosphate

uridyltransferase of *E. coli* and *Streptococcus pneumoniae*, and it contains approximately nine rungs with one external loop between seventh and eighth rung (Brown et al. 1999; Olsen and Roderick 2001; Sulzenbacher et al. 2001). No external loops have been observed in the two isoforms of the type-II LβH antifreeze proteins (Graether et al. 2000; Jenkins and Pickersgill 2001; Leinala et al. 2002a; Leinala et al. 2002b). Each β-stand contains small, uncharged residues (V, A, S, T, C) and conserved larger hydrophobic residues (L, I, V) that face the interior of the LβH to create a hydrophobic core (Jenkins and Pickersgill 2001; Govaerts et al. 2004; Iengar et al. 2006; Kajava and Steven 2006). These constraining requirements at interior positions of the β-strand are presumed to have limited sequence divergence of LβH proteins throughout evolution (Parisi and Echave 2001).

Known type-I LβH proteins are composed of a single LβH domain that is capped at either the N- or C-terminus with β-helical domains. All known type-I LβH proteins exist as native trimers, with adjacent monomers interacting at the surface of the LβH domains. In contrast, known type-II LβH proteins exist as a single LβH domain protein where the likely functional unit is a monomer (Leinala et al. 2002a; Leinala et al. 2002b).

Although oligomerization is thought to increase the overall stability of LβH proteins, no experimental evidence has been provided.

Recent modeling studies have proposed β-helical type folds as possible solutions for the structure of misfolded proteins associated with prion and Huntington's diseases (Govaerts et al. 2004; Stork et al. 2005; Yang et al. 2005; Langedijk et al. 2006; Merlino et al. 2006). These studies have converged on the β-helical architecture due to the structural features that these folds share with the unresolved structure of the longer chain amyloids. For example, biophysical studies have indicated that the longer chain amyloids are β-rich with extensive cross-β structure, a feature that could be accounted for by end-to-end polymerization of β-helical-like subunits. The LβH fold is highly regular and symmetrical with little variability in shape and size over the length of the domain (Zheng et al. 2007). These features also have led to the suggestion that LβH fold may be used as a building block for nanotubular structures with application in nanotechnology (Haspel et al. 2006; Haspel et al. 2007; Zheng et al. 2007). However, the accuracy of modeling studies employing LβH folds in the fields of amyloid and nanotechnology research has been limited by the relative absence of information pertaining to the sequence and structural

features of this relatively rare protein fold. In the present study, we have re-examined 9 available LβH structures with the aim of more rigorously defining structural features of the LβH. We show that there are strict residue preferences at β-helix turn regions, in addition to the highly conserved hydrophobic residues at the β-helix core. MD simulations using simple LβH models confirm that backbone hydrogen bonds and the hydrophobic core provide structural stability to the LβH fold. The stability of the type-I LβH structure relative to the smaller type-II LβH structure provides a possible explanation for the abundance of type-I LβH over type-II LβH in the genomes. A survey on the currently available proteomic database shows the existence of the LβH fold in all genomes.

Position dependent amino acid propensities of the LβH

In order to evaluate the amino acid propensities of the repeating feature of the LβH fold, sequence fragments of type-I LβH and type-II LβH rungs were extracted from the 9 known crystal structures of LβH proteins. Incomplete rungs located at the top and bottom of the LβH domains were excluded from the analysis since these capping sequences often contained sequences that do not follow the usual sequence pattern associated with LβH's. The residues that were part of the external loop regions and/or deviated from the structural alignments were also excluded from the analysis. Figure 1.1 shows the resulting amino acid propensities at each position of type-I LβH and type-II LβH. As reported elsewhere, the residues at the B_1^i and B_3^i positions are oriented toward the core of the LβH, and as such are limited to uncharged and hydrophobic residues in the proteins characterized to date at atomic level. Additionally, given the limited available volume of the interior of the LβH turn, B_1^i residues are limited to smaller side chains (V, A, S, T, C). Residues at positions B_2^o and B_4^o face the outside of the β-

helix domain and represent a wide range of residues, including charged to aromatic side chains.

Glycine and proline in LβH turns

The amino acid residue propensities of the β-helix turn regions of the LβH are summarized in Table 1. The type-I LβH turn consists of four residues (B_4^o, T_1, T_2, and B_1^i) and often resembles the classical type-II β-turn while the type-II LβH turn consists of three residues (B_4^o, T, and B_1^i). Hydrogen bonds are often observed in type-I LβH turns between the backbone carbonyl oxygen of the B_4^o residue and amide hydrogen of the B_1^i residue in type-I LβH. This same hydrogen bonding was not observed in type-II LβH turns. According to the current analysis, both type-I LβH and type-II LβH turn regions contain a high incidence of glycine. Glycine is highly conserved at the B_4^o position in the type-I LβH turn, most likely due to steric constraints as the B_4^o side-chain orients toward the β-strand of the following rung (Iengar et al. 2006). Glycine also has a relatively high occurrence at type-I LβH T_2 position and at type-II LβH T position while almost no glycines were observed at any other positions.

The occurrence of proline was limited to position T_1 of the type-I LβH. Some prolines were observed at T_2 and B_4^o positions but only in cases where an external loop preceded the T_2 residue, thus allowing the necessary flexibility to accommodate proline or an incomplete rung located at the edge of the LβH domain. The location of proline in the LβH domain is also conserved with approximately 56% of prolines located at the top or bottom rungs of the LβH domain.

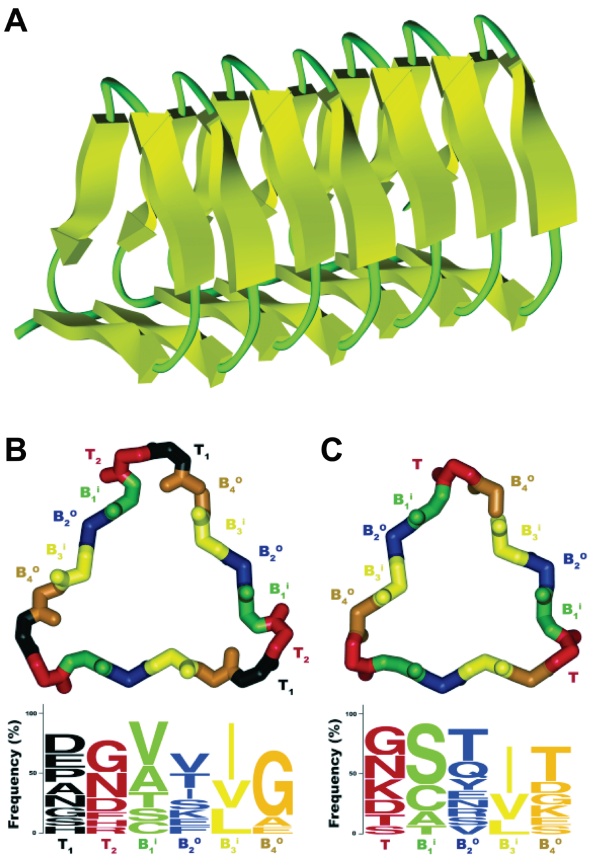


Figure 1.1. Side and cross-section view of LβH structures: (A) The left-handed parallel β-helix (LβH) of 1HM9 (segment 260-379). (B) Type-I LβH, and (C) Type-II LβH, labeled T (turn) and B (β-sheet) with superscript "i" for a residue facing *inside* and "o" for a residue facing *outside.* The position dependent residue propensities are depicted as sequence logos. Percent amino acid residue frequencies >5% are shown.

Table 1. Sequence statistics of LβH turn regions

A. Position dependent amino acid residue propensities of type-I and type-II LβH folds.

	Position in LβH rung			
Type I	B_4^O	T_1	T_2	B_1^i
	GLY 55%	ASP 16%	GLY 27%	VAL 36%
		PRO 12%	ASN 21%	ALA 22%
		GLU 12%	ASP 12%	THR 14%
		ASN 10%		SER 11%
		ALA 10%		CYS 10%
Type II	B_4^O	T	B_1^i	
	THR 29%	GLY 24%	SER 51%	
	GLY 10%	ASN 20%	CYS 22%	
	LYS 10%	LYS 17%	ALA 12%	
	ASP 10%	ASP 12%		
		THR 10%		

B. Distribution of proline and glycine (%)

	Position in LβH rung					
Type I	T_1	T_2	B_1^i	B_2^O	B_3^i	B_4^O
Proline	100	0	0	0	0	0
Glycine	9	30	0	1	0	60
Type II	T	B_1^i	B_2^O	B_3^i	B_4^O	
Proline	0	0	0	0	0	
Glycine	71	0	0	0	29	

C. Location of prolines in β-helix domain

β-helix rungs	72% (39/54)
Top & Bottom rungs	56% (30/54)
Middle rungs	17% (9/54)
Loop region	28% (15/54)

[a] Amino acid residue frequencies with >10% are listed.

[b] Percentage of proline and glycine residues observed in each position. * a and b were calculated based on 87 Type-I LβH and 23 Type-II LβH turn regions.

[c] Percentage calculated from 54 prolines that were found in b-helix domain of 8 type-I LβH proteins, including the incomplete rungs at top and bottom of type-I

Distribution and prevalence of the LβH fold

To date, there are only 9 LβH proteins of known structure, thus limiting the amount of available proteomic data with which to assess the prevalence and distribution of residues in this fold. A couple of human proteins have been predicted to adopt the LβH fold (Parisi et al. 2004), and the Pfam database (Finn et al. 2006) categorizes a number of putative LβH proteins, based on the ubiquitous hexapeptide motif. Nonetheless, it remains unclear how frequently the LβH fold might occur in the genomes. In order to accurately predict the prevalence of the LβH fold in the genomes, the proteomic data from the UniProt database was reorganized based on NCBI taxonomic categories (Wheeler et al. 2000) and processed to remove all redundant sequence data using a pair-wise comparison within the same specie group. A total of 144,549 redundant sequences were eliminated from 4,135,679 protein sequences using this approach.

The prediction of LβH folds was performed using amino acid sequence patterns as well as structural constraints. LβH's contain highly conserved residues that are constrained by alternating residue positions oriented toward the LβH core or the outside (see discussion above). This

repeating motif intrinsically exists in all known type-I LβH structures. The sequence pattern of the type-II LβH is less obvious and dominated by the functional TXT motif of the spruce budworm antifreeze proteins, therefore making this fold more difficult to detect. The LβH sequence profiles were constructed by a multiple sequence alignment of sequence fragments of known LβH proteins, as described above. Using these LβH profiles, the revised UniProt proteomic data was examined using the hmmsearch program (Finn et al. 2006) to identify proteins with matching sequence patterns. At the end of each search round, proteins with E value < 0.1 were selected to build a new profile for a subsequent search round. The process was repeated iteratively until no new sequences were identified. For the type-I LβH, this iterative search converged after the 5^{th} search round.

The candidate LβH proteins were subsequently examined using structural constraints based on the features of known LβH's. These criteria included: 1) the predicted LβH sequence should contain at least two predicted LβH rungs, the minimal stable unit, 2) the side chain volume of interior residues should not exceed 420 $Å^3$ and 400 $Å^3$ per rung for type-I LβH and type-II LβH respectively (maximum van der Waals side-chain volume calculated from known LβH protein and maximum core volume

estimated by CASTp was ~382 $Å^3$ and ~420 $Å^3$ for LβH-I LβH and ~360 $Å^3$ and ~ 400 $Å^3$ for LβH-II LβH respectively), 3) the location and length of the external loop should be consistent with known structures, such that no more than one external loop should occur in a single rung and the loop should range from 1 to 50 residues in length, 4) proline is only allowed at β-helix turn positions at T_1 and T_2. From the 5539 initial candidates identified by sequence pattern, 662 proteins were eliminated using these structural constraints, as shown in Figure 1.2A. The remaining 4877 predicted LβH proteins were categorized by NCBI taxonomy data as shown in Figure 1.2B, and represented as the percentage of LβH protein occurrences in the genomes. The crude estimate of LβH distribution in the genomes was determined by calculating the ratio of predicted LβH proteins to the total number of proteins in the proteomic data. These findings were consistent with the taxonomic distribution of known LβH structures in the Protein Data Bank (PDB) (Berman et al. 2000), and Pfam database records. However, estimating the average occurrence of predicted LβH proteins per species in each taxonomic category suggested that the actual taxonomic distribution of LβH might be largely equivalent across the genomes. In the case of type-II

LβH protein prediction, no novel proteins were found other than the 38 homologs of spruce budworm antifreeze proteins.

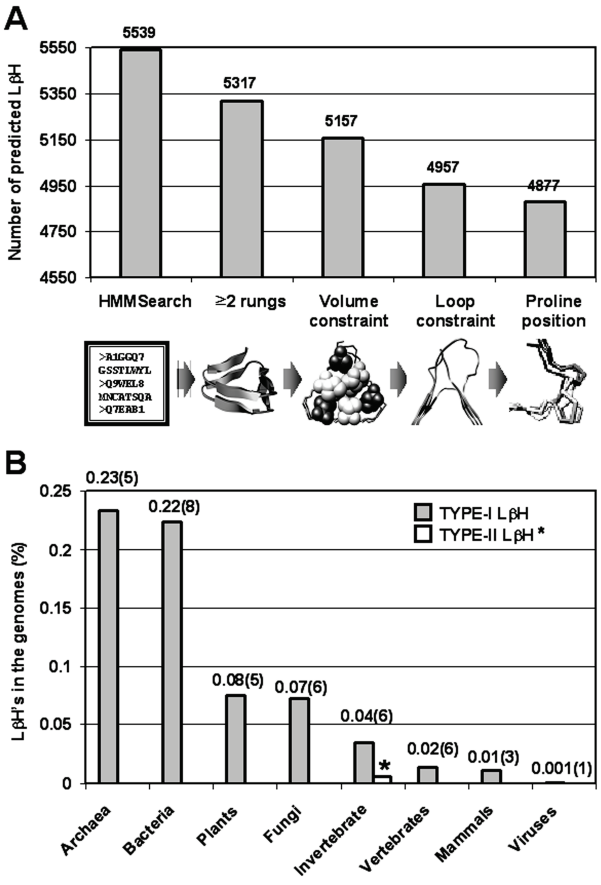


Figure 1.2. Genome-wide search for LβH domains using sequence patterns and structural constraints: (A) Histogram of the total number of predicted LβH's based on sequence profile and the use of structural criteria to filter the predictions. (B) Percentage of predicted LβH over the total number of proteins in each division. Average numbers of LβH predicted per specie in each division are shown in parentheses. * LβH Type II was predicted in 38 antifreeze proteins of the insect spruce budworm and closely related species.

LβH loop regions and domain Size

In an attempt to further characterize structural components of the LβH fold, the external loop regions and LβH domain size were examined from known and predicted LβH proteins. For the predicted LβH proteins, the loop regions were established based on sequence content and location relative to the predicted β-helix domains. The predicted loops were categorized based on length. The occurrence of loops of various lengths was calculated, and plotted along with observed data from known LβH proteins (Figure 1.3A). The observed and predicted loop distribution was comparable with respect to loop length and occurrence. The domain size of the predicted LβH proteins was computed by calculating the number of rungs that each predicted LβH protein contains, as shown in Figure 1.3B. Based on this distribution, the most abundant size of LβH domain was estimated to be four to six rungs. The analysis of the occurrence and distribution of loops and domain size, suggests that most LβH domains are composed of four to six consecutive LβH rungs, interrupted by external loops of variable length (1-40 residues).

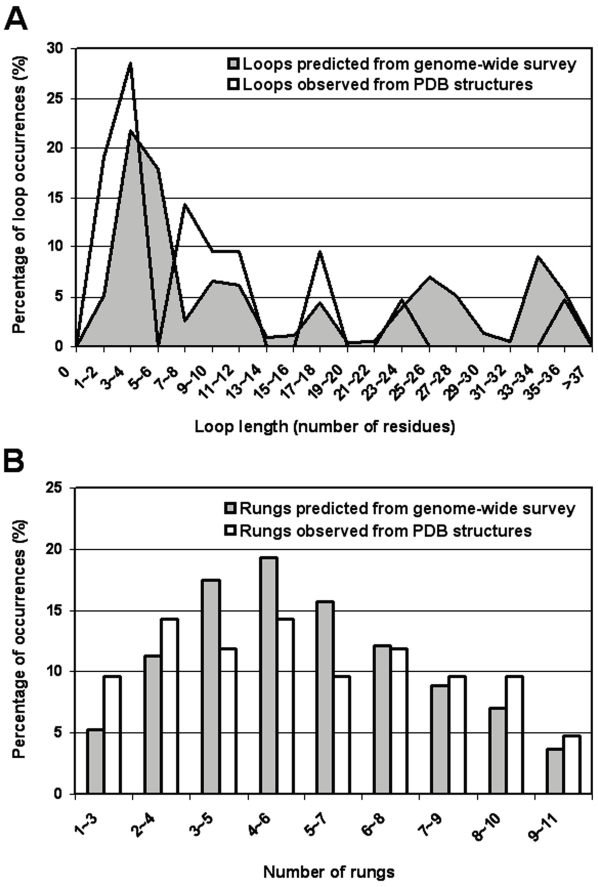


Figure 1.3. The distribution of LβH loop regions and LβH domain: (A) Occurrences of LβH external loops and their sizes. Loop sizes and occurrences predicted from the genome-wide search (un-filled line) and observed from 14 PDB structures (filled line) were overlaid. (B) The size of predicted (shaded bars) and observed (open bars) LβH domains based on the number of rungs. Both graphs (A) and (B) were plotted on the x-axis of a sliding window interval due to difficulties of defining the start and end of the LβH domain.

Discussion

Residue propensities in LβH turns: glycine and proline

We have focused on the occurrence of glycine and proline residues to account for their possible roles in the LβH structure. β-helix turn regions include a high propensity for glycine at position T_2 of type-I LβH and position T of type-II LβH. This feature can be explained by the left-handed α-helical conformation (α_L) adopted at these positions, allowing the β-strand to propagate in a new direction and hence facilitate folding of the β-helix. Proline also has a unique propensity at the β-helix turn regions in type-I LβH of being constrained to only the T_1 position. Since β-strand positions (B_1^i, B_2^o, B_3^i, and B_4^o) require backbone-backbone hydrogen bond formation with the β-strand of the following rungs, proline would undermine the structural stability of the fold. The extremely low occurrence of proline and glycine at other positions suggests that these substitutions are structurally prohibited. These unique residue propensities of LβH turns may be used to identify or distinguish the rung structures of type-I LβH and type-II LβH.

In addition to the positional propensity of proline in the LβH rung structure, we also observed a high occurrence of proline at either the top and

bottom regions of β-helix domain. This feature may explain a possible role for prolines in folding of the LβH. The introduction of prolines at the edge of the β-helix domain may decrease conformational entropy, thus facilitating folding of the LβH domain or serve to terminate or cap the repetitive β-helical structure. In the case of type-II LβH, disulfide bridges formed by cysteines located inside of the β-helix were observed on the top and bottom regions of LβH domain. Their role in the stability and folding of LβH may be similar to that of prolines in type-I LβH's. It can be speculated that the folding process of a LβH may not be thermodynamically favorable without these structural features to restrict conformational flexibility.

Genomic prevalence of the LβH fold

To date, there are no known human or mammalian proteins that have been documented to incorporate a LβH fold. The largest proportion of predicted type-I LβH proteins are of bacterial or microbial origins. It is an intriguing question as to whether LβH-like protein folds exist in other genomes given that LβH-like structures have been proposed as models of misfolded human proteins associated with disease. Initial speculation was

that since all known type-I LβH's were microbial in origins, and the misfolded form of disease related proteins is toxic to mammalian cellular environments, the LβH fold may only exist in a narrow-range of species that can tolerate its unique structural feature. Our genome-wide survey of LβH suggests that the LβH fold probably exists evenly across the genomes, rather than occurring in a narrow range of species or taxonomic categories. Our analysis predicted the human proteins; dynactins p25 and p27 subunits, eukaryotic translation initiation factor 2B epsilon and gamma subunits, and GDP-mannose pyrophosphorylase A and B as containing a LβH domain. These results are consistent with other predictions of human LβH proteins (Parisi et al. 2004). However, to date, these structural predictions have not been confirmed experimentally. The notion that LβH proteins exist in the human proteome may support the idea that the LβH fold could serve as a building block for the fibrilization of misfolded proteins associated with human disease (e.g. prion and Huntington's disease). However, experimental evidence supporting the fibrilization of the β-helix fold is limited (Schuler et al. 1999). If it is assumed that the fibrilization of LβH domains is possible, then native soluble LβH proteins may avoid this outcome by the incorporation of structural elements to restrict self-assembly

or their ability to self-assemble is an essential part of their function. For example, all known type-I LβH proteins contain α-helical domains at either the N- or C-terminal of the LβH domains. If self-assembly of the LβH fold were possible via exposed β-faces, these α-helical domains would cap propagation and hence inhibit fibrilization. Further experimental studies are required in order to elucidate the self-assembly of the LβH fold and define a possible role in human misfolding diseases and nanotechnology.

Materials and Methods

Construction of LβH position-specific sequence profile

Eight of 14 known type-I LβH structures [PDB ID codes: 1LXA, 1TDT, 1KRR, 1OCX, 1XAT, 1HV9, 1QRE, 1SSQ] and one type-II LβH structure (1M8N) along with sequences from 12 homolog proteins were used to construct LβH sequence alignments and build a position-specific sequence logo. The three-dimensional structures of the LβH domains were superimposed using the InsightII software (Accelrys San Diego, CA), and the sequences that comprise only a complete LβH rung were extracted to build a multiple alignment of the LβH sequences. The rungs that contained

external loop regions and rungs that are located at the top and bottom of the LβH domain were excluded since they contained abnormal sequence composition. For amino acid propensities of LβH turn regions, the sequences of type-II LβH turns observed in a subset of type-I LβH proteins were extracted and included in the sequence alignment of type-II LβH turns. The positions of the LβH were defined using the following nomenclature: T_1 (1^{st} turn residue), T_2 (2^{nd} turn residue), B_1^i (1^{st} β-strand residue, facing inside), B_2^o (2^{nd} β-strand residue, facing outside), B_3^i (3^{rd} β-strand residue, facing inside), and B_4^o (4^{th} β-strand residue, facing outside).

Proteomic database and data preparation

The complete non-redundant proteomic sequences for all of the organisms examined in this study were obtained from the Universal Protein Knowledgebase (UniProt) (Consortium 2007) consortium database, and the corresponding taxonomic data were obtained from the National Center for Biotechnology Information (NCBI) (Wheeler et al. 2000). Although, the UniProt database is considered as a non-redundant database, it still contains redundancies due to sequences of subspecies or strains. In order to reduce

the redundancies caused by subspecies, all the sequences were reorganized based on taxonomic categories, and all the subspecie sequences within the same species were grouped together. Using the FASTA program (Pearson 1994a; b), the identical (>98% sequence identities) sequences within the same taxonomic group were removed.

Genome-wide search for LβH folds

The HMMER software package (Eddy 1998) was initially used to predict LβH folds across the genomes. From the multiple sequence alignments of the LβH domain, hidden Markov profiles were created using the hmmbuild program available as part of the HMMER package. The hmmsearch of HMMER was performed iteratively against the prepared data set (above) with E value < 0.1. The predicted domain sequences of each LβH candidate protein were grouped together and aligned against the LβH position-specific sequence profile to identify LβH features that included; (1) the number of rungs, (2) residues that face toward the inner core of the LβH, (3) the external loop regions, (4) residues at the β-helix turn regions. Each candidate protein was scored and filtered based on the number of rungs that it contains, inner core volume estimated from the interior residue side chain

van der Waals volume, length and the number of external loops, and the occurrence of prolines at β-helical turns. The volume of interior residues was calculated as the sum of side-chain volumes of the interior residues for each complete rung.

References

Accelrys, I. San Diego, CA.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Brown, K., Pompeo, F., Dixon, S., Mengin-Lecreulx, D., Cambillau, C., and Bourne, Y. 1999. Crystal structure of the bifunctional N-acetylglucosamine 1-phosphate uridyltransferase from Escherichia coli: a paradigm for the related pyrophosphorylase superfamily. *Embo J* **18:** 4096-4107.

Consortium, T.U. 2007. The Universal Protein Resource (UniProt). *Nucleic Acids Res* **35:** D193-197.

Eddy, S.R. 1998. Profile hidden Markov models. *Bioinformatics* **14:** 755-763.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., et al. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res* **34:** D247-251.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. 2004. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101:** 8342-8347.

Graether, S.P., Kuiper, M.J., Gagne, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. 2000. Beta-helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* **406:** 325-328.

Haspel, N., Zanuy, D., Aleman, C., Wolfson, H., and Nussinov, R. 2006. De novo tubular nanostructure design based on self-assembly of beta-helical protein motifs. *Structure* **14:** 1137-1148.

Haspel, N., Zanuy, D., Zheng, J., Aleman, C., Wolfson, H., and Nussinov, R. 2007. Changing the charge distribution of beta-helical-based nanostructures can provide the conditions for charge transfer. *Biophys J* **93:** 245-253.

Iengar, P., Joshi, N.V., and Balaram, P. 2006. Conformational and sequence signatures in beta helix proteins. *Structure* **14:** 529-542.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

Kajava, A.V., and Steven, A.C. 2006. Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv Protein Chem* **73:** 55-96.

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Leinala, E.K., Davies, P.L., Doucet, D., Tyshenko, M.G., Walker, V.K., and Jia, Z. 2002a. A beta-helical antifreeze protein isoform with increased activity. Structural and functional insights. *J Biol Chem* **277:** 33349-33352.

Leinala, E.K., Davies, P.L., and Jia, Z. 2002b. Crystal structure of beta-helical antifreeze protein points to a general ice binding model. *Structure* **10:** 619-627.

Merlino, A., Esposito, L., and Vitagliano, L. 2006. Polyglutamine repeats and beta-helix structure: molecular dynamics study. *Proteins* **63:** 918-927.

Olsen, L.R., and Roderick, S.L. 2001. Structure of the Escherichia coli GlmU pyrophosphorylase and acetyltransferase active sites. *Biochemistry* **40:** 1913-1921.

Parisi, G., and Echave, J. 2001. Structural constraints and emergence of sequence patterns in protein evolution. *Mol Biol Evol* **18:** 750-756.

Parisi, G., and Echave, J. 2004. The structurally constrained protein evolution model accounts for sequence patterns of the LbetaH superfamily. *BMC Evol Biol* **4:** 41.

Parisi, G., Fornasari, M.S., and Echave, J. 2004. Dynactins p25 and p27 are predicted to adopt the LbetaH fold. *FEBS Lett* **562:** 1-4.

Pearson, W.R. 1994a. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* **24:** 307-331.

Pearson, W.R. 1994b. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* **25:** 365-389.

Raetz, C.R., and Roderick, S.L. 1995. A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science* **270:** 997-1000.

Schuler, B., Rachel, R., and Seckler, R. 1999. Formation of fibrous aggregates from a non-native intermediate: the isolated P22 tailspike beta-helix domain. *J Biol Chem* **274:** 18589-18596.

Simkovsky, R., and King, J. 2006. An elongated spine of buried core residues necessary for in vivo folding of the parallel beta-helix of P22 tailspike adhesin. *Proc Natl Acad Sci U S A* **103:** 3575-3580.

Stork, M., Giese, A., Kretzschmar, H.A., and Tavan, P. 2005. Molecular dynamics simulations indicate a possible role of parallel beta-helices in seeded aggregation of poly-Gln. *Biophys J* **88:** 2442-2451.

Sulzenbacher, G., Gal, L., Peneff, C., Fassy, F., and Bourne, Y. 2001. Crystal structure of Streptococcus pneumoniae N-acetylglucosamine-1-phosphate uridyltransferase bound to acetyl-coenzyme A reveals a novel active site architecture. *J Biol Chem* **276:** 11844-11851.

Wheeler, D.L., Chappey, C., Lash, A.E., Leipe, D.D., Madden, T.L., Schuler, G.D., Tatusova, T.A., and Rapp, B.A. 2000. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **28:** 10-14.

Yang, S., Levine, H., Onuchic, J.N., and Cox, D.L. 2005. Structure of infectious prions: stabilization by domain swapping. *FASEB J* **19:** 1778-1782.

Zheng, J., Zanuy, D., Haspel, N., Tsai, C.J., Aleman, C., and Nussinov, R. 2007. Nanostructure design using protein building blocks enhanced by conformationally constrained synthetic residues. *Biochemistry* **46:** 1205-1218.

Chapter 2

Stability analysis of the left-handed β-helical fold using molecular dynamics simulations

Structural stability of the LβH fold

While MD simulations are unable to calculate the absolute stability of a folded protein structure, they can be useful in studying the relative stability of related structures. We performed 10 ns MD simulations at 300 K on simple type-I LβH models, ranging from a 1-rung model (18 residues) to a 5-rung model (90 residues) (see Figure 2.1A) in explicit solvent (for details, see Materials and Methods). An analysis of positional root-mean-squared deviations (RMSD) relative to the starting models indicated that all the model systems had reached equilibrium after approximately 2 ns (Figure 2.1B). The average RMSD relative to the starting models were plotted against the number of rungs (Figure 2.1C). The relative stability of the LβH structures estimated by RMSD calculation showed a large stability difference between the 1-rung and 2-rungs models. However, the subsequent addition of rungs to the 2-rung model to generate the 3-, 4-, and 5-rung models did not show significant additional stability contribution in the system. Table 2 summarizes the average RMSD calculated for C_α and all atoms, showing that the C_α RMSD measure was the sufficient measure for the structural deviation of LβH MD simulations. Secondary structure content was determined by the DSSP algorithm (Kabsch and Sander 1983).

The secondary structure content of a 1-rung LβH model after 10 ns showed an increase in random coil and a decrease in β-sheet content, providing an estimate of how structural elements changed over time. However, the DSSP algorithm was not sufficiently accurate to determine the secondary structure changes to the 3-, 4-, 5-rung LβH models. According to the DSSP algorithm, these models gained β-sheet content with a corresponding loss of turn content, while the actual structures were shown to be relatively well maintained by visual inspection and Ramachandran plot analysis (data not shown). Changes to the interstrand backbone-backbone hydrogen bonding network between initial and final structures provided another measure of structural deviation after the 10 ns MD simulations. While the 2-, 3-, 4-, and 5-rung LβH models maintained their hydrogen bonding network within the standard deviation, the 1-rung LβH model gained hydrogen bonds during the MD simulation as result of collapse of LβH core to form a two-stranded antiparallel β-sheet (Figure 2.1A).

We also sought to understand any positional contribution to the stability of the LβH fold by starting from a 1-rung LβH model (18 residues) and "growing" the LβH fold to a 2-rung LβH model (38 residues) in two residue increments. The relative stability of individual models was

measured by RMSD after a 10 ns MD simulation at 300 K. The starting 18-residue model (1-rung LβH model) contained a single hydrogen bond at the β-helix turn region between two adjacent β-strands. The incremental addition of two residues to yield the 20-residue model formed the first hydrogen bond between backbones of parallel β-strands. As expected, there was a notable stability gain with the creation of a complete rung. The subsequent addition of residues showed a gradual increase in stability as the number of hydrogen bonds between rungs increased (Figure 2.2).

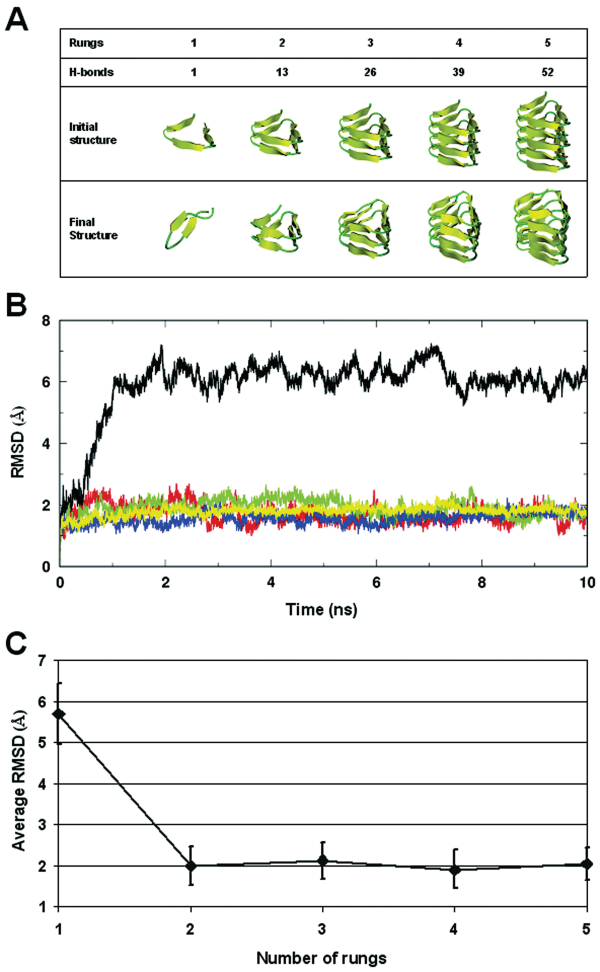


Figure 2.1. MD simulation of LβH models with increasing numbers of rungs: (A) Initial structures and final structures after a 10 ns MD simulation (B) Backbone RMSD of LβH models relative to their initial structures as a function of simulation time at 300 K. 1, 2, 3, 4 and 5-rung LβH's are shown in black, red, green, blue and yellow, respectively. (C) Average backbone RMSD of the last 2 ns interval. Error bars are calculated from five independent simulations.

Table 2. Structural statistics from the MD simulation of LβH

A. Average RMSD with respect to the starting LβH structures (Å)

	Backbone (C_α)	Backbone (all atoms)
1-rung LβH	5.87 (0.31)	6.78 (0.58)
2-rung LβH	1.90 (0.30)	2.73 (0.30)
3-rung LβH	2.06 (0.18)	2.85 (0.27)
4-rung LβH	1.76 (0.17)	2.47 (0.21)
5-rung LβH	1.95 (0.16)	2.67 (0.13)

B. Secondary structure elements content (%)

	β-sheet	Turn / Bend	Coil
LβH-WT	33	33	33
1-rung LβH	15 (16)	36 (8)	49 (14)
2-rung LβH	59 (8)	21 (7)	21 (5)
3-rung LβH	55 (5)	23 (4)	20 (3)
4-rung LβH	65 (11)	18 (6)	17 (6)
5-rung LβH	81 (11)	8 (5)	11 (7)

C. Interstrand backbone-backbone hydrogen bonds

	1 rung	2 rung	3 rung	4 rung	5 rung
LβH-WT	1	13	26	39	52
LβH-MD	4 (1)	13 (1)	23 (4)	36 (3)	49 (2)

The analysis was performed on results of five independent trials MD. Average values are reported with standard deviations in parentheses.

[a] The backbone RMSD values were calculated with respect to the initial structures, averaged over the 8-10 ns interval of the MD trajectories.

[b] The secondary structure content was calculated using the DSSP algorithm.

[c] Backbone-backbone hydrogen bonds between strands were calculated using InsightII software.

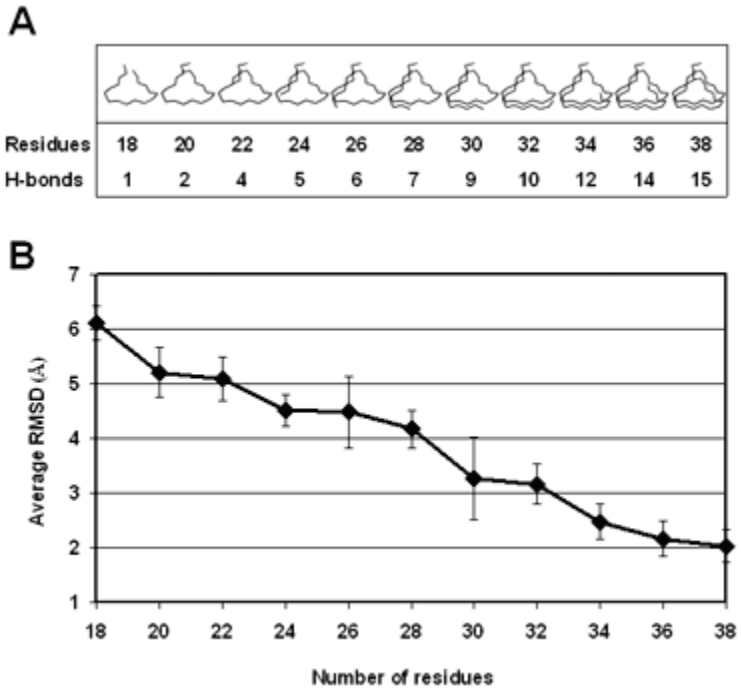


Figure 2.2. MD simulation of elongating LβH models. (A) Structures of LβH models with an increasing number of residues and corresponding interstrand backbone-backbone hydrogen bonds. (B) Average backbone RMSD from 8 to 10 ns MD simulation time interval. Error bars are calculated from five independent simulations at 300 K.

The relative stability contribution of the LβH hydrophobic core was also explored using MD simulations. In order to conduct this calculation, interior residues (at B_3^i positions) of type-I LβH (LβH-WT) were substituted with glycine, alanine, or valine in order to build LβH-GLY, LβH-ALA, and LβH-VAL models, respectively (Figure 2.3A). The cavity volumes of LβH-GLY, LβH-ALA, LβH-VAL and LβH-WT were estimated by CASTp(Binkowski et al. 2003). Conversely, the side chain volume of the interior residues for each model was calculated as the sum of the Van der Waals volume of individual amino acid side chains, minus the volume of glycine (Creighton 1993). MD simulations of these models were performed at 300 K for 10 ns, and average RMSD were calculated based on the starting models (Figure 2.3B). LβH-VAL showed a significant decrease in RMSD values, indicative of a comparable stability relative to LβH-WT. The relatively high RMSD value of LβH-ALA indicated that the defective packing of the hydrophobic core significantly decreased the stability of the LβH structure. As shown in the final structure of LβH-ALA (Figure 2.3A), the architecture of the β-helix fold transformed into a two strand per rung structure resulting from the collapse of the β-helix core. However, the destabilizing effect of the hydrophobic core could be overcome by

elongation of β-helix rungs such that the stability of a 5-rung LβH-ALA model was comparable to that of a 5-rung LβH-WT structure (Figure 2.3B). A 5-rung model of LβH-GLY showed relatively high RMSDs, indicating that glycine substitutions significantly decreased the structural stability of the LβH due to the hydrophobic core defect and excessive backbone flexibility. An examination of root-mean-square fluctuation (RMSF) of C□ before and after MD simulations indicated significant movement at core residues of the LβH-GLY model (Figure 2.3D).

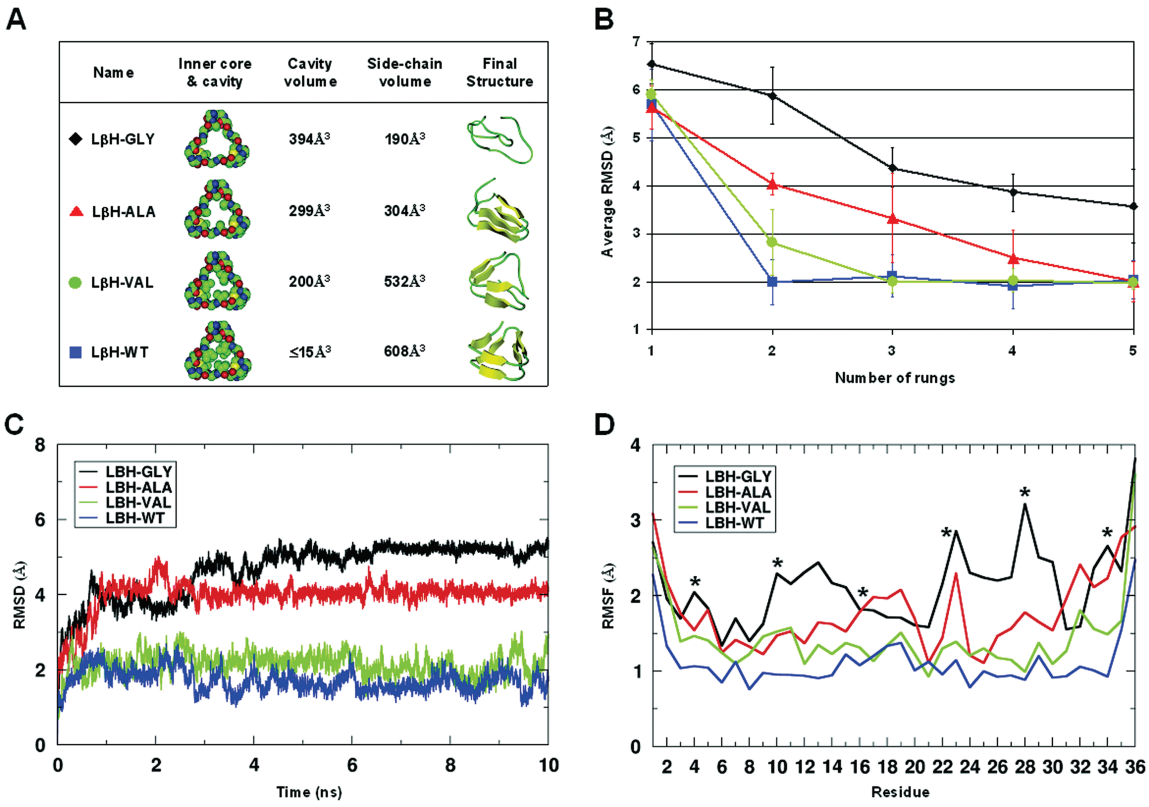


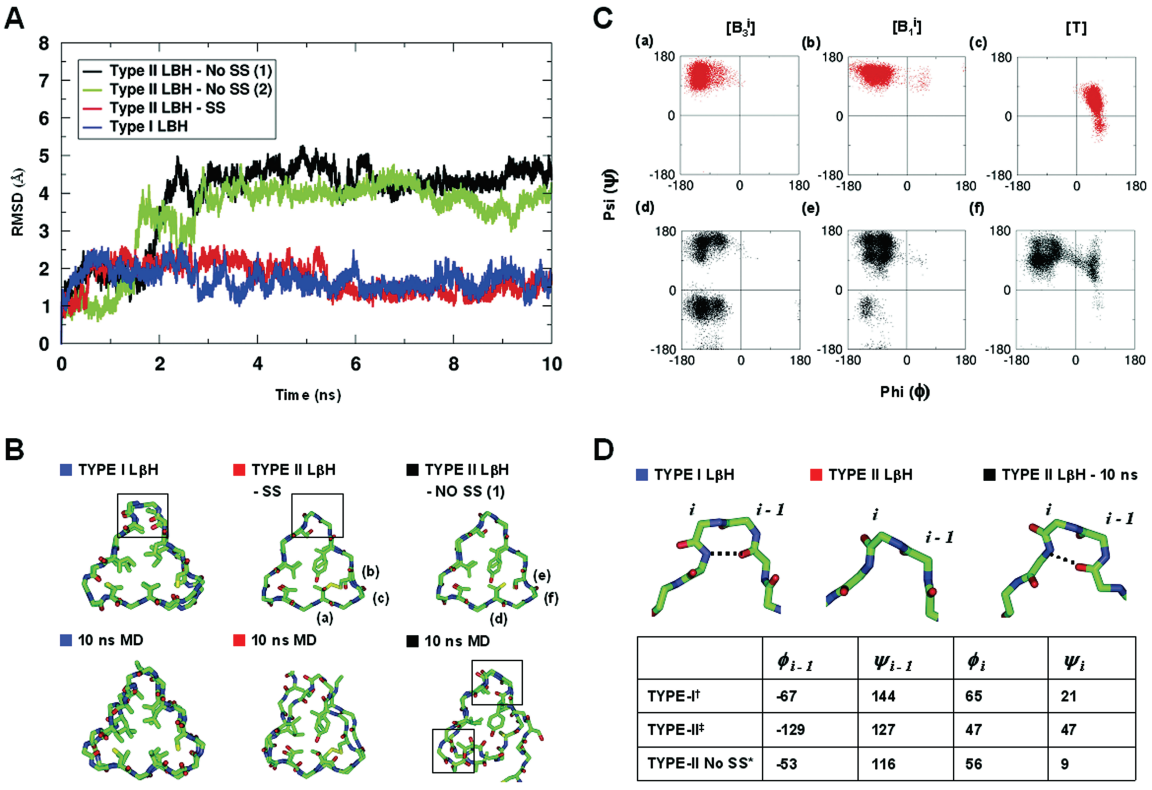
Figure 2.3. MD simulations of LβH models with varying cavity sizes: (A) LβH-GLY, LβH-ALA, LβH-VAL models with residues at B_3^i positions substituted with glycine, alanine and valine, respectively. LβH-WT is the wild type LβH model based on 1LXA (residues 120-155). For 2-rung LβH models, the cavity volumes of the interiors were calculated by CASTp[36] and the side-chain volumes were calculated as the sum of side-chain volumes of the interior residues. Side-chain volumes were defined as the Van der Waals volume of individual amino acids minus the volume of glycine[37]. Final

structures of LβH models after 10 ns MD simulation at 300 K are shown. (B) Average backbone RMSD of the last 2 ns interval from five independent trials of MD simulation for 10 ns at 300 K are plotted for each model. (C) Backbone RMSD and (D) Residue-based Cα RMSF of 2-rung LβH models for LβH-GLY, LβH-ALA, LβH-VAL and LβH-WT are shown in black, red, green and blue, respectively. * indicates the interior residues at position B_3^i.

Stability of type-I LβH vs. type-II LβH

The relative stability of type-I LβH and type-II LβH was examined by MD simulations using 2-rung models. Unlike type-I LβH, known type-II LβH proteins contain disulfide bonds between cysteines at positions B_3^i and B_1^i of successive rungs. The hydrophobic core of type-II LβH is less regularly packed with aromatic residues and disulphides inside the β-helix than that of type-I LβH. The stability of a 2-rung model of a type-II LβH with a B_3^i-B_1^i disulfide bridge was compared with a 2-rung model of type-I LβH using MD simulation (Figure 2.4). The disulfide bridged type-II LβH model showed comparable stability to the type-I LβH model in 10 ns MD simulation. In order to understand the stability contribution made by the disulfide bridge to the type-II LβH, two rungs of type-II LβH that did not contain disulfide bonds were compared with a 2-rung type-I LβH model and a 2-rung disulfide bridged type-II LβH model using MD simulation. Figure 2.4A shows the RMSD analysis of the three model systems. The Cα atoms of the type-I LβH rungs and disulfide bridged type-II LβH rungs after 10 ns MD simulation remained within 2 Å of the starting structure, while the Cα atoms of the type-II LβH rungs with no disulfide bridges deviated from the

initial position by more than 4 Å and lost β-helical architecture. Figure 2.4B & C show the difference in movements of backbone and backbone torsion angles for residues located at the corner involved with the disulfide bridge. Large movements of the backbone Cα's were observed in the type-II LβH lacking the disulphide bridge. In addition, an examination of the 10 ns structure of the type-II LβH lacking the disulfide bridge showed that some of type-II LβH turns transformed into a type-I-LβH-like turn, with hydrogen bonds observed between the amide hydrogen of the B_1^i residue and backbone carbonyl oxygen of a residue located two residues upstream from the B_1^i residue (Figure 2.4D). Our findings suggest that; 1) the disulfide bonding seen in type-II LβH is a major contributor to the stability of this fold, 2) in the absence of a disulfide bridge, the type-II LβH is significantly less stable than the type-I LβH. These findings provide a reasonable explanation for the prevalence of type-I LβH over type-II LβH proteins.



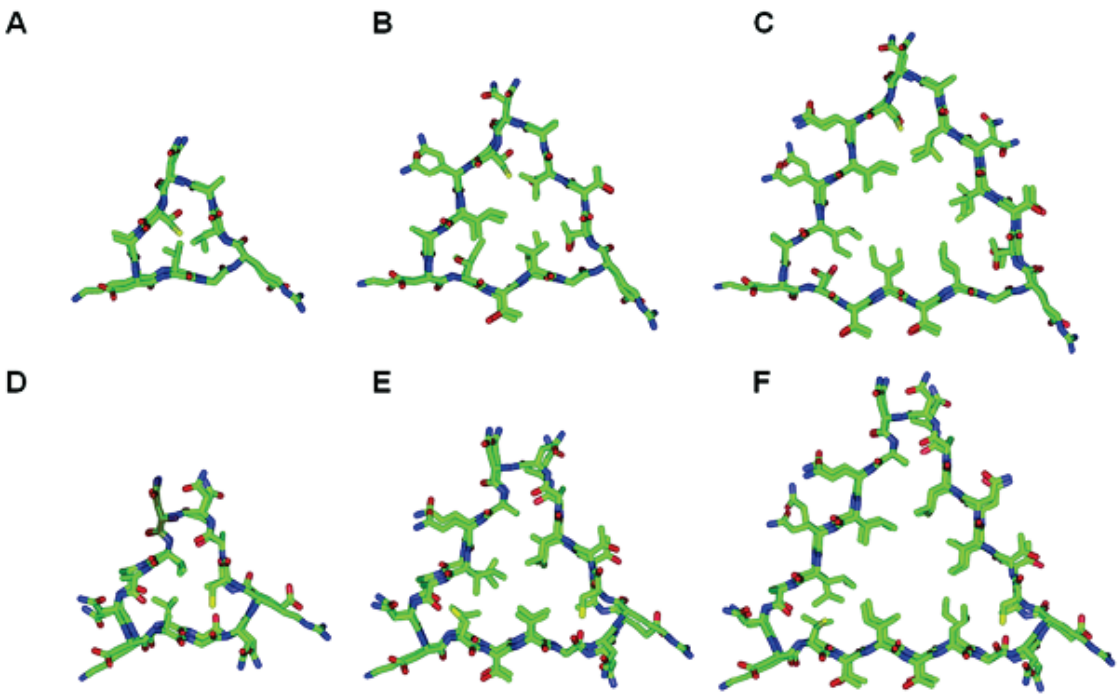
	ϕ_{i-1}	ψ_{i-1}	ϕ_i	ψ_i
TYPE-I†	-67	144	65	21
TYPE-II‡	-129	127	47	47
TYPE-II No SS*	-53	116	56	9

Figure 2.4. Stability analysis of type-I LβH vs. type-II LβH: (A) 10 ns MD simulations of 2-rung models of type-I LβH and type-II LβH with and without the disulfide bridge at 300 K. Two types of type-II LβH without the disulfide bridge were examined. Type-II LβH – No SS (1) model used two rungs of 1M8N lacking the disulfide bridge. Type-II LβH – No SS (2) model used the same template as the Type-II LβH – SS model, but the disulfide bridge was removed by substituting a cysteine residue with a serine residue at position B_1^i. (B) Initial and final structure of type-I LβH, type-II

LβH with and without the disulfide bridge. (C) Ramachandran plot analysis showing the distribution of the main chain dihedral angles of three residue positions: residues at position B_3^i (a & d) and B_1^i (b & e) and residue at position T (c & f). Type-II LβH with and without the disulfide bridge was shown in red and black respectively. (D) Turn regions of Type-I LβH, type-II LβH, and type-II LβH without the disulfide bridge after 10 ns MD simulation. † the values were taken from[2]. ‡ the mean values of dihedral angles were calculated from 15 type-II LβH turn regions of 1M8N. * the mean values were calculated from the final structures of MD from five independent trials.

Structural variations of the LβH fold

Although there are variations of the LβH in nature such as the triple-stranded β-helix and left-handed β-roll as well as the right-handed β-helices, no other types of LβH with type-I or –II architecture have been documented beyond the 9 LβH proteins studied here. In order to assess the possibility of a LβH fold with different calibers, we built and examined simple models: three residues per β-strand (9 residues per rung), four residues per β-strand (12 residues per rung), seven residues per β-strand (21 residues per rung) and eight residues per β-strand LβH (24 residues per rung) based on the type-I and type-II LβH templates (Figure 2.5). A structural examination of these LβH models compared to type-I LβH (18 residues per rung) and type-II LβH (15 residues per rung) showed that while the number of interstrand backbone-backbone hydrogen bonds per residue remained constant as the LβH size increased, the cavity volume and conformational constraint of LβH core would make the 18 residue per rung and 15 residues per rung fold (i.e type-I and type-II LβH) the most sensible fit to the LβH architecture.



G

2-rung Model	Tripeptide LβH (hypothetical model)	Pentapeptide LβH (Type-II LβH: 1M8N)	Heptapeptide LβH (Type-II LβH: 1M8N)
Number of residue / rung	9	15	21
Translation	4.77 Å / rung (0.53 Å / residue)	4.79 Å / rung (0.32 Å / residue)	4.93 Å / rung (0.23 Å / residue)
T^{aL} residues / rung	3 (3/9, 33%)	3 (3/15, 20%)	3 (3/21, 14%)
β-strand residues / rung	6 (6/9, 67%)	12 (12/15, 80%)	18 (18/21, 86%)
Number of H-bonds *(Backbone-backbone interstrand)*	6 (0.33 H-bonds / residue)	12 (0.40 H-bonds / residue)	18 (0.43 H-bonds / residue)
Cavity volume *(glycine as interior residues)*	70 $Å^3$	377 $Å^3$	892 $Å^3$

2-rung Model	Tetrapeptide LβH (hypothetical model)	Hexapeptide LβH (Type-I LβH: 1LXA)	Octapeptide LβH (hypothetical model)
Number of residue / rung	12	18	24
Translation	4.81 Å / rung (0.40 Å / residue)	4.87 Å / rung (0.27 Å / residue)	4.88 Å / rung (0.20 Å / residue)
T^{aL} residues / rung	3 (3/12, 25%)	3 (3/18, 17%)	3 (3/24, 12.5%)
β-strand residues / rung	9 (9/12, 75%)	15 (15/18, 83%)	21 (21/24, 87.5%)
Number of H-bonds *(Backbone-backbone interstrand)*	9 (0.38 H-bonds / residue)	15 (0.42 H-bonds / residue)	21 (0.44 H-bonds / residue)
Cavity volume *(glycine as interior residues)*	184 $Å^3$	448 $Å^3$	949 $Å^3$

Figure 2.5. Structural variations of the LβH fold. In order to assess the possibility of a LβH fold with different calibers, we built and examined simple models: (A) three residues per β-strand model (9 residues per rung, tripeptide LβH), (B) type-II LβH, five residues per β-strand model (15 residues per rung, pentapeptide LβH), (C) seven residues per β-strand model (21 residues per rung, heptapeptide LβH), (D) four residues per β-strand model (12 residues per rung, tetrapeptide LβH), (E) type-I LβH, six residues per β-strand model (18 residues per rung, hexapeptide LβH), (F) eight residues per β-strand model (24 residues per rung, octapeptide LβH). Models were built using type-I LβH (1LXA) and type-II LβH (1M8N) as templates. (G) The number of interstand backbone-backbone hydrogen bonds per residue was calculated using the InsightII software[21]. The cavity volumes of the LβH core with glycine as interior residues were calculated by CASTp[34]. A structural examination of these LβH models compared to type-I LβH (18 residues per rung) and type-II LβH (15 residues per rung) showed that while the number of interstrand backbone-backbone hydrogen bonds per residue remained constant as the LβH size increased, the cavity volume and conformational constraint of LβH core would make the 18 residue per rung

and 15 residues per rung fold (i.e type-I LβH and type-II LβH) the most sensible fit to the LβH architecture.

Discussion

Structural stability of the LβH

Key features of the LβH fold that have received interest in modeling studies of amyloid fibril and nanotubular structures are the inherent compactness of the fold and the availability of accessible β-faces to initiate and sustain self-assembly. Previous modeling studies have suggested that the 2-rung structure of the LβH (~36 residues) was the minimal unit required to create a stable nuclei for seeded fibrilization of β-helices (Govaerts et al. 2004; Langedijk et al. 2006), and it has also been used as a building block in the design of the nanotubular fibers(Haspel et al. 2006; Haspel et al. 2007; Zheng et al. 2007).

The structural stability of the LβH and the role of hydrogen bonds and the hydrophobic core have been previously reviewed (Jenkins and Pickersgill 2001; Iengar et al. 2006). The current report systematically examined changes in stability as a result of elongation of the β-helix rungs and defects to the hydrophobic core. We have shown that the 2-rung LβH is likely to be the minimal stable unit. An essential determinant of global stability was shown to be the hydrophobic core, such that incorrect packing,

or packing defects rendered the 2-rung LβH unstable. Based on the relative stability comparison between two-rung models of type-I LβH and type–II LβH, it is conceivable that type-I LβH is a more sensible fit as a building block of self-assembling structures.

Conclusion

Recent modeling studies have suggested LβH-like folds may be possible structural solutions to the misfolded isoforms of proteins involved in neurodegenerative diseases. However, it is not clear whether disease-related proteins such as prion proteins or polyglutamine rich sequences can adopt LβH-like folds in a relatively dehydrated state since they do not include the distinctive patterns that are found in known LβH proteins. Since the LβH hydrophobic core is critical to the structural stability of the fold, the inclusion of alanine and glycine rich sequences that are a feature of some amyloidogenic proteins does not seem compatible with the LβH architecture. While some theoretical studies proposed the possibility of hydrophilic or charged residue inclusions in the LβH core such as models of polyglutamines, further studies that incorporate the impact of the dehydrated state will be necessary to probe the structural compatibility of LβH with those residues that are unprecedented in documented structures.

Materials and Methods

Model construction of simple LβH

The simple LβH models were built on the scaffold of UDP *N*-acetylglucosamine acyltransferase (1LXA) and spruce budworm antifreeze protein (1M8N). The amino acids and coordinates from residues 120-135, 120-155 of 1LXA (type-I LβH) and 35-64, 45-74 of 1M8N (type-II LβH) were used as templates to build simple models of LβH with different numbers of rungs and residues using the InsightII software(Accelrys San Diego, CA). While interior residues were kept as the original template, the outer residues were replaced with the commonly occurring residues in native structures in order to build a consistent set of models. Models of LβH-GLY, LβH-ALA, and LβH-VAL (see Figure 2.3A) were built with residues at all B_3^i position of type-I LβH model substituted with glycine, alanine and valine, respectively. The side chain positions of the LβH models were subsequently optimized by using SCWRL 3.1 (Canutescu et al. 2003). The models were optimized by energy minimization using the GROMACS 3.1.3 package (Van Der Spoel et al. 2005). The final models of simple LβH's contained only complete rungs of β-helix with no abnormal turns or external loops.

MD simulation of LβH models

All simulations were performed with the GROMACS software package (Van Der Spoel et al. 2005), using the GROMMOS 43a3 force field(Daura et al. 1998). Simple LβH models were solvated individually in cubic boxes filled with water molecules (Eisenberg and McLachlan 1986). A Single Point Charge (SPC) water model was used for the solvent molecules in the simulation (Berendsen et al. 1981). Sodium ions and chlorine ions were used to electro-neutralize the system. Solutes, solvent and counterions were coupled independently to reference temperature baths at 300 K (Berendsen et al. 1984), and the pressure was maintained by coupling the system weakly to an external pressure bath at one atmosphere (Langedijk et al. 2006). Bond lengths were constrained by the LINCS procedure (Hess et al. 1997) and non-bonded interactions were evaluated using twin-range cut-off of 0.8 and 1.4 nm for Lennard-Jones and Coulomb potentials. The long-range electrostatic interactions beyond the cut-off were treated with the generalized reaction field model, using a dielectric constant of 54 (Langedijk et al. 2006). The integration time step was set to 0.002 ps and the trajectory coordinates and energies were stored at 0.5 ps intervals.

The analysis was performed using the built-in programs of GROMACS software package.

References

Accelrys, I. San Diego, CA.

Berendsen, H.J., Postma, J.P., van Gunsteren, W.F., and Hermans, J. 1981. *Interaction models for water in relation to protein hydration*. Reidel Publishing Company, Dordrecht, pp. 331-342.

Berendsen, H.J., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81:** 3684-3690.

Binkowski, T.A., Naghibzadeh, S., and Liang, J. 2003. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res* **31:** 3352-3355.

Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L., Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12:** 2001-2014.

Creighton, T.E. 1993. *Proteins: Structures and Molecular Properties*, Second ed. Freeman, New York.

Daura, X., Mark, A.E., and van Gunsteren, W.F. 1998. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **19:** 535-547.

Eisenberg, D., and McLachlan, A.D. 1986. Solvation energy in protein folding and binding. *Nature* **319:** 199-203.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. 2004. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101:** 8342-8347.

Haspel, N., Zanuy, D., Aleman, C., Wolfson, H., and Nussinov, R. 2006. De novo tubular nanostructure design based on self-assembly of beta-helical protein motifs. *Structure* **14:** 1137-1148.

Haspel, N., Zanuy, D., Zheng, J., Aleman, C., Wolfson, H., and Nussinov, R. 2007. Changing the charge distribution of beta-helical-based nanostructures can provide the conditions for charge transfer. *Biophys J* **93:** 245-253.

Hess, B., Bekker, H., Hermans, J., Berendsen, H.J., and Fraaije, J.G.E.M. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18:** 1463-1472.

Iengar, P., Joshi, N.V., and Balaram, P. 2006. Conformational and sequence signatures in beta helix proteins. *Structure* **14:** 529-542.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

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

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. 2005. GROMACS: fast, flexible, and free. *J Comput Chem* **26:** 1701-1718.

Zheng, J., Zanuy, D., Haspel, N., Tsai, C.J., Aleman, C., and Nussinov, R. 2007. Nanostructure design using protein building blocks enhanced by conformationally constrained synthetic residues. *Biochemistry* **46:** 1205-1218.

Chapter 3

In-vivo folding study of the left-handed β-helical fold

Introduction

The parallel β-helix is a repetitive fold where the repeating unit is a β-helical coil formed by segments of β-strands (Jenkins and Pickersgill 2001; Iengar et al. 2006; Choi et al. 2008). Each rung of the canonical β-helix consists of two to three β-strands interrupted by turn or loop regions (Iengar et al. 2006; Kajava and Steven 2006; Simkovsky and King 2006). The β-helical rungs are aligned to form a cross-β structure such that elongated β-strands connected by hydrogen bonds lie parallel to the helical axis (Govaerts et al. 2004; Simkovsky and King 2006; Choi et al. 2008). Structural repetition of coils creates a cylindrical hydrophobic core. The hydrophobic core of the β-helical proteins is characterized by buried stacks of similar side chains (Jenkins and Pickersgill 2001; Simkovsky and King 2006; Choi et al. 2008). While right-handed β-helices (RβH) are generally characterized by β-strands connected by variable length turns and loops, LβH is more rigid and repetitive than the right-handed β-helix variant (Iengar et al. 2006).

It has been proposed that misfolded proteins associated with neurodegenerative diseases such as prion disease may adopt the LβH fold

(Govaerts et al. 2004; Yang et al. 2005; Langedijk et al. 2006). Recent solid-state nuclear magnetic resonance studies of the fungal HET-s prion protein showed that the misfolded amyloid conformation may adopt an architecture that is structurally unrelated to the native conformations, but that is similar to β-helix or β-solenoid folds (Hennetin et al. 2006; Kajava and Steven 2006; Wasmer et al. 2008). Hence, it is possible that mammalian prion proteins may also adopt β-helical architecture in the misfolded state. Misfolded amyloid proteins share common structural characteristics, even when the native proteins are evolutionarily or structurally unrelated. Amyloid fibrils are generally unbranched, protease-resistant filaments with dominant β-sheet structures organized in a cross-β fashion, in which the β-strands run perpendicular to the fibril axis (Serpell and Smith 2000; Jimenez et al. 2002; Murali and Jayakumar 2005; Serpell et al. 2007). Previous prion amyloid modeling studies have converged on the β-helical architecture due to the structural features that β-helical folds share with the unresolved structure of longer chain amyloids. It has also been observed that expression of the isolated β-helical domain of the RβH protein, P22 tailspike protein, readily forms amyloid-like fibers (Schuler et al. 1999). Therefore, understanding the role of amino acid sequence in the folding of β–helices

may provide insight into how misfolded proteins form elongated β–sheet structures.

E. coli LpxA is the first example of a protein where LβH is the predominant secondary structure (Raetz and Roderick 1995). LpxA is a soluble, cytoplasmic protein that catalyzes the first step in the biosynthesis of lipid A, the hydrophobic anchor of lipopolysaccharides in Gram-negative bacteria (Galloway and Raetz 1990; Raetz 1990; Mohan et al. 1994; Sorensen et al. 1996; Wyckoff and Raetz 1999). Lipid A is required for the growth of *E. coli* and most other Gram-negative bacteria, and is also necessary for maintaining the integrity of the outer membrane as a barrier to toxic chemicals (Galloway and Raetz 1990; Raetz 1990; Vaara 1993b; a; Onishi et al. 1996). LpxA is composed of a single β-helical domain that is capped at the C-terminus with an α-helical domain. It is a trimeric protein formed from identical subunits that interact at the surfaces of the β-helical domains. The β-helical domains of LpxA are highly regular and symmetrical with little variability in shape and size over the length of the domain (Zheng et al. 2007). The 10 rungs (or coils) of the β-helix of LpxA are composed of 30 repeating hexapeptide motifs ([LIV]-[GAED]-X_2-[STAV]-X) (Raetz and Roderick 1995). Three repeats (18 amino acid

residues) make up one rung of the β-helix. Each rung of the canonical β-helix consists of three flat and untwisted parallel β-strands connected by either a one- or two-residue turn or a long excursion loop region (Jenkins and Pickersgill 2001; Hennetin et al. 2006; Iengar et al. 2006; Kajava and Steven 2006; Choi et al. 2008). Each β-strand contains small, uncharged residues (V, A, S, T, C) and conserved larger hydrophobic residues (L, I, V) that face the interior of the LβH to create a hydrophobic core (Jenkins and Pickersgill 2001; Iengar et al. 2006; Kajava and Steven 2006; Choi et al. 2008). These constraining requirements for the interior positions of the β-strand are presumed to have limited sequence divergence of the LβH proteins throughout evolution (Parisi and Echave 2001). Despite growing interest in the LβH as a possible structure of prion amyloid, there has been little structural characterization of the LβH fold. The accuracy of amyloid modeling studies employing the LβH fold has been limited by the relative absence of information pertaining to the sequence and structural features of this relatively rare protein fold.

In the present study, we investigated the folding and stability of the LpxA protein to examine the structural requirements of the LβH fold by altering the sequence in the β-helical domain. The *in-vivo* system developed

to study β-helical folding takes advantage of the strong correlation between the structural integrity of the β-helical domain of LpxA and *E. coli* growth. Site-directed mutagenesis was used to investigate the effect of mutations in the β-helical domain. We also conducted a new threading of PrP onto LpxA and designed the recombinant LpxA-PrP protein, in which the PrP fragment was incorporated in the β-helical domain of LpxA.

In-vivo functional assay to study folding of β-helices in LpxA

Given the structure of the *E. coli* LpxA protein [Protein Data Bank ID code 1LXA] (Raetz and Roderick 1995; Williams et al. 2006), the supersusceptibility of SM101 strain *E.coli* to antibiotics (Vuorio and Vaara 1992; Vaara 1993a; b) and results of *in-vitro* LpxA activity studies (Odegaard et al. 1997; Wyckoff and Raetz 1999), it was apparent that the structural integrity of the β-helical domain of LpxA is closely correlated with LpxA function and *E. coli* growth (Figure 3.1). Recently, in *E. coli,* the active site of LpxA has been identified and the structure with bound ligands has been described (Williams et al. 2006; Williams and Raetz 2007). Mutagenesis and structural studies of LpxA showed that the bottom region of the β-helical domain interacts directly with the active site of LpxA (Figure 3.1a & b). According to previous studies, the *E. coli* mutant SM101 (*lpxA2*, Strep^r), which contains the *LpxA* mutant allele (*lpxA2*) with a mutation at G189S, has very low levels of LpxA activity and defective lipid A biosynthesis (Galloway and Raetz 1990; Mohan et al. 1994; Odegaard et al. 1997). It has also been shown that SM101 is temperature-sensitive (Galloway and Raetz 1990; Mohan et al. 1994). Even at its permissive

temperature (30°C), SM101 displays a reduced lipid A content, and hypersensitivity to antibiotics such as novobiocin, rifampin, and erythromycin that are normally excluded by the outer membrane (Vuorio and Vaara 1992; Vaara 1993b; a; Odegaard et al. 1997). Based on our Ramachandran plot analysis of the LpxA structure, the G189S mutation may prevent correct folding of the β-helical domain of LpxA or destabilize the structure because glycine is the only favorable residue for position 189. When SM101 is transformed with plasmids containing the wild type *lpxA* gene, normal growth is recovered (Galloway and Raetz 1990).

To determine if SM101 could be used to study LpxA β-helical folding *in-vivo*, we constructed pJC1 (His-Tag LpxA) and pJC2 (His-Tag LpxA mutant containing a single tyrosine) plasmids as described in the "Material and Methods". pJC1 bore the gene for wild type LpxA, while pJC2 contained the gene for LpxA where all tyrosine residues except one were substituted with either phenylalanine or histidine (Y66F, Y77F, Y219F, Y223F, and Y243H). This LpxA variant has a relatively uncomplicated spectroscopic signature facilitating the characterization of the folded and unfolded states. Negative control mutants were made by site-directed mutagenesis at H125A and I86R of *lpxA*. H125 is an important residue in the LpxA active site. Mutation of H125 almost completely eliminates LpxA

activity (Wyckoff and Raetz 1999; Williams and Raetz 2007). The I86 residue is located in the 6th rung of the β-helical domain in the hydrophobic core of a β-helix (Figure 3.1a). Based on our examination of the structure, a substitution at I86 with a large, charged residue such as arginine could promote improper folding or destabilization of the β-helical domain and thereby decrease LpxA activity. Hence, H125A was used as a functional negative control mutant, and I86R was used as a structural negative control mutant. Both mutants were used to test whether an LpxA functional assay could be used as an *in-vivo* folding assay for β-helical folding.

In order to achieve increased levels of expression of the *lpxA* gene from pJC2 transfected bacteria (pET vectors), SM101 was lysogenized with bacteriophage DE3 using lambda as the helper phage. The resulting strain, SM101 (DE3), was made electrocompetent in order to increase the efficiency of transformation. pJC1, pJC2, pJC2-H125A, pJC2-I86R plasmids were transformed into SM101 (DE3), and incubated at 30°C. Ampicillin and streptomycin were used to select for bacteria from the SM101 strain that were transformed with the plasmid. Diluted overnight cultures from selected colonies of each transformant and SM105 (isogenic wild type strain of SM101) were inoculated in 96 well plates and grown at 30°C in medium with ampicillin, streptomycin, and novobiocin (0 - 256

μg/ml). Novobiocin was used to examine antibiotic susceptibility and growth of bacteria from the SM101 strain, transformed with the wild type and mutant plasmids. OD_{600} values were recorded after 12 hours to measure the bacterial growth. These values were plotted against novobiocin concentration. Introduction of the wild type *lpxA* gene allowed *E. coli* cells containing the *lpxA2* mutation to grow under the same range of novobiocin concentrations as SM105 (data not shown). As shown in Figure 3.2a, a single tyrosine mutant construct, pJC2, also showed growth similar to wild-type *E. coli*. Therefore, pJC2 was used as a wild type LpxA control in the present study. In contrast, pJC2-H125A and pJC2-I86R showed significant reductions in growth and hypersensitivity to novobiocin (MIC of 4 μg/ml versus 256 μg/ml for wild type). This result was consistent with previous studies of SM101 (Vuorio and Vaara 1992; Vaara 1993a).

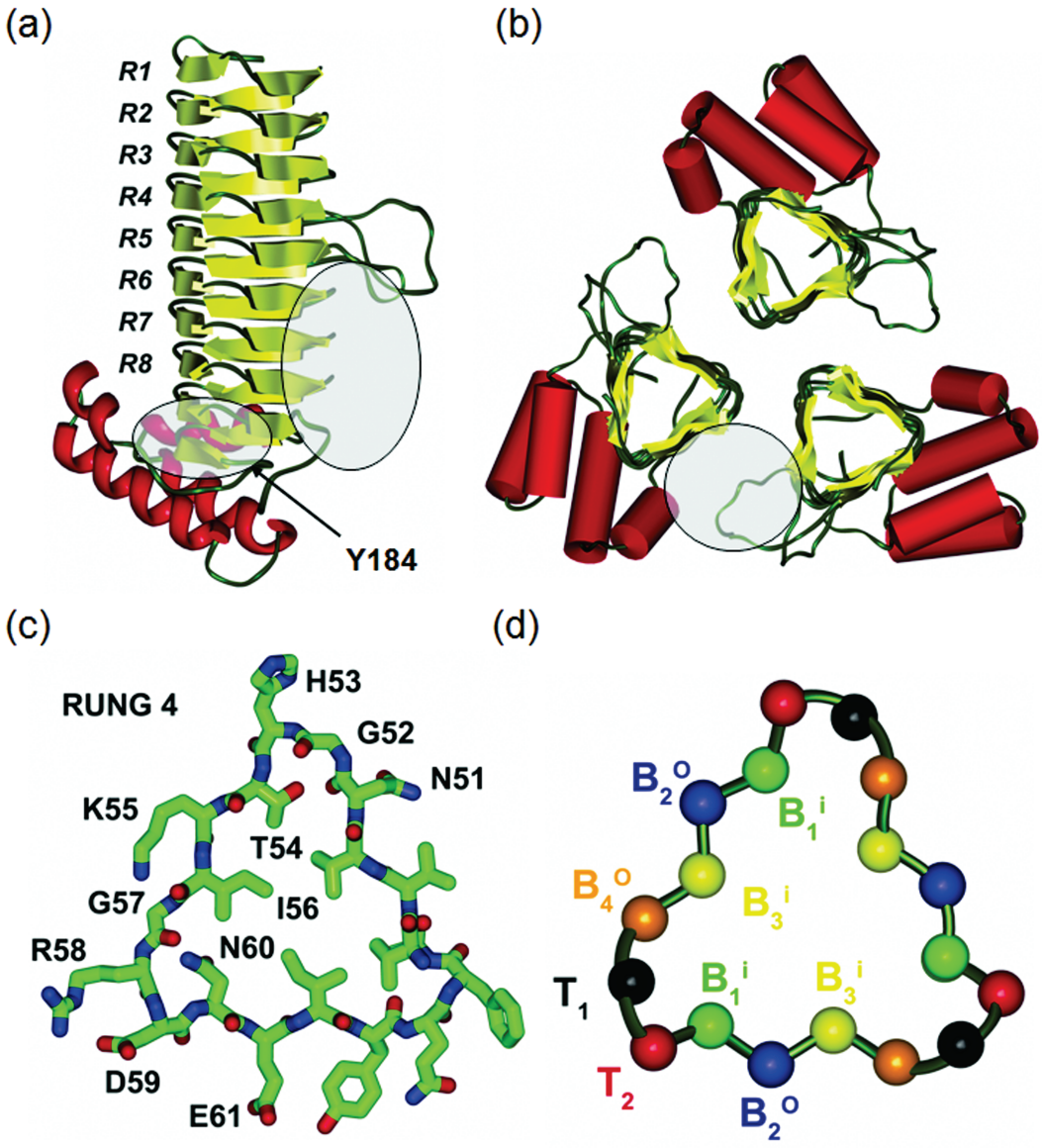

(a)
R1
R2
R3
R4
R5
R6
R7
R8
Y184
(b)
(c)
RUNG 4
H53
G52
N51
K55
T54
G57
I56
R58
N60
D59
E61
(d)
B_2^o
B_1^i
B_4^o
B_3^i
T_1
B_1^i
B_3^i
T_2
B_2^o

Figure 3.1. Monomeric and trimeric forms of the LβH protein, UDP-*N*-acetylglucosamine acyltransferase (LpxA). (a) LpxA monomer structure (PDB ID: 1LXA) with labels at each rung. β-helical, loop, and α-helical domains are shown in yellow, green, and red ribbons respectively. The ligand binding site is indicated by the shaded circle. A tyrosine residue (Y184) at the bottom region of β-helical domain is indicated. (b) Cross-sectional view of the LpxA trimer. The ligand binding site is indicated by the shaded circle. (c) Cross-sectional view of the fourth rung of the LpxA β-helical domain. The residues of the identified targeted region for the LpxA *in-vivo* folding assay (residues 51-62) are indicated. (d) Cross-sectional view of the LpxA β-helical domain. The turn (T) and β-sheet (B) with the superscript "i" for a residue facing inside and "o" for a residue facing outside.

Mutagenesis target selection

The N-terminal half of the β-helical domain most likely functions as a structural support and interface for trimeric formation based on structural analysis. Our initial mutagenesis and deletion studies of the β-helical domain showed that the N-terminal region of the LpxA β-helical domains may not be critical for LpxA functional activity and growth of *E. coli*. In contrast, the C-terminal half of the β-helical domain is directly involved with the active site such that slight changes in this region may reduce the LpxA activity significantly. Thus, in order to reduce the chance of false positive or false negative results, the ideal target mutation region for studying the stability and folding of the β-helical domain needs to be located away from the active site but not close to the N-terminus. To identify the target site, we conducted an arginine scan of the β-helical domain by site-directed mutagenesis at hydrophobic residues in the β-helical core, located in rungs 1 to 7 (Figure 3.1a). An *in-vivo* folding assay with novobiocin was conducted at 30°C with the LpxA mutants: I2R, I20R, I38R, I56R, I86R, V111R, V129R, and V147R along with controls as described earlier (Figure 3.2a). The log LD_{50} (Lethal Dose, 50%) was calculated from the bacterial growth curve plotted against the novobiocin concentration. Log LD_{50} was

then plotted against the location of mutagenesis sites referenced by rung number (Figure 3.2b). The arginine scan of the hydrophobic core β-helical domain showed that LpxA can be fully or partially active even with a disruption of up to two rungs (I38R) of β-helical domain. However, an arginine mutation at I56 almost completely eliminated the activity of LpxA. Therefore, based on our arginine scan and structural analysis of the active site, we identified the region between rungs 3 and 4 (residues 51-62) as the ideal target region for studying the folding and stability of the β-helical region.

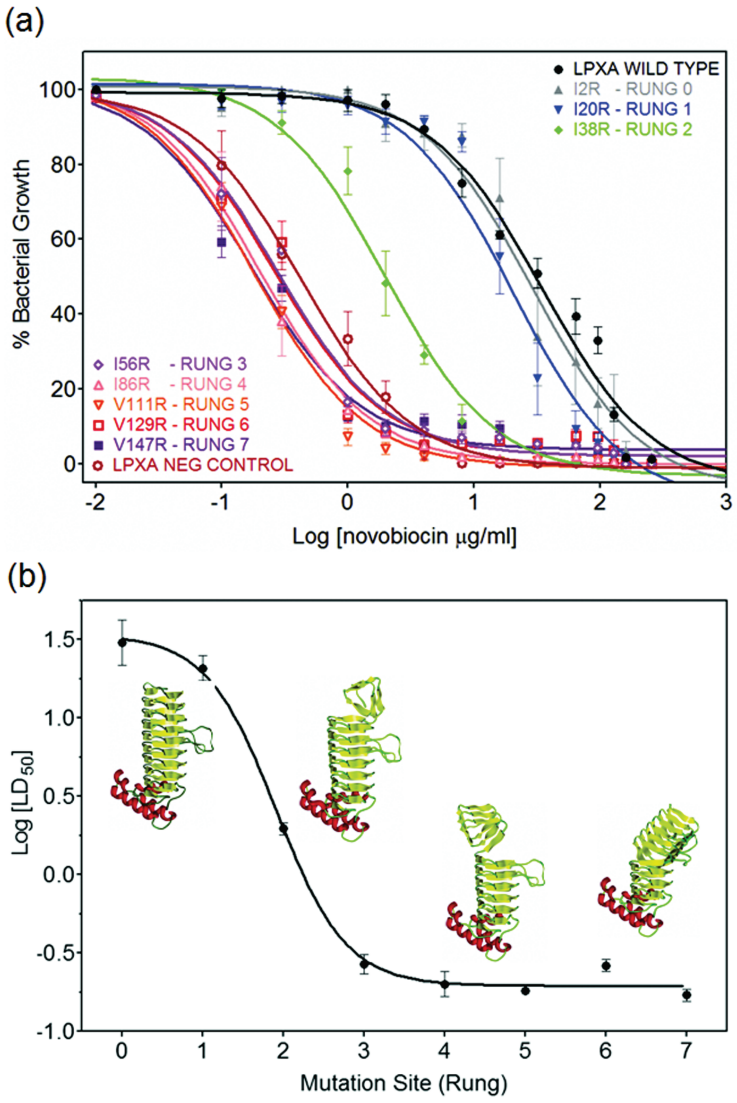

(a)
LPXA WILD TYPE
I2R - RUNG 0
I20R - RUNG 1
I38R - RUNG 2
I56R - RUNG 3
I86R - RUNG 4
V111R - RUNG 5
V129R - RUNG 6
V147R - RUNG 7
LPXA NEG CONTROL
% Bacterial Growth
Log [novobiocin μg/ml]
(b)
Log [LD50]
Mutation Site (Rung)

Figure 3.2. LpxA *in-vivo* folding assay for wild type LpxA and arginine LpxA mutant proteins at 30°C. (a) LpxA folding assay (novobiocin: 0 - 256 μg/ml) showing bacterial growth of the SM101 *E. coli* strain transformed with wild type LpxA, LpxA-H125A, or LpxA arginine mutants (I2R, I20R, I38R, I56R, I86R, V111R, V129R, and V147R). Hydrophobic residues at the B^{i}_{3} position were mutated in the LpxA arginine mutants. (b) Average log [LD_{50}] of each arginine LpxA mutant for three trials of the LpxA *in-vivo* folding assay, showing the large difference between the LpxA-I20R and LpxA-I56R mutants.

Site-directed mutagenesis in the hydrophobic core of the β-helical domain

In order to study the effect of non-hydrophobic residues in the core of the β-helical region on LpxA folding, we conducted site-directed mutagenesis at residue I56, substituting with alanine, asparagine, glutamine, glycine, glutamate, and arginine. SM101 were transformed with mutant LpxA and the *in-vivo* folding assay was performed at 30°C and 37°C. Wild type LpxA and negative control (LpxA-H125A) mutants were also characterized in the *in-vivo* folding assay. At 30°C, substitution of I56 with alanine did not significantly affect the *in-vivo* LpxA activity, but glycine, asparagine, and glutamine mutations showed lower LpxA activity than the wild type. While the glutamate mutant showed significantly decreased levels of LpxA activity, the arginine mutant showed complete loss of LpxA activity. This effect was an even greater than for the negative control (Figure 3.3a). At 37°C, wild type, alanine, asparagine, glutamine, and glycine mutations showed similar levels of LpxA activity within the range of standard deviation, but their LpxA activities at 37°C were decreased by two- to four-fold compared to those at 30°C. The glutamate and arginine mutants showed complete loss of LpxA activity at 37°C (Figure 3.3b).

To verify the LpxA *in-vivo* folding assay results at the level of protein expression, LpxA mutants were individually overexpressed from a pET plasmid in SM101 (DE3) at 30°C. *E. coli* lysate samples were separated by electrophoresis on an SDS polyacrylamide gel, and overexpressed LpxA monomer bands were visible. As can be seen in Figure 3.3c, LpxA from lysates of the wild type LpxA and LpxA-H125A bacteria, as well as the alanine LpxA mutant bacteria was found mostly in the soluble fractions. Less intense LpxA bands were observed from the soluble fractions of lysates from the asparagine and glutamine mutant bacteria. LpxA from the glycine and arginine mutant bacteria were found mostly in the insoluble fraction, indicating that the arginine mutation had a detrimental effect on folding of the β-helix. Although the expression level of soluble glycine LpxA mutant was surprisingly lower than expected from the LpxA *in-vivo* folding assay, protein expression of the LpxA mutants was consistent with the LpxA *in-vivo* folding assay overall.

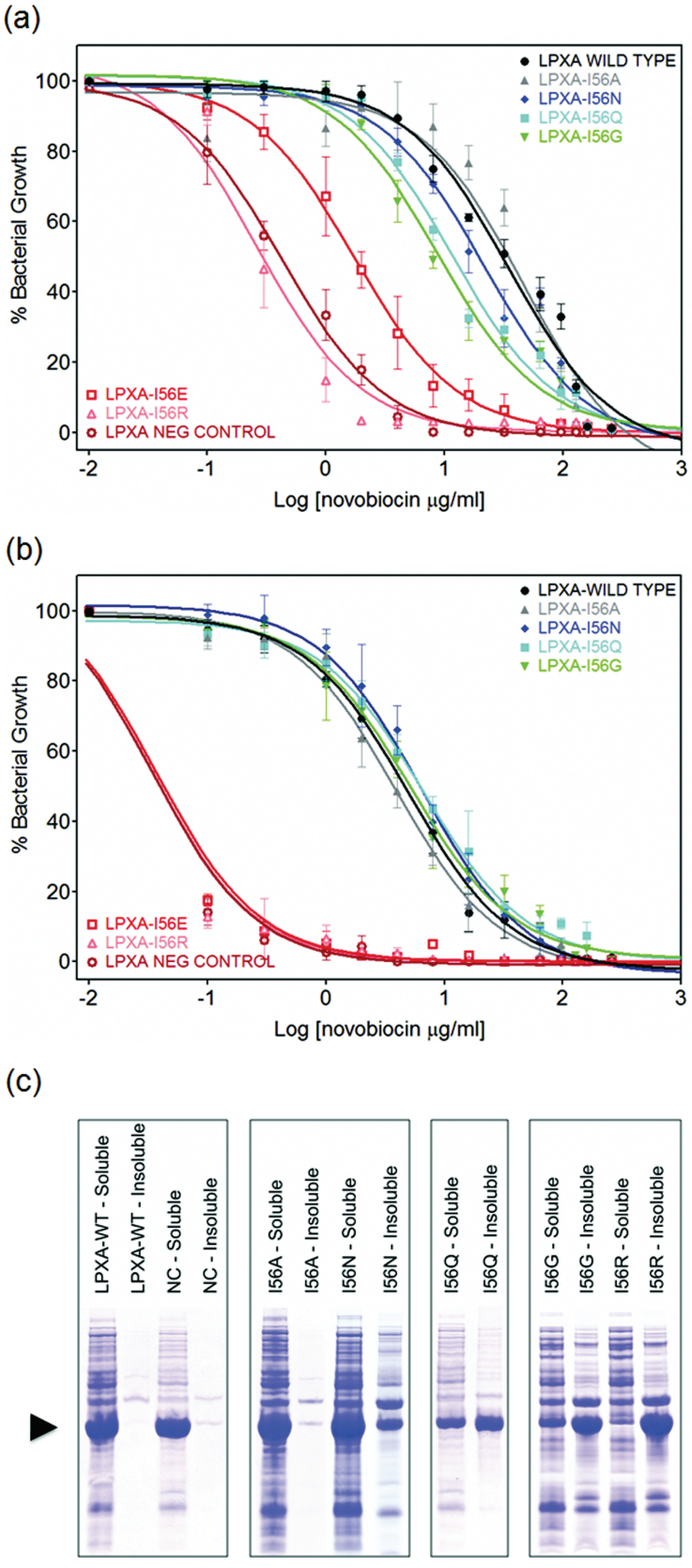

(a)
LPXA WILD TYPE
LPXA-I56A
LPXA-I56N
LPXA-I56Q
LPXA-I56G
LPXA-I56E
LPXA-I56R
LPXA NEG CONTROL
% Bacterial Growth
Log [novobiocin µg/ml]
(b)
LPXA-WILD TYPE
LPXA-I56A
LPXA-I56N
LPXA-I56Q
LPXA-I56G
LPXA-I56E
LPXA-I56R
LPXA NEG CONTROL
% Bacterial Growth
Log [novobiocin µg/ml]
(c)
LPXA-WT - Soluble
LPXA-WT - Insoluble
NC - Soluble
NC - Insoluble
I56A - Soluble
I56A - Insoluble
I56N - Soluble
I56N - Insoluble
I56Q - Soluble
I56Q - Insoluble
I56G - Soluble
I56G - Insoluble
I56R - Soluble
I56R - Insoluble

Figure 3.3. *In-vivo* folding assay for LpxA mutant proteins with mutations in the hydrophobic core. LpxA folding assay, showing bacterial growth of the SM101 *E. coli* strain, transformed with LpxA containing one of the following mutations: (a) At 30°C, LpxA Wild Type (LD_{50} = 34 ± 1.5 μg/ml), LpxA Negative Control (LD_{50} = 0.42 ± 1.0 μg/ml), I56A (LD_{50} = 46 ± 2.3 μg/ml), I56N (LD_{50} = 22 ± 1.8 μg/ml), I56Q (LD_{50} = 12 ± 1.5 μg/ml), I56G (LD_{50} = 8.6 ± 1.7 μg/ml), I56E (LD_{50} = 1.7 ± 1.6 μg/ml), and I56R (LD_{50} = 0.27 ± 1.3 μg/ml). (b) At 37°C, LpxA Wild Type (LD_{50} = 5.0 ± 1.6 μg/ml), LpxA Negative Control (LD_{50} = 0.034 ± 1.1 μg/ml), I56A (LD_{50} = 3.9 ± 1.5 μg/ml), I56N (LD_{50} = 6.4 ± 1.9 μg/ml), I56Q (LD_{50} = 6.5 ± 1.8 μg/ml), I56G (LD_{50} = 5.2 ± 1.7 μg/ml), I56E (LD_{50} = 0.037 ± 1.0 μg/ml), and I56R (LD_{50} = 0.034 ± 1.2 μg/ml). (c) SDS polyacrylamide gel electrophoresis of soluble and insoluble fractions of overexpressed LpxA cellular lysates is shown. Arrow indicates monomer bands of LpxA and LpxA mutant proteins.

Proline mutagenesis at the turn regions of the β-helical domain

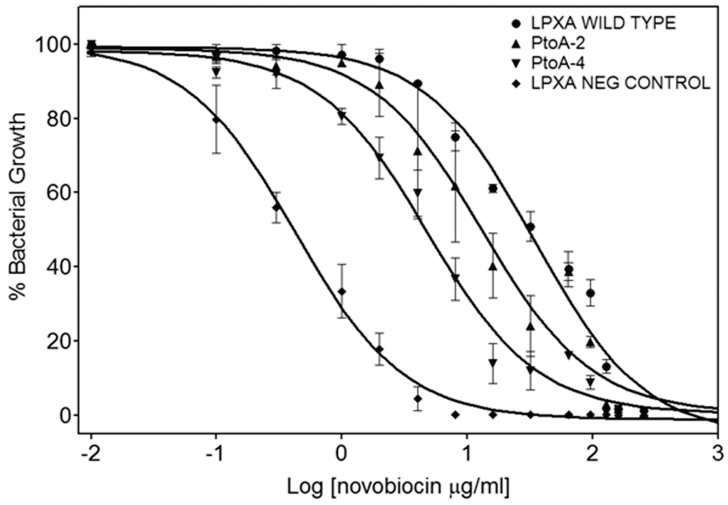
We also explored the effect of proline mutations on LpxA activity, irrespective of their solvent accessibility within the protein structure. In our previous structural analysis of the LβH, we reported a high occurrence of proline at the T_1 position in either the top or bottom regions of the β-helix, suggesting that proline residues in these positions may influence folding of the β-helix (Choi et al. 2008). To investigate this hypothesis for the LpxA β-helix, two proline residues (P28 and P34) and four proline residues (P10, P28, P34, and P183) at T_1 positions found in the β-helical domain were replaced with alanine by site directed mutagenesis. These were labeled PtoA-2 and PtoA-4 respectively. The P182 residue, which is located in the unusual position of B^o_4, was excluded due to its deleterious effect on LpxA folding when mutated to glycine or alanine (data not shown). Results from the *in vivo* folding assay showed that the activity of the PtoA-4 mutant was significantly lower than wild type LpxA, but still greater than the negative control, LpxA-H125A while the PtoA-2 mutant showed relatively greater activity than the PtoA-4 mutant (Figure 3.4a).

Unfolding and refolding of LpxA monitored by tyrosine fluorescence

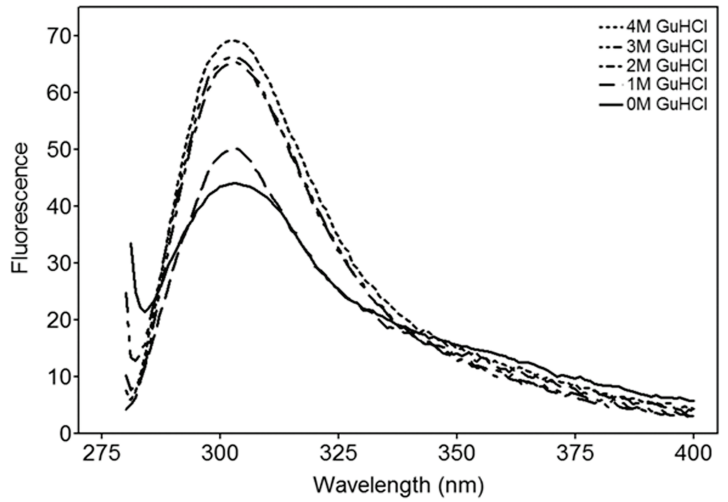
The thermodynamic stability of the wild type and proline LpxA mutant was probed by tyrosine fluorescence spectroscopy. Since LpxA contains several tyrosine and no tryptophan residues, all but a single tyrosine residue (Y184) were replaced with phenylalanine or histidine in order to simplify the fluorescence spectrum during folding and unfolding. Fluorescence emission in the near-ultraviolet range reflects the environment of the tyrosine residue, thereby giving a measure of the tertiary structure of a protein (Creighton 1997). Upon excitation at 280 nm, maximal fluorescent emission was observed at 303 nm (Figure 3.4b). Addition of the denaturant, guanidine hydrochloride (GuHCl), increased the fluorescent emission intensity. To determine whether LpxA unfolding induced by a chemical denaturant is reversible, guanidine hydrochloride-induced folding and unfolding of LpxA was monitored by measuring the tyrosine fluorescence spectra (Figure 3.4c). Both denaturation and renaturation occurred at approximately 1.5 M GuHCl. Only a single state transition was observed and no trimer assembly was detected, indicating that the change in fluorescence spectra reflects changes of the monomer structure.

To determine whether proline mutations in the β-helical domain affect the thermodynamic stability of LpxA, the unfolding transition of the PtoA-2 and PtoA-4 mutants were also monitored over a range of GuHCl concentrations. As shown in Figure 3.4c, denaturation of the PtoA-2 and PtoA-4 mutants occurred at approximately 1.3 M GuHCl and 1.1 M GuHCl respectively.

(a)


LPXA WILD TYPE
PtoA-2
PtoA-4
LPXA NEG CONTROL
% Bacterial Growth
Log [novobiocin μg/ml]

(b)


4M GuHCl
3M GuHCl
2M GuHCl
1M GuHCl
0M GuHCl
Fluorescence
Wavelength (nm)

(c)

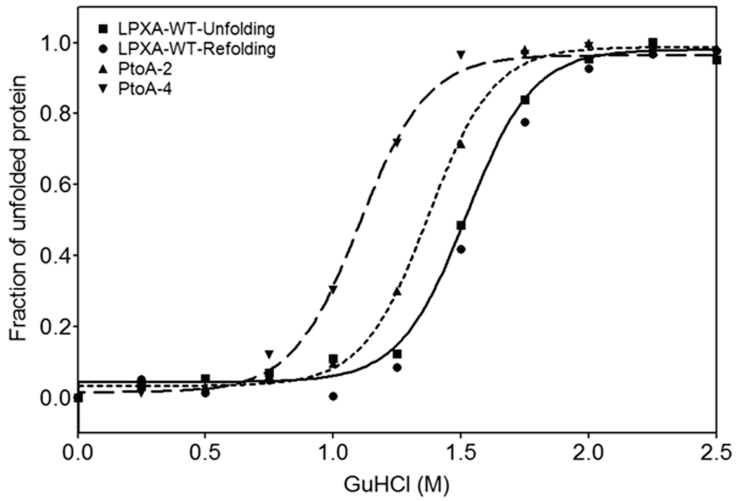

LPXA-WT-Unfolding
LPXA-WT-Refolding
PtoA-2
PtoA-4
Fraction of unfolded protein
GuHCl (M)

Figure 3.4. *In-vivo* folding assay and spectroscopic characterization of wild type LpxA and LpxA proline mutant proteins. (a) The functional activity of LpxA mutants with two and four prolines replaced by alanine residues in the β-helical domain, PtoA-2 (LD_{50} = 13 ± 3.2 μg/ml) and PtoA-4 (LD_{50} = 4.8 ± 1.7 μg/ml) at 30°C. (b) Fluorescence emission spectra of LpxA in its native state (solid line) and in its GuHCl-denatured states (dotted line) were recorded in 100 mM sodium phosphate, pH 7.5, containing 0 to 4 M GuHCl. The dash-dotted line represents the difference between the native and denatured protein spectra. (c) The folding and unfolding transition of wild type LpxA and LpxA-mutant proteins. Transitions were measured using fluorescence (emission at 303 nm) after 30 min incubation. Uncorrected fluorescence emission spectra were recorded with a Spex Fluoromax (excitation at 275 nm, spectral bandwidth of 3 nm (excitation) and 3 nm (emission)). All spectra were recorded at 25°C and are buffer-corrected. The protein concentrations were 100 μg/ml.

Residue tolerance in the β-helical domain

A selected target region of the β-helical domain (residues 51-62) was further studied by determining the amino acid residue tolerance at each position (Figure 3.1c & d). Based on our *in-vivo* and *in-vitro* folding analysis of LpxA mutants, we selected concentrations of novobiocin (4 μg/ml) to measure viable LpxA activity at 37°C and 30°C. We performed site-directed mutagenesis at all six β-helix positions (T_1, T_2, B^i_1, B^o_2, B^i_3, and B^o_4) to identify allowed and disallowed residues at specific positions of the β-helix. Two different residues per position type were selected for mutagenesis. LpxA mutants were transformed into SM101 (DE3) and grown under novobiocin selection (4 μg/ml) at 37°C and 30°C. Table 3 summarizes the allowed and disallowed amino acid residues for each position with lists of residues that are observed at each position of the β-helix based on sequence analysis of the LβH (Choi et al. 2008). While proline residues are restricted to only T_1 and T_2 positions, as expected from the previous study (Choi et al. 2008), glycine and hydrophilic residues such as asparagine and glutamine, that were not seen in the β-helical core were shown to be tolerated at B^i_1 and B^i_3 positions.

Table 3. Amino acid residue tolerance of LpxA LβH domain from the *in-vivo LpxA* folding assay

Type	Position	Known[a]	Allowed[b]	Now Allowed[c]
T_1	G52, R58	P,D,G,A,E,N,S,H,K,T,Y,M,F,R,Q,I,C	L,V,W	
T_2	H53, D59	N,G,D,P,F,R,E,K,Q,H,S,Y,T,L,A,M, W,C	I,V	
B^i_1	T54, N60	V,A,T,S,C,I,N,L,F	G,Q,M	R,K,W,Y,H,E, D,P
B^o_2	K55, E61	V,T,I,S,E,Y,K,F,H,R,M,N,W,L,Q,A, G,C	D	P
B^i_3	I56, I62	I,V,L,A,F,T,M,C	G,N,Q,S	R,K,W,Y,H,E, D,P
B^o_4	G57, N51	G,E,A,D,H,V,N,S,I,K,R,M,T,L,Y,Q, C,F,W		P

[a]Values were taken from the multiple sequence alignment of LpxA and other LβH proteins (Choi et al. 2008) and listed in order of occurrence frequency. Amino acid residues newly observed to be [b]tolerated, and [c]not tolerated at each position of β-helix domain of LpxA. *In-vivo* LpxA folding assay was conducted at 30 ºC and 37 ºC with 8 μg/ml of novobiocin as a selection concentration.

Discussion

Glycine and hydrophilic residues in the hydrophobic core of the β-helix

The correlation between the structural integrity and LpxA activity as measured by bacterial growth allows LpxA to be used as a unique model system for studying the folding and stability of the LβH *in-vivo* and *in-vitro*. We have shown that *in vivo* folding of the LpxA β-helical domain can be studied by measuring bacterial growth in the presence of antibiotics such as novobiocin. An arginine scan of the hydrophobic core of the β-helical region was used to identify the target region for LpxA site-directed mutagenesis. Mutations were made in this target region and *in-vivo* folding assays were performed. The results showed that the LpxA β-helical domain tolerated disruption of up to two rungs (residues 1-38). However, an arginine mutation at the third or fourth rung resulted in deleterious effects on the growth of *E. coli* and presumably LpxA functional activity. These regions are not directly involved in the active site or ligand binding, making them a rational target site for structural studies of the β-helix.

Site-directed mutagenesis of the identified target region and subsequent *in-vivo* folding assays showed that the LpxA β-helix can tolerate substitution of residues in the hydrophobic core. While sequence analyses of LβH proteins showed strong conservation of hydrophobic residues such as isoleucine, leucine and valine at B^{i}_{3}, the *in-vivo* folding study showed that a single substitution with hydrophilic residues such as asparagine and glutamine are still tolerated by the β-helix. In addition, a single substitution with glycine at B^{i}_{3} showed a decrease in bacterial growth over the range of antibiotic concentrations examined. However, some viable LpxA activity remained. Results of the *in vivo* assay were also confirmed by SDS-PAGE electrophoresis of overexpressed mutant LpxA protein. While the alanine mutation did not alter the level of soluble protein expression, asparagine, glutamine and glycine mutations resulted in increased levels of insoluble aggregate formation as compared to the wild type, suggesting that the decreased LpxA activity in the *in vivo* assay may reflect decreased β-helix formation or sequestration. Decreased levels of β-helix formation may occur due to destabilization of the β-helical domain caused by gaps created in the core of the β-helix resulting from the glycine mutation or steric hindrance and unfavorable side chain interaction caused by the asparagine and glutamine mutations. These results are consistent with alanine site-

directed mutagenesis study of the RβH, P22 tailspike protein (Simkovsky and King 2006). However, our results revealed that LpxA can still tolerate a point mutation with residues such as glycine and glutamine at B^{i}_{3} positions that were thought be critical in β-helical folding.

The effect of proline mutations on the thermodynamic stability of the LβH protein

The thermodynamic stability of the β-helix can be estimated by monitoring the intrinsic tyrosine fluorescence spectrum of proteins with a single tyrosine residue in the β-helical domain. LpxA proline mutants where two to four proline residues in the β-helical domain were replaced with alanine were used to study the effect of prolines on β-helical folding and stability. As previously described (Choi et al. 2008), prolines in the β-helical domain are mostly located at the top and bottom region of the domain, leading us to speculate that they may play an important role in initiation or termination of β-helical folding. Our result shows that β-helical folding occurs without the proline residues, but that the LpxA activity is significantly reduced. Based on the LpxA sequence and structure, the proline

residues targeted for alanine substitution were not located in the active site, nor are they expected to be involved in trimer formation. Therefore, the reduced LpxA activity of the proline mutants can be explained by the decreased thermodynamic stability of the β-helix as indicated by unfolding transitions monitored using tyrosine fluorescence spectroscopy in the presence of a chemical denaturant. Proline mutations of the β-helix may also have an effect on efficiency of β-helical folding. However, in order to accurately determine the folding efficiency of the β-helix, folding kinetics of the wild type and other mutant proteins need to be investigated further.

Residue tolerance of the β-helical domain

β-helical proteins, especially LβH proteins, contain distinct patterns of alternating hydrophobic and non-hydrophobic residues. This sequence feature is associated with unique structural characteristics of the β-helix. Previous studies of the LβH structure have shown relatively strong sequence conservation at each position of the β-helical fold (Jenkins and Pickersgill 2001; Iengar et al. 2006; Choi et al. 2008). However, β-helical proteins may tolerate slight structural deviations or substitutions of amino acid residues at

complementary positions that maintain structural integrity. To investigate the residue tolerance of the β-helix, point mutations were introduced at each position in an identified target region (residues 51 – 62) by site-directed mutagenesis. The *in-vivo* folding assay was then used to determine whether the mutation affected the β-helical domain. Since the target region is not located at the trimeric interface, the point mutations are not expected to affect LpxA trimer formation. All amino acid residues at turn positions (T_1 and T_2) and all amino acid residues, except proline, at outward positions (B^o_2 and B^o_4) were shown to be tolerated as expected from previous studies (Iengar et al. 2006; Choi et al. 2008). However, at the interior of the β-helix, unexpected residues such as glycine and glutamine were observed at B^i_1 and B^i_3 positions. The results of this study suggest that the β-helical structure can tolerate exceptions at interior positions without complete destabilization of the structure. Under relatively dehydrated conditions such as those associated with amyloid formation, it is possible that stacks of these unexpected residues such as glutamine could be tolerated owing to side chain-side chain hydrogen bonding. This could have implications for the polyglutamine repeat diseases.

Materials and Methods

Materials

Bacterial strains SM105 (*lpxA*, Strep^r) and SM101(*lpxA2*, Strep^r) (Galloway and Raetz 1990) were obtained from the *E. coli* Genetic Stock Center, Yale University (New Haven, CT). Antibiotics and guanidine hydrochloride were purchased from Sigma (St. Louis, MO). Restriction enzymes (NdeI, BamHI, and DpnI) and Quick Ligation Kit were purchased from New England BioLabs (Ipswich, MA). Pfu turbo DNA polymerase, XL1-Blue supercompetent cells, and QuikChange Site-Directed Mutagenesis Kit were purchased from Stratagene (La Jolla, CA). DH5α cells were purchased from Invitrogen (San Diego, CA). Lambda DE3 Lysogenization Kit was purchased from Novagen (Darmstadt, Germany). 96-well, flat bottom culture plates were purchased from Corning Costar (Corning, NY). QIAprep Spin MiniPrep kits, QIAquick Gel Extraction Kit, and the Ni-NTA purification system were purchased from Qiagen (Hilden, Germany). Primers for mutagenesis and the *lpxA-PrP* recombinant gene were custom-made by Integrated DNA Technologies (Coraville, IA). pTO1(Odegaard et al. 1997), the vector containing wild type *E coli. lpxA*, was kindly provided by Dr. C. R. H. Raetz (Duke University Medical Center, Durham, NC).

Bacterial strains, transformation, and growth conditions

SM101(DE3) were prepared from SM101 using the Lambda DE3 lysogenization kit and made electrocompetent for transformation (Ausubel 1987). Cells were grown in LB medium, containing 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl, or on LB plus 15 g/L agar, at 30°C (Ausubel 1987). Transformants were plated on LB agar containing appropriate antibiotics at the following final concentrations: 100 μg/ml ampicillin, 30 μg/ml streptomycin.

Plasmid construction

pTO1, containing wild type *E. Coli lpxA*, was previously constructed using the vector pET23c as described previously (Odegaard et al. 1997). The *lpxA* gene was re-cloned into pET23c for expression in *E. coli* as follows: PCR primers were designed from the DNA sequence of *E. coli lpxA* to generate LpxA protein carrying a C-terminal fusion of six histidine residues (His-Tag) to enable one-step purification of the expressed protein. *Nde*I and *BamH*I sites were incorporated into these oligonucleotide primers to facilitate cloning. We used the pET expression system in which genes are driven by the T7 promoter to obtain high levels of protein expression. The

primer sequences were as follows: forward primer 5' GAA GGA GAT ATA CAT ATG ATT GAT AAA 3' and reverse primer 5' ATT CGG ATC CCC ACG AAT CAG ACC GCG 3'. *Nde*I and *BamH*I sites are underlined. Using *pTO1* DNA as the template, a PCR reaction was carried out with the forward and reverse primers, using *Pfu* polymerase. The cycle was: 94°C for 30 sec, 50°C for 30 sec, and 65°C for 2 min. This cycle was repeated 25 times. The PCR reaction product was separated on a 0.8% agarose gel, and the expected DNA product at approximately 800 bp was identified. DNA from this fragment was eluted using the QIAquick Gel Extraction Kit and subsequently digested with *Nde*I and *BamH*I. The digested DNA fragment was ligated into pET-23c(+) that had been also digested with *Nde*I and *BamH*I. The ligation mixture was transformed into *E. coli* DH5α cells with selection by 100 μg/ml ampicillin. Colonies were picked and grown for DNA preparation. Positive recombinants were screened by digestion with *Nde*I and *BamH*I to excise the 800 bp insert. This led to the identification of the pJC1 (His-Tag LpxA) plasmid.

Site-directed mutagenesis

LpxA mutants were made by using the PCR-based QuikChange mutagenesis procedure. Five tyrosine residues of LpxA were subjects to site-directed mutagenesis in order to make the single tyrosine recombinant LpxA. The resulting vector that contained the single tyrosine wild type LpxA clone, pJC2, encoded the full-length LpxA protein with the following mutations: Y66F, Y77F, Y219F, Y223F, and Y243H. For all other LpxA mutants, methylated pJC2 DNA was used as a template DNA for site-directed mutagenesis. A PCR reaction was carried out with forward and reverse primers, using *Pfu* polymerase. The cycle was: 94°C for 30 sec, 55°C for 1 min, and 62°C for 10 min. This cycle was repeated 18 times. Plasmid DNA was purified using QIAprep Spin MiniPrep kits. DNA sequencing of an 800 bp region containing the targeted codon confirmed mutant clones.

LpxA *in-vivo* folding assay

Previously prepared SM101 (DE3) electrocompetent cells were transformed with pJC2 vectors, containing the wild-type and mutant LpxA clones by using a Gene Fulser Xcell Electroporator (BioRad, Hercules, CA).

Transformants were plated on LB agar containing appropriate antibiotics (100 μg/ml ampicillin and 30 μg/ml streptomycin) and incubated at 30°C. Colonies were picked and inoculated in LB medium (100 μg/ml ampicillin and 30 μg/ml streptomycin) and incubated for 12 hours at 30°C. 2 μl of diluted overnight inoculums were mixed in each well of 96-well plates, prepared with LB medium (100 μg/ml ampicillin, 30 μg/ml streptomycin and 0 - 256 μg/ml novobiocin). The 96-well plates were incubated on a shaker at 30°C for 12 hours and at 37°C for 6 hours. The OD_{600} was recorded using a SpectraMax M5 microplate reader (Molecular Device, Sunnyvale, CA).

Protein expression and purification

LpxA and LpxA mutant proteins were expressed in SM101 (DE3) cells. LpxA transformed cells were grown in LB medium (100 μg/ml ampicillin and 30 μg/ml streptomycin) at 30 °C until OD_{600} = 0.8. Cells were induced with 0.5 mM of isopropyl β-D-1-thiogalatopyranoside (IPTG) and grown for another 12 hours. Cell suspensions were centrifuged and cell pellets were collected and lysed. Cell lysates were centrifuged at 15000 rpm for 30 min to collect the soluble fraction. The pellets were collected and

resuspended in 8 M urea as insoluble fractions. For the *in-vitro* folding and unfolding study, the soluble proteins were purified using the Ni-NTA purification system and size-exclusion chromatography. The wild type LpxA protein eluted as a 90 kDa protein, indicating that LpxA mostly exists as a trimer in solution. The collected protein samples were separated on SDS polyacrylamide gels for verification of the LpxA monomer (28 kDa).

In-vitro folding and unfolding (tyrosine fluorescence spectroscopy)

Folding transitions were measured in a Spex FluoroMax spectrofluorometer (HORIBA Jobin Yvon Inc, Edison, NJ) with a thermostated cell holder using plastic fluorescence cuvettes. The excitation wavelength was 280 nm and the emission wavelengths were 280 - 400 nm. For unfolding transitions, cold buffer (100 mM sodium phosphate, pH 7.5, containing 0.2 M NaCl and 1 mM EDTA) was first added to a small volume of concentrated protein solution. Then, a cold, concentrated guanidine hydrochloride solution containing 100 mM sodium phosphate, pH 7.5, 0.2 M NaCl, and 1 mM EDTA was added (final volume 1 ml). For refolding transitions, 4 M guanidine hydrochloride solution was added and the samples

were incubated for 30 min. Refolding was initiated by rapid dilution with buffer containing 0.2 M NaCl and 1 mM EDTA.

References

Ausubel, F.M. 1987. Current protocols in molecular biology, pp. 2 v. (loose-leaf). Greene Publishing Associates; J. Wiley, order fulfillment, Brooklyn, N. Y. Media, Pa.

Choi, J.H., Govaerts, C., May, B.C., and Cohen, F.E. 2008. Analysis of the sequence and structural features of the left-handed beta-helical fold. *Proteins* **73:** 150-160.

Creighton, T.E. 1997. *Protein structure : a practical approach*, 2nd ed. IRL Press at Oxford University Press, Oxford ; New York, pp. xxiii, 383 p.

Galloway, S.M., and Raetz, C.R. 1990. A mutant of Escherichia coli defective in the first step of endotoxin biosynthesis. *J Biol Chem* **265:** 6394-6402.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. 2004. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101:** 8342-8347.

Hennetin, J., Jullian, B., Steven, A.C., and Kajava, A.V. 2006. Standard conformations of beta-arches in beta-solenoid proteins. *J Mol Biol* **358:** 1094-1105.

Iengar, P., Joshi, N.V., and Balaram, P. 2006. Conformational and sequence signatures in beta helix proteins. *Structure* **14:** 529-542.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. 2002. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99:** 9196-9201.

Kajava, A.V., and Steven, A.C. 2006. Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv Protein Chem* **73:** 55-96.

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Mohan, S., Kelly, T.M., Eveland, S.S., Raetz, C.R., and Anderson, M.S. 1994. An Escherichia coli gene (FabZ) encoding (3R)-hydroxymyristoyl acyl carrier protein dehydrase. Relation to fabA and suppression of mutations in lipid A biosynthesis. *J Biol Chem* **269:** 32896-32903.

Murali, J., and Jayakumar, R. 2005. Spectroscopic studies on native and protofibrillar insulin. *J Struct Biol* **150:** 180-189.

Odegaard, T.J., Kaltashov, I.A., Cotter, R.J., Steeghs, L., van der Ley, P., Khan, S., Maskell, D.J., and Raetz, C.R. 1997. Shortened hydroxyacyl

chains on lipid A of Escherichia coli cells expressing a foreign UDP-N-acetylglucosamine O-acyltransferase. *J Biol Chem* **272:** 19688-19696.

Onishi, H.R., Pelak, B.A., Gerckens, L.S., Silver, L.L., Kahan, F.M., Chen, M.H., Patchett, A.A., Galloway, S.M., Hyland, S.A., Anderson, M.S., et al. 1996. Antibacterial agents that inhibit lipid A biosynthesis. *Science* **274:** 980-982.

Parisi, G., and Echave, J. 2001. Structural constraints and emergence of sequence patterns in protein evolution. *Mol Biol Evol* **18:** 750-756.

Raetz, C.R. 1990. Biochemistry of endotoxins. *Annu Rev Biochem* **59:** 129-170.

Raetz, C.R., and Roderick, S.L. 1995. A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science* **270:** 997-1000.

Schuler, B., Rachel, R., and Seckler, R. 1999. Formation of fibrous aggregates from a non-native intermediate: the isolated P22 tailspike beta-helix domain. *J Biol Chem* **274:** 18589-18596.

Serpell, L.C., Benson, M., Liepnieks, J.J., and Fraser, P.E. 2007. Structural analyses of fibrinogen amyloid fibrils. *Amyloid* **14:** 199-203.

Serpell, L.C., and Smith, J.M. 2000. Direct visualisation of the beta-sheet structure of synthetic Alzheimer's amyloid. *J Mol Biol* **299:** 225-231.

Simkovsky, R., and King, J. 2006. An elongated spine of buried core residues necessary for in vivo folding of the parallel beta-helix of P22 tailspike adhesin. *Proc Natl Acad Sci U S A* **103:** 3575-3580.

Sorensen, P.G., Lutkenhaus, J., Young, K., Eveland, S.S., Anderson, M.S., and Raetz, C.R. 1996. Regulation of UDP-3-O-[R-3-hydroxymyristoyl]-N-acetylglucosamine deacetylase in Escherichia coli. The second enzymatic step of lipid a biosynthesis. *J Biol Chem* **271:** 25898-25905.

Vaara, M. 1993a. Antibiotic-supersusceptible mutants of Escherichia coli and Salmonella typhimurium. *Antimicrob Agents Chemother* **37:** 2255-2260.

Vaara, M. 1993b. Outer membrane permeability barrier to azithromycin, clarithromycin, and roxithromycin in gram-negative enteric bacteria. *Antimicrob Agents Chemother* **37:** 354-356.

Vuorio, R., and Vaara, M. 1992. The lipid A biosynthesis mutation lpxA2 of Escherichia coli results in drastic antibiotic supersusceptibility. *Antimicrob Agents Chemother* **36:** 826-829.

Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319:** 1523-1526.

Williams, A.H., Immormino, R.M., Gewirth, D.T., and Raetz, C.R. 2006. Structure of UDP-N-acetylglucosamine acyltransferase with a bound antibacterial pentadecapeptide. *Proc Natl Acad Sci U S A* **103:** 10877-10882.

Williams, A.H., and Raetz, C.R. 2007. Structural basis for the acyl chain selectivity and mechanism of UDP-N-acetylglucosamine acyltransferase. *Proc Natl Acad Sci U S A* **104:** 13543-13550.

Wyckoff, T.J., and Raetz, C.R. 1999. The active site of Escherichia coli UDP-N-acetylglucosamine acyltransferase. Chemical modification and site-directed mutagenesis. *J Biol Chem* **274:** 27047-27055.

Yang, S., Levine, H., Onuchic, J.N., and Cox, D.L. 2005. Structure of infectious prions: stabilization by domain swapping. *FASEB J* **19:** 1778-1782.

Zheng, J., Zanuy, D., Haspel, N., Tsai, C.J., Aleman, C., and Nussinov, R. 2007. Nanostructure design using protein building blocks enhanced by

conformationally constrained synthetic residues. *Biochemistry* **46:** 1205-1218.

Chapter 4

Design and engineering of a recombinant left-handed β-helical protein with inclusion of a prion fragment

Recombinant LpxA with inclusion of PrP fragment in the β-helical domain

Based on the observed residue tolerance at each position of the LpxA β-helix, we threaded fragments of mouse PrP (residues 104-143) onto LpxA to determine whether it is possible for this PrP sequence to fold into a β-helix. This region has been shown to form β-rich structures in isolation and as part of larger PrP peptides under a variety of conditions. Unlike the previous comprehensive modeling studies of the misfolded form of prion protein (PrP^{Sc}), we only focused on threading the short alanine- and glycine-rich region onto the LpxA sequence. From previous studies (Govaerts et al. 2004; Choi et al. 2008), the criterion for β-helix threading includes proline residues at the T_1 and T_2 positions since they are the only allowed positions in the β-helical domain. Threading also requires relatively small residues such as alanine and serine at B^i_1 due to steric hindrance, and relatively large hydrophobic residues at B^i_3 for the hydrophobic core. However, the residue tolerance of the β-helix determined in this study allowed us to place glycine and alanine at the B^i_1 and B^i_3 positions. Glycine residue restriction at some B^o_4 positions was also considered. The threading result (Figure 4.1a) shows

a reasonable sequence alignment of mouse PrP sequence (PrP residues 104-143) with the LpxA sequence (LpxA residues 22-59) with exception of a short loop (PrP residues 134-135). While glycine and alanine residues were placed at several B^{i}_{1} and B^{i}_{3} positions, proline residues are placed at T_1 positions in order to meet the criteria.

To test the hypothesis of β-helical modeling for PrP^{Sc} and determine whether this PrP fragment can adopt the β-helical fold in a structural context that fafvors this fold, we synthesized the fragment of the pseudo prion gene from the threading result, and incorporated it into the *lpxA* gene to yield LpxA-PrP recombinant protein (Figure 4.1a & b). Several residues were changed from the wild type mouse PrP sequence to the LpxA sequence due to their involvement at the trimeric interface (F29 and H47 at T_2) and structural restriction (G39 at B^{o}_{4}, T42 at B^{i}_{1}, T54 at B^{i}_{3}, and I56 at B^{i}_{3}). Although these residues were observed at the corresponding positions at other rungs or in other LβH proteins, they were not compatible with the selected target region of LpxA based on our preliminary studies (data not shown). The synthesized pseudo prion gene was subcloned into the corresponding region of *lpxA* in pJC2 plasmid, and the resulting construct was then transformed into SM101 (DE3) and tested for LpxA activity. The mutant LpxA-V48R+T54R was constructed and used as a negative control.

It contains arginine mutations in the region where the PrP sequence was inserted to ensure that the PrP inserted region is sensitive to the *in-vivo* folding assay. As shown in Figure 4.1c, the LpxA-PrP recombinant protein showed partial activity that was significantly lower than the wild type based on bacterial growth under a range of novobiocin concentrations. However, the recombinant protein showed greater LpxA activity than the negative controls and comparable activity to the PtoA-4 mutant. These results suggest that folding of the LpxA-PrP recombinant protein is not as favorable as wild type LpxA, but that β-helical folding may still be possible. Overexpression of LpxA-PrP under IPTG induced expression conditions resulted in insoluble aggregation. Therefore, efforts to obtain enough LpxA-PrP for biophysical studies have not been successful thus far.

(a)

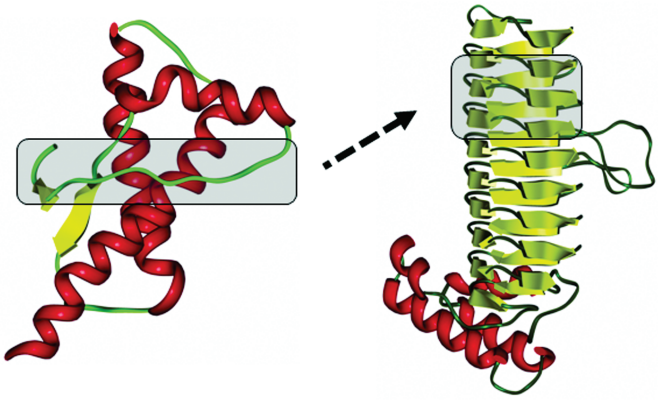
Threading of mouse PrP 104-143 on LPXA

```
LPXA 16-33 (RUNG 2)   EGASIG ANAHIG PFCIVG
MoPrP 104-115                PKTNLK HVAGAA
LPXA-PrP                     PKTNLK HfAGAA

LPXA 34-51 (RUNG 3)   PHVEIG EGTVLK SHVVVN
MoPrP 116-135         AAGAVV GGLGGY MLGSAM SR
LPXA-PrP              AAGAVg GGtGGY MhGSAM

LPXA 52-69 (RUNG 4)   GHTKIG RDNEIY QFASIG
MoPrP 136-143         PMIHFG ND
LPXA-PrP              PMtHiG ND
```

(b)



(c)

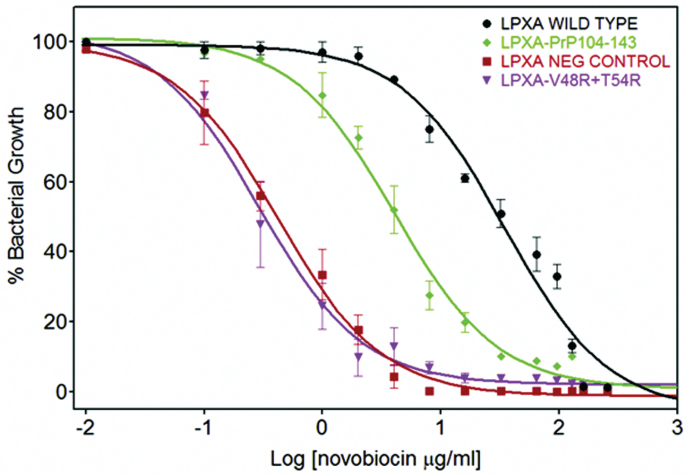


Figure 4.1. Threading the PrP sequence onto the LpxA protein and *in-vivo* folding assay of the LpxA-PrP recombinant protein. (a) Threading of mouse PrP (residues 104-143) onto LpxA (residues 22-59). LpxA-PrP is a synthesized pseudo-PrP gene sequence. Mutations are indicated by lower-case letters. Inserted loop of PrP (PrP residues S134-R135) are indicated in italics. (b) Native PrP structure (PDB ID: 1QM0, residue 125-228). The PrP region (residue 124-143) used in the LpxA threading is indicated by the shaded rectangle. LpxA monomer structure (PDB ID: 1LXA) with the target region for PrP fragment insertion in the β-helical domain is indicated by the shaded rectangle. (c) LpxA *in-vivo* folding assay of LpxA-PrP (LD_{50} = 4.1 ± 1.3 μg/ml) and the control mutant, LpxA-V48R+T54R (LD_{50} = 0.29 ± 1.2 μg/ml) at 30°C.

Discussion

LpxA-PrP hybrid recombinant protein supports the β-helical model of PrP^{Sc}

While it is still unclear if there is a general structural theme that defines amyloid fibril structure, increasing evidence indicates that the mammalian prion protein may adopt β-helical architecture in the misfolded state during amyloid formation (Govaerts et al. 2004; Yang et al. 2005; Langedijk et al. 2006; Wasmer et al. 2008). The inherent compactness and availability of accessible β-faces to initiate and sustain self-assembly make β-helical folds good candidates for the structural unit of the prion amyloid fibril. Previous modeling studies of PrP^{Sc} showed that the protein could adopt two to four rungs of the β-helical fold (Govaerts et al. 2004; Yang et al. 2005; Langedijk et al. 2006). However, questions have been raised on the validity of these β-helical models of PrP^{Sc} because PrP contains sequence features that do not match sequences of known β-helical proteins. While the insolubility of PrP^{Sc} has limited structural studies to test the models by classical techniques, such as x-ray diffraction and NMR spectroscopy, the short fragment known to be prone to amyloid formation

can be threaded and experimentally tested using a recombinant LβH protein containing the PrP amyloid-prone fragment. This recombinant protein was used to investigate the possibility of β-helical structure in the misfolded prion protein conformation.

Although results from the *in-vivo* functional assay showed that the LpxA-PrP recombinant protein had LpxA activity comparable to other LpxA mutants, it may not provide sufficient evidence to conclude that the PrP fragment misfolds into a β-helix. First, the synthesized pseudo-prion gene contains several residues that are different from the wild type prion gene. Wild type LpxA residues were maintained for trimeric interface residues, a conserved glycine residue at B^{o}_{4} position, and several interior residues. Although the PrP residues at these positions are thought to be tolerated in the LβH fold based on the current and previous studies, they seem to be critical in folding of the LpxA protein based on our preliminary studies. Residues involved in the trimeric interaction may not be critical to β-helical folding, but they may destabilize the formation of LpxA trimer, leading to a false result for the *in-vivo* folding assay. The B^{o}_{4} position can accommodate a range of different residues, but it is often highly conserved as a glycine residue due to a steric hindrance. Isoleucine and phenylalanine have been

observed at interior positions B^i_1 and B^i_3, however, the packing of these residues with other wild type residues in the targeted region of LpxA β-helical domain may not be compatible. Interpretation of the *in-vivo* folding assay results for LpxA-PrP is still somewhat problematic as the recombinant protein did not exhibit wild-type like LpxA activity. LpxA activity was completely lost in the arginine negative control, which is expected to disrupt the same region of the LpxA β-helix as where the PrP fragment was incorporated. This result suggests that LpxA-PrP β-helical folding may occur but in a less stable form than in wild type LpxA. Overexpression of the LpxA-PrP recombinant protein was attempted under several conditions, but insoluble aggregates are often formed under the induced conditions, hampering biophysical characterization of the LpxA-PrP recombinant protein. Further studies are necessary to explore expression conditions that will allow examination of the structure and biophysical characteristics of the LpxA-PrP recombinant protein, and may lead to a better understanding of PrP structure.

We attempted to craft a PrP sequence that becomes relatively insoluble when it adopts a β-rich structure onto a sequence that is water soluble when it forms a trimeric LβH fold. However, even if PrP naturally

adopts β-helical architecture in the misfolded state, the exact structure could deviate from known β-helical proteins, such as fungal HET-s prion protein or could disrupt the solubility of the chimeric construct. Therefore, it may be difficult to directly test the full-length β-helical model of PrP^{Sc} on known soluble β-helical proteins that facilitate spectroscopic characterization. However, despite the discrepancies between the actual PrP gene and the synthesized pseudo-prion gene, the LpxA-PrP recombinant protein, containing the PrP fragment can still provide important experimental evidence that PrP may adopt β-helical structures.

References

Choi, J.H., Govaerts, C., May, B.C., and Cohen, F.E. 2008. Analysis of the sequence and structural features of the left-handed beta-helical fold. *Proteins* **73:** 150-160.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. 2004. Evidence for assembly of prions with left-handed beta-helices into trimers. *Proc Natl Acad Sci U S A* **101:** 8342-8347.

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319:** 1523-1526.

Yang, S., Levine, H., Onuchic, J.N., and Cox, D.L. 2005. Structure of infectious prions: stabilization by domain swapping. *FASEB J* **19:** 1778-1782.

Chapter 5

Molecular modeling of the misfolded insulin monomeric subunit

Introduction

Protein misfolding under conditions that destabilize the native state and subsequent fibrillar aggregation of the aberrant conformers are involved in the pathogenic processes of many neurodegenerative and systemic diseases, such as Alzheimer's, the prion diseases, and type II diabetes (Prusiner et al. 1998; Dobson 2001; 2003). Misfolded fibrillar proteins share common structural characteristics even when the native proteins are evolutionarily or structurally unrelated. This observation suggests that the tendency of misfolded proteins to form ordered aggregates, known as amyloids, is a generic property of polypeptide chains (Sunde et al. 1997; Dobson 2003; Serpell et al. 2007). Amyloid fibrils are generally unbranched, protease-resistant filaments with dominant β-sheet structures organized in a cross-β fashion, in which the β-strands run perpendicular to the fibril axis (Serpell and Smith 2000; Jimenez et al. 2002; Murali and Jayakumar 2005; Serpell et al. 2007). Electron microscopy (EM), cryo-EM, and atomic force microscopy (AFM) have revealed that the protofilament units are compact, β-strand repeats with typical diameters ranging 15 to 40 Å (Jimenez et al. 1999; Jimenez et al. 2002; Khurana et al. 2003). Although amyloid fibrils share common structural properties, their overall morphology can vary

depending on the precursor protein, conditions of fibrillization, and the number and arrangement of protofilaments (Goldsbury et al. 1997; Jimenez et al. 1999; Bouchard et al. 2000; Jimenez et al. 2002).

Over the last few decades a great deal of effort has been invested in trying to gain a structural understanding of the mechanisms of protein misfolding and subsequent amyloid fibril formation. For example, recent X-ray crystallography studies of microcrystals formed from the short peptides of amyloid-related proteins have revealed a distinctive packing in the β-structure core (Nelson et al. 2005; Sawaya et al. 2007; Wiltzius et al. 2008). EM and solid-state nuclear magnetic resonance (NMR) studies of $A\beta_{1\text{-}40}$ and the fungal HET-s prion protein showed that the misfolded amyloid conformation may adopt an architecture that is structurally unrelated to the native conformations, but similar to β-structures previously observed in nature, namely β-solenoid folds (Jenkins and Pickersgill 2001; Petkova et al. 2002; Luhrs et al. 2005; Hennetin et al. 2006; Kajava and Steven 2006; Sachse et al. 2008; Wasmer et al. 2008). The limitation to studies of misfolded proteins by traditional biophysical methods has led to the use of molecular modeling to probe plausible structural solutions.

Insulin, a small hormone protein consisting of two polypeptide chains, adopts an α-helical conformation in its native state. The insulin sequence is well conserved among mammalian species, with few variations. Two polypeptide chains are linked by two interchain and one intrachain disulfide bridges (Blundell et al. 1971; Brange et al. 1997a). Although insulin does not appear to be directly involved with any known human amyloid diseases, native insulin does readily convert to an inactive fibrillar form under a wide range of conditions (Brange et al. 1997a). In one clinical study, amyloid fibril–like deposits containing insulin were found at sites of insulin injections in a diabetic patient (Dische et al. 1988). An interesting feature of insulin is that its three disulfide bridges are retained in the in vitro and ex vivo fibrillar form (Brange et al. 1997a; Jimenez et al. 2002; Devlin et al. 2006). Thus, these disulfide bonds must constrain the possible conformational rearrangements during the α-helix to β-sheet transition (Devlin et al. 2006). This conformational constraint makes insulin a unique model system for studying protein misfolding and subsequent amyloid fibrillization.

Native insulin exists as a zinc-coordinated trimer of dimers under physiological conditions or alternatively as a dimer or tetramer in a zinc-free environment (Bryant et al. 1993; Brange et al. 1997a; Devlin et al. 2006).

Under conditions of low pH and high temperature, insulin readily dissociates into monomers, and subsequently assembles into fibrils or aggregates (Brange et al. 1997a; Jimenez et al. 2002). Insulin fibril formation was first observed in the 1940s, with the nucleation and elongation process first being reported by Waugh (Waugh 1946; Waugh et al. 1950; Koltun et al. 1954). In 1972, a fiber diffraction study by Burke et al. showed the cross-β structural pattern of insulin (Burke and Rougvie 1972). Studies using NMR have demonstrated that the native insulin structure is disrupted prior to fibrillization (Hua and Weiss 2004). Recent mass spectrometry, AFM, hydrogen exchange and small-angle X-ray scattering (SAXS) studies have shown the formation of small oligomeric species consisting of up to 12 insulin molecules. These species act as an elongating unit, that further assemble into larger irreversible aggregates and ultimately into protofilaments and mature fibrils (Nettleton et al. 2000; Ahmad et al. 2005; Jansen et al. 2005; Vestergaard et al. 2007). Studies using Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopy indicate that the initial aggregates retain their predominantly α-helical structure, but that there is a subsequent conversion to a β-sheet structure as the fibrillization process continues (Bouchard et al. 2000).

In the present study, we modeled a β-rich insulin monomer using either a β-roll or β-helix, and assessed the suitability of each of these models as possible subunits of fibrillar insulin. We examined each model based on available biophysical data, and studied the relative stability of the model structures using molecular dynamics (MD) simulations. Fibrillar insulin models were constructed by using each of the subunits and fiber diffraction patterns were simulated. By comparing the simulated fiber diffraction patterns with those obtained experimentally from insulin fibers, we were able to assess the suitability of each model.

Molecular modeling of the monomeric subunit of the insulin fibril

In modeling the insulin monomeric subunit, we threaded (Bowie et al. 1991) the insulin sequence onto existing β-solenoid folds found in the PDB database (Figure 5.1). As mentioned above, one of the unique features of insulin is its three disulfide bridges within chains A and B, all of which are retained in the fibrillar state (Brange et al. 1997a; Devlin et al. 2006). Previous mutagenesis and deletion experiments have shown that the C-terminal region of chain B (residues 23-30) is not required for insulin fibrillization (Brange et al. 1997a; Brange et al. 1997b). With these structural constraints in mind, the threading of insulin chains A and B onto a β-roll or β-helix produced relatively simple and unique solutions. Few threading solutions could satisfy the disulfide bridge constraints, structural features of the β-solenoid folds, and the available experimental data, including mutagenesis, isomer and single chain experiment, FTIR, CD, and cryo-EM (Brange et al. 1997a; Bouchard et al. 2000; Jimenez et al. 2002; Hong and Fink 2005; Devlin et al. 2006; Hong et al. 2006; Huang et al. 2006; Choi et al. 2008). In the β-helix model, the insulin sequence was threaded into two rungs of β-helix with three short β-sheets per rung, forming a

triangular shape with an opening at one corner. While chain B of the β-helix model contains typical β-helical turns as observed in nature, chain A contains an unusual four-residue turn with a disulfide bridge, connecting the adjacent β-sheets. In the β-roll model, insulin was also threaded into two rungs, with each rung comprising one short and one long β-sheet connected by tight turns and a disulfide bridge. A comparison of the cross-sectional views of the native (Figure 5.1A) and β-solenoid insulin model structures (Figure 5.1B & C) showed that a structural transition from the α-helical to the β-solenoid structure without complete unfolding and refolding would be possible, with the disulfide bridges intact. Although the types of disulfide bridges introduced into the β-solenoid models are not commonly found in β-solenoid proteins, they are structurally feasible. However, introducing disulfide bridges at the turn regions of the insulin β-helix model resulted in slight distortions to the β-sheets.

Both insulin models satisfied the sequence and structural features of the β-roll and β-helix folds, such that residues facing the inner core of the structures were mostly similar hydrophobic or small polar residues that stacked on top of each other (Hennetin et al. 2006; Kajava and Steven 2006; Choi et al. 2008). In particular, the inner-core residues of the β-helix model

were well packed, with similar residues stacking in two chains. The inner-core packing of the β-roll model was less consistent than that of the β-helix model, with glutamine and glutamate residues residing in the core; however, these residues are occasionally found in native β-roll proteins. In addition to structural constraints, the β-solenoid insulin models may also explain the role of the C-terminal region of the B chain in fibrillization. In both models, the C-terminal region (B chain: residues 23–30) did not fit onto the β-solenoid structures due to side-chain constraints. It therefore made logical sense to place this stretch of sequence outside of the β-solenoid core.

While the β-helix model was relatively symmetrical with dimensions of 20-30 Å × 30-40 Å, the β-roll model had a long rectangular shape with a width of 15-25 Å and a length of 35-45 Å (Figures 5.1B and C). Although the C-terminal region of chain B is not shown in the figures and was not used to calculate the dimensions, the size estimates were in fair agreement with the approximate dimensions of the insulin protofilament. For example, a cryo-EM study showed that the protofibrillar insulin was composed of relatively flat β-sheets with typical sizes ranging from 30-40 Å (Jimenez et al. 2002). Based solely on dimensional requirements the β-helix model was the better of the two β-solenoid models.

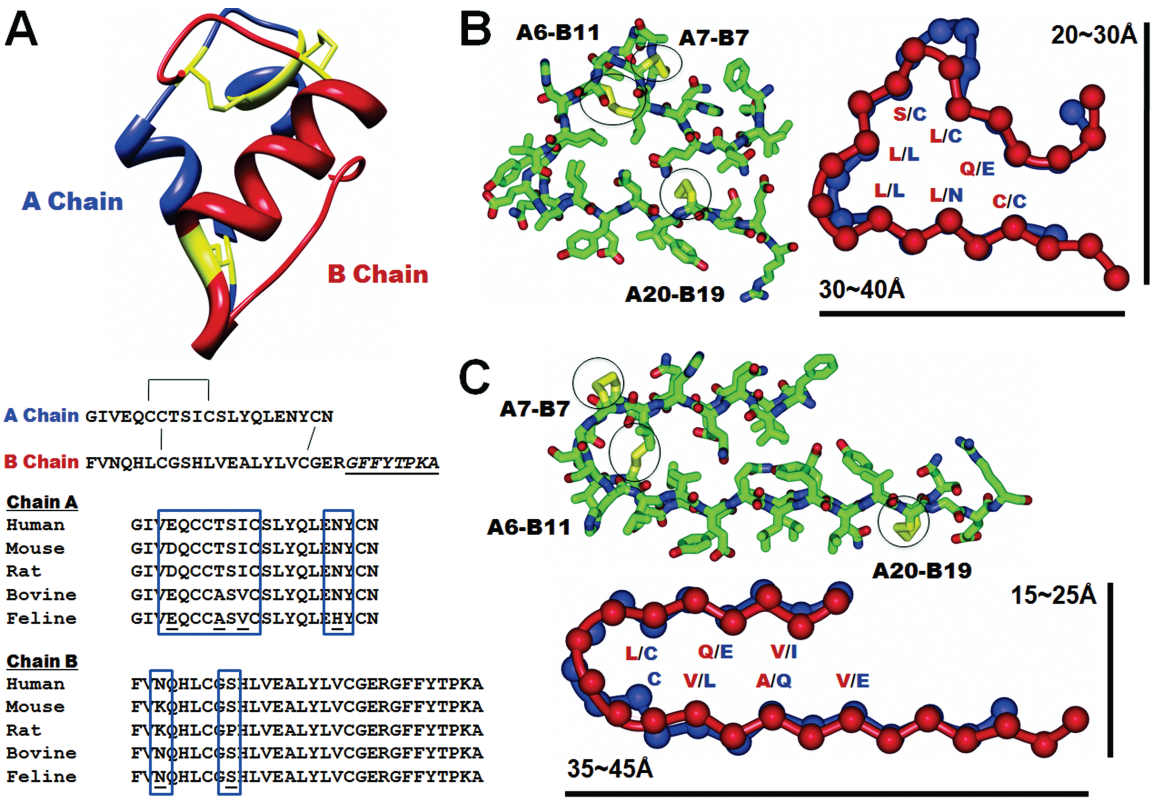


Figure 5.1. Insulin monomer structure and β-solenoid–based monomeric subunit models for the insulin amyloid fibril. (**A**) Insulin monomer structure (PDB ID: 1GUJ) with sequences of chains A and B. One intrachain and two interchain disulfide bridges are shown with yellow stick figures in the structure and by connecting lines in the sequences. The C-terminal region of chain B, which is not involved in amyloid fibrillization, is in italic font and underlined. Multiple sequence alignment of the insulin sequences from five

different mammalian species is shown, with sequence variations shown in the boxes. Cross-sectional view of insulin β-solenoid monomeric models, including (**B**) the C-terminally truncated β-helix insulin monomeric subunit model based on a template from 1HV9 (segment 293-323), and (**C**) the C-terminally truncated β-roll insulin monomeric subunit model based on a template from 1VH4 (segment 230-274). Stick figures are shown with the side-chains, and disulfide bridges are depicted in yellow. Schematic figures with the amino acid residues facing inside are shown with approximate dimensions of the structures. Chains A and B are shown in blue and red, respectively.

Structural stability of the β-solenoid models of fibrillar insulin subunit

We conducted a 10-ns MD simulation using the β-roll and β-helical insulin monomers with intact disulfide bridges under acidic conditions, using a range of temperatures (300-375 K) to emulate the experimental conditions of *in vitro* insulin fibrillization (see Material and Methods). Analysis of the positional root-mean-squared deviations (RMSDs) of the backbone relative to the starting structures indicated that each model reached equilibrium after ~ 4 ns (Figure 5.2). The relative stabilities of the model structures estimated by RMSD revealed no identifiable differences among the β-roll models at the three different temperatures. However, the β-helix models showed an increase in RMSD of ~ 1.5 Å when the temperature was increased by 30 to 45 K. This result shows that the β-roll model is relatively more stable than the β-helix model during temperature increase. The final structures after a 10-ns MD simulation from each model system also revealed that, while the β-roll models maintained their initial fold architecture relatively well, the β-helix model displayed an early stage

disruption of the β-helical fold at 345 K and a structural transition from a β-helical to a β-roll–like structure at 375 K.

The average RMSDs for C_α and for all atoms, changes in secondary structure contents, and interchain backbone-backbone hydrogen bonds were calculated from five independent trials for both β-solenoid models (Table 5). While the interchain backbone-backbone hydrogen bonds before and after 10-ns simulation showed little differences between the β-roll and β-helix models, the average RMSDs of C_α and for all atoms of the two models consistently showed significantly different stabilities at 375 K. The average change in the secondary structural content, as calculated by the DSSP algorithm, showed a gain in coil structure, with a relatively large standard deviation for the separate trials of the β-helix model compared to the β-roll model.

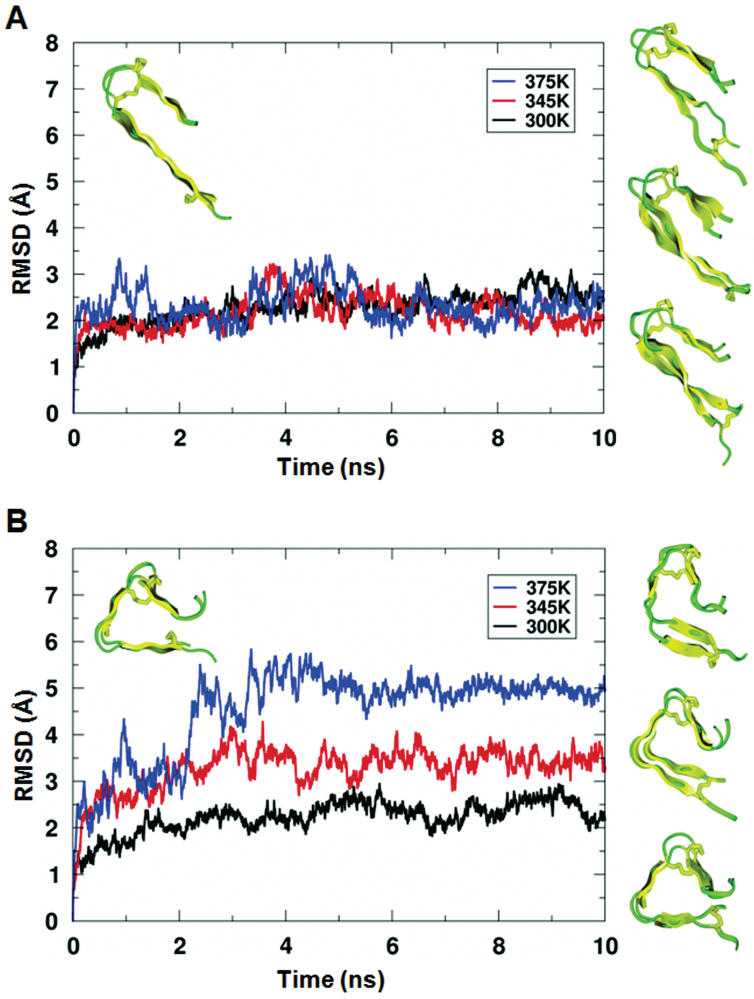


Figure 5.2. Molecular dynamics simulation of the β-roll and β-helix insulin monomeric models with increasing temperature in acidic conditions. Backbone RMSDs of the β-roll (**A**) and β-helix (**B**) models relative to their initial structures as a function of simulation time at 300 K (black), 345 K (red), and 375 K (blue). The initial structure is shown at the top left, and the final structures are shown on the right panel in the order of increasing temperature from bottom to top.

Table 5. Structural statistics from the MD simulation of the β-roll and β-helix insulin models at 375 K.

A. Average RMSD with respect to the starting structures[a] (Å)

	Backbone (C_α)	Backbone (all atoms)
β-roll model	2.62 (0.38)	3.63 (0.43)
β-helix model	4.73 (0.35)	5.76 (0.26)

B. Average change in secondary structural content[b] (%)

	β-sheet	Turn/Bend	Coil
β-roll model	- 4.0 (6.5)	+ 7.0(2.7)	- 3.0 (7.6)
β-helix model	- 2.0 (11.9)	- 1.5(5.5)	+ 3.5 (9.1)

C. Interchain backbone-backbone hydrogen bonds[c]

	Starting Structure	After 10 ns
β-roll model	19	12.0 (3.5)
β-helix model	13	10.4 (1.5)

Values are reported as mean (S.D.) of five independent MD simulations. [a]Backbone RMSDs were calculated with respect to the initial structures, averaged over the 8-10 ns interval of the MD trajectories. [b]The secondary structural content was calculated using the DSSP algorithm. [c]Backbone-backbone hydrogen bonds between chains were calculated using InsightII software.

The relative stability of a single chain from either the β-roll or β-helix insulin model was also explored using MD simulations in order to investigate the fibrillization of a single insulin chain, as is observed *in vitro* (Hong and Fink 2005; Devlin et al. 2006; Hong et al. 2006). To emulate the experimental conditions of single insulin chain fibrillization, we conducted an MD simulation at 345 K in acidic conditions for 10 ns. The relative stability of chains A and B from a β-roll and a β-helix model were compared using the RMSD of C_α calculated relative to the starting models (Figure 5.3). Although the simulation of isolated chains of the β-roll model showed that they were less stable than the two-chain model due to absent interchain hydrogen bonds and disulfide bridges, they still maintained β-roll architecture and showed improved stability relative to the β-helix model. Additionally, a model of chain A, which consists of one intrachain disulfide bridge, showed improved relative stability compared to a model of chain B, suggesting a possible role for the disulfide bridge in stabilizing the structure during fibrillization. This result is consistent with our previous study on the stability of the β-helical fold (Choi et al. 2008).

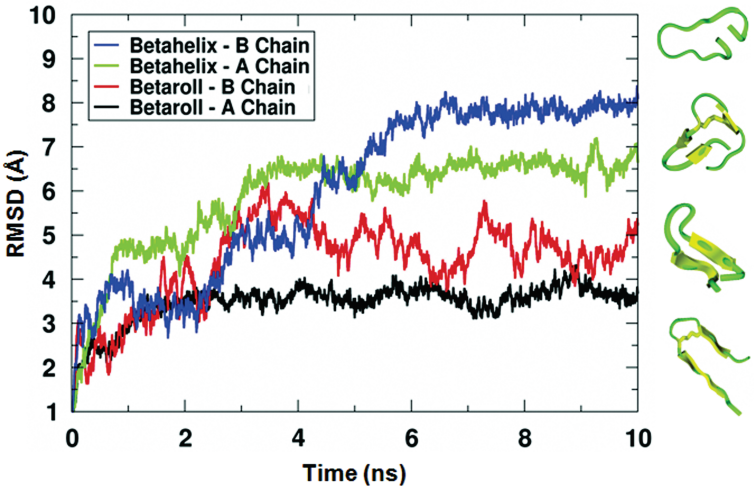


Figure 5.3. Molecular dynamics simulation of single chains from the β-roll and β-helix monomeric models at 345 K under acidic conditions. Backbone RMSDs of isolated chains A and B from the β-roll model and isolated chains A and B from the β-helix model relative to their initial structures as a function of simulation time are shown in black, red, green, and blue, respectively. The final structures are shown on the right panel.

Discussion

β-solenoid models satisfy experimental constraints of misfolded insulin

Insulin is a unique model system with which to model amyloid fibrils, given the conformational restrictions imposed on the protein by its intrinsic disulfide bridges. We have shown here that both the β-roll and β-helix models are possible solutions for the insulin amyloid fibril structure. These β-solenoid structures provide the characteristic cross-β pattern and β-signatures seen in fiber diffraction studies, FTIR, and CD studies of fibril protein aggregates. In addition, previous studies on the fibrillization of insulin isomers and single chain analogues are consistent with our β-solenoid models. Based on the approximate cross-sectional dimension, both the β-roll and β-helix models were comparable to the insulin fibril, but the β-helix model was a better fit to the estimated diameters of the insulin protofilament.

The β-solenoid models of the fibrillar insulin subunit also agreed well with studies characterizing the fibrillization of insulin isomers and single chains (Devlin et al. 2006; Hong et al. 2006; Huang et al. 2006). For

example, a single-chain insulin analogue in which the B chain C-terminus was tethered to the A chain N-terminus was shown to lack fiber-specific α to β transition. This may be explained by the β-solenoid model, where such an insulin sequence may not be easily compatible with the sequence and structural requirements of the fold. Conversely, insulin structural isomers with alternative pairings of disulfide bridges, which were shown to form amyloid fibrils, are compatible with both the β-roll and β-helix models (data not shown). These disulfide isomers retain the A20-B19 pairing, while the A6-A11 and A7-B7 pairings are alternated with one another. Although these alternative disulfide pairings are not ideally suited to the insulin β-solenoid models shown in Figure 5.1, they can be tolerated by using different registries at the turn regions of both models.

Compatibility between the native and isomeric insulin and the β-solenoid models also agrees with cross-seeding experiments using wild-type and isomeric insulin (Huang et al. 2006). Furthermore, the β-solenoid models agree with single-chain insulin fibrillization experiments. One rung of the β-solenoid structure formed by the isolated single chain can further aggregate into homogeneous protofilaments, with morphologies and dimensions different from those of the wild-type insulin fibrils but retaining

the ability to cross-seed with wild-type insulin (Devlin et al. 2006; Hong et al. 2006).

A multiple sequence alignment of mammalian insulins showed that some species contain point mutations, such as histidine at A18 for feline and proline at B9 for rat that may prevent folding to a β-helical structure while allowing folding to a β-roll structure (Figure 5.1). Follow-up experiments on the fibril-forming tendency of insulins from these species or cross-seeding studies with human insulin may provide additional information as to whether the β-roll or β-helix fold is the better modeling solution.

Structural stability of the insulin β-solenoid monomer models

While research as to whether oligomeric species or mature amyloid fibrils act as toxic agents in the pathogenic process is still on-going, increasing evidence indicates that the monomeric units of misfolded amyloid proteins may act as a templating or assembling units for the larger intermediate aggregate species (Collins et al. 2004; Toyama et al. 2007; Vestergaard et al. 2007). Some key features of β-roll and β-helix folds that make them good candidates as monomeric structural units of the insulin

fibril are the inherent compactness of their folds and the availability of accessible β-faces to initiate and sustain self-assembly. Previous studies of the β-helices have shown that the 2-rung unit of the β-helical fold is likely to be the minimal stable unit, and the packing of the hydrophobic core is the essential determinant of global stability (Choi et al. 2008). Additionally, we have previously shown that native disulfide bridges may help in the stabilization or folding of the β-helix fold (Choi et al. 2008). Molecular models of β-solenoid insulin indicate that this protein can adopt two full rungs of β-rolls or β-helix folds and accommodate well-packed side-chains in its core. However, it is difficult to evaluate the absolute thermodynamic stability of the monomeric unit, in order to evaluate its possible role in the nucleation and elongation steps of amyloid fibrillization. Molecular dynamics simulations can be used effectively to evaluate the relative stability of different insulin monomeric models compared to their initial structure under fibrillogenic conditions. Although the actual protein stability may deviate from this relative stability estimation, such simulations can provide insight into how similar models of the same protein would behave under the same fibrillogenic conditions. Based on our simulation of two different β-solenoid insulin models at different temperatures, it is likely that the β-roll model is the more stable structure as a monomeric unit.

Materials and Methods

Molecular modeling of the insulin monomer

The β-roll and β-helix models of C-terminally truncated human insulin (A chain: 1-21 and B chain: 1-22) were built using the iron transporter stabilizer protein SufD (Protein Data Bank [PDB] ID code: 1VH4) and the C-terminal domain of *N*-acetyglucosamine 1-phosphate uridyltransferase GlmU (PDB ID code: 1HV9) as scaffolds. For the β-roll model, the structural coordinates from residues 254-274 and 230-249 of 1VH4 were copied and reassembled onto the scaffold so as to follow the selected threading of the insulin A and B chains, respectively (see Results). For the β-helix model, the coordinates from residues 309-323 and 293-312 of 1HV9 were used for the A and B chains, respectively. The backbone chains were refined by relocating C_α of the cysteine residues (A chain: residues 6, 7, 11, and 20 and B chain: residues 7 and 19) to satisfy the disulfide bond constraints. The side chains of the insulin sequence were subsequently placed using InsightII software (Accelrys 2000). For the β-helical model, the loop region of chain A (residues 7-10) was built using PLOP (Jacobson et al. 2004). The side chain positions of both models were subsequently optimized using SCWRL 3.1(Canutescu et al. 2003). Both

models were optimized using two rounds of energy minimization using the GROMACS 3.1.3 package (Van Der Spoel et al. 2005).

Molecular dynamics simulation of monomeric insulin models

All simulations were performed with the GROMACS software package (Van Der Spoel et al. 2005), using the GROMMOS 43a3 force field(Daura et al. 1998), as described in previous studies (Langedijk et al. 2006; Choi et al. 2008). The model structures used in the simulations were the C-terminally truncated β-roll and β-helix insulin. Models were solvated individually in octahedron boxes filled with water molecules (Eisenberg and McLachlan 1986). A single point charge water model was used for the solvent molecules in the simulation (Berendsen et al. 1981). Sodium ions were used to electroneutralize the system. Solutes, solvent, and counterions were coupled independently to reference temperature baths at 300 K, 345 K, and 375 K (Berendsen et al. 1984), and the pressure was maintained by weakly coupling the system to an external pressure bath at 1 atm (Langedijk et al. 2006). Bond lengths were constrained by the LINCS procedure(Hess et al. 1997) and nonbonded interactions were evaluated using twin-range cut-offs of 0.8 and 1.4 nm for the Lennard-Jones and Coulomb potentials.

Long-range electrostatic interactions beyond the cut-offs were treated with the generalized reaction field model, using a dielectric constant of 54 (Langedijk et al. 2006). The integration time step was set to 0.002 ps and the trajectory coordinates and energies were stored at 0.5 ps intervals. To emulate the acidic, fibrillogenic condition of pH ~2, two histidine residues, one glutamate residue, and the C-terminus were protonated. Analysis was performed using the built-in programs of the GROMACS software package (Van Der Spoel et al. 2005).

References

Accelrys. 2000. InsightII, San Diego, CA.

Ahmad, A., Uversky, V.N., Hong, D., and Fink, A.L. 2005. Early events in the fibrillation of monomeric insulin. *J Biol Chem* **280:** 42669-42675.

Berendsen, H.J., Postma, J.P., van Gunsteren, W.F., and Hermans, J. 1981. *Interaction models for water in relation to protein hydration*. Reidel Publishing Company, Dordrecht, pp. 331-342.

Berendsen, H.J., Postrma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81:** 3684-3690.

Blundell, T.L., Cutfield, J.F., Cutfield, S.M., Dodson, E.J., Dodson, G.G., Hodgkin, D.C., Mercola, D.A., and Vijayan, M. 1971. Atomic positions in rhombohedral 2-zinc insulin crystals. *Nature* **231:** 506-511.

Bouchard, M., Zurdo, J., Nettleton, E.J., Dobson, C.M., and Robinson, C.V. 2000. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* **9:** 1960-1967.

Bowie, J.U., Luthy, R., and Eisenberg, D. 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253:** 164-170.

Brange, J., Andersen, L., Laursen, E.D., Meyn, G., and Rasmussen, E. 1997a. Toward understanding insulin fibrillation. *J Pharm Sci* **86:** 517-525.

Brange, J., Dodson, G.G., Edwards, D.J., Holden, P.H., and Whittingham, J.L. 1997b. A model of insulin fibrils derived from the x-ray crystal structure of a monomeric insulin (despentapeptide insulin). *Proteins* **27:** 507-516.

Bryant, C., Spencer, D.B., Miller, A., Bakaysa, D.L., McCune, K.S., Maple, S.R., Pekar, A.H., and Brems, D.N. 1993. Acid stabilization of insulin. *Biochemistry* **32:** 8075-8082.

Burke, M.J., and Rougvie, M.A. 1972. Cross- protein structures. I. Insulin fibrils. *Biochemistry* **11:** 2435-2439.

Canutescu, A.A., Shelenkov, A.A., and Dunbrack, R.L., Jr. 2003. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* **12:** 2001-2014.

Choi, J.H., Govaerts, C., May, B.C., and Cohen, F.E. 2008. Analysis of the sequence and structural features of the left-handed beta-helical fold. *Proteins* **73:** 150-160.

Collins, S.R., Douglass, A., Vale, R.D., and Weissman, J.S. 2004. Mechanism of prion propagation: amyloid growth occurs by monomer addition. *PLoS Biol* **2:** e321.

Daura, X., Mark, A.E., and van Gunsteren, W.F. 1998. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **19:** 535-547.

Devlin, G.L., Knowles, T.P., Squires, A., McCammon, M.G., Gras, S.L., Nilsson, M.R., Robinson, C.V., Dobson, C.M., and MacPhee, C.E. 2006. The component polypeptide chains of bovine insulin nucleate or inhibit aggregation of the parent protein in a conformation-dependent manner. *J Mol Biol* **360:** 497-509.

Dische, F.E., Wernstedt, C., Westermark, G.T., Westermark, P., Pepys, M.B., Rennie, J.A., Gilbey, S.G., and Watkins, P.J. 1988. Insulin as an amyloid-fibril protein at sites of repeated insulin injections in a diabetic patient. *Diabetologia* **31:** 158-161.

Dobson, C.M. 2001. Protein folding and its links with human disease. *Biochem Soc Symp***:** 1-26.

Dobson, C.M. 2003. Protein folding and misfolding. *Nature* **426:** 884-890.

Eisenberg, D., and McLachlan, A.D. 1986. Solvation energy in protein folding and binding. *Nature* **319:** 199-203.

Goldsbury, C.S., Cooper, G.J., Goldie, K.N., Muller, S.A., Saafi, E.L., Gruijters, W.T., Misur, M.P., Engel, A., Aebi, U., and Kistler, J. 1997. Polymorphic fibrillar assembly of human amylin. *J Struct Biol* **119:** 17-27.

Hennetin, J., Jullian, B., Steven, A.C., and Kajava, A.V. 2006. Standard conformations of beta-arches in beta-solenoid proteins. *J Mol Biol* **358:** 1094-1105.

Hess, B., Bekker, H., Hermans, J., Berendsen, H.J., and Fraaije, J.G.E.M. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18:** 1463-1472.

Hong, D.P., Ahmad, A., and Fink, A.L. 2006. Fibrillation of human insulin A and B chains. *Biochemistry* **45:** 9342-9353.

Hong, D.P., and Fink, A.L. 2005. Independent heterologous fibrillation of insulin and its B-chain peptide. *Biochemistry* **44:** 16701-16709.

Hua, Q.X., and Weiss, M.A. 2004. Mechanism of insulin fibrillation: the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate. *J Biol Chem* **279:** 21449-21460.

Huang, K., Maiti, N.C., Phillips, N.B., Carey, P.R., and Weiss, M.A. 2006. Structure-specific effects of protein topology on cross-beta assembly: studies of insulin fibrillation. *Biochemistry* **45:** 10278-10293.

Jacobson, M.P., Pincus, D.L., Rapp, C.S., Day, T.J., Honig, B., Shaw, D.E., and Friesner, R.A. 2004. A hierarchical approach to all-atom protein loop prediction. *Proteins* **55:** 351-367.

Jansen, R., Dzwolak, W., and Winter, R. 2005. Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. *Biophys J* **88:** 1344-1353.

Jenkins, J., and Pickersgill, R. 2001. The architecture of parallel beta-helices and related folds. *Prog Biophys Mol Biol* **77:** 111-175.

Jimenez, J.L., Guijarro, J.I., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M., and Saibil, H.R. 1999. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *Embo J* **18:** 815-821.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. 2002. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99:** 9196-9201.

Kajava, A.V., and Steven, A.C. 2006. Beta-rolls, beta-helices, and other beta-solenoid proteins. *Adv Protein Chem* **73:** 55-96.

Khurana, R., Ionescu-Zanetti, C., Pope, M., Li, J., Nielson, L., Ramirez-Alvarado, M., Regan, L., Fink, A.L., and Carter, S.A. 2003. A general

model for amyloid fibril assembly based on morphological studies using atomic force microscopy. *Biophys J* **85:** 1135-1144.

Koltun, W.L., Waugh, D.F., and Bear, R.S. 1954. An X-Ray Diffraction Investigation of Selected Types of Insulin Fibrils. *Journal of the American Chemical Society* **76:** 413-417.

Langedijk, J.P., Fuentes, G., Boshuizen, R., and Bonvin, A.M. 2006. Two-rung model of a left-handed beta-helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* **360:** 907-920.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D., and Riek, R. 2005. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc Natl Acad Sci U S A* **102:** 17342-17347.

Murali, J., and Jayakumar, R. 2005. Spectroscopic studies on native and protofibrillar insulin. *J Struct Biol* **150:** 180-189.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. 2005. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435:** 773-778.

Nettleton, E.J., Tito, P., Sunde, M., Bouchard, M., Dobson, C.M., and Robinson, C.V. 2000. Characterization of the oligomeric states of

insulin in self-assembly and amyloid fibril formation by mass spectrometry. *Biophys J* **79:** 1053-1065.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. 2002. A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **99:** 16742-16747.

Prusiner, S.B., Scott, M.R., DeArmond, S.J., and Cohen, F.E. 1998. Prion protein biology. *Cell* **93:** 337-348.

Sachse, C., Fandrich, M., and Grigorieff, N. 2008. Paired beta-sheet structure of an Abeta(1-40) amyloid fibril revealed by electron microscopy. *Proc Natl Acad Sci U S A* **105:** 7462-7466.

Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., et al. 2007. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447:** 453-457.

Serpell, L.C., Benson, M., Liepnieks, J.J., and Fraser, P.E. 2007. Structural analyses of fibrinogen amyloid fibrils. *Amyloid* **14:** 199-203.

Serpell, L.C., and Smith, J.M. 2000. Direct visualisation of the beta-sheet structure of synthetic Alzheimer's amyloid. *J Mol Biol* **299:** 225-231.

Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., and Blake, C.C. 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273:** 729-739.

Toyama, B.H., Kelly, M.J., Gross, J.D., and Weissman, J.S. 2007. The structural basis of yeast prion strain variants. *Nature* **449:** 233-237.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. 2005. GROMACS: fast, flexible, and free. *J Comput Chem* **26:** 1701-1718.

Vestergaard, B., Groenning, M., Roessle, M., Kastrup, J.S., van de Weert, M., Flink, J.M., Frokjaer, S., Gajhede, M., and Svergun, D.I. 2007. A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. *PLoS Biol* **5:** e134.

Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. 2008. Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319:** 1523-1526.

Waugh, D.F. 1946. A Fibrous Modification of Insulin. I. The Heat Precipitate of Insulin. *Journal of the American Chemical Society* **68:** 247-250.

Waugh, D.F., Thompson, R.E., and Weimer, R.J. 1950. Assay of insulin in vitro by fibril elongation and precipitation. *J Biol Chem* **185:** 85-95.

Wiltzius, J.J., Sievers, S.A., Sawaya, M.R., Cascio, D., Popov, D., Riekel, C., and Eisenberg, D. 2008. Atomic structure of the cross-{beta} spine of Islet Amyloid Polypeptide (Amylin). *Protein Sci*.

Chapter 6

Molecular modeling and fiber diffraction analysis of the insulin amyloid fibril

Molecular models of insulin fibrils

We modeled the insulin fibril using both the β-roll and β-helix insulin subunit models as building blocks. 23 different insulin fibril models were constructed with three different architectural schemes. A sample of 14 of these fibril models are shown in Figure 6.1. According to a previous cryo-EM study, two to six protofilaments can be intertwined to make a larger fibril with approximate dimensions of 30 × 40 Å for a single protofilament (Jimenez et al. 2002). Since the dimensions of the β-solenoid models of the insulin subunit ranged from ~20-40 Å in diameter, the parallel stacking of monomer subunits along the fibril axis can fit into the cross-sectional area of a single protofilament. Based on the measured layer line spacings calculated from the diffraction patterns of insulin amyloid fibrils (Jimenez et al. 2002), the protofilament models were built by translating the β-roll and β-helix monomeric subunits along the axis of the fibril and rotating them around the axis. We then constructed fibril models, comprised of two, four, and six β-roll and β-helix protofilaments, by generating symmetrical protofilaments (Figs. 6.1A & B).

A previous modeling study based on SAXS data suggested that the helical oligomeric species intertwine into a dense protofilament, and that the

two protofilaments twist around each other to form a protofibril (Vestergaard et al. 2007). Based on the estimated dimensions of insulin oligomeric species observed by SAXS and the tentative model of the insulin fibril proposed in these studies, we constructed five symmetric models using trimeric, tetrameric, pentameric, hexameric, and octameric arrangements of a β-helix insulin subunit as putative insulin protofilament models. The β-roll subunit model was similarly examined, but failed to provide reasonable solutions for the insulin protofilament (data not shown). In contrast to the proposed model by Vestergaard et al.[31], which has eight-fold symmetry formed by eight helical oligomers intertwined to form a protofilament of an average diameter of ~ 50 Å, the diameters of the β-helix protofilament models ranged from 50 Å for a trimeric arrangement to 85 Å for an octameric arrangement. Considering the dimensions of the insulin protofilament calculated from cryo-EM (30-40 Å), the only sensible model was the three-fold symmetric, trimeric model, in which three β-helix–based protofilaments intertwined with one another. A pair of three-fold symmetrical β-helical protofilaments twisted around each other to form a protofibril with a diameter of ~ 100 Å, was consistent with both the cryo-EM and SAXS studies.

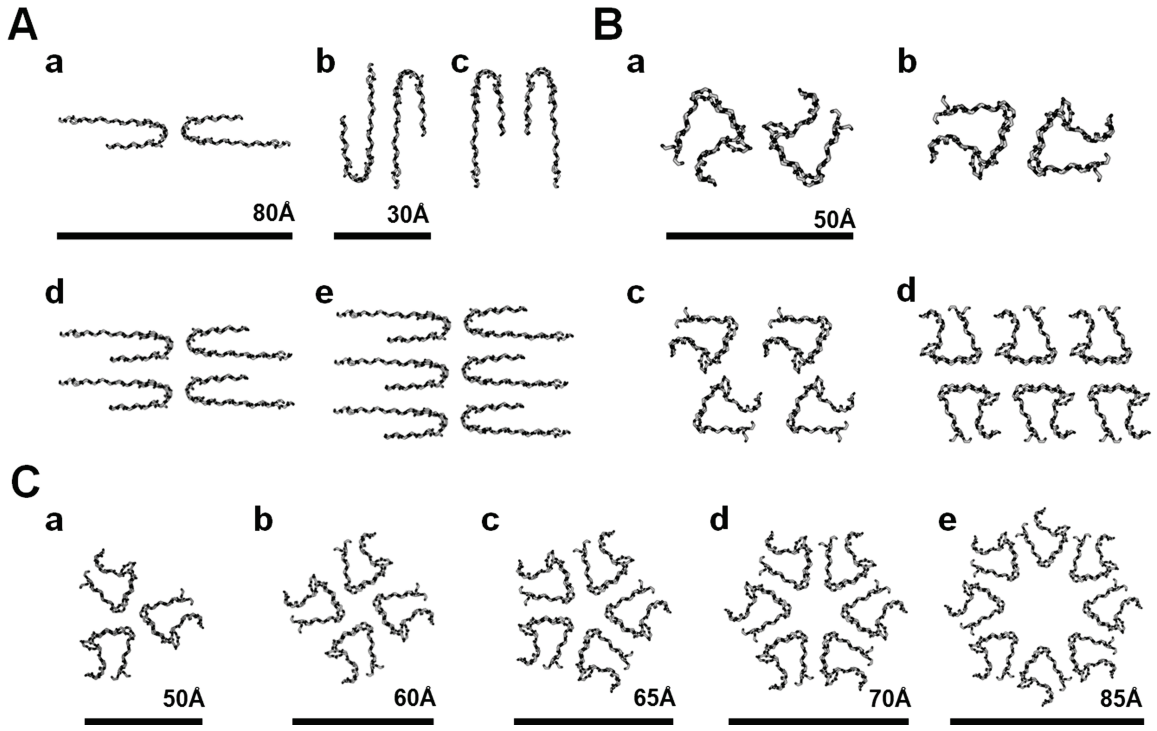


Figure 6.1. Quaternary structure arrangements to simulate insulin amyloid fibrils. (**A**) Cross-sectional view of the β-roll–based insulin fibril models. (a-c) Two-protofilament models, (d) four-protofilament model, and (e) six-protofilament model. (**B**) β-helix-based models. (a-b) Two-protofilament models, (c) four-protofilament model, and (d) six-protofilament model. Fibril models were constructed based on the measured layer line spacings calculated from the diffraction patterns of insulin amyloid fibrils (Jimenez et al. 2002). (**C**) The multifold-symmetric models of a β-helix–based

protofilament: (a) trimeric, (b) tetrameric, (c) pentameric, (d) hexameric, and (e) octameric models. Models were constructed based on the helical periodicity calculated from the insulin oligomeric species observed by SAXS (Vestergaard et al. 2007).

Simulated vs. experimental fiber diffraction patterns of the insulin fibril

We simulated fiber diffraction patterns using the β-solenoid insulin fibril models. We simulated the diffraction patterns for: 1) β-roll and β-helix fibril models, 2) two, four, and six protofilament arrangements, and 3) multifold symmetric arrangements of β-helix–based models. The simulated diffraction patterns of all the fibril models were constructed and compared against the experimental patterns (Figure 6.2 – 6.5). For comparison, the meridional and equatorial reflection profiles of the simulated diffraction patterns were compared against the experimental diffraction patterns of the insulin amyloid fibril.

The simulated diffraction patterns of all the models showed a strong meridional reflection arc at ~ 4.8 Å. In contrast, the intensity and range of the meridional reflection at ~ 9.6 Å varied between the models. The diffraction patterns of all the β-roll fibril models consistently showed a relatively weak, single peak at the ~ 9.6 Å region. This would be expected from fibril models constructed using the two-rung β-solenoid structure as assembling subunits. However, the β-helix fibril models gave diffraction

patterns with multiple peaks of various intensities at ~ 10 Å (Figure 6.4). This may reflect some irregularity and distortion in the β-sheets caused by introducing disulfide bridges at the turns of the β-helix subunit model. Although it is difficult to simulate the exact location and intensity of the meridional reflections seen in experimental diffraction patterns, the β-roll fibril models showed a reasonable fit to the meridional reflections (Figure 6.2 & 6.3).

For the equatorial reflections, a wide range of variations appeared in the number, intensity, and location of arcs among the different fibril models (Figure 6.3 – 6.5). The experimental diffraction patterns revealed four major equatorial peaks at ~ 33 Å, ~ 15 Å, ~ 10 Å, and ~ 4.8 Å, which are believed to reflect the spacing between the β-sheets along the cross-section of the amyloid fibrils (Figure 6.2B). Thus, we focused on identifying fibril models that would generate comparable reflection arcs at the ~ 33 Å, ~ 15 Å, and ~ 10 Å regions by exploring two-, four-, and six-protofilament arrangements of the β-roll and β-helix models (Figure 6.3 & 6.4). A model fibril composed of six-protofilament β-roll model subunits (Figure 6.1Ae) generated the most comparable triplet of the equatorial reflection arcs (Figure 6.2). Other fibril models failed to meet these requirements.

We also simulated and searched the fiber diffraction patterns from the multifold symmetric arrangements of the β-helix fibril models, which were built based on the previous SAXS study data, in order to explore the possibility of alternative arrangements of insulin protofilaments (Figure 6.5). However, none of the simulated diffraction patterns of these β-helix fibril models resulted in comparable reflections at the regions discussed above.

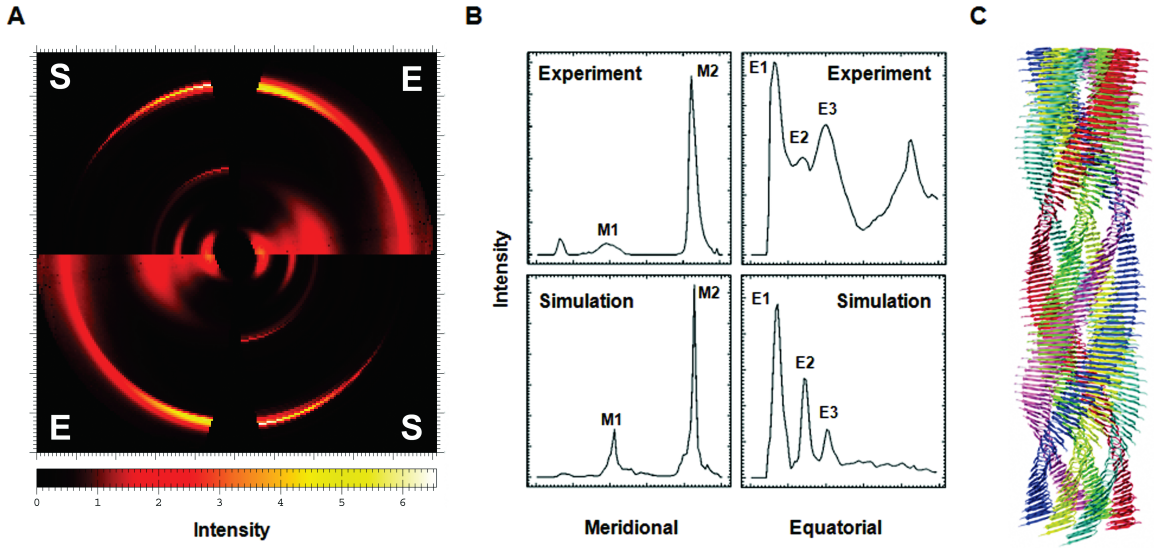


Figure 6.2. Experimental and simulated diffraction patterns of an insulin amyloid fibril and a six-protofilament β-roll–based insulin fibril model. (**A**) Quadrant view of the experimental insulin X-ray fiber diffraction pattern (E) (Vestergaard et al. 2007) and the simulated fiber diffraction patterns of a six-protofilament insulin fibril model based on the β-roll monomeric subunit model as a building block, constructed by DISORDER (S) (Borovinskiy 2006). The image was prepared by FIT2D (Hammersley 1997). (**B**) Meridional and equatorial reflection profiles of the experimental and simulated diffraction patterns. Two meridional reflection arcs at ~ 9.6 Å and ~ 4.8 Å are labeled as M1 and M2, respectively, and three equatorial

reflection arcs at ~ 33 Å, ~ 15 Å, and ~ 10 Å are labeled as E1, E2, and E3, respectively. (**C**) Tentative model of the six-protofilament β-roll–based insulin amyloid fibril. The image was prepared by the UCSF Chimera package (Pettersen et al. 2004).

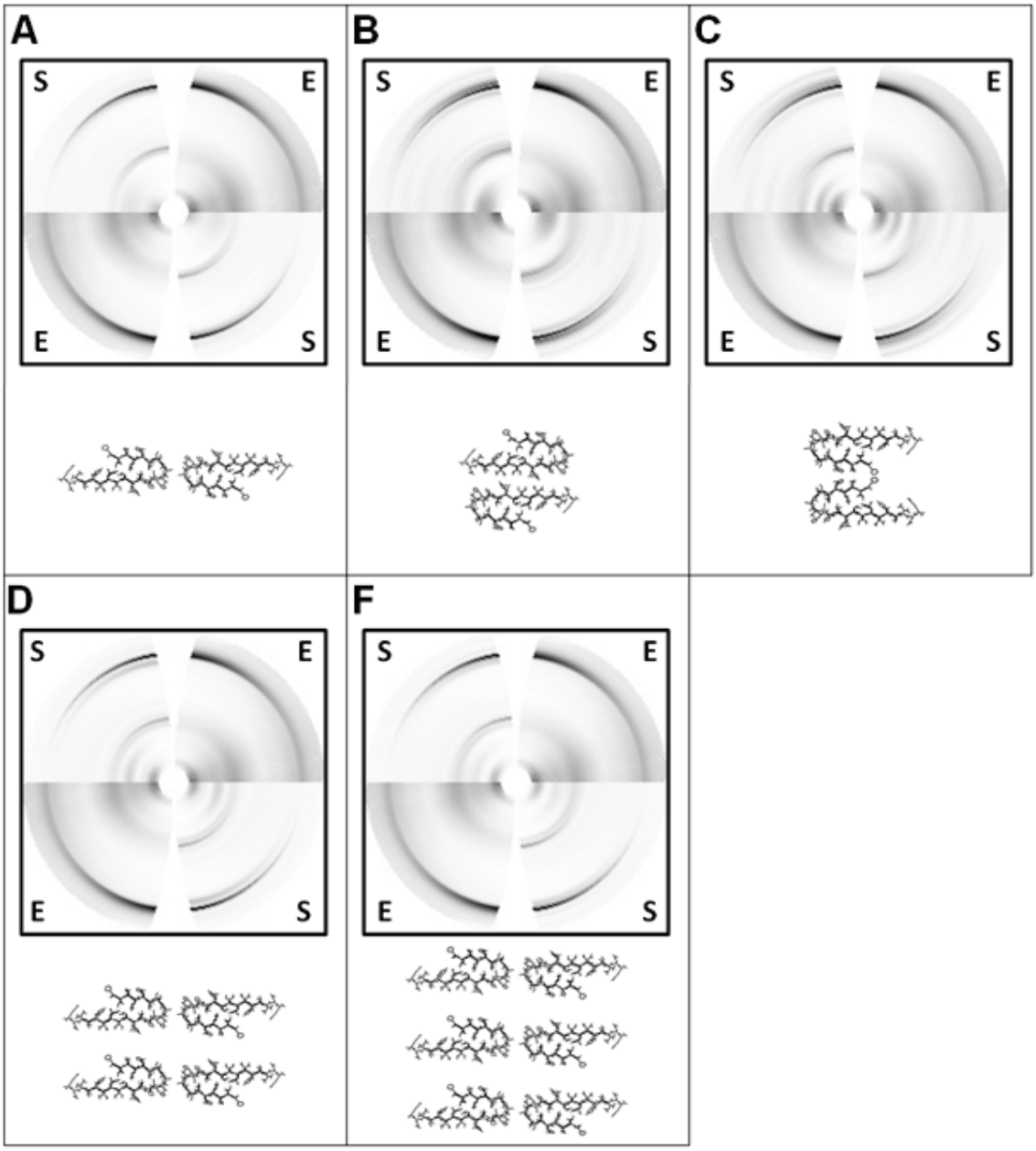


Figure 6.3. Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of β-roll–based insulin fibril models. (**A-F**) Each panel shows the cross-section of a model with a quadrant view of the experimental insulin X-ray fiber diffraction pattern (E) and the simulated fiber diffraction pattern (S).

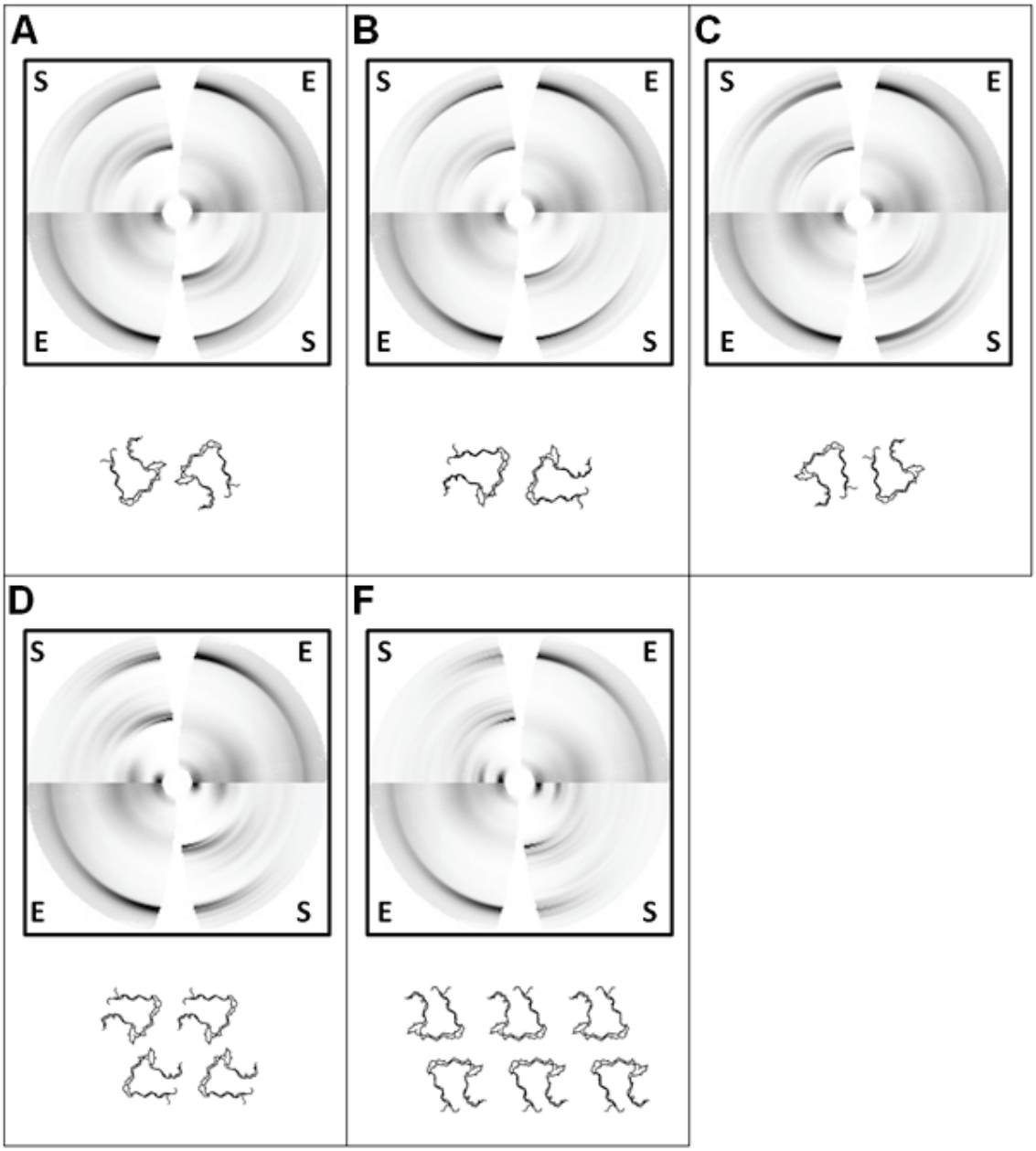


Figure 6.4. Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of β-helix–based insulin fibril models. (**A-F**) Each panel shows the cross-section of a model with a quadrant view of the experimental insulin X-ray fiber diffraction pattern (E) and the simulated fiber diffraction pattern (S).

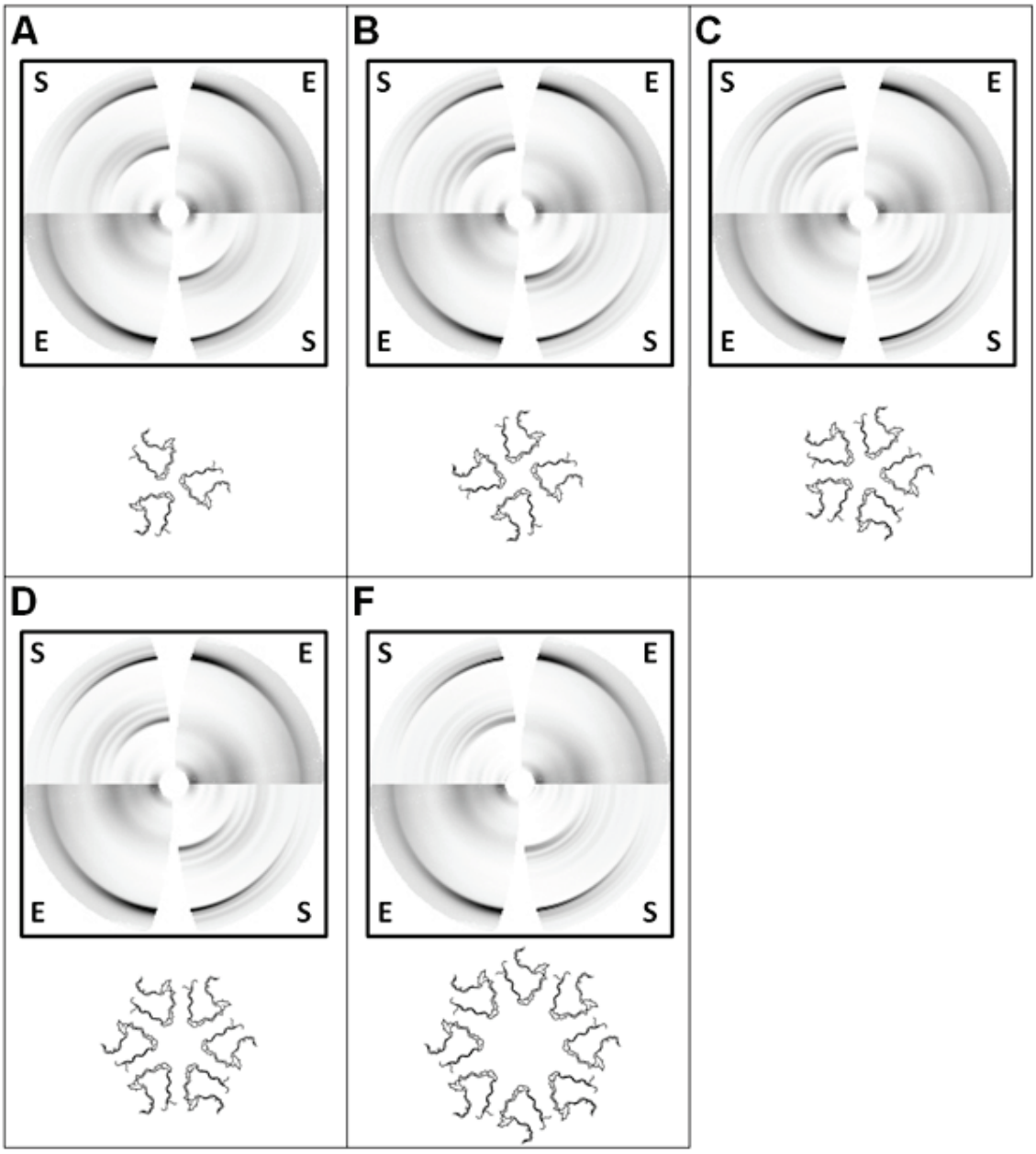


Figure 6.5. Experimental diffraction pattern of an insulin amyloid fibril and simulated diffraction patterns of multifold-symmetric β-helix–based insulin fibril models. (**A-F**) Each panel shows the cross-section of a model with a quadrant view of the experimental insulin X-ray fiber diffraction pattern (E) and the simulated fiber diffraction pattern (S).

Discussion

Molecular models of the insulin fibril and a comparison of simulated and experimental fiber diffraction patterns

While X-ray fiber diffraction studies can provide the characteristic cross-β pattern structure that is a signature of amyloid fibrils, this technique has not been effective in determining the structural arrangement of amyloid fibrils, despite the recent improvements of high resolution fiber diffraction. Diffraction images at best can provide gross structural information, subject to the quality of the fiber under study. For example, subtle variations in the diffraction image often result from differences between sample preparations. One approach to harness the information contained within diffraction images is to start with fibril models and generate theoretical fiber diffraction patterns through computer simulations and compare them to the experimental pattern in order to validate the models. In this way the most likely fibril structure can be identified from many thousands of possible structural solutions. Although the comparison of simulated and experimental fiber diffraction patterns is not able to provide explanations for unknown experimental factors that may affect the fiber diffraction, it can be

useful in studying the relative difference in diffraction patterns generated by similar fibrillar structures.

Insulin fibrils with different morphologies can be formed depending on the arrangement and number of protofilaments involved (Jimenez et al. 2002). In an attempt to acquire a more detailed understanding of the fibril structure and its arrangement, we used simple monomeric β-solenoid models to build complex fibril models by varying the number of protofilaments and their arrangement. Using a computational method developed to simulate the diffraction pattern of disoriented fibrils, we tested whether the diffraction patterns from the constructed β-solenoid fibril models were comparable to the experimental patterns. An exhaustive search for the best-fitting model through the enumeration of different models, arrangements, spacings between protofilaments, and degrees of disorientation resulted in the selection of the six-protofilament β-roll fibril model. The final fibril model is consistent with previous studies showing six intertwining protofilaments or three protofibrils formed by two protofilaments (Jimenez et al. 2002; Vestergaard et al. 2007).

Although the comparison of fiber diffraction patterns showed that the six-protofilament β-roll insulin fibril model agreed well with the

experimental data, there were two major discrepancies. First, a model could not be identified that would generate a meridional arc at ~ 33 Å and an equatorial arc at ~ 4.8 Å (Figure 6.2B). Second, there was a large difference in the intensities of the equatorial arcs. As shown in Figure 6.2A, the experimental diffraction pattern showed relatively stronger and wider peaks, especially at ~ 10 Å and ~ 4.8 Å, than any of our simulated patterns. While differences between the actual insulin fibril and the fibril model are major concerns, these may have resulted from disorientation of the amyloid fibrils due to heterogeneities arising in fibril. The physical nature of amyloid fibril disorientation is still poorly understood and, therefore, difficult to account for in computational simulations. Also, the flexible C-terminus of the B chain, which was not included in the present modeling study, may affect the diffraction signal. Further studies are necessary to probe how these unknown factors can be applied in the search for a better-fitting fibril model, in order to understand the nature of the discrepancies between the simulated and experimental data.

Conclusions

Solid-state NMR, fiber diffraction, EM, and AFM studies of $A\beta_{1-40}$ and the fungal HET-s prion protein have shed light on the structural study of the amyloid fibril. Amyloid fibrils share common structural architectures resembling the β-solenoid fold such as β-rolls and β-helices. However, studies have also suggested morphological diversity among the different amyloid fibrils, which may reflect structural diversity of the monomeric, structural unit. To further understand the detailed structural arrangements of the amyloid fibril, we used several computational tools to model and investigate the insulin monomeric subunit and fibril. We have shown that the β-solenoid is a likely architecture that insulin may adopt in the fibrillogenic state, although further studies are necessary to examine the discrepancies between the theoretical models and experimental data revealed when comparing the fiber diffraction patterns.

Materials and Methods

Molecular modeling of the insulin fibril and fiber diffraction simulations

Molecular models of the fibrils were built by translating the β-roll and β-helix monomeric units along the fibril axis and rotating them around the axis. The degree of the rotation angles varied depending on the helical period of the fibril, which is caused by β-sheet twisting. As a result, two or more protofilaments formed a fibril by winding around each other. All insulin fibril models were constructed using InsightII (Accelrys 2000) and DISORDER (Borovinskiy 2006), based on structural constraints, such as the helical period, number of protofilaments in the fibril, and the relative positions and orientations of the monomers in the fibril cross-section (Jimenez et al. 2002; Vestergaard et al. 2007). The simulation and analysis of fiber diffraction patterns were conducted using FIT2D (Hammersley 1997), FibreFix (Rajkumar et al. 2005), and DISORDER (Borovinskiy 2006). For every fibril model, the diffraction patterns, with varying degrees of disorientation between the fibrils, were simulated. The value of the disorientation parameter varied from 10-25°. The meridional and equatorial profiles of the simulated and experimental fiber diffraction patterns were

then compared. An image of the insulin fiber diffraction pattern image was kindly provided by Dr. Minna Groenning. This image was previously captured using an insulin fiber grown from human insulin (5 mg/ml) in 20% acetic acid (pH 2.0), with 0.5 M sodium chloride at 45 °C. X-ray synchrotron data were collected at 20 °C; λ=1.3 Å, and a 350-nm specimen:detector distance during 30 s exposure time (Vestergaard et al. 2007). The diffraction pattern was preprocessed to subtract the background noise and to build a quadrant view image using the FibreFix software package (Rajkumar et al. 2005).

References

Accelrys. 2000. InsightII, San Diego, CA.

Borovinskiy, A. 2006. DISORDER (http://fibernet.vanderbilt.edu/software/disorder).

Hammersley, A. 1997. FIT2D: An Introduction and Overview. ESRF, Grenoble, France.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. 2002. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99:** 9196-9201.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25:** 1605-1612.

Rajkumar, G., Al-Khayat, H., Eakins, F., He, A., Knupp, C., and Squire, J. 2005. FibreFix - A New Integrated CCP13 Software Package. *Fibre Diffraction Review* **13:** 11-18.

Vestergaard, B., Groenning, M., Roessle, M., Kastrup, J.S., van de Weert, M., Flink, J.M., Frokjaer, S., Gajhede, M., and Svergun, D.I. 2007. A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. *PLoS Biol* **5:** e134.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

12-12-2008

Date