CHARACTERIZATION OF THE EARLY INTERACTIONS ON THE FOLDING PATHWAY OF THE ILEAL LIPID-BINDING PROTEIN BY $^{19}$F-NMR

A Thesis in
Integrative Biosciences

by
Heather Kenney Basehore

© 2007 Heather Kenney Basehore

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

August 2007
Committee page

The thesis of Heather Kenney Basehore has been reviewed and approved* by the following:

Ira J. Ropson  
Associate Professor of Biochemistry and Molecular Biology  
Thesis Adviser  
Chair of Committee

Michael Pishko  
Associate Professor of Chemical Engineering

Cara-Lynne Schengrund  
Professor of Biochemistry and Molecular Biology

Michael F. Verderame  
Associate Professor of Medicine, Associate Dean for Graduate Studies

Peter J. Hudson  
Chair, Integrative Biosciences Program  
Director, Huck Institutes of the Life Sciences

*Signatures are on file in the Graduate School.
Abstract:

Determining whether two structurally related proteins fold via similar mechanisms is an important question in structural biology. Intracellular lipid-binding proteins (iLBP’s) are a large family of proteins that share a common β-barrel structure in spite of very low sequence similarities. One member of this family, intestinal fatty acid-binding protein (IFABP) has been shown to fold by a mechanism of specific hydrophobic collapse in a core structural region. These kinetic folding intermediates differ from intermediates observed for the structurally related protein, ileal lipid-binding protein (ILBP). There is some evidence that both proteins initiate folding by the same mechanism, but that stable folding intermediates do not accumulate to the same level for both. The proposed folding mechanism of IFABP is supported by equilibrium unfolding studies by $^{19}$F-NMR. Residues in contact with each other in the native protein retain elements of structure at denaturant concentrations much higher than the unfolding midpoint concentrations. This is manifested in the $^{19}$F-NMR spectra in the form of NMR spectral changes that differ in response to denaturant depending on the residue’s role in the folding pathway. Similar analysis of ILBP demonstrates that some residues in this protein also display non-uniform behavior by $^{19}$F-NMR with increasing denaturant, although the residues responsible for this behavior were not identified in the initial analysis. By labeling the protein with $^{19}$F one residue at a time, assignment of the peak resonances for each $^{19}$F-Phe in ILBP have now been made. The results of these $^{19}$F-NMR experiments reveal that some of the interactions that occur early in folding may be conserved, but the specific nature of the initiating reactions differs for these two related proteins.
Table of Contents:

List of Figures..............................................................................................................vi
List of Tables...............................................................................................................viii
List of Abbreviations.................................................................................................ix
Acknowledgements....................................................................................................x

Chapter 1: Introduction and Review of the Current Literature.................................1
   A. Introduction: The Protein Folding Problem.......................................................1
      a. Energetics of Protein Folding...............................................................3
      b. The Presence of Intermediates...........................................................5
      c. Hydrophobic Collapse.......................................................................8
   B. Comparing Folding Mechanisms in Structurally Related Proteins.................10
      a. SH3 Domain Proteins........................................................................11
      b. Proteins L and G..............................................................................11
      c. Flavodoxin-like Proteins....................................................................12
      d. Intracellular Lipid-Binding Proteins..................................................13
   C. NMR in the Study of Protein Folding..............................................................14
      a. NMR Methods for Biomolecules.........................................................15
      b. $^{19}$F-NMR of Proteins......................................................................16
      c. Site-Specific Labeling for $^{19}$F-NMR...............................................28
   D. The Intracellular Lipid-Binding Proteins.........................................................30
      a. Model System for Folding Studies.....................................................34
      b. Kinetic and Equilibrium Data for iLBP’s..........................................35
   E. Conclusions......................................................................................................40
Chapter 2: Hypothesis

Chapter 3: Materials and Methods

A. Protein Source and Expression
   a. Wild-type ILBP
   b. Fluorine-labeled Protein
   c. Site-Specific Fluorine Incorporation

B. Purification

C. Mutagenesis

D. Reagents

E. Mass Spectrometry

F. Equilibrium Measurements by Fluorescence

G. NMR

Chapter 4: Results

A. Incorporation of $^{19}$F into ILBP
   a. Verification of $^{19}$F-Incorporation by Mass Spectrometric Analysis
   b. Effects of $^{19}$F on Equilibrium Unfolding by Fluorescence Spectroscopy

B. $^{19}$F-NMR Analysis of $^{19}$F-labeled ILBP
   a. $p$-F-Phe-ILBP
   b. 3-F-Tyr ILBP
   c. Single-Site $^{19}$F Incorporation at ILBP Phenylalanine Residues

C. Chemical Shift Analysis of Collective $^{19}$F-NMR Data
Chapter 5: Discussion

A. $^{19}$F-NMR Equilibrium Unfolding Study of ILBP
   a. Chemical Shift Analysis
   b. Phe2, Phe6, Phe34, Phe79, and Phe94
   c. Phe17
   d. Phe47 and Phe63
   e. Explanation of Chemical Shift
   f. Interpretation of Results and Comparison to Previous Results for IFABP

B. Conclusions
   a. Folding of ILBP
   b. The Role of Intermediates in Folding
   c. Aggregation
   d. Summary

References
List of Figures:

Figure 1.1: NMR Spectra of Conformations in Slow, Intermediate, and Fast Exchange……………………………………………………………20

Figure 1.2: Structure of IFABP with Aromatic Residues Highlighted………………..24

Figure 1.3: Representative Structure of an iLBP with ligand………………………….32

Figure 3.1: Schematic of pRO.hILBP(6His) cloning……………………………………46

Figure 3.2: Representative SDS-PAGE gel………………………………………………49

Figure 4.1: Representative MALDI-TOF Spectra…………………………………………60

Figure 4.2: Unfolding of ILBP and p-F-Phe-ILBP by Fluorescence…………………..62

Figure 4.3: $^{19}$F-NMR Spectra of p-F-Phe-labeled ILBP…………………………………….64

Figure 4.4: Total Relative Peak Area and Individual Peak Shifts for p-F-Phe-ILBP…65

Figure 4.5: $^{19}$F-NMR Spectra for 3-F-Tyr-ILBP………………………………………….67

Figure 4.6: Total Relative Peak Area and Individual Peak Shifts for p-F-Tyr-ILBP…68

Figure 4.7: Native State Spectra for Single-Site p-F-Phe-labeled ILBP Variants……..70

Figure 4.8: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe2…………………73

Figure 4.9: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe6…………………74

Figure 4.10: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe17……………75

Figure 4.11: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe34……………76

Figure 4.12: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe79……………78

Figure 4.13: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe94……………79

Figure 4.14: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe47……………81

Figure 4.15: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe63……………82

Figure 4.16: Chemical Shifts of Single-Labeled ILBP Variants…………………………..84
Figure 5.1: Structure of ILBP ................................................................. 88

Figure 5.2: Structures of IFABP and ILBP ............................................. 100
List of Tables:

Table 3.1: Oligonucleotide Primers for Site-Directed Mutagenesis……………………….54
Table 4.1: Slopes of the Best-Fit Lines for Unfolded Chemical Shifts……………………..85
Table 5.1: Local Primary Structure of Aromatic Residues in ILBP…………………………..94
List of Abbreviations:

$^{19}$F-NMR - fluorine nuclear magnetic resonance

1D – one-dimensional

3D – three-dimensional

3-F-Tyr – 3-fluoro-tyrosine

6-F-Trp – 6-fluoro-tryptophan

CRABP I – cellular retinoic acid-binding protein I

CRBP II – cellular retinol-binding protein II

DNA – deoxyribonucleic acid

E. coli – Escherichia coli

IFABP – intestinal fatty acid-binding protein

ILBP – ileal lipid-binding protein

iLBP – intracellular lipid-binding protein

NMR – nuclear magnetic resonance

OD$_{600}$ – optical density at 600nm

PCR – polymerase chain reaction

*p*-F-Phe – *para*-fluoro-phenylalanine

PheRS - phenylalanyl tRNA synthetase

ppm – parts per million

RNA – ribonucleic acid

SH3 – *src* homology 3

tRNA$^{Phe}$ – phenylalanyl tRNA
Acknowledgements:

The work presented here is the culmination of many years of study, determination, and creative thinking, combined with dedicated work to complete the experiments and interpretations given. No scientist can list all who have assisted in their work along the way, but an attempt to thank those who truly stand out is the least that can be offered in the way of an expression of the immense gratitude that I feel toward those people who have encouraged me along this path. Dr. Ropson and Dr. Dalessio have provided me with not only excellent instruction in the theories and techniques used in this work, but also generously given me a place to work where I was free to think creatively and explore new ideas. This freedom was extended as well by the members of my thesis committee, and has been instrumental in shaping my views on the process of scientific research. I am grateful to the members of the Biochemistry and Molecular Biology department at the medical center, faculty, staff, and students alike, for their endless willingness to discuss projects, provide ideas and reagents, and make me feel welcome and at home in their department.

None of this could have been accomplished without my excellent support system, starting with my parent’s encouragement from a very young age to learn and explore, and extending all the way through the friendships I have been lucky enough to make as a graduate student. The most shining examples of this support system along the way have been my family, Mom, Dad, Rachael, Adam, Philip, and Grandma, and especially my husband Brad. Together with my closest friends, Stephanie, Jackie, Katie, and Kristi, they have provided me with the encouragement and strength to continue and succeed, even at the times when I wasn’t sure I could. I thank all of you sincerely as I continue along the path that you have helped me to forge.
Chapter 1: Introduction and Review of the Current Literature

A. Introduction: The Protein Folding Problem

Proteins are essential to all processes of life. From cell division to neural signaling, the proteins encoded by the genome are the workhorses, catalysts, and directors of virtually every biological event. Despite extensive research on these important molecules, biologists still do not understand how proteins perform such a vast array of functions. The DNA sequence contains the information to encode the protein. DNA is transcribed to RNA, which spells out each amino acid with a specific three-letter codon that is then translated into the protein chain. The mechanism is elegant in its simplicity, but “DNA ⇒ RNA ⇒ protein,” the central dogma of molecular biology, is not nearly the whole story. The next logical step in understanding the processes of life is to understand how the sequence of amino acid residues becomes a functional protein. It is clear that the peptide sequence is sufficient to dictate the three-dimensional structure, and that this 3D structure is the key to the protein’s function. The question is, how is the structure determined by the sequence?

The rules of atomic interactions dominate the field of chemistry. These well-defined rules govern the types of inter-atomic and molecular interactions that are allowed, preferred, or disallowed. Chemists have been carefully analyzing the atomic properties of different elements for hundreds of years. As long as the molecules are rather small, for example, a typical organic molecule of molecular weight 200g/mol, the rules behind their composition, shape, and properties are relatively straightforward. As the molecules become much larger, for example, a small protein with a molecular weight of 10,000 g/mol, there are many more
interactions to consider, and the rules governing the molecular properties are much more difficult to define. Usually, the protein structure is determined by X-ray crystallography or by NMR before any attempt is made to draw conclusions about the interactions that dictate the shape. Much progress has been made in this way, and a discussion of the emergent themes follows in the sections below. However, as a set of underlying rules governing protein folding is established, it becomes more important to determine the physical properties behind the rules (Clarke 2001).

While a protein’s ability to fold into its correct three-dimensional structure is certainly crucial for proper biological function, misfolding is likewise implicated in a number of diseases. The amyloid diseases are most commonly associated with misfolded proteins. These include Alzheimer’s disease, Parkinson’s disease, type II diabetes, and the spongiform encephalopathies. Even more diseases can be attributed to protein folding errors when mutations that alter or decrease the stability of a particular protein are present. Some examples of this include the cystic fibrosis trans-membrane conductance regulator protein, which is mutated in cystic fibrosis patients, and the p53 protein, which is implicated in a number of cancers when mutated (See Dobson 2000, and Raso 2000, for details on the above protein folding diseases). While the role of the misfolded protein may be different in each disease, a better understanding of the misfolding process will lead to a better understanding of the disease and to strategies for treatment, cure, and prevention. Understanding misfolding can only follow better awareness of the mechanisms of correct protein folding. By careful determination of the steps that lead from a linear polypeptide chain to a functional three-dimensional molecule, improvements will be made both in protein engineering and in understanding disease.
a. **Energetics of protein folding**

For several decades after the discovery of the central dogma, protein folding had been assumed to follow a series of steps, with each “step” bringing the polypeptide chain closer to the native structure in a restricted set of events occurring on a defined pathway. Observation of stable folding intermediates seemed to provide evidence for these pathways. The model was based on the Levinthal paradox, the assertion that the amount of time it would take a protein to sample all possible conformations in order to achieve the folded structure is much longer than the age of the universe, on the order of $10^{27}$ years (see Zwanzig 1992). Since protein folding often occurs on a timescale of seconds or less, the folding mechanism cannot simply be a random occurrence, thus a series of defined events made sense. However, improvements in both experimental and theoretical methods have contributed much in the past few decades toward more comprehensive theories about protein folding (Dobson 2000), and schemes more complex than a simple series of steps are more likely to be responsible for the protein fold, as discussed below.

One might suggest that other cellular factors catalyze and stabilize the folding reaction. While there are certainly chaperones and other cellular components that both increase the rate of folding and reject improperly folded proteins, these are not responsible for the folding of the majority of soluble proteins, as many proteins are capable of folding into their native states *in vitro*. Christian Anfinsen, Stanford Moore, and William H. Stein were awarded the Nobel Prize for Chemistry in 1972, for proving that the amino acid sequence is in itself sufficient to dictate the final three-dimensional conformation of a protein (Anfinsen 1973). The work by Anfinsen’s group, in particular, showed that the ribonuclease
protein could regain function after complete denaturation and renaturation. Since the native protein has four disulfide bonds, the ability of the protein to regain function indicated that all four bonds had re-formed correctly, in spite of the fact that there are 105 possible pairings that the eight sulfhydryl groups could have made by random collision. These and other studies eventually lead to the formation of the “thermodynamic hypothesis,” which posits that the three-dimensional structure of a native protein under physiological conditions is the structure in which the free energy of the whole system is lowest. This means that the native conformation is governed by the sum of inter-atomic interactions, and thus the native state is determined by the amino acid sequence (Anfinsen 1973). There must, therefore, be some energetic mechanism that leads from the nascent peptide chain to the correct three-dimensional structure, with physiological buffer conditions and the primary amino acid sequence being sufficient for the reaction to occur.

The current view of the folding process envisions an energetic landscape with minima located in funnel-shaped valleys. Here the unfolded peptide reaches the minimum energy state, i.e., the native state, by searching the landscape for energy minima. Since each lower energy state is located within the funnel, the number of remaining possible conformations is more and more restricted the farther the protein “falls” down the landscape. Native-like contacts within the protein chain are presumed to be more stable than nonnative contacts, and so as native contacts are formed, the availability of other non-native contacts is reduced and the protein is driven to the native three-dimensional structure (reviewed in Jahn 2005). Many unfolded conformations exist at the top of the funnel and collapse to a single native state conformation at the bottom. The drive to reduce the number of unfolded conformations and collapse to the native state defines the funnel shape (Dill 1999).
b. The Presence of Intermediates

Equating the energy landscape of a protein folding reaction to a funnel allows one to picture how, as the disorder of the unfolded chain is reduced, the conformational space accessible to the polypeptide is also reduced (Dobson 2004). For small proteins, this results in a picture of an energy landscape that looks very much like a simple funnel, with only two states, the ensemble of unfolded protein conformations at the top of the funnel and the folded native-state protein at the bottom. Experiments on the folding of larger proteins, however, often reveal more species than simply the unfolded and completely folded states observed in small proteins, suggesting the presence of many metastable minima on the energy landscape. Often these partially folded states represent intermediate species defined by regions of the protein that successfully form native contacts at an earlier time than the rest of the structure. If a protein routinely encounters one or more of these minima upon folding, and accumulates there to such a degree to be detectable, it can be said that this represents an intermediate form of the folding protein. Many proteins smaller than 100 amino acids appear to fold via a two state mechanism, without passing through any detectable intermediate forms on folding. The presence of intermediates along the folding pathway of proteins larger than 100 residues may be explained by the tendency of larger chains to collapse upon themselves in aqueous solution. In order to reorganize the collapsed chain, there may be higher-energy barriers, resulting in the accumulation of intermediates (Jahn 2005). In other cases, the minima may represent cases in which the protein has formed a significant number of non-native contacts, and becomes temporarily trapped in a misfolded conformation (Dobson 2004). Regardless of the origins of these states, these intermediates are of great interest since they represent the
series of events necessary for the protein to fold into its final three-dimensional structure. It is important, however, to be able to determine whether the intermediate state being observed represents early native contacts, a so-called on-path intermediate, or a state trapped in a misfolded form, an off-path intermediate.

The observation of folding intermediates for some but not all proteins has lead many to speculate on the role of the intermediates in the process. Questions have been raised as to whether proteins that appear to fold by a two-state mechanism really encounter intermediates. Folding will appear to be two-state if an intermediate is less stable than the unfolded form, since no kinetic intermediate will be observed if it does not accumulate (Clark 1997, Creighton 1996). The observation that some intermediates appear to be on-path while others appear to be non-productive raises further questions as to their roles. Three reasons a non-productive intermediate might form include incorrect proline isomerization, formation of non-native but stable interactions, or the presence of separate folding pathways (Clarke 1997). Studies on sperm whale myoglobin variants lead to the conclusion that evolutionary selection of residues involved in a folding intermediate in this protein are driven by the need to destabilize the intermediate as much as possible (Isogai 2006). Others have questioned whether formation of intermediates increases the rate of folding or inhibits folding by creating kinetic traps. The fact that small, single domain proteins fold with simple two-state kinetics without forming intermediates provides the basis for the argument that intermediates may actually slow the folding of larger proteins (Jackson 1998). However, a recent analysis of the kinetics of intermediate formation from the unfolded state, and the kinetics of native state formation from the intermediate, as well as the non-local contacts involved, reveals a strong correlation between the three properties (Kamagata 2006). The
implication from these studies is that the two kinetic rates are strongly correlated because they involve organization of the same non-local clusters, meaning that non-local native-like interactions are key to intermediate formation and thus the intermediate is critical for proper folding to the native state.

Fawzi and coworkers (2005) found differences in the rate of protein aggregation to correlate with the presence of folding intermediates. They describe how the presence of an early intermediate in protein G may be protective against aggregate formation, in contrast to the structurally related protein, protein L, which aggregates more rapidly and does not exhibit the early intermediate. They propose that some proteins have evolved to form early folding intermediates as a protective measure against aggregation (Fawzi 2005). The authors further propose that structurally related proteins may have evolved different folding pathways in response to differences in protein interaction partners as a protective mechanism against aggregation. A more detailed discussion of these proteins is included below.

The formation of protein aggregates is of special concern due to the devastating human diseases characterized by the presence of aggregated protein. Alzheimer’s and Parkinson’s diseases, among others, are known as “amyloidoses” due to the presence of amyloid fibrils, aggregated protein super-structures. An interesting characteristic of the amyloid fibrils is that the overall appearance and molecular structure of the protein fibrils is similar in all amyloid diseases, regardless of the protein that serves as the precursor species for amyloid formation. The precursor proteins vary widely among the different diseases and show no sequence or structural similarity in their native forms. Indeed, the ability to form the amyloid type of aggregation may be an inherent property of all protein species, yet the amyloid fibril always has the same cross-β structure and the ability to bind the dye Congo
red. It has been suggested that the presence of off-path folding intermediates, or the accumulation of partially folded proteins, can lead to the formation of aggregates (Chiti 2006, Jahn 2005). Because the fibrils have mainly β-sheet topology, studies of the folding mechanism of β-sheet proteins are particularly relevant to these diseases.

c. Hydrophobic Collapse

One hypothesis describing the initiation of protein folding is that hydrophobic residues collapse on each other to exclude water from their environment, and native contacts emerge from the collapsed structure. A recent study of the Trp-cage miniprotein TC5b revealed that there is residual structure in the unfolded state due to hydrophobic collapse, with long range inter-residue contacts (Mok 2007). Theoretical studies using φ-value analysis of the small α−β protein NTL9 further emphasized the importance of the hydrophobic interaction in folding (Anil 2005). Here the authors studied the φ-value analyses of variants in which a series of mutations had been made at different hydrophobic sites, and determined that increasing hydrophobicity increased the folding rate, without a measurable impact on the stability of this protein. They determined that hydrophobic interactions are a driving force in formation of the folding transition state. Many studies have been done in an attempt to determine whether hydrophobic collapse is non-specific, or whether it involves the early formation of native contacts between specific hydrophobic residues.

The influential review by Kauzmann (1959) forms a foundation for the well-accepted concept that hydrophobic collapse is entropically driven, but the force responsible for folding remains unclear. Kauzmann’s model posits that the large free energy change necessary for a protein to fold could be a combination of the free-energy changes from non-polar side chains
moving from the solvent-exposed surface to the interior of the folded protein, similar to moving a non-polar solute from water to an organic solvent. In 1945, Frank and Evans proposed that water must experience an ordering effect in the presence of a nonpolar molecule, thus clustering of nonpolar residues in a protein would reduce water contact, resulting in a reduction in ordering and an increase in entropy (-ΔS) for the solvent. Although this model was accepted for over thirty years, it was not clear what the contribution of each side-chain is to the free-energy change, nor whether this is sufficient to drive folding (Ben-Naim 2003). Experimental evidence for this theory remains minimal, and recent work has suggested a more complex source for the entropic force than just solvent ordering (Finney 2003). More recent models combine Kauzmann’s with models that include formation of a specific hydrophobic core in the folding reaction (Brylinski 2006). Regardless of the model, there is little dispute that clustering of hydrophobic amino acid side chains provides stabilization for the native fold. The uncertainty lies in whether the hydrophobic collapse occurs before formation of secondary structure, or concurrent with secondary structure formation, and whether the initial hydrophobic contacts in the folding pathway are specific and native-like, or nonspecific (Baldwin 2002).

Monte Carlo simulations have shown that portions of the protein can form secondary structure in the unfolded state to provide a basis for subsequent folding events to occur (Srinivasan 1999). Physical data from intestinal fatty acid-binding protein (IFABP) were included in the analysis, which followed earlier work by Ropson (1992) showing that the tryptophan at position 82 is located at a site that maintains some structure at high denaturant concentrations, suggesting an early folding event at a site containing two phenylalanines and a leucine in addition to the tryptophan. In a study designed to examine the effects of single-
site mutations on the structure of IFABP, it was found that mutation of a hydrophobic residue had a much greater effect on folding than mutation of a solvent-exposed polar residue (Rajabzadeh 2003). Further experiments supporting the hypothesis of hydrophobic collapse in formation of native contacts early in the folding of this protein are described in more detail below. Equilibrium and kinetic studies of goat α-lactalbumin (GLA) variants in which each of the four tryptophan residues were mutated to phenylalanines, revealed that the tryptophan in the hydrophobic core is structured in the intermediate state (Chedad 2005). Studies in lysozyme have shown that there are networks of hydrophobic contacts under denaturing conditions, although these may involve formation of native and non-native contacts (Klein-Seetharaman 2002). Both native and nonnative hydrophobic interactions may play a role in the early stages of folding, and the nature of hydrophobic collapse remains an area of intense research.

B. Comparing Folding Mechanisms in Structurally Related Proteins

Recently, results from computational studies of Apo-Azurin protein demonstrated that the most stable regions of the native structure are those that form interactions early in folding (Chen 2006). The protein folding process can be partitioned into stages, although these stages are not necessarily sequential or temporally separated. Folding occurs by specific or nonspecific chain collapse, formation of secondary and tertiary structure, and desolvation of the chain as it folds to lower energy conformations. These events apparently depend on the particular protein sequence under investigation, which means it is probably useless to try to
determine a ‘universal’ mechanism of folding for all protein sequences, but is more informative to determine whether similar proteins fold by a similar mechanism (Ferguson 2003). A variety of structurally related protein families have been studied in an attempt to answer this question.

a. **SH3 Domain Proteins**

The SH3 domain proteins are approximately 60 residues and fold to the same β-sandwich fold with five β-strands organized in two orthogonal β-sheets. Many of these proteins had been found to fold via a two-state mechanism, with structural conservation in the transition state (Grantcharova 2001). The structurally homologous src and α-spectrin SH3 domains have a sequence identity of only 36% but share the same native fold. By comparing variants of the two proteins, it was determined that the same transition state is structurally conserved (Martinez 1999, Grantcharova 1998). More sensitive experiments have recently lead to the suggestion that the SH3 domain proteins fold through a previously undetected intermediate state, but the intermediate is also suggested to be conserved within members of the family (Korzhnev 2006). NMR studies have further supported the hypothesis that SH3 domain proteins fold by the same mechanism (Bezsonova 2006). Therefore, experimental and theoretical investigation has lead to the conclusion that the family of SH3-domain proteins folds via the same pathway, in spite of low sequence similarity.

b. **Proteins L and G**

IgG-binding proteins L and G have very low sequence identity, yet fold into the same structure containing four β-strands and one a-helix. These small proteins have approximately
60 amino acids, yet little detectable sequence homology (Grantcharova 2001). Protein L folds via a two-state mechanism in which the N-terminal β-hairpin has native-like structure at the transition state, but no stable intermediate is observed (Kim 2000). Protein G has a folding intermediate and the β-hairpin near the C-terminus is structured at the rate-limiting step (Park 1999, Shimada 2002). A theoretical study revealed that sequence-specific factors cause the N-terminal β-hairpin to be more stable in protein L, and the C-terminal hairpin to be more stable in protein G (Islam 2004). Thus, it appears that the presence of a β-turn is required in the folding transition state, but the hairpin that serves this function is different for each protein (McCallister 2000). Although it was thought that the two proteins would fold via the same mechanism due to their structural similarities, it appears that the subtle sequence differences between the two proteins result in different paths to the native state.

c. Flavodoxin-like Proteins

Nine different protein superfamilies share the common flavodoxin-like fold (Brenner 1997). There is very little sequence similarity between members of the family, and the proteins have a variety of different functions, but all share one of the most common fold topologies. The structure is a five-stranded parallel β-sheet with α-helices at either side. Early studies of flavodoxin-like folding identified a stable folding nucleus near the N-terminus that was proposed to be a common feature of these proteins (Steensma 1998). Two separate intermediates were later observed in folding of at least three flavodoxin-like proteins. In a study comparing the nature of the two intermediate states among apoflavodoxin, CheY, and cutinase, it was found that the structure of the off-pathway intermediate is conserved, although the intermediate shows different levels of stability in the three proteins.
These authors also conclude that the on-pathway intermediate appears to be conserved as well. However, theoretical analysis of the three proteins was used to illustrate that alternative pathways exist, which correspond to nucleation occurring from opposite sides of the central $\beta$-strand (Nelson 2006). Others have determined that two flavodoxin-like proteins do not share the same transition state structure (Bueno 2006). Thus the question of whether the structurally-related flavodoxin-like protein fold via the same mechanism remains a subject of discussion.

d. Intracellular Lipid-Binding Proteins

The intracellular lipid-binding protein family has been useful in analyzing the relationship between 3D structure and folding mechanism, since the more than 200 members of the family share a common tertiary structure but very little sequence identity, as described below. Although it was thought that proteins with the same structure would fold via the same mechanisms, one of the first studies to test this hypothesis in a primarily $\beta$-sheet protein family found that this may not be the case. Burns et al studied the folding pathways of three iLBP’s, IFABP, CRABP I (cellular retinoic acid binding protein I), and CRBP II (cellular retinol binding protein II), and determined that, although all folded through at least one kinetic intermediate state, the nature of the intermediates differed. CRABP I has a molten-globule-like kinetic unfolding intermediate, with native-like structure but perturbed tertiary structure, while IFABP and CRBP II have a kinetic unfolding intermediate with little secondary structure (Burns 1998). A later study concluded that rat ILBP also has a molten-globule-like intermediate (Dalessio 2000). These studies support the hypothesis that different sequences can follow different paths to the native state. In a computational analysis of these
four proteins, differences in predicted nucleation centers were found for each protein, supporting the experimental evidence for observed differences in folding (Nikiforovich 2002).

The proteins in the iLBP family are larger than those in the SH3 family, the protein L or protein G proteins, or the flavodoxins. In the studies of these three small protein families, different conclusions have been drawn about the role of the intermediate states and the conservation of the folding pathways between the structurally related proteins. Improved analytical techniques continue to result in the observation of previously unknown states on the folding pathways of all proteins, and results of analyses based on mutations may skew the results. For these reasons, continued analyses of related proteins are necessary to determine whether proteins with very different sequences follow the same pathways or use the same mechanisms to achieve the same 3D structures.

### C. NMR in the Study of Protein Folding

Nuclear magnetic resonance (NMR) is a technique that structural biologists are relying on for detailed information about the structure of biological macromolecules. A testament to the usefulness of this method is the ever-increasing number of chemical and biological scientists incorporating NMR into their research on the 3D structures of biomolecules, without being specifically trained as NMR spectroscopists (Cavenaugh 1996, preface). The use of NMR in determination of 3D protein structures has resulted in an enormous contribution to the known structure database, since crystallization is not required for NMR structure determination. Additionally, NMR remains one of the few methods for
characterizing the dynamic structure of unfolded or partially folded proteins (See Dyson 2004 for review).

**a. NMR Methods for Biomolecules**

Although new developments continue to be made for use of NMR in structural biology, the techniques are all founded on the basic tenants of nuclear magnetic resonance spectroscopy (see Hore 1995). Proton NMR (\(^1\)H-NMR) is very common due to the presence of \(^1\)H in virtually all organic molecules, the high natural abundance of the \(^1\)H isotope, and the large gyromagnetic ratio that results in a stronger NMR signal for protons than for all other relevant nuclei. In spite of these advantages, more advanced techniques are required for protein studies. NMR spectra become increasingly more complicated as the number of observed nuclei in the molecule increases, and 1D-\(^1\)H-NMR cannot be easily interpreted for peptides larger than approximately 20 amino acids. Fortunately, biological enrichment of proteins with other appropriate nuclei, namely \(^13\)C, and \(^15\)N, is straightforward. Heteronuclear NMR experiments can be used to determine structural properties of proteins up to 30kDa or larger (Cavenaugh 1996, Ch7). Incorporation of \(^13\)C or \(^15\)N into recombinant proteins can be accomplished using simple biological manipulation of the organisms used to produce the protein. Enrichment of bacterial growth media with \(^13\)C-glucose and/or \(^15\)N-ammonium salt results in incorporation of the isotope into the nascent proteins. There are many examples in the literature of the use of NMR spectroscopy in analysis of protein structures. Many of these studies are complicated by the need to assign resonances, by the complexity of peaks for multiple, partially-folded protein conformations, and by the amount of time needed to collect complex spectra.
NMR is invaluable in studies of native protein structure, but also remains one of the few methods available for characterization of unfolded or partially folded states (Dyson 1998, Dyson 2004). Partially folded or unfolded proteins cannot be crystallized due to their dynamic nature, but NMR can be used to characterize these structures (Kumar 2004). Equilibrium folding studies using NMR have been very successful (see Dyson 2004 and Kumar 2004 for review). However, assignment of resonances for proteins in partially folded states presents considerable difficulty, the difficulty increasing with protein size. A typical protein with only 100 amino acids can contain around 1000 $^1$H, 600 $^{13}$C, and 150 $^{15}$N which need to be assigned to their specific residue and location in order to interpret the structural data from these types of experiments. In a typical folding experiment, each of these nuclei could have three or more resonances corresponding to their native, intermediate, and unfolded states, all of which are in some type of exchange with each other. Thus one of the main drawbacks to many types of NMR experimental methods for folding studies is in the overwhelming number of peaks observed. Further developments in the field are made to simplify the observed spectra and the assignment of NMR resonances to their correct respective nuclei. One of the most straightforward ways to simplify the data is to limit the number of nuclei being observed in the protein, which can be done very efficiently by using $^{19}$F-NMR.

b. $^{19}$F-NMR of Proteins

$^{19}$F-NMR is a powerful method for observing site-specific structural changes along the protein folding pathway (Shu 2004). There are several reasons to use $^{19}$F-NMR in the analysis of protein structure and folding. Although $^1$H, $^{13}$C, and $^{15}$N NMR are all powerful
techniques, the amount of information contained in these types of spectra for proteins can be daunting, since each NMR peak must be assigned to its corresponding amino acid residue for each state. \(^{19}\text{F}\)-NMR is an appealing alternative because individual species of amino acids can be studied one at a time. Fluorine does not naturally occur in a protein molecule, therefore in order to use \(^{19}\text{F}\)-NMR, the fluorine must be specifically incorporated at individual residues. The methods for this incorporation are described in more detail below, but in general the most common technique for incorporation of fluorine into recombinant protein, as with \(^{13}\text{C}\) or \(^{15}\text{N}\), is to supplement the bacterial growth media used for protein synthesis with the isotope. Since fluorine is not a natural element present in proteins, the media must be supplemented with an amino acid analogue that has been chemically labeled with fluorine in its structure, and one must use a bacterial strain auxotrophic for that amino acid to prevent the synthesis of unlabeled amino acid. The major benefit of this technique is that only the selected amino acid will be fluorine labeled with no background signal. It is most common for the labeled amino acid to be a phenylalanine, tryptophan, or tyrosine, since fluorinated analogues of aromatic amino acids are commercially available at low cost and are not immediately toxic to \(E.\ coli\).

Advantages of using \(^{19}\text{F}\)-NMR in protein analysis are that fluorine has minimal effects on the protein structure, is exquisitely sensitive to its local environment in NMR studies, and the \(^{19}\text{F}\) isotope not only has 100% natural abundance, but the NMR signal is nearly as sensitive as \(^{1}\text{H}\) (83%) (Danielson 1999). The fluorine atom is small, only slightly larger than the hydrogen atom it replaces. Although the C-F bond has a larger dipole moment, minimal structural perturbations are observed when the \(^{1}\text{H}-^{19}\text{F}\) substitution is made (Frieden, 2003). The sensitivity of fluorine to its local environment and to local shielding
effects usually results in 1D NMR spectra with clearly separated peaks for each fluorinated amino acid in the protein. The natural abundance and sensitivity of $^{19}$F-NMR allow for lower protein concentrations to be used in NMR experiments than are required for other methods. This is especially important in experiments involving proteins that tend to be unstable or to aggregate at high concentrations, or that are difficult to obtain in large quantities.

$^{19}$F-NMR can also be used to calculate rate constants and populations in cases where different protein conformations exist in equilibrium and the NMR peaks have different chemical shifts. For example, if two NMR peaks represent two separate conformations in slow exchange with each other, the areas of the two peaks can be used to determine dissociation constants (Frieden 2004, Frieden 1993). Additionally, $^{19}$F-NMR has the distinction from other types of spectroscopic experiments that the information is related to protein side chains and not the peptide backbone. Since it is likely that the side chain conformations are crucial to the native structure, the steps related to stabilization of the side chains are critical to understanding folding (Frieden 2003). NMR has also been used to analyze the side-chain behavior upon ligand binding. Li and Frieden studied holo-IFABP by $^{19}$F-NMR and determined that the side chains become less ordered on ligand binding (Li 2005b). Finally, the use of $^{19}$F-NMR in protein folding studies has been informative for many different model systems. The 1D spectral peaks in equilibrium folding studies show different characteristics in the native versus the unfolded state. By collecting NMR spectra at different denaturant concentrations, the behavior of the individually labeled residues, i.e., whether the site has native, intermediate, or unfolded conformation, can be determined.

Several possible scenarios can be detected by $^{19}$F-NMR folding studies (Frieden 1993). First, $^{19}$F in a simple N → U two state transition would show one NMR peak for the
native and one peak for the unfolded state. The relative areas of the peaks in this case correspond to the relative populations of the two conformations at a given urea concentration. If the two conformations are in slow chemical exchange and the chemical shift and line width of the two peaks would not change with denaturant. The second scenario is a transition involving formation of one or more intermediate states. Here, one peak would be observed for the native conformation in the absence of denaturant, and one peak would be assigned to the unfolded state at the highest level of denaturant, but over the range of denaturation the peak characteristics would differ depending on the nature of the intermediate state. An intermediate in slow chemical exchange with either the native or unfolded state would appear as a separate peak, and there would be some peak broadening between the two exchanging forms. The relative populations of each state could again be determined based on the peak areas for native, intermediate, and unfolded resonances. An intermediate exchanging with the native or unfolded state on an intermediate time scale would have only one peak in the presence of the two conformations, and this peak would have significantly broadened line width, in some cases broad enough to flatten completely into the baseline. Finally, an intermediate exchanging with the native or unfolded state on a fast time scale would show only one peak with no significant broadening occurring at a chemical shift determined by the population-weighted average of the conformations (Frieden 1993). The unfolded state peaks for protein $^{19}$F-NMR are a special example of the last case, in which many unfolded conformations coexist in rapid exchange and define the unfolded state peak. Examples of spectra for these different scenarios are given in figure 1.1.
Figure 1.1: NMR Spectra of Conformations in Slow, Intermediate, and Fast Exchange

Three representative series of NMR spectra demonstrating conformations in (A) slow, (B) intermediate, and (C) fast chemical exchange. Panel A represents the NMR spectra for IFABP-Trp6, panel B represents IFABP-Trp82, and panel C represents ILBP-Phe47. Panels A and B were reproduced with permission from Ropson 2006.
In an early study, Ropson (1992) used $^{19}$F-NMR and site-directed mutagenesis to analyze the unfolding behavior of two tryptophan residues in IFABP. Later studies examined the behavior of all aromatic residues in this protein (Ropson 2006, Li 2005, Li 2005). Experiments examining the four tryptophans in murine adenosine deaminase have also been done using $^{19}$F-NMR (Shu 2004). For these proteins, the distribution of aromatic residues throughout the structures has allowed fluorine incorporation at these sites to be informative probes for different structural regions. By collecting NMR spectra under different levels of denaturation, the regions labeled with the probe can be analyzed for different unfolding behaviors. In all of these cases, it has been observed that the residues being probed by the $^{19}$F-NMR experiments exhibit non-uniform behavior in response to denaturant. In some cases, this behavior has been assigned to individual residues, but in order to assign the resonances in these proteins labeled at multiple sites with fluorine, mutations had to be made that might have perturbed the folding mechanism. Recent advances in protein engineering using a 21st amino acid codon system have allowed for site-specific incorporation of fluorine without dramatic mutations, described in more detail below.

Li and Frieden used the site-specific $p$-$F$-Phe incorporation approach described by Furter (1998) to characterize the eight Phe residues in IFABP (Li 2005b). They were able to assign the eight resonances observed in the fully $p$-$F$-Phe-labeled protein to their respective residues, and observed that all eight did not behave in the same way. For example, they noted that Phe68 and Phe93 exhibited broader line widths than the other peaks in the native state, which they attributed to the presence of two separate conformations for each of the residues in the native protein. The broad peaks could be deconvoluted into two peaks at all concentrations of denaturant, with one peak for each residue decreasing with increasing
denaturant. The decreasing peaks persist to a level of denaturant much higher than the midpoint concentrations for unfolding determined by optical methods, and therefore represent some unfolded-like intermediate state. Residues Phe68 and Phe93 show NMR spectra that change in response to denaturant in an almost identical fashion, which may indicate that the environments that they experience during unfolding are the same or very similar (Li 2005). They also observed that there is a loss in intensity without a substantial change in line width for the peaks in the fully \( p \)-F-Phe-labeled IFABP between concentrations of 3.5 and 6 M urea. This denaturant concentration is higher than the point at which the protein appears to be unfolded by other methods. The explanation for this loss of intensity is that intermediates similar to the unfolded state are present and exchanging at an intermediate rate (Li 2005, Ropson 2006).

Based on their previous studies, Li and Frieden propose that turns between \( \beta \)-strands act as nucleation sites for folding initiation (Li 2005, Hodson 2001, Ropson 1992). In the single \( p \)-F-Phe incorporation studies, they observe that at least two phenylalanines have native-like peaks that persist at concentrations of denaturant at which the protein is mostly unfolded by other methods. This may indicate a core region of the protein that maintains native-like structure when the majority of the protein is unfolded. Their conclusions suggest that clusters of hydrophobic side chains remain organized up to the point of global unfolding, and that some clusters definitely persist up to higher concentrations of denaturant (Li, 2005).

The single-site \( p \)-F-Phe-incorporation studies of IFABP by Li and Frieden (2005 2005) mainly characterized the sidechain behavior of the native apo- and holo-states. Ropson et al. further examined the behavior of the aromatic side chains in the unfolded state of IFABP (2006). In these studies, \(^{19}\text{F-NMR} \) spectra of IFABP labeled at either both of the two
tryptophans, all four of the tyrosines, or all eight of the phenylalanines, were collected over a
range of urea concentrations. Lineshape analysis revealed that several residues continue to
show changes in their spectral properties, even at concentrations of denaturant at which the
protein is known to have undergone global unfolding. The interpretation of these results is
that some but not all of the labeled residues, those with NMR peaks that still show changes in
the high urea concentration range, are involved in some residual structure. These sites could
indicate folding nucleation sites if they are residues that are in the same structural regions in
the native state. Figure 1.2 represents the IFABP structure with each of the aromatic residues
highlighted. Those residues that were determined to be part of this nucleation site are
illustrated by sticks and lines in the ribbon diagram.
Figure 1.2: Structure of IFABP with Aromatic Residues Highlighted

IFABP structure with aromatic residues highlighted in color. The residues determined to be involved in the initiation site, Phe68, Tyr70, Trp82, and Phe93, are illustrated as sticks, while F47 is shown as lines. All phenylalanine residues are highlighted in light purple, and all tyrosines in dark blue. Note: All ribbon diagrams of protein structures presented here were created using the MacPymol software.
Ropson’s approach to assigning the peaks in the $^{19}$F-NMR spectra to their corresponding residue differed from Frieden’s site-directed incorporation studies described above. In order to assign the two tryptophan peaks to the correct residues, Ropson designed two mutant IFABP proteins in which each tryptophan was conservatively mutated to a tyrosine, leaving only one tryptophan in each mutant. By comparing the mutant NMR spectra to those of the wild-type protein, the peaks could be assigned to the correct residue. Likewise the four tyrosines were mutated to phenylalanines one at a time to assign the tyrosine resonances, and two of the eight phenylalanine resonances were assigned by mutation to alanine (Ropson 2006).

The results of these studies reveal characteristics of the partially folded intermediate state in IFABP. First, of the two tryptophans in the protein, Trp6 showed only two distinct NMR peaks, one for the native state and one for the unfolded state. At urea concentrations near the midpoint, both peaks were visible. This indicates that this residue is in an environment that effectively unfolds via a two-state mechanism. However, Trp82 $^{19}$F-NMR spectra were different. For this residue, the native state peak disappeared near the midpoint urea concentration, but no peak for the unfolded state was observed until much higher urea concentration. This residue therefore experiences an environment which passes through at least one intermediate state which is in intermediate exchange with the unfolded state (Ropson 1992, Ropson 2006).

The results for the 3-F-Tyr-labeled IFABP proteins showed similar variations in NMR spectra. Analysis of the protein labeled at all four tyrosine residues revealed that there is missing total peak intensity in the range of denaturant between the midpoint for unfolding and the concentration for total unfolding. The peaks in this range show lineshape
characteristics for only three and not four labeled residues, indicating that one of the four tyrosines shows the same “missing intensity” behavior as Trp82, or that one of the tyrosines is also in intermediate exchange. By collecting $^{19}$F-NMR spectra for each of the four tyrosine \to phenylalanine mutants, it was determined that Tyr70 is the residue responsible for this behavior, with the interpretation that Tyr70 is also in an environment that passes through at least one additional conformation during unfolding. In the crystal structure for IFABP, Tyr70 and Trp82 are spatially close together (see figure 1.2) (Ropson 2006).

The eight phenylalanines in IFABP are distributed throughout the tertiary structure. NMR analysis of the $p$-F-Phe-labeled protein revealed peak characteristics similar to those for the 6-F-Trp and 3-F-Tyr labeled proteins. Missing intensity was again observed over the range of denaturant concentrations between the midpoint for unfolding and the denaturant concentration at which the protein is completely unfolded, with the missing intensity here corresponding to at least four of the eight phenylalanine residues. The peak assignments had already been confirmed using site-specific $p$-F-Phe incorporation (Li 2005b), so the phenylalanine \to alanine mutants constructed by Ropson were used to determine the characteristics of the missing intensity in proteins with only seven phenylalanine residues. The mutations confirmed that Phe68 and Phe93 are two of the four phenylalanine responsible for the missing peaks. The missing intensity for Phe68, Phe93, Tyr70, and Trp82 is attributed to the presence of an additional conformation of the protein that is present over the range of denaturant concentrations where this behavior is observed. The lack of missing intensity for the remaining aromatic residues indicates that these are not participating in this equilibrium intermediate state. This intermediate protein conformation is in exchange with the unfolded state on an intermediate timescale (Ropson 2006). When two conformations are present and
exchange at a slow rate, the NMR spectra show two broad peaks, corresponding to the two conformations. At fast exchange rates, a single, sharper peak is observed that is the population-weighted average of the frequencies of the two different conformations. However, when the rate of exchange between the two conformations is on an intermediate timescale, a single, broad peak is observed, which can be so broad in some cases that it disappears into the baseline (Frieden 1993). The “missing intensity” peaks in the IFABP spectrum are attributed to the presence of two structural conformations in intermediate exchange (Ropson 2006).

The four residues positively identified to participate in the intermediate close to the unfolded state in IFABP (Trp82, Tyr70, Phe68, and Phe93,) are all neighbors in the native state of IFABP (see figure 1.2). This is an important observation, because it appears that native-like contacts persist in the intermediate, indicating the presence of a folding initiation site. Since the residues all exhibit the same behavior upon unfolding, they are likely to be participating in the same structure. Thus, native-like contacts are maintained late in the folding pathway, at a point at which much of the protein has no native structure. This provides support for the model that these four residues are involved in the formation of an early structure in IFABP folding (Ropson 2006). Ropson notes that at least two other phenylalanines show similar behavior and may therefore participate in the early folding intermediate. Of the remaining phenylalanines, Phe2, Phe47, and Phe62 are located near the same hydrophobic cluster. In their original model, Ropson and Frieden had proposed Phe47, Phe62, Phe68, and Phe93 to be involved in this hydrophobic cluster (Ropson 1992). However, Li and Frieden’s NMR study showed that Phe47 may not be involved, based on its narrow line width, minimal chemical shift change, and observation that the residue’s
movement appears to be independent of Phe62 and Phe68 (Li 2005b). Therefore the remaining phenylalanines involved in the early intermediate have yet to be positively assigned and it remains to be proven whether Phe47 is involved in the hydrophobic initiation site in IFABP.

c. Site-Specific Labeling for $^{19}$F-NMR

As indicated above, there are two potential strategies that can be used to unambiguously assign the observed resonances from the NMR experiment to the corresponding amino acid. The first is to use site-directed mutagenesis to change one amino acid at a time with a conservative replacement, prepare the mutated protein in the labeled media environment, and collect the NMR spectra. Assignment of the missing resonance can be made to the mutated residue. This strategy, although effective for many proteins (eg, Ropson 1992, Ropson 2006), has several potential pitfalls. There is always the possibility, for instance, of disrupting the protein structure upon even a conservative mutation. In addition, even if the global structure does not appear to have been altered, a mutation could potentially influence the folding mechanism by altering the environment of a folding initiation site. Finally, $^{19}$F-NMR is exceptionally sensitive to local environment, so even minor changes in the environment of the remaining labeled residues could result in significant changes in the appearance of the native state spectrum. This can cause difficulties in assigning the resonances, with the difficulty increasing as the number of labeled residues increases.

Another strategy is to incorporate the fluorinated analogue at only one residue at a time, as in the study by Li described above (2005) which is based on the system (Furter 1998), that utilizes a “21st” amino acid incorporation system. This involves a site-specific
mutation to a suppressor codon, and co-expression of the mutated gene with a tRNA / tRNA synthetase pair that translate the suppressor codon as the desired analogue. The system developed by Furter (1998) has been designed to incorporate the p-F-Phe analogue at the site of an Amber stop codon (TAG). This is accomplished by co-transforming *E. coli* with the amber-substituted recombinant gene and the genes for the phenylalanyl-amber suppressor-tRNA (tRNA\textsubscript{Phe}) and phenylalanine tRNA synthetase (PheRS) from *Saccharomyces cerevisiae*. The bacterial strain chosen must be resistant to p-F-Phe, since most wild type strains will incorporate the analogue in place of any phenylalanine, which will result in all phenylalanines in the protein being labeled. In the resistant strain, a mutated PheRS does not recognize p-F-Phe in its substrate binding site (Furter 1998) and thus charges the tRNA\textsubscript{Phe} with only unlabeled phenylalanine. Wild-type yeast also incorporate p-F-Phe into their proteins, so use of the yeast PheRS without modification, along with the suppressor tRNA\textsubscript{Phe}, allows incorporation of the analogue at the suppressor codon site. The *E. coli* will thus not incorporate the analogue at any other site, and by supplementing the media with an excess of the analogue compared to unlabeled phenylalanine, it can be ensured that the yeast machinery incorporates p-F-Phe at the amber codon in the majority of the expressed protein. There is some level of background “contamination” labeling with p-F-Phe at other Phe sites, which is most likely due to cross-species charging of the *E. coli* tRNA\textsubscript{Phe} by the yeast PheRS (Furter 1998), but the extent of the background label is minimal compared to the Amber-encoded site.

Other labs are focusing on using *E. coli* phenylalanyl-tRNA synthetase to incorporate amino acid analogues into nascent proteins during synthesis in bacteria. There are four main requirements for incorporating an unnatural amino acid analogue into nascent protein using
nonsense codon suppression. These include sufficient quantities of the analogue, a tRNA charged with the analogue, a gene including the nonsense codon, and a system for translation. With these four reagents, over 100 different amino acid analogues have been incorporated into many different proteins, whether in cell-extract or in cell-intact translation (England 2004). Using such a system, one group was able to engineer proteins containing \( p \)-bromo-, \( p \)-iodo-, \( p \)-cyano-, \( p \)-ethynyl-, \( p \)-azido-phenylalanine, all using the same mutant tRNA synthetase with an enlarged substrate binding site (Kirshenbaum, 2002). Jackson, et al, have described a system for incorporation of a fluorophenylalanine analogue into proteins in \textit{E. coli} with no background contamination (Jackson 2007). Other groups have developed similar systems for protein production in eukaryotic systems, for example, Chin et al have developed a 21st amino acid codon system for analogue incorporation in yeast (Chin 2003).

\textbf{D. The Intracellular Lipid-Binding Proteins}

The folding behavior of \( \beta \)-sheets has not been as well characterized as that of \( \alpha \)-helices since it has been difficult to study \( \beta \)-sheets in isolation (Eaton 2000), though recently more focus has been placed on this important aspect of protein folding. One model for the formation of \( \beta \)-sheets is that they are initiated at turns between \( \beta \)-strands, after which the hydrogen bonding reactions can quickly “zip up” the remainder of the fold (Muñoz 1998). One group examined this hypothesis using short peptides known to form stable \( \beta \)-hairpin structures (Dyer 2004). The hypothesis that turn initiation begins upon collapse of non-local hydrophobic core regions was tested by comparing \( \beta \)-hairpin peptides, and the folding rate increased significantly when the loop connecting the hydrophobic cluster was made shorter.
(Dyer 2004). Other models for protein folding also incorporate collapse of hydrophobic residues in initiation of the fold, and will be discussed in more detail below. Thus hydrophobic interactions closing turns in hairpin loops may occur as an initiating reaction in folding.

In addition to the NMR studies described above, the intracellular lipid-binding proteins (iLBPs) have been studied extensively in terms of structure and folding dynamics. The members of this family share a common three-dimensional structure dominated by a β-barrel containing ten antiparallel β-strands with a small portal gated by a short helix-turn-helix motif (see figure 1.3). The ligand-binding site is located within the β-barrel. The presence of a large gap sufficient to interrupt hydrogen bonding between the fourth and fifth β-strand is a common feature, and the secondary structural elements of all members of the family differ very little (Banaszak 1994). Although the functions of these proteins are not completely understood (Montoudis 2006), they are proposed to enhance the aqueous solubility of hydrophobic ligands and function in transfer and cytoplasmic diffusion of fatty acids and lipids (Chmurynzska 2006). Early studies of low molecular weight lipid binding proteins were done to elucidate the role of soluble proteins as lipid carriers (reviewed in Ono 2005). This large family of related proteins includes several different tissue-specific fatty acid- or lipid-binding proteins.
Figure 1.3: Representative Structure of an iLBP with ligand

Structure of IFABP (PDB ID: 2IFB) illustrating the β-barrel structure of the iLBP’s with the ligand-binding cavity capped at one end by two short α-helices. The bound fatty acid ligand is shown as a space-filling model in blue, and is located between the two β-sheets.
The iLBP’s can be divided into four subfamilies based on the ligands they bind and their relationship to each other. The gene structure for all iLBP’s is conserved, with three introns of variable lengths in identical positions separating the four exons (Haunerland 2004). The human iLBP’s are between 126 and 134 amino acids in length, and share a sequence similarity of only 22-73% (Zimmerman, 2002). When invertebrate and other vertebrate species are included in the analysis the sequence identity is as low as 7% (Marcelino 2006). In spite of the sequence differences, however, the β-barrel structure is conserved in all family members. Phylogenetic analysis of the vertebrate iLBP’s lead to the conclusion that a common ancestral gene for the family probably arose in animals after the kingdom diverged from plants and fungi. The first gene duplication probably occurred approximately 930 million years ago, with subsequent duplications giving rise to the distinct iLBP family members currently known (Schaap 2002, Chmurynzska 2006). Present day genes have specific ligand-binding properties based on the protein type (Schaap 2002).

In addition to serving as model systems for β-barrel protein studies, there is clinical relevance to studying this protein family. Variants of the family have been developed as probes for fatty acids in serum, which can be markers for several different diseases (Richieri 1999, Azzazy, 2006). Other members of the family are potentially markers that can be used to detect conditions like cardiac injury (Azzazy 2006), renal cell carcinoma (Teratani 2007), obesity and metabolic syndrome (Xu 2006), or lung damage (Lachmann 2006 ), among others. The utility of this protein family as both biomarkers and as agents for detection of other biomarkers has resulted in a the significant amount of research on the iLBP proteins. Structural studies of the iLBP family contribute to these clinical developments.
a. Model System for Folding Studies

The iLBP proteins provide an excellent model system for the folding of β-sheet proteins for several reasons. The structure is mostly β-sheet, and thus provides a system for studying this important protein fold. The proteins are readily available and recombinant protein can be obtained in high quantities, since they are amenable to expression and purification from bacteria. The proteins are relatively stable under laboratory conditions and not exceptionally sensitive to different buffers. Finally, the vast bank of information already available on several family members allows for direct comparisons to be made between related proteins.

Structural studies of iLBP family members reveal that all share the common β-barrel structure in spite of dramatic sequence differences. Family members have been well characterized in terms of stability, ligand binding properties, and folding mechanism. An interesting approach that can be taken with this family is to compare the folding mechanisms of different members and determine the influence of a common structure on the folding behavior of different proteins. Studies on the folding mechanism of different members of a family of proteins test the hypothesis that structural topology drives the mechanism of folding (Burns 2001). Systematically analyzing the folding mechanisms of related structures provides insight as to whether the folding mechanism is conserved along with native conformation. Burns and Ropson have found evidence that the folding pathways may not necessarily be conserved, since they were able to observe intermediates by kinetic methods in the folding pathways of some iLBP family members that were not present in others (Burns
It has been suggested that the observed differences in iLBP folding are due to a combination of global and local sequence influences (Gunasekaran 2001).

The proteins provide a model for \( \beta \)-sheet proteins due to the high amount of \( \beta \)-sheet structure. The influence of the short helix-turn-helix motif is not expected to have a large effect on folding, since removal of the motif does not significantly disrupt the \( \beta \)-barrel structure of one family member (Kim 1996, Ogbay 2003). A helix-less variant of IFABP in which this region was replaced with a short Gly linker was generated to examine the functional and structural role of the helices. Remarkably, the overall structure of the variant was nearly superimposable with the \( \beta \)-barrel of the wild-type protein (Kim 1997). Although the variant had some slight differences in stability and ligand-binding properties, retention of the \( \beta \)-barrel structure indicates that the helices do not play a significant role in folding of IFABP. However, removal of the helices in another member of the family resulted in an unstable protein (Kouvatsos 2006), which may indicate that all members of the family are not stabilized by the same interactions.

b. Kinetic and Equilibrium Data of iLBP’s

Early equilibrium and kinetic studies of IFABP suggested that the folding mechanism is not two state and encounters at least one intermediate on the folding pathway (Ropson 1990). \(^{19}\)F-NMR analysis of the two tryptophans in the protein revealed that Trp82 participates in an intermediate that has little secondary structure (Ropson 1992). Further kinetic analysis revealed that the protein indeed folds via at least two intermediate states, the initial being formed very rapidly and involving burial of some hydrophobic residues, and the
second formed as the hydrophobic core becomes more compact and native-like (Ropson 1997, Yeh 2001).

The folding mechanism of IFABP was compared to ileal lipid binding protein (ILBP) to determine whether these two structurally related proteins fold via the same mechanism (Dalessio 2000). Studies up to that point had shown that small proteins with similar structures had similar folding mechanisms, but a comparison of these larger proteins containing both α-helices and β-sheets, and higher sequence divergence, had not been done. The results of the comparison of the two ILBP’s concluded that the two proteins do not, in fact, appear to fold by the same mechanism by fluorescence kinetic studies. ILBP and IFABP were compared by fluorescence and circular dichroism (CD) spectroscopy, and the following conclusions were made. The unfolding data for both proteins fit the model

$$N \Rightarrow I \Rightarrow U$$

with N indicating the native state, I a folding intermediate populated along the folding pathway, and U the unfolded state. However, the intermediate observed for ILBP had properties similar to the native state protein, but the intermediate observed for IFABP did not show secondary structure elements. Similarly, refolding in ILBP fit a model

$$U \Rightarrow I \Rightarrow N$$

with the intermediate showing native-like secondary structure. The refolding data for IFABP, however, fit the model

$$D \Leftrightarrow U \Leftrightarrow I \Leftrightarrow N$$

With D representing a collapsed denatured state, and the intermediate here appears to form rapidly and to have structural features similar to the intermediate observed during unfolding (Dalessio 2000). The characteristics of the observed intermediates indicate that the IFABP
intermediate forms early in the folding reaction, since it does not show elements of secondary structure. The intermediate for ILBP must occur at a later stage of folding, after some secondary structure has formed. The differences in the observed intermediates were attributed to variation in a hydrophobic cluster region in the two proteins, in which IFABP contains more hydrophobic residues than does ILBP. In further experiments, the hydrophobic regions of these two proteins were effectively swapped, adding a hydrophobic residue to this core region in ILBP and removing one in IFABP. The result of swapping the core regions was an apparent swap in observed folding mechanism (Dalessio 2005). However, since the observation of a folding intermediate requires that the state accumulate to a detectable amount, it may still be the case that the two structurally related proteins pass through all of the same intermediate states, but accumulate intermediates to different degrees at different stages of folding. In fact, an intermediate in ILBP similar to the early folding intermediate in IFABP was observed (Dalessio 2005), as described below.

More recent folding equilibrium data on these proteins is discussed in detail below, and compared to the results observed in the kinetic studies. Both types of folding studies have revealed differences in the types of intermediates observed on the folding pathways of related iLBP family members. However, in a comparison of the pair-wise interactions of residues of CRABP I with sequences of 52 other iLBP family members, Gunasekaran, et al., determined that interactions between hydrophobic residues are highly conserved. This is in spite of the fact that only eight amino acids outside of the helical region have significant sequence conservation among these family members (Gunasekaran 2004). This could indicate that some of the hydrophobic interactions play a conserved critical role in folding,
although conserved interactions could also be important for overall structure and stability, as well as in the functional roles of the proteins.

The IFABP and ILBP folding mechanism comparison is incomplete. Dalessio (2000) previously compared equilibrium and kinetic folding and unfolding data for rat IFABP and rat ILBP. The equilibrium studies by fluorescence and circular dichroism spectrometry showed that both proteins fit two-state folding models with no intermediate observed in either. However the kinetic data revealed at least one intermediate for each. IFABP was shown to progress first through a denatured state which was not determined to be on the folding pathway, and through an on-path intermediate that lacks secondary structure, placing this state early in the folding pathway. The intermediate observed for ILBP has secondary structure and has molten-globule like properties. In a later study, the same authors note that improved instrumentation resulted in the appearance of an early intermediate in ILBP similar to the one observed for IFABP (Dalessio 2005). The authors offer several explanations for the differences observed in the folding pathways of these two related proteins (Dalessio 2000). First, it may be the case that, since the tryptophan reporters are in dissimilar regions in the two proteins, they are revealing information about different regions of the structure. It was suggested that structurally homologous reporter elements might reveal both intermediates in both proteins. They tested this possibility in CRBP II, which has similar unfolding behavior to IFABP, and CRABP I, which is similar to ILBP. When mutations were made to place tryptophan reporters at identical locations in CRBP II and CRABP I, the observed kinetic mechanisms did not change, indicating that the location of the tryptophan is not likely to be responsible for the different behaviors observed (Dalessio 2000). A second
possibility is that the two proteins indeed each pass through both intermediates, proposing the mechanism

\[ U \Rightarrow I_1 \Rightarrow I_2 \Rightarrow N \]

for both proteins, but the energetic stability of the two intermediates differs. As noted previously, an intermediate must be stable enough to accumulate in order to be observed by kinetic studies. Lastly, they present the possibility that the two proteins fold via different mechanisms. The energy landscape model certainly permits a wide array of possible pathways between unfolded and native state, and the sequence differences between these two proteins could dictate different pathways that encounter different metastable intermediate states.

To determine the role of the hydrophobic core residues in the folding pathway differences observed between IFABP and ILBP, Dalessio et al. (2005) generated a set of mutant proteins which altered the number of hydrophobic residues in the core. Their hypothesis was that more hydrophobic residues are buried in the IFABP hydrophobic core than in the ILBP core, leading to the increased stabilization of the early folding intermediate. Adding a hydrophobic residue to the ILBP core should thus result in the observation of the early intermediate without observation of the molten globule-like intermediate, which is what was observed. Further, replacement of a hydrophobic residue in IFABP with a more hydrophilic residue similar to the core of ILBP should result in accumulation of the molten-globule intermediate, which was also observed. The results of this study support the hypothesis that both proteins follow similar folding pathways, with different intermediates being observed as a result of differences in the intermediate energetic stability. However, the hydrophobic – hydrophilic substitutions are rather dramatic, and the results could also
indicate that these substitutions altered the folding pathways. To prove that both proteins follow the same folding pathway, experiments in which the hydrophobic cores are not modified must be designed, especially since formation of the early intermediate is proposed to be caused by specific hydrophobic collapse in this region.

To address the role of the hydrophobic core region in formation of the early intermediate without mutating the corresponding residues, $^{19}$F-NMR equilibrium folding studies were carried out for IFABP (Li 2005, Li 2005, Ropson 2006). The experiments were designed to identify residues that form interactions under strongly denaturing conditions and determine whether these residues are also in contact in the native protein, as described above. The $^{19}$F-NMR analysis identified at least four aromatic residues whose NMR spectra continue to show changes above denaturant concentrations where other sites have stabilized. This indicates that during unfolding, while many of the aromatic residues are already in their unfolded state, these four retain some structure and constitute some intermediate state. This intermediate occurs closer to the unfolded state than the first folding intermediate observed in the kinetic studies. The four residues identified as participating in this intermediate are close together in the native state, and, further, are located in the hydrophobic core region of the protein. This supports the hypothesis that native-like interactions in the hydrophobic core occur early in the folding pathway in IFABP.

**E. Conclusion**

To summarize, there exists a gap in understanding how a protein folds from its primary amino acid sequence to a well-defined tertiary structure. There are many examples in the literature of studies of protein folding aimed at determining the rules governing folding
behavior. These studies include kinetic analyses, where folding is observed in real time, and equilibrium studies, where different stages of the folding reaction are inferred by observations made under varying levels of denaturation. In larger proteins, conformations other than the native and unfolded state are frequently observed. These intermediate conformations appear to be stages along the folding pathway that limit the search through structural space and direct the protein to the correct structure. In other cases intermediates represent structures in which stable non-native contacts have been formed and thus represent misfolded kinetic traps. When on-path intermediates are observed early in the folding pathway, they provide information about the reactions that initiate the protein fold. These interactions are of interest because they limit the further contacts that can be made as the protein folds. Identification of these initiating reactions has implications not only in the field of structural biology, but can be useful in developing treatments for conditions caused by accumulation of misfolded proteins. Many of the misfolding diseases involve dissimilar proteins accumulating in the same misfolded form. This is perplexing since the original proteins do not share sequence or native fold similarity. This leads one to question whether there is a common misfolding pathway that very different proteins can follow to the same final form. An extension of this question is whether families of structurally similar proteins with different primary structures fold to the same native state via the same pathways. The iLBP family has been examined in several studies aimed at addressing this question.

Kinetic analyses of two related members of the iLBP family, IFABP and ILBP, have provided evidence that both proteins pass through an intermediate state early in the folding pathway, and that stability of the early intermediate differs for the two proteins. An early intermediate has been characterized for IFABP and it has been determined that formation of
early native-like contacts upon specific hydrophobic collapse results in stability of this intermediate conformation. The work presented in this dissertation has been done to analyze the $^{19}$F-NMR behavior of each of the eight Phe residues in ILBP one at a time to determine the role of the hydrophobic core in folding of this protein. It will be demonstrated that the $^{19}$F-NMR behaviors of the aromatic residues differ from those observed in IFABP, but that there is evidence to support a role for at least some degree of native hydrophobic collapse early in folding.
Chapter 2: Hypothesis

Based on the similarities in structure between members of the iLBP family, one might suggest that all proteins of this type use the same mechanism to achieve the final three-dimensional structure. However, marked differences have been illustrated in the character of the intermediate states of IFABP compared to ILBP, with the former showing a folding intermediate similar to the unfolded state, and the latter showing a stable intermediate closer to the native state. The methods used for detection of these intermediates were sensitive to formation of local structure, but unable to illustrate characteristics of individual residues upon folding. Utilizing $^{19}$F-NMR with proteins labeled at individual amino acids with fluorine, it has been shown that the behavior of certain hydrophobic residues in IFABP confirm the presence of the intermediates seen by optical methods, and supports a mechanism of hydrophobic collapse in the hydrophobic core region early in the folding reaction as an initiating site in folding. It is hypothesized here that the analogous residues in ILBP will show some characteristic of early structure indicating that the folding initiation site is evolutionarily conserved, but the stability of the initiating site is not sufficient to cause accumulation of an intermediate until later in the reaction, when more native contacts have been made.
Chapter 3: Materials and Methods

A. Protein source and expression

All studies used recombinant proteins expressed and purified from *E. coli*, utilizing the protocols described below. The human ILBP proteins and mutants were modified at the N-terminal end with a histidine tag for ease of purification [hILBP(6His)]. The tag added the sequence MRGSMHHHHHHHGGSNS upstream of the initial methionine. Purification and expression of all protein mutants, regardless of $^{19}$F label incorporation, was done using the same plasmid background. The bacterial strains used for protein production varied depending on the type of fluorine labeling, with a wild-type strain, BL21(DE3), used for unlabeled protein, the phenylalanine- and tyrosine-auxotrophic strain DL39(DE3) for complete labeling of phenylalanine or tyrosine residues, and the K10F6Δ(DE3) strain, specifically resistant to $p$-F-Phe incorporation due to a mutated phenylalanyl tRNA synthetase, for the single-site labeled proteins.

a. Wild-type ILBP

The plasmid pRO.hILBP(6His) was constructed using the plasmid pRO.148, which was obtained from Dr. J. Bann in the lab of Dr. Carl Frieden at Washington University, St. Louis, MO (Frieden 2004). The plasmid pRO.148 contains the yeast phenylalanine synthetase gene (PheRS) engineered into the commercial plasmid pQE.16 (Qiagen) as a PvuII restriction site insert. PQE.16 contains the mouse dihydrofolate reductase (DHFR) gene. In order to sub-clone hILBP(6His) in place of DHFR, it was first necessary to recover the pQE.16 vector separate from the PheRS gene, since there were no appropriate restriction
sites in the complete pRO.148 plasmid. Since the pQE.16 sequence and the PheRS gene sequence were both approximately 4000 base pairs, it was not possible to separate the two pieces by digestion with PvuII alone. Instead, the pRO.148 plasmid was digested with PvuII and BglII, according to the enzyme manufacturer’s protocol (Promega), which allowed for isolation of the ~4000 base-pair PheRS gene fragment. Digesting the pRO.148 plasmid with PvuII and EcoRV similarly resulted in isolation of the pQE.16 fragment. The two fragments were isolated by agarose gel electrophoresis (see Sambrook 1989, for standard protocol) and the 4000 base-pair bands purified using the QIAquick gel extraction kit (Qiagen). The pQE.16 thus isolated was then ligated into a circular form by blunt-end ligation at the PvuII sites, following manufacturer’s recommendations (Promega), and amplified by transforming the ligation product into XL1-Blue cells, taking advantage of the ampicillin-resistance gene in the pQE.16 plasmid. The hILBP gene was obtained as a gift from Dr. Cistola at Washington University School of Medicine, St. Louis, MO, and sub-cloned in our lab into the pET28b(+) vector for introduction of the histidine-tag at the N-terminus. This was done using PCR to isolate the hILBP gene from the original vector, using primers designed to introduce BamHI site at the 5’-end and a SacI site at the 3’-end. The PCR product was then digested and ligated into the corresponding sites in the pQE.16 vector. This ligation replaced the DHFR gene with the hILBP gene, and the vector was re-named pQE.16-hILBP. This plasmid was digested with PvuII, and the PheRS insert blunt end-ligated back in, forming the vector named pRO.hILBP(6His) (See figure 3.1).
Figure 3.1: Schematic of pRO.hILBP(6His) cloning

Schematic of cloning strategy to replace DHFP gene with hILBP(6His) in pRO.148 plasmid. Resultant vector pRO.hILBP contains hILBP(6His) gene, PheRS gene, and is ampicillin-resistant.
It was later determined that the six-histidine tag on ILBP in the pRO.hILBP(6His) vector may cause undesired protein aggregation, and the linker region was changed to match another tag in use in the laboratory that did not have this problem (modified tag sequence: MHVDPMHHHHHHHGGGSNS). This was accomplished by generation of the primers IJR089 and IJR090 (see Table 3.1), and site-directed mutagenesis PCR, as described below. For ease of comparison, however, ILBP residue numbering will correspond to the wild-type, untagged protein throughout the discussion.

**b. Fluorine-labeled Protein**

The methods for incorporation of $^{19}$F-amino acid analogues have recently been outlined in detail (Frieden, 2004 – methods enzymol). The two separate approaches used here include either analogue incorporation at all locations for a particular amino acid, or single-site incorporation. For $^{19}$F-analogue incorporation in place of a given amino acid species, the bacterial growth media is supplemented with the analogue, and a strain of *E. coli* auxotrophic for the particular amino acid is used. This results in incorporation of the analogue at each corresponding site in nascent proteins.

In order to produce ILBP labeled with $^{19}$F at each phenylalanine site, the plasmid containing the wild-type gene, pRO.hILBP(6His), is transformed into chemically competent DL39(DE3) strain of *E. coli* by standard protocol. Briefly, a fresh colony of DL39(DE3) cells was cultured in liquid media overnight, diluted 1:50 the next morning in fresh media, and grown to an OD$_{600}$ of approximately 0.8. The cells were collected, washed in 50mM CaCl$_2$ once, and re-suspended in 50mM CaCl$_2$, 15% glycerol. Plasmid DNA was added to the cells, which were then heat-shocked at 42° for 30 seconds, allowed to recover in fresh media for
one hour at 37°, and selectively plated. This protocol was used for all subcloning and bacterial transformations. The DL39(DE3) bacterial strain is auxotrophic for phenylalanine, which ensures that the analogue will be incorporated into nascent proteins. A colony from a freshly-transformed culture of the bacteria is grown to saturation, usually overnight, in 2xYT rich media. The culture is then diluted 1:50 into minimal media and cell growth monitored hourly by optical density at 600nm (OD$_{600}$). When the OD$_{600}$ reaches 0.8 – 1.0, the cells are harvested by centrifugation and re-suspended in minimal media containing p-F-Phe in place of unlabeled phenylalanine. Isopropyl β-D-thiogalactopyranoside (IPTG) is added to the media to a final concentration of 1mM to induce protein production. Cells are grown for 3 – 4 hours or until the cells are no longer dividing, as determined by OD$_{600}$. At the completion of growth, cells are again harvested by centrifugation and stored at −80° C. To determine that protein production has been induced, 500µl samples are removed from the cultures at one-hour intervals post-induction and analyzed by SDS polyacrylamide gel electrophoresis. An induction gel is shown in figure 3.2.

Full incorporation of $^{19}$F-Tyr analogue or of $^{19}$F-Trp is accomplished essentially by the same protocol as the $^{19}$F-Phe. The DL39(DE3) cell line is also auxotrophic for Tyr, so this is the cell line used for $^{19}$F-Tyr incorporation. All cultures are grown in minimal defined media, with the appropriate analogue substituted for the unlabeled amino acid before induction. For all cell lines described, bacteria are cultured at 37° in a shaking incubator. For cultures of 300-500ml, cultures are grown in 2L baffled flasks and shaken at a minimum of 300rpm for maximum aeration of the cultures.
Figure 3.2: Representative SDS-PAGE gel

A typical SDS-PAGE gel showing induction of protein production and purification of the protein by using the His-tag protocol outlined. Lanes:
(A) pre-induction sample of bacterial culture
(B) post-induction sample
(C) column fraction on addition of 10mM imidazole buffer
(D-J) column fractions on addition of 25-100mM imidazole buffer
(MW) low molecular weight marker.

Gel shows that a significant level of induction was achieved by the expression method outlined, and that the purification scheme resulted in little or no contamination with other protein species.
c. Site-Specific Fluorine Incorporation

Single-site fluorine incorporation was done by co-expressing the ILBP(6His) gene with the PheRS and tRNA\textsuperscript{Phe}, essentially as described (Frieden 2004). The strategy differs from incorporation of an analogue at all sites for the particular amino acid in that the cell line chosen is not auxotrophic for the amino acid but resistant to incorporation of the analogue. The system used here for incorporation of \(\rho\)-F-Phe at individual sites is based on the work of Furter (1998). Briefly, K10F6\(\Delta\)(DE3) cells are co-transformed with the pRO.hILBP(6His) plasmid or one of the mutant variants of this plasmid, as well as the pRO.117 plasmid, also obtained from Dr. Frieden’s lab. The pRO.117 plasmid contains the yeast amber-suppressor tRNA\textsuperscript{Phe}, which is necessary for translation of the amber codon as \(\rho\)-F-Phe. Freshly transformed cells are cultured as above with the described media (Frieden 2004).

B. Purification

Recombinant proteins are purified under denaturing conditions in one step by taking advantage of the six histidine residues engineered onto the N-terminus of hILBP. The protocol is based on that recommended by the manufacturer for use with the His-SELECT Ni-NTA resin (Sigma). Cell pellets are stored at -80° before purification. To release the proteins from the cells, the frozen cell pellets are re-suspended in approximately 1.5x their mass of 8M urea buffer (8M urea, 50mM NaH\textsubscript{2}PO\textsubscript{4}, 100mM NaCl, 2mM imidazole, pH 8.0) and homogenized. The homogenate is then transferred to a one-quart size zip-top freezer bag and immersed in a dry ice/ethanol bath until the cells are frozen. The bag is quickly transferred to a 37°C water bath to thaw the cells. The freeze-thaw cycle is repeated for a
total of five cycles. The slurry is then transferred to a centrifuge tube and centrifuged for 20 min at 12,000 x gravity. The supernatant is transferred to a second centrifuge tube and the pellet re-suspended in fresh 8M urea buffer, and both tubes spun again. The combined supernatants contain the protein, with little possibility of any protein being trapped in inclusion bodies in the pellets due to the denaturing buffer conditions since the proteins are denatured.

The resin used for His-tag purification is the His-SELECT Ni-NTA resin from Sigma. A column is prepared from this resin by transferring 20mL resin to a 35mL plastic syringe, adapted with tubing and cap fittings and attached to a peristaltic pump. To equilibrate the column, the same 8M urea buffer is used to wash the column until the pH of the eluate matches the pH of the buffer. The combined supernatants are loaded onto the column, which is then washed with 8M urea buffer until the OD$_{280}$ of the eluate has stabilized. At this point, a gradient mixer is used to apply a steady gradient of buffer decreasing from 8M to 0M urea to the column, with a total gradient volume of approximately 150mL. The column is then thoroughly rinsed with 0M urea buffer (“wash buffer;” 50mM NaH$_2$PO$_4$, 100mM NaCl, 2mM imidazole, pH=8.0) to ensure complete removal of all denaturant. To remove any non-specific protein contaminants still bound to the resin, approximately two column volumes of wash buffer containing 10mM imidazole and then two column volumes containing 25mM imidazole are applied to the column. This is followed by a wash of buffer containing 100mM imidazole until the OD$_{280}$ stabilizes near zero, and a final wash with 250mM imidazole to ensure complete removal of all protein. Aliquots of the fractions collected from the column are run on a 4-20% gradient SDS-PAGE gel to determine the fractions containing pure protein. These fractions are pooled together and further processed for storage.
Protein concentrations are determined by optical density at 280 nm. The molar absorption coefficient for the histidine-tagged proteins was estimated using the calculation

\[ \varepsilon_{280}(M^{-1}cm^{-1}) = (\# \text{Trp})(5500) + (\# \text{Tyr})(1490) + (\# \text{Cystine})(125) \]

as described (Pace 1995). The value for hILBP(6His) is 11,460 M$^{-1}$cm$^{-1}$. Based on the calculated molecular weight of 16176.3 g/mol for wild-type protein containing the His-tag, the molar absorption coefficient was used to calculate the extinction coefficient of 0.708 mL/mg. Using a cuvette with a pathlength of one centimeter, the following calculation is then applied to determine the protein concentration in solution in units of milligrams protein per milliliter of solution:

\[ \text{OD}_{280} = (0.708 \text{ ml/mg}) \times c \]

or

\[ \text{Concentration (mg/mL)} = \frac{\text{OD}_{280}}{0.708} \]

To determine purity, the absorbance at 260 nm was also recorded, with \(\frac{\text{OD}_{280}}{\text{OD}_{260}}\) ratios of 1.5-2.5 considered to be pure.

C. Mutagenesis

Site-directed mutagenesis is performed according to the protocol available from the Quick-Change Site-Directed Mutagenesis kit (Stratagene). This technique utilizes a high-fidelity DNA polymerase in a thermal cycling reaction in which two complementary oligonucleotides are used to replicate the entire plasmid. The oligonucleotides are designed to span the region of the gene containing the target mutation site, with the sequences of the primers differing from the plasmid only at the desired mutation site. After thermal cycling
through a typical PCR protocol, the reaction is treated with a restriction enzyme, DpnI, that digests only methylated DNA, such that parental plasmid isolated from a dam+ bacterial strain will be digested, but the newly generated plasmids containing the mutation will not. A portion of the reaction is then transformed into a high transformation efficiency cloning bacteria, and isolated colonies analyzed for expression of the mutated plasmid. In some cases, a silent mutation could be introduced such that mutated plasmids also contained a change in a restriction site. In these cases, potentially mutated plasmids are analyzed by restriction digestion to diagnose the presence of the mutation. In the other cases, samples can only be verified by gene sequencing. In all cases, correct gene sequences are verified before proceeding with protein purification and experiments. A list of the oligonucleotide primers used to generate the mutants described here is found in table 3.1. The sequence identification codes for the correct mutant plasmids are also listed. All primers listed were generated by Anne Stanley in the Penn State College of Medicine, Macromolecular Core Facility. All sequencing reactions were performed by the Molecular Genetics Core Facility at the College of Medicine. The identification codes listed are those used by the core facility and recorded in the laboratory.
<table>
<thead>
<tr>
<th>Mutation in hILBP(6His)</th>
<th>Primer ID’s and sequences</th>
<th>Correct sequence ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe2Amber</td>
<td>IJR075 - CGAATTCCATGGCTTAGACCGGCAAGT&lt;br&gt; IJR076 – CGAACTTGGCCGGTCTAAGCCATGGAAT</td>
<td>IR229</td>
</tr>
<tr>
<td>Phe6Amber</td>
<td>IJR065 - GGCAAGTAGGAGATGGAGAGTGAG&lt;br&gt; IJR066 – CATCTCTCTACTTGGCCGGTCAAAGC</td>
<td>IR228</td>
</tr>
<tr>
<td>Phe17Amber</td>
<td>IJR067 - GATGAGTAGATGAAGCTCCTTGGGATATCC&lt;br&gt; IJR068 – GCTTCATCTACCTCATCATAATTTCTTC</td>
<td>IR189</td>
</tr>
<tr>
<td>Phe34Amber</td>
<td>IJR084 - CCCGCAACTAGAAGATCCTACCGG&lt;br&gt; IJR085 - CCCTGACGATCTTCTAGTTGCGGG</td>
<td>IR221</td>
</tr>
<tr>
<td>Phe47Amber</td>
<td>IJR077 - GGGCAAGGACTAGACTTGGTCC&lt;br&gt; IJR078 - GGGACCAAGTCTAGTGCTGCC</td>
<td>IR214</td>
</tr>
<tr>
<td>Phe63Amber</td>
<td>IJR073 - CCATGACCAACAAAGTACTGTGGGCAAGG&lt;br&gt; IJR074 - CTTGACCAACAGTCTACTTGGTGCTAGG</td>
<td>IR233</td>
</tr>
<tr>
<td>Phe79Amber</td>
<td>IJR079 - GGGCAAGACGAGTGAGAGGCCACTGTC&lt;br&gt; IJR080 - GCACAGTGGCCTCTACGTCTGCCC</td>
<td>IR222</td>
</tr>
<tr>
<td>Phe94Amber</td>
<td>IJR081 - GGTGTGAATTAGCCCAACTATCACCAG&lt;br&gt; IJR082 - CTGCCGATTAGTGCTGGCTACACC</td>
<td>IR240</td>
</tr>
<tr>
<td>His Tag</td>
<td>IJR089 - GGAGAAAATTAACCTATGCACGTGACGCCCATCACC&lt;br&gt; IJR090 - GGCGATGGCTCATTGGCTACGCTAGTTAATTTCTCC</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.1: Oligonucleotide Primers for Site-Directed Mutagenesis**

Oligonucleotide primers used for site-directed mutagenesis of the pRO.hILBP(6His) plasmid to incorporate the Amber codon (TAG) at the specified Phe residue. Primer and sequence identification codes are those used by the Molecular Genetics Core Facility and in the laboratory. Primers IJR089/IJR090 altered the histidine tag for improved soluble protein yield.
D. Reagents

All chemicals used were of reagent grade. p-F-Phe, 3-F-Tyr, and 6-F-Trp were obtained from Sigma. 30mM stocks of p-F-Phe and 10mm stocks of 3-F-Tyr and 6-F-Trp were prepared, sterile-filtered through a 0.22µM nylon membrane, and stored at 4° until use. Ultra pure grade urea was obtained from Amresco (Solon, Ohio). Stocks of 10M urea were prepared ahead of time and stored at -20°. Urea concentrations were determined by refractive index at 25° using a Milton-Roy Abbe-3 refractometer as previously described (Pace 1986). Briefly, refractive index of buffer was subtracted from that of the urea stock, and the difference used to calculate the urea concentration by using the equation described by Pace. Reagents for mass spectrometry matrix preparation were provided by the core facility.

E. Mass Spectrometry

To verify the correct protein size and to test for fluorine incorporation, purified proteins were analyzed by MALDI-TOF. Based on the amino acid sequence of the hILBP(6His) protein, the wild-type, unlabeled protein size is calculated to be 16176.3 g/mol. Each fluorine incorporated into the structure increases the molecular weight by 18 g/mol. The MALDI-TOF analyses were performed using a reflectron, delayed extraction MALDI-TOF mass spectrometer (Voyager DE-Pro, Perspective BioSystems / Applied BioSystems) at the Penn State College of Medicine, Mass Spectrometry and Proteomics Core Facility. Matrix was freshly prepared according to the core facility protocol by washing 10mg sinapinic acid (3,5-Dimethoxy-4-hydroxycinnamic acid) with hexane and then diluting with 300µl 1%TFA,
200µl dH₂O, and 500µl acetonitrile, mixing, then centrifuging to clarify. This is similar to the protocol published by Moore (1997), with the variation of using sinapinic acid in place of α-cyano-4-hydroxycinnamic acid. Protein samples were prepared by diluting stock protein solutions in standard buffer, pH 8.0, to 10mM in water, pH 8.0. 1µl samples of protein were spotted onto clean MALDI plates and allowed to dry. The dried spots were overlaid with approximately 1µl of the freshly prepared matrix, and dried again. At least three data points were collected for each protein spot.

F. Equilibrium Measurements by Fluorescence

Sample preparation for the fluorescence measurements was done by preparing stock solutions of 0.1 mg/ml protein in either 9M urea buffer (10nM Na₂HPO₄, 15mM NaH₂PO₄, 75mM NaCl, 0.1mM EDTA, 9M urea) or standard buffer (10nM Na₂HPO₄, 15mM NaH₂PO₄, 75mM NaCl, 0.1mM EDTA), both at pH 8.0. The samples at all urea concentrations measured were then prepared by dilution of the stock solutions into each other using a Hamilton titrator programmed to prepare 300ul samples from 0M and 9M stock solutions. All samples were equilibrated at 25° for a minimum of one hour before use. To observe fluorescence, 300ul samples were transferred to a 0.1cm cuvette placed in the Photon Technology International spectrophotometer with FeliX32 software package and thermostated cell at 25° and incubated for at least two minutes to ensure temperature equilibration. The excitation wavelength was set at 290nm with a 5nm band pass and the emission collected over the range of 305-400nm with a 6nm band pass. Emission at 360nm was recorded and used in all analyses, although emission at 340nm and 350nm was also
recorded and analyzed to confirm the observations. Data were analyzed by nonlinear least-squares fits generated using KaleidaGraph (Synergy Software) as described (Burns 1998). The equation used to fit the curves was adapted from Santoro and Bolen (Bolen 1988) as described (Burns 1998).

**G. NMR**

Samples for $^{19}$F-NMR were prepared by dialyzing the purified protein against a dilute NMR buffer, pH 8.0. Buffer dilutions were determined based on stock protein concentration to allow approximately 8.3mg of protein to be lyophilized and re-suspended in a final volume of 1mL at a 1x buffer concentration (standard NMR buffer: 10nM Na$_2$HPO$_4$, 15mM NaH$_2$PO$_4$, 75mM NaCl, 0.1mM EDTA, pH 8.0). Two such aliquots of each labeled protein were prepared, and lyophilized stocks stored at $-20^\circ$ until use. On the day of data collection, the stocks were re-suspended in either 1mL H$_2$O (10% D$_2$O) or 1mL 8M urea (10% D$_2$O). After data collection at 0M and 8M urea, samples were recovered and serially diluted into each other to obtain the desired urea concentrations. Samples were equilibrated for at least 5min at 25$^\circ$ before data collection. NMR data were collected essentially as described for IFABP (Ropson 2006). A Bruker Avance 500 digital NMR spectrometer was used in conjunction with a $^{19}$F/$^1$H dual channel probe capable of $^{19}$F detection with $^1$H decoupling. Sweep width was set at 35ppm, and at least 256 transients were collected for each sample. A delay times of 2.5 seconds and acquisition time of 1 second were chosen to ensure accurate integration based on separate experiments to determine T1 and T2 relaxation times. 6-F-Trp was used as the reference standard in all experiments, and all reported chemical shifts were in
parts per million (ppm) from this reference. Data analysis was performed using the MacNUTS NMR Utility Transfer Software package from Acorn NMR. Relative integrated peak intensities were calculated based on a standard concentration of 0.25mM 6-F-Trp for each sample.
Chapter 4: Results

A. Incorporation of $^{19}$F into ILBP

a. Verification of $^{19}$F Incorporation by Mass Spectrometric Analysis

In order to verify that the fluorine nucleus had been incorporated into the proteins before NMR analysis, samples of purified protein were compared by MALDI-TOF mass spectrometry. Incorporation of fluorine in place of hydrogen increases the molecular weight by 18 g/mol for each fluorine incorporated. Analysis by SDS-PAGE had been completed on each sample before MS analysis to determine purity, so samples were compared for differences in molecular weight without calibrating the instrument for exact molecular weight. As shown in the spectra in figure 4.1, the observed molecular weight for a protein containing only one fluorine, F34$p$-F-Phe, was 27.18 g/mol greater than unlabeled protein. Labeling at each of the four tyrosine residues resulted in 81.21 g/mol higher than unlabeled, and labeling at each of the eight phenylalanine residues results in 147.84 g/mol higher than unlabeled protein. These values corresponded well with the expected increases of 18 g/mol for one fluorine, 72 g/mol for four fluorines, and 144 g/mol for eight fluorines. The differences in molecular weight for the single-site $^{19}$F-incorporation variants were all slightly higher than the expected value of 18 g/mol, which was most likely due to the partial background labeling of other Phe residues, the “contamination” labeling described by Furter (1998). This is also observed in the NMR spectra, as described below. In no case were the values for the fluorine-containing proteins less than for the unlabeled protein, and as such the preparations were considered suitable for use in $^{19}$F-NMR experiments.
Figure 4.1: Representative MALDI-TOF Spectra

MALDI-TOF spectra for (A) hILBP(6His); (B) hILBP(6His)-Phe34-p-F-Phe; (C) 3-F-Tyr-hILBP(6His); and (D) p-F-Phe-hILBP(6His). Molecular weights of the prominent peak are indicated. Differences in molecular weight are as expected for the degree of $^{19}$F-labeling.
b. Effects of $^{19}$F on Equilibrium Unfolding by Fluorescence Spectroscopy

Equilibrium unfolding was characterized by fluorescence spectroscopy. The unfolding profiles the wild-type protein containing the N-terminal histidine tag and the protein labeled with $p$-F-Phe were compared. Data were analyzed as previously described (Burns 1998). Both proteins fit best to a two state folding model. Incorporation of the $p$-F-Phe label at all eight Phe sites in ILBP slightly destabilized the protein, reducing the midpoint for denaturation from $3.54 \pm 0.02$ M urea for the unlabeled protein to $3.10 \pm 0.03$ M urea for the labeled ILBP. This is in contrast to previous observations for IFABP in which incorporation of $^{19}$F into aromatic residues resulted in slight increases in the urea denaturation midpoint, although in both proteins the influence of the label is small (See Ropson 2006 for example). However the $\Delta G_{H_2O}$ (the extrapolated stability of the protein in the absence of denaturant) and $m_G$ (the dependence of $\Delta G$ on denaturant concentration) values for ILBP were both decreased in response to $^{19}$F-incorporation, as in IFABP. Ropson (2006) suggests that these changes may be due to the increased hydrophobicity of the fluorinated phenylalanine residues compared to their unlabeled counterparts. The equilibrium intermediate may be slightly stabilized by this increase in hydrophobicity, which could decrease the cooperativity of the unfolding transition for the $^{19}$F-labeled protein (Ropson 2006).
<table>
<thead>
<tr>
<th>Protein</th>
<th>Midpoint (M)</th>
<th>ΔG_{H2O} (kcal mol$^{-1}$)</th>
<th>m_G (kcal mol$^{-1}$ M$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hILBP(6His)</td>
<td>3.53 ± 0.02</td>
<td>7.81 ± 0.61</td>
<td>-2.21 ± 0.17</td>
</tr>
<tr>
<td>P-F-Phe-hILBP(6His)</td>
<td>3.14 ± 0.03</td>
<td>4.66 ± 0.34</td>
<td>-1.49 ± 0.11</td>
</tr>
</tbody>
</table>

*Figure 4.2: Unfolding of ILBP and p-F-Phe-ILBP by Fluorescence*

Unfolding curve for hILBP(6His) with no label (open symbols) and labeled with p-F-Phe (closed symbols) fit to a two-state model (solid lines). The midpoint for unfolding in the absence of $^{19}$F is 3.54M, and with the p-F-Phe label at all eight Phe sites is 3.10M. Midpoint, ΔG_{H2O} and m_G values and errors are listed.
B. ¹⁹F-NMR Analysis of ¹⁹F-labeled ILBP

a. p-F-Phe-ILBP

Unfolded ¹⁹F-NMR spectra for ILBP uniformly labeled with fluorine at each of the eight phenylalanine residues are shown in figure 4.3. The eight residues show well-resolved peaks in the native spectra that converge into four peaks at 4M urea and resolve into six peaks by 8M urea. The tentative native peak assignments made by site-specific fluorine incorporation are listed. No native peaks are observed at urea concentrations above 4M. Analysis of the total areas for all peaks at high denaturant concentrations indicates that there is no significant change in the integrated peak intensity of the unfolded peaks between urea concentrations of 4M and 8M (see figure 4.4). This suggests that the only conformation present in significant concentrations is that of the unfolded protein, with no intermediate conformation in slow or intermediate chemical exchange with the native or unfolded state. However, it is clear from the spectra that all peaks do not show the same type of chemical shift behavior. Some combination of the short-range interactions with neighboring atoms and the local electric fields is responsible for changes in chemical shift behavior due to shielding (Gerig 2001). Analysis of the electrostatic field of each p-F-Phe due to the local sequence does not reveal differences that account for the chemical shift behavior, as discussed below. It is therefore likely that the chemical shift differences observed among the eight Phe residues are a result of differences in inter-atomic interactions experienced by each Phe due to residual tertiary contacts at high denaturant concentrations.
$^{19}$F-NMR spectra of native 4FPhe-ILBP(6His). The chemical shift is from 6FTrp. All native-like peaks have shifted to unfolded conformation by 4M urea. The chemical shift is from 6FTrp. At least two peaks show much greater peak shift behavior than the other peaks.
**Figure 4.4: Total Relative Peak Area and Individual Peak Shifts for p-F-Phe-ILBP**

Bar graph shows the total relative peak area for the sum of all unfolded-like peaks over the indicated urea concentration range. There is very little change in the total area above 4M urea, at which point no native-like peaks are observed, which indicates that there are no conformations in slow- or intermediate-exchange with the unfolded conformations. The lines indicate the peak shift behavior of the observed resonances. Two peaks (B and E) appear to show different chemical shifts than the other four peaks, but due to the presence of overlapping peaks at all concentrations, the differences in chemical shift are difficult to determine.
b. 3-F-Tyr-ILBP

The $^{19}$F-NMR spectra for unfolded ILBP labeled with fluorine at each of the four tyrosine residues are shown in figure 4.5. The four native state peaks are completely absent at urea concentrations above 4.1M. Three peaks are visible in the unfolded spectra, with the center peak having approximately twice the area and intensity of the two other peaks, representing two overlapping peaks. The areas of the three peaks do not change significantly after the disappearance of the native state peaks, indicating that there is no other conformation present in slow or intermediate exchange with the unfolded conformation. Chemical shift analysis indicates that at least one of the three peaks exhibits slightly different shift behavior. The most up-field peak continues to shift slightly between urea concentrations of 4M and 8M, as in table 85. This behavior indicates that at least one of the four tyrosine residues experiences a different environment upon urea unfolding than the other three residues. These observations are not similar to those reported for the four tyrosine residues in IFABP. The resulting $^{19}$F-NMR spectra for the native and the unfolded states of 3FTyr-ILBP are very similar to those for 3FTyr-IFABP, however no tyrosines in ILBP shows the “missing intensity” behavior reported for IFABP over the range of 4M to 8M denaturant (Ropson 2006).
Figure 4.5: $^{19}$F-NMR Spectra for 3-F-Tyr-ILBP:

(A) $^{19}$F-NMR spectra of 3FTyr-ILBP(6His) at the indicated urea concentrations. The chemical shift scale is in ppm from 6FTrp. Peaks that are present in the native state have shifted completely above 4M urea. (B) Spectra of the unfolded state peaks of 3FTyr-ILBP(6His) at the indicated concentrations of urea. The chemical shift scale is in ppm from 6FTrp. The most upfield peak shows slightly different shift behavior at high denaturant concentrations than the other peaks.
Corrected relative peak area and peak shift for hILBP(f-Tyr)

\[ y = 0.0068x + 15.301 \quad R^2 = 0.9124 \]
\[ y = 0.0033x + 15.196 \quad R^2 = 0.9292 \]
\[ y = 0.0026x + 15.117 \quad R^2 = 0.9037 \]

Figure 4.6: Total Relative Peak Area and Individual Peak Shifts for p-F-Tyr-ILBP:

Chemical shifts of the three tyrosine peaks plotted against denaturant concentration. Slopes of the line, shown as the coefficient of \( x \) in the equations, indicate degree of shift over the [urea] range. Peak B represents the largest of the three peaks, which accounts for two tyrosines whose spectra overlap.
c. Single-Site $^{19}\text{F}$ Incorporation at ILBP Phenylalanine Residues

Incorporation of 4FPhe at single sites within the structure according to the protocol outlined above resulted in yields of approximately 10-20mg protein per liter of media. The single-site labeled proteins showed one major $^{19}\text{F}$-NMR peak, with some slight background labeling of the remaining phenylalanine residues. This is the “contamination” described in the original publication of this system (Furter 1998), and, as pointed out by Li (2005), is actually useful in the assignment of the peaks in the fully-labeled spectra. $^{19}\text{F}$-NMR spectra were collected for each of the single-site variants over a range of urea concentrations, and the resulting spectra aligned reasonably well with the spectra observed for the protein labeled with fluorine at all eight phenylalanine sites. Each of seven native spectra for the single-site fluorine labeled proteins are shown in figure 4.6 aligned with the native spectrum for fully $p$-$^{19}\text{F}$-Phe-labeled-ILBP. No native spectrum was collected for the protein labeled with $p$-$^{19}\text{F}$-Phe at Phe6. Figure 4.7 shows the overlaid single-site native state spectra in alignment with the fully labeled $p$-$^{19}\text{F}$-Phe-ILBP spectrum. Phe47 and Phe63 appear to be slightly de-shielded in the single site preparation compared to the protein labeled with fluorine at all eight sites, since the single-labeled peaks are shifted downfield from the corresponding peaks in the fully-labeled protein. The fluorine nucleus is exceptionally sensitive to its local environment, and these two residues are in close proximity to each other in the native state, which explains the additional shielding provided by the presence of two fluorine nuclei in this region. It is possible that the Phe2 peak is also downshifted in the single-site labeled protein compared to the wild-type, which could also be explained by the close proximity of the Phe2 aromatic ring to those of Phe47 and Phe63. Further experiments would be required to definitively assign these native state resonances.
Figure 4.7: Native State Spectra for Single-Site p-F-Phe-labeled ILBP Variants

Native-state spectra for each of seven single-site p-F-Phe-labeled ILBP variants, aligned with the native-state spectrum for p-F-Phe-ILBP. Spectra are shown individually (A) and overlapped to reproduce the fully-labeled spectrum (B). Chemical shift is from 6-F-Trp. No spectra are shown for the protein labeled individually at Phe6.
Phe2 and Phe6 are located on β-strand A (see figure 5.1). Phe2 is close to the N-terminus of the wild-type protein, but in the variant of ILBP used in these studies, the native N-terminus is blocked by the His-tag sequence. If Phe2 were close to the end of the protein, it might be expected to be solvent-exposed more quickly than the other residues with increasing urea concentration, and behave more like a free amino acid. However when this was tested for IFABP-Phe2, it was found that this residue behaved similarly to other phenylalanines (Li 2005). In the ILBP studies presented here, Phe2 and Phe6 behave similarly, in that neither appears to show significant changes in $^{19}$F-NMR peak shape in response to increasing levels of urea, similar to the results for IFABP. For this reason, the His-tag is assumed to have little influence on the unfolding behavior of the nearby phenylalanine residues. Integration of the $^{19}$F-NMR peaks for Phe2 and Phe6 reveals that no intermediate conformation in slow or intermediate exchange with the unfolded state is present, and the minimal chemical shift changes for these two residues indicates that no significant intermediate in fast exchange with the unfolded state is present either. These results imply that neither Phe2 nor Phe6 maintains residual structure on unfolding.

Two of the Phe residues, Phe17 and Phe34, are structurally located within the helix-turn-helix motif of ILBP. No native-like resonance for either Phe17 or Phe34 was observed at urea concentrations above 4M. Phe17 is located within the center of the first helix. Interestingly, the unfolded resonance for this residue shifted more over the range of 4M to 8M urea than any other phenylalanine residue. On the other hand, the peak shifts for Phe34 were minimal over this range. Phe34 is located in the short linker region between the helices and β-strand B. No other aromatic residue is present in the helical region, although three of the four tyrosines (Tyr14, Tyr53, and Tyr119) in the structure are located in the β-barrel
close to the helices. As described above, it does not appear that Phe17 has any interaction with these tyrosine after native structure has been denatured, since no tyrosine shows similar peak shift behavior as Phe17. Phe34 has less chemical shift over the range of high denaturant concentrations than any other residue analyzed. Therefore, the interaction between the helical region and the β-barrel is not maintained in an early folding intermediate, although some structural element may be present within the first helix in the region of Phe17.
Figure 4.8: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe2.

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shifts very little over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Figure 4.9: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe6.
The unfolded spectra are depicted over a range of denaturant concentrations. The major peak shifts very little over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Figure 4.10: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe17

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shifts significantly over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Figure 4.11: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe34

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shifts very little over the range of urea concentrations. The minor peaks in the 6M and native spectra are artifacts.
Previous analysis of IFABP has revealed the presence of a folding initiating site within a specific hydrophobic core region of the structure (See Ropson 2006 for eg.). Phe79 and Phe94 are in the structurally homologous region in ILBP. Phe94 corresponds to IFABP-Phe93. Phe79 is located at the N-terminus of $\beta$-strand F, and is in close proximity to Phe94. In previous $^{19}$F-NMR studies of IFABP, residues in this region continued to show changes in line shape and peak characteristics over the range of high denaturant concentrations, contributing to the hypothesis that this region maintains some elements of native-like tertiary structure late in unfolding and may therefore be an initiating site for folding. In the current analysis of ILBP, however, neither of these two residues showed the characteristics of residual structure at high denaturant concentrations. The unfolded peaks for Phe79 and Phe94 are shown in figures 4.12 and 4.13. The IFABP experiments indicated that Phe93, (analogous to Phe94 in ILBP) as well as three other nearby IFABP aromatic residues, Trp82, Tyr70, and Phe68, have an additional conformation in intermediate exchange with the unfolded state over the range of 4M to 6M urea, where the protein is approximately 75% unfolded (Ropson 2006). This was manifested as a loss of intensity of the unfolded $^{19}$F-NMR peaks for these residues over this denaturant range, indicating the presence of the intermediate. In contrast, no such phenomena were seen for the Phe94 or Phe79 residues in ILBP upon incorporation of fluorine at these sites. No native-like peaks were observed for ILBP-Phe94 or ILBP-Phe79 above 4M urea, and the unfolded state peaks showed very slight changes between 4M and 8M denaturant, indicating that these two residues do not appear to participate in any residual structure at high denaturant concentrations.
Figure 4.12: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe79

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shifts significantly over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Figure 4.13: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe94

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak does not shift over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Phe47 and Phe63 also correspond to residues that cluster in the hydrophobic core region of IFABP. Other than Phe17, these two residues show the most significant changes in $^{19}$F-NMR resonance over the high range of denaturant concentrations. Although there is no evidence supporting the presence of two separate conformations in either slow or intermediate exchange for these two residues, there must be some factor influencing the environment of Phe47 and Phe63 differently than the other Phe residues. As discussed below, analysis of the local primary structure did not reveal obvious differences in the local sequences surrounding these residues that can account for the observed chemical shift differences. The changes in chemical shift could be due to these two residues experiencing different types of intramolecular tertiary contacts resulting in different levels of solvent exposure in response to urea, affecting the chemical shifts. Both residues are located in the hydrophobic core analogous to the core region in IFABP, however neither IFABP homologue was definitively identified to be involved in the early folding intermediate (Ropson 2006). ILBP-Phe47 and Phe63 do not exhibit exchange on an intermediate time scale like the IFABP core residues, but the changes in chemical shift at high denaturant concentrations indicate the presence of some population of conformations in the region of these two residues that is not present in other structural regions early in the ILBP folding pathway.
Figure 4.14: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe47

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shows significant shift over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
Figure 4.15: $^{19}$F-NMR spectra for ILBP labeled with p-F-Phe at Phe63

Native spectrum is shaded in gray, with the unfolded spectra over a range of denaturant concentrations expanded above. The major peak shows significant shift over the range of urea concentrations. The minor peaks are due to background labeling of other phenylalanine residues.
C. Chemical Shift Analysis of Collective $^{19}$F-NMR Data

Plotting the chemical shift of the $^{19}$F-NMR peaks for each of the single-site labeled proteins over the range of 4M to 8M urea reveals a linear relationship. The steeper the slope of the line, the more the resonances shift. If each phenylalanine were experiencing the same type of environment, the slopes could be expected to overlay, which would indicate that all regions of the protein examined by the $p$-F-Phe probes unfold uniformly. However, as shown in figure 4.16, the slopes are greater for residues Phe17, Phe47, and Phe63 than for the other phenylalanine. The chemical shifts from 4M to 8M urea are illustrated in figure 4.16 for each of the single-site labeled proteins. Although they have not been assigned, the slopes for the three distinct tyrosine peaks are shown in figure 4.6. The slopes of the linear fits to each line are listed in increasing order in table 4.1. In order to exclude solvent effects, spectra were collected over the same urea concentration range for solutions containing equimolar amounts of $p$-F-Phe and the reference 6-F-Trp. The chemical shift for the free amino acid between 4M and 8M urea was less than 0.005ppm.
Figure 4.16: Chemical Shifts of Single-Labeled ILBP Variants

Chemical shift peaks plotted against urea concentration for each of the eight individually $p$-F-Phe-labeled ILBP proteins over the range of 4M to 8M urea. Slopes of the best fit lines are listed in Table 4.1.
Table 4.1: Slopes of the Best-Fit Lines for Unfolded Chemical Shifts

Slopes of the best-fit lines for plots of chemical peak shift against urea concentration for the range of 4M to 8M urea, as shown in figures 4.16 and 4.6. Slopes are listed in increasing order. One of the tyrosine peaks represents two tyrosine residues whose spectra overlap.
Chapter 5: Discussion

A. $^{19}$F-NMR Equilibrium Unfolding Study of ILBP

Preliminary $^{19}$F-NMR studies of ILBP lead to the observation that chemical shift changes in response to urea are not uniform for all residues. This observation corresponded with available data on the related IFABP folding behavior, in which certain aromatic residues which are in close proximity in the native state show non-uniform changes in response to denaturant. In order to determine whether the ILBP residues with this behavior are also neighbors in the native state, and thus whether a folding initiation mechanism is conserved between these two proteins, further work was needed to identify and monitor the aromatic residues in ILBP individually for response to denaturant by NMR. The peaks that continue to change at high denaturant concentrations potentially represent residues that are involved in an early folding intermediate. This early folding intermediate has not been well characterized in kinetic studies for ILBP, however the other spectral methods require that an intermediate be stable enough to accumulate to a higher concentration than is required by $^{19}$F-NMR. Previous $^{19}$F-NMR studies in IFABP had revealed that residues involved in formation of an early folding intermediate also showed non-uniform changes in response to urea in $^{19}$F-NMR spectra. The IFABP residues exhibiting the non-uniform behavior were assigned to the protein’s hydrophobic core, which has been proposed to be the location of the folding initiation site. To determine whether the potential intermediate in the ILBP is formed by interactions in the hydrophobic core of this structurally homologous protein, the experiments presented here were designed to assign the $^{19}$F-NMR resonances to their respective residues.
$^{19}$F-NMR spectra of proteins labeled with only one $p$-F-Phe are well-aligned with the spectra of the protein labeled at multiple Phe and confirm that the chemical shifts over the range of high denaturant concentration differ among the phenylalanine residues.
Figure 5.1: Structure of ILBP

Diagram of human ILBP (PDB ID: 1o1u) with β-strands labeled A-J and helices labeled I and II.
a. Chemical shift analysis

The chemical shifts for each of the single-site $p$-F-Phe-labeled proteins are plotted against urea concentration (see figure 4.16). A similar plot for each of the three separate peaks for 3-F-Tyr is shown in figure 4.6, although these are not assigned to their corresponding residues. There is a linear relationship between chemical shift and urea concentration over this denaturant range, with the slopes of the lines indicating the degree of chemical shift. Table 4.1 lists the absolute values of the slopes of the 11 lines in increasing order. Lower values indicate that the peak shifted very little, and higher slope values indicate a greater degree of chemical shift over the high range of denaturant concentration. It is noteworthy that all peaks do not have the same degree of chemical shift behavior, which indicates that different residues experience different environments on unfolding. As a control, $^{19}$F-NMR spectra of free $p$-F-Phe in solution containing the same 6-F-Trp reference standard were collected, and the chemical shift also showed a linear relationship to urea concentration. The slope of the line for the reference plot is 0.0002, indicating virtually no change in chemical shift over the range of 4M to 8M urea, which excludes instrumental effects as causative of the chemical shift behavior for the ILBP residues.

b. Phe2, Phe6, Phe34, Phe79, and Phe94

Five of the Phe residues have little chemical shift over the range of 4M to 8M urea (see table 4.1). Phe34 has the least amount of chemical shift in this range. Phe6 and Phe94, as well as three of the Tyr residues, have almost identical chemical shift behavior, but only slightly more than Phe34. Phe79 and Phe2 have slightly higher chemical shifts, and these two residues are similar in shift behavior to each other. The integrated peak intensity for Phe79
appears to increase moderately in peak area with increasing denaturant concentration, which could indicate the presence of another conformation in chemical exchange at this site, however line width analysis reveals that the peaks are not broadened at the lower concentrations, and no other peak was observed in these spectra, which eliminates chemical exchange on both the intermediate and slow exchange scales. Further, analysis of the total integrated peak intensity for the protein labeled with $^{19}$F at all Phe sites shows that total peak area for the combined eight resonances does not change over this range, so the observation for Phe79 is most likely an artifact and not indicative of an alternative protein conformation at this site. Based on the chemical shift behaviors, these five Phe residues appear to unfold uniformly and experience the same degree of solvent exposure in response to urea, indicating that they are not likely to be involved in any region of residual structure early in ILBP unfolding.

c. Phe17

Phe17 shifted more than any other aromatic residue in response to denaturant concentration. Phe17 is located within the first $\alpha$-helix and is in close proximity to Tyr14 and Tyr119 in the native state. Experiments published on variants of IFABP in which the helical region has been removed (Ogbay 2003, Kim 1996) revealed that the helices in IFABP have little effect on formation or stability of the $\beta$-barrel structure. Two separate variants were engineered, one in which the helices were replaced with a long Gly-Ser motif ($\Delta$17-SG), and the second in which the helices were substituted with a two-Gly linker ($\Delta$27-GG). Both variants had as much $\beta$-sheet content as the wild-type protein. Slight differences in ligand binding properties were observed, but the only overall structural differences between the
variants and wild-type IFABP were in the helical region (Ogbay 2003). Because of the structural similarities between the wild-type and helix-less IFABP proteins, the helices were suggested to play little if any role in the mechanism of the β-barrel folding, and ILBP-Phe17 was not expected to show any residual structure early in folding.

A recent study of the similar helix-less mutation in ILBP showed that the helices influence this structure much differently than IFABP. Here the two helices were replaced with a four-residue Gly-Ser linker, similar to the Δ27-GG mutation in IFABP, but the resultant protein was unfolded under physiological conditions (Kouvatsos 2006). Addition of ligand to the helix-less protein resulted in recovery of the structure, but the apo-protein conformation had NMR and CD properties similar to a disordered peptide. To date, no studies on the influence of the helices on the folding mechanism of ILBP have been published. The observation by Kouvatsos that removal of the helices results in a disordered apo-structure suggests that the helical region in ILBP plays a significant role in the stability and possibly in the folding of this protein. The behavior of Phe17 upon unfolding presented here lends support to this suggestion. However, this is not supported by the observations for the tyrosine residues or for Phe34. Three of the tyrosines are in the β-barrel in close proximity to the helices, and Phe34 is in the linker region between the helices and the second β-strand. If it were the case that Phe17 were involved in an early folding intermediate, as might be suggested by the chemical shift behavior, it is likely that other nearby aromatic residues would show similar behavior. Yet Phe34 shows less chemical shift than any aromatic residue over the high denaturant range, and of the four tyrosines in the protein, only one shows greater than average shift (See table 4.1). Therefore Phe17 does not appear to be interacting with any of these residues at high urea concentrations. Of all aromatic residues in
ILBP, however, Phe17 is the only one in a helix. The conserved helical region in CRABP I was shown to form secondary structure independently when a short peptide corresponding to this region was analyzed (Sukumar 1997), however helical secondary structure is not expected to persist at high urea concentrations. The roles of Phe17 and of the \( \alpha \)-helical region in ILBP folding remain unclear.

d. Phe47 and Phe63

ILBP-Phe47 is homologous to IFABP-Phe47, which was originally proposed to be involved in the initiation site (Ropson 1992) but more recent evidence suggested that it may not be appropriate to include this residue in the hydrophobic core responsible for folding initiation (Li 2005). In those studies, it was demonstrated that the resonance for Phe47 has a more narrow line width than for the other phenylalanine residues in IFABP, which indicates a higher degree of motional freedom than the other phenylalanines, and that the characteristics of the chemical shift during denaturation are also different than those of phenylalanines in the hydrophobic core. This suggests that the changes in Phe47 in IFABP are independent of the changes in the hydrophobic core (Li 2000b).

ILBP-Phe63 is a homologue for IFABP-Phe62, which is also suggested to participate in the early folding intermediate in IFABP (Ropson 1992, Li 2005Biochem). The chemical shift changes between 4M and 8M urea in the ILBP experiments are greater for Phe63 and Phhe47 than for any other Phe, with the exception of Phe17. This could be explained by residual structure persisting in this region at denaturant concentrations for which the majority of the structure has become unfolded. The two residues are in close contact in the native protein, so any structure that they both participate in could be a native-like structure.
Evidence of their interaction is seen in the native state spectra, in which the single-site labeled protein peaks are significantly de-shielded in comparison to the corresponding peaks in the individual $p$-F-Phe-labeled ILBP spectra. The de-shielding is a consequence of the close proximity of the two fluorine groups in the fully-labeled protein being eliminated in the single-site labeled proteins. Interaction between these two residues is conserved across all members of the iLBP family (Gunasakaran 2004). This interaction may play a role in initiation of protein folding, and may represent an evolutionarily conserved folding mechanism. It remains to be proven, though, that the two residues are in contact in the early folding intermediate for ILBP or IFABP.

e. Explanation of Chemical Shift

Chemical shift in protein $^{19}$F-NMR spectra is caused by local electrostatic field and non-covalent interactions with nearby atoms. The fluorine environment in a labeled protein changes significantly from the native to the denatured state, with the major influences on fluorine shielding coming both from changes in electric field, and from intermolecular interactions with the solvent or with remaining contacts to other atoms in the protein (Gerig 2001, Danielson 1996). The changing solvent environment and alterations in the interactions with other protein atoms upon unfolding results in the dramatic chemical shift between the native and unfolded forms. As solvent exposure is not expected to differ for a $^{19}$F on a phenylalanine that is fully solvent-exposed, the changes in chemical shift observed at high urea concentrations are due either to the electrostatic environment of the local primary structure or to longer-range intramolecular interactions.
To determine whether the differences in chemical shift observed in ILBP were influenced by electrostatic field or by atomic interactions, the electrostatic environment of the local primary structure at each phenylalanine and tyrosine residue was examined (table 5.1). The electrostatic field farther than two residues away is not expected to have a significant effect; therefore sequences two amino acids up- and down-stream of each aromatic residue were analyzed for polarity and charge. No obvious relationship between local electric field and chemical shift emerges from the analysis. The residue that exhibits the greatest change in chemical shift over the range of 4M to 8M urea, Phe17, is surrounded by more charged residues than the other phenylalanine sites, however Tyr14 also has a similar charge environment as well as more polar neighbors, yet no tyrosine resonance shows as much shift as Phe17. In an earlier study to determine the influence of primary structure on $^{19}$F-NMR signal in response to urea, Ropson generated five-amino acid peptides corresponding to the residues two amino acids up- and downstream from the tryptophan residues in IFABP. $^{19}$F-NMR analysis of the 5-mer peptides revealed no difference in the dependence of the chemical shift on urea concentration due to local sequence effects (Ropson, unpublished results). Although analysis of the 5-mers for ILBP listed in table 5.1 was not done, there is no evidence to suggest that the behavior of these peptides should be different than that observed for the IFABP peptides. It is unlikely that differences in the primary structure surrounding each phenylalanine or tyrosine residue account for the differences in the dependence of chemical shift on urea concentration, which implicates long range intra-molecular interactions as more likely to cause the observed chemical shift behavior.
<table>
<thead>
<tr>
<th>Residue</th>
<th>5-mer sequence</th>
<th>Electrostatics</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe34</td>
<td>RNFKI</td>
<td>P+ P F P+ Np</td>
<td>0.0018</td>
</tr>
<tr>
<td>Phe6</td>
<td>GKFEM</td>
<td>Np P+ F P- Np</td>
<td>0.0031</td>
</tr>
<tr>
<td>Phe94</td>
<td>VNFPN</td>
<td>Np P F Np P</td>
<td>0.0033</td>
</tr>
<tr>
<td>Phe79</td>
<td>KTFKA</td>
<td>P+ P F P+ Np</td>
<td>0.0053</td>
</tr>
<tr>
<td>Phe2</td>
<td>MAFTG</td>
<td>Np Np F P Np</td>
<td>0.0056</td>
</tr>
<tr>
<td>Phe63</td>
<td>NKFTV</td>
<td>P P+ F P Np</td>
<td>0.0074</td>
</tr>
<tr>
<td>Phe47</td>
<td>QDFTW</td>
<td>P P- F P Np</td>
<td>0.0104</td>
</tr>
<tr>
<td>Phe17</td>
<td>DEFMK</td>
<td>P- P- F Np P+</td>
<td>0.0158</td>
</tr>
<tr>
<td>Tyr14</td>
<td>KNYDE</td>
<td>P+ P Y P- P-</td>
<td></td>
</tr>
<tr>
<td>Tyr53</td>
<td>QHYSG</td>
<td>P P+ Y Np Np</td>
<td></td>
</tr>
<tr>
<td>Tyr97</td>
<td>PNYHP</td>
<td>Np P Y P+ P</td>
<td></td>
</tr>
<tr>
<td>Tyr119</td>
<td>VTYER</td>
<td>Np P Y P- P+</td>
<td></td>
</tr>
<tr>
<td>Trp49</td>
<td>FTWSQ</td>
<td>Np P W Np P</td>
<td>0.0054 **</td>
</tr>
</tbody>
</table>

**Table 5.1: Local Primary Structure of Aromatic Residues in ILBP**

The 5-mer peptide sequences surrounding the aromatic residues in ILBP, along with a description of the electrostatic character of the residues. +/- indicate a positively or negatively charged residue. ‘Np’ indicates a non-polar and ‘P’ a polar residue. The chemical shift behavior is represented by the slopes listed in Table 4.1 for Phe residues. No slopes are listed for tyrosine residues since they have not been assigned, however the slopes for the three tyrosine peaks observed are 0.0026, 0.0033, and 0.0068.

** The Trp49 data is an unpublished result from Boyer and Ropson.
f. Interpretation of Results and Comparison to Previous Results for IFABP

The data presented here for ILBP do not indicate the presence of multiple conformations in slow or intermediate exchange with the unfolded state at urea concentrations above 4M, concentrations at which the protein is shown to be unfolded by fluorescence measurements. However, simple visual inspection of the spectra of \( p\)-F-Phe-ILBP and 3-F-Tyr-ILBP reveal that not all residues exhibit the same types of resonance changes as denaturant concentration increases. The observation that some resonances exhibit more dramatic chemical shifts than others indicates that the corresponding residue side chains experience different solvent exposure, potentially due to stabile long-range interactions with other atoms in the molecule for those residues. The residual structure observed in IFABP is located in the hydrophobic core region of the protein. The homologous region in ILBP contains Phe79, Phe94, Phe63, and Tyr97 (see figure 5.2). Additionally, Phe47 may be included in this region for ILBP, although it is not clear whether the homologous residue in IFABP is definitely included in the initiation site (Li 2005, Ropson 2006). While the residues corresponding to ILBP-Phe79 and ILBP-Phe94 showed more than one conformation in the \(^{19}\)F-NMR analysis of IFABP, neither residue showed this behavior in ILBP, with negligible change in peak area and chemical shift over the range of urea concentrations from 4M to 8M.

Phe47 and Phe63 may come together earlier in folding than other regions of the protein. The two residues are both at the ends of \( \beta \)-strands and physically close together in the native state, but are not close to the same \( \beta \)-turns. The turns between \( \beta \)-strands have been presented as a possible site for folding initiation (Muñoz 1998), but in this case two residues on neighboring loops are observed together. This may support a slightly different model for
folding based on formation of large loops (Berezovsky 2001). The authors of that study propose that standard size loops of 25-30 amino acids are formed at the initial stages of folding when the ends of the large loops come together. The final shape of the molecule is then stabilized by secondary contacts resulting in the formation of helices, sheets, and other loop-to-loop contacts. Although the sequence between Phe47 and Phe63 is shorter than that proposed in this model, the model provides an interesting perspective on how folding could be initiated in ILBP. β-strands C and D could be considered to be independent units with the ends coming together at the earliest stage of folding. The scene would then be set for the other hydrophobic residues in the core of the protein to become organized in their correct native contacts and the remaining secondary structure to form rapidly. The two β-strands between Phe47 and Phe63 are analogous to two segments of IFABP that have a strong intrinsic bias in their sequence to form β-strands (Srinivasan 1999, Nikiforovich 2002). The perspective that protein folding is similar to folding a carpenter’s rule presented by Srinivasan (1999) may apply to ILBP, in that β-strands C and D coming together could be compared to two hinges in a rule being bent to contact each other without being dependent on the region in between.

There is also the possibility that the two turn regions that contain Phe63 and Phe47 form independently early in the folding pathway. The chemical shift behavior of the two residues, though greater than that of most other residues observed, is not exactly the same for the two residues. It could be the case that the two hairpin turns form independently of each other, consistent with the hypothesis of turn formation in folding initiation (Muñoz, 1998), and with the observation by Li (2005b) that IFABP-Phe47 folds independently of IFABP-Phe-62. Mutations that disrupt the hydrophobic core of IFABP in the regions of residues 63-
66 affect the folding and stabilization of the protein (Kim 1997, Rajabzadeh 2003), and IFABP-Leu64 has been shown to retain residual structure at high denaturant concentrations (Kim 1997, Hodson 2001). The turn between strands D and E, near Phe63, is highly conserved in the iLBP family (Rotondi 2003). It could be the case that this high degree of conservation indicates an important role for this turn region in folding initiation for the family, which is supported by the data here. However, when all turns in IFABP were analyzed by mutation, the three most C-terminal turns were found to have the most effect on stability, and that mutation in any of the eight turns still resulted in protein with native-like structure by CD criteria (Kim 1998). One of the most de-stabilizing mutations was in the half-turn between the two β-sheets (Kim 1997). The authors suggested that disruption of this turn resulted in a disrupted hydrophobic core region. The IFABP data from Kim et al. (1997, 1998) do not rule out the possibility that the B-C and D-E hairpin turns form early in ILBP folding.

Earlier kinetic analysis of ILBP folding in our lab had indicated that the protein folds via a simple, reversible mechanism and encounters an on-path intermediate that retains all of its native secondary structure (Dalessio 2000). In a more recent study using improved instrumentation, a second intermediate with no remaining secondary structure was observed in the unfolding pathway by stopped-flow fluorescence (Dalessio 2005). This intermediate appears to have some tertiary contacts to other residues and no secondary structure. As this intermediate occurs early in the folding pathway, Phe17, Phe47, and/or Phe63 may be involved in formation of this type of intermediate state. There may be a relationship between an early intermediate observed in IFABP by 19F-NMR and the structure observed here in the Phe63/Phe47 region of ILBP.
The results of $^{19}$F-NMR analysis of IFABP support the hypothesis that early native contacts form in the hydrophobic core region of this protein (Ropson 2006). The residual structure observed in the IFABP spectra, however, has very different spectral characteristics than that of ILBP. In the IFABP experiments, NMR peaks flattened into baseline over the denaturant range at which the residual structure is proposed to exist. This phenomenon was interpreted as the presence of a stable intermediate conformation in exchange with the unfolded state conformation on an intermediate time scale. No such intermediate is observed in the ILBP experiments. Two protein conformations in slow exchange on the NMR time scale are presented as two separate peaks in a 1D spectrum. If there are two conformations in fast exchange, the spectra displays only one peak at the average resonance (Frieden 1993). The peaks observed for unfolded proteins represent multiple conformations in rapid exchange with each other, appearing at a resonance corresponding to a population-weighted average for the many conformations (Ropson 2006). The fact that this average resonance continues to show changes for Phe17, Phe47, Phe63, and one of the tyrosines in ILBP indicates that some populations of conformations are changing, in spite of the fact that the protein appears by optical methods to be unfolded. Those populations that continue to change at high denaturant concentration may represent nascent elements of structure in specific regions of the protein. As these are all close to the unfolded state, the conformations could represent regions of the protein that fold earlier than others.
Figure 5.2: Structures of IFABP and ILBP

Diagrams of (A) ILBP (PDB ID: 1o1u) and (B) IFABP (PBD ID: 2ifb) with the phenylalanine residues highlighted in light purple and the tyrosine residues highlighted in dark blue. The residues shown in this study to be involved in early structure in ILBP and by Ropson (2006) to be involved in early structure in IFABP are illustrated as sticks.
B. Conclusions

In spite of the structural similarities between the two proteins, the intermediate states that accumulate in folding of ILBP differ from those of IFABP. IFABP appears to fold through an early intermediate state that involves hydrophobic collapse and formation of native-like tertiary contacts in the hydrophobic core of the protein. The similar hydrophobic core region in ILBP does not appear to collapse to the same degree, although at least two of the six aromatic residues in the core region continue to experience structural changes at high denaturant concentrations. The presence of an early folding intermediate for ILBP has been suggested by kinetic studies using optical methods (Dalessio 2005). However, optical methods require that the intermediate accumulate to a significant degree to be detected, and the intermediate state has not been well-characterized. In an effort to describe the interactions forming this early intermediate, the studies presented here were undertaken to determine whether the observed NMR behavior could be attributed to the aromatic residues in the hydrophobic core of the protein. If so, evidence for a conserved mechanism of folding initiation between ILBP and IFABP would have been presented. However, the unusual NMR behavior that was observed for IFABP and attributed to specific aromatic residues was not observed for ILBP, and as such the nature of the early intermediate in ILBP appears to be somewhat different from that of IFABP. However, localization of two residues involved in early structure in ILBP folding to the hydrophobic core region suggest that some degree of hydrophobic collapse may be a conserved mechanism for the initiation of folding in the ILBP family. The lack of participation by other core residues in this intermediate state may be a
result of differences in the hydrophobic residues near the D-E turn region between the IFABP and ILBP.

### a. Folding of ILBP

The kinetic data published for several members of the family indicate that there is an initial collapse event in folding that occurs before formation of secondary structure. For some members of the family, a stable intermediate is observed at this stage, and native-like hydrophobic interactions may have already formed. This conclusion for IFABP is based on the observation that hydrophobic residues in contact in the native state maintain some elements of structure near the unfolded state. Trp82, Tyr70, Phe68, and Phe93 are located in the IFABP hydrophobic core and have been shown by NMR to be involved in formation of an early structure. IFABP-Phe68 is substituted by Ser69 in ILBP, resulting in fewer hydrophobic residues in the ILBP hydrophobic core, and also in decreased stability of the early intermediate state. The $\beta$-turn (D-E) near position 68/69 is similar between the two proteins and contains one of the few residues conserved throughout $>90\%$ of the iLBP family (glycine) (Banaszak 1994, Dalessio 2000). One recently proposed mechanism for $\beta$-sheet formation involves collapse of the extended unfolded structure followed by turn formation, which is then stabilized by formation of native hydrophobic contacts between adjacent $\beta$-strands (Du 2006). This model posits that the inter-strand hydrophobic contacts serve to stabilize the fold by preventing unfolding of the hairpin. This may be an appropriate model for folding initiation in iLBP’s. Evolutionary conservation of the D-E turn may be due to the role of the turn in folding initiation. The hairpin could be stabilized in IFABP-like proteins due to the interaction between Phe62 and Phe68, and therefore allow for accumulation of an
intermediate to such a degree to be observable by optical methods. Further, the reduced hydrophobicity for ILBP in this region of native structure may cause instability of the region after turn formation, explaining why no stable intermediate is observed. The presence of the Phe68/Phe62 interaction in IFABP could also be expected to stabilize the hydrophobic interactions with Phe93, Tyr70, and Trp82. Since the corresponding residue to IFABP-Phe68 is a serine in ILBP, this hydrophobic interaction is not present, and could explain why ILBP-Phe79 and Phe94, hypothesized to be members of the hydrophobic core, do not appear to be involved in formation of an early folding intermediate. Similarly, the turn between strands B and C, near Phe47 in both proteins, is the only other turn region with a residue (Gly44) that is conserved in >90% of iLBP’s (Banaszak 1994, Dalessio 2000). Folding initiation could occur concurrently at both turns, but not involve immediate interaction between the two nearby phenylalanines (Phe47/Phe63 in ILBP, Phe47/Phe62 in IFABP) since they are not part of the same hairpin structure. This explains the observation for both proteins that Phe47 appears to be involved in some structure early in folding, but may behave independently of IFABP-Phe62 (Li 2005) or ILBP-Phe63.

b. The Role of Intermediates in Folding

Part of the debate about the role of intermediates in protein folding involves the question of whether intermediates form native or non-native contacts. Support for the existence of both types of intermediate has been presented in the literature. For example, recent experiments in apomyoglobin revealed the presence of native and non-native early interactions in kinetic folding intermediates in the same protein (Nishimura 2006). Even theoretical studies of iLBP’s have suggested that some regions have an intrinsic bias for a
non-native secondary structure (Srinivasan 1999), although no experimental evidence for this type of intermediate has been found. Characterization of the intermediates in IFABP revealed native-like interactions early in folding (Li 2005, Ropson 2006). The early interactions indicated by the results of the ILBP studies presented here indicate that some residual structure persists after the majority of the protein is unfolded. To determine whether this structure composes a native-like intermediate, it must be determined whether there are remaining contacts that are present in the native state. The early intermediate state supported by these experiments may contain native-like contacts between Phe62 and Phe47, since these are involved in some structure at high denaturant concentration and are in contact in the native state.

c. Aggregation

A wide variety of different proteins have been shown to form the same types of aggregates. A common global architecture is likely to exist among disease causing aggregates, since studies of different amyloid-forming proteins so far have revealed a common macromolecular structure in the aggregated forms (Barnaby 2006). Although not associated with a disease state, variants of iLBP family members have been shown to aggregate into amyloid-like structures (eg, Ignatoya 2005). Early folding events are proposed to be critical for aggregation to occur, as fully or partially unfolded ensembles on the folding pathway are more vulnerable to aggregation (Chiti, 2006). Determining whether structurally-related proteins with low sequence similarity fold via the same mechanism is critical to understanding whether the wide variety of different proteins involved in amyloid diseases misfold in the same way.
Although the studies presented here show that two structurally related proteins do not have similarly stable folding intermediates, they do support a common mechanism for folding between the two proteins by demonstrating that residues in homologous hydrophobic cores are involved in early structures, and suggest that variable sequences can attain the same fold via similar mechanisms. This is relevant to the study of misfolding diseases, since widely variable sequences misfold into very similar structures. Two approaches suggested for therapeutic development for these diseases are stabilization of the native protein to prevent aggregation, or de-stabilization of the misfolded proteins to enhance clearing (Barnaby 2006). The differences in the early intermediate in IFABP and ILBP appear to represent different stabilities of a conformation caused by the same type of interactions. The results of the ILBP study presented here are encouraging in this respect, as they indicate that treatments developed for one type of misfolding disease may be appropriate for use in treating any of the 20 or more such conditions.

d. Summary

$^{19}$F-NMR coupled with site-specific $^{19}$F- incorporation into proteins is a powerful method for analyzing folding behavior one residue at a time during the protein-folding pathway. This technique has been used here to identify regions of the human ileal lipid-binding protein that are involved in the earliest structures that occur in folding. It is not possible to observe these transient structures by other methods due to the dynamic nature of the protein at the early stages of folding. In comparing the results here with those obtained for a structurally related protein, it is suggested that the two proteins form early folding conformations in homologous regions of their structures. The differences in these structures
have been suggested to be due to the presence or absence of a specific hydrophobic interaction, in agreement with previously published models. These studies are relevant to the protein folding field, as well as to the larger medical research field, as they relate folding mechanisms in two structurally similar proteins with dissimilar amino acid sequences.

Although the work in ILBP presented here provides strong argument in favor of a folding mechanism with similarities to that of the related IFABP pathway, the differences observed are perhaps more striking. One of the most intriguing observations is the unusual behavior of Phe17, the helical phenylalanine that responds more dramatically to denaturant than any other phenylalanine. The role of the α-helical region of ILBP deserves more in-depth analysis, as both the results presented here on ILBP-Phe17, and the recent publication from Kouvelos (2006), are in stark contrast to the available data on the role of the helices in IFABP.

Further studies on the hydrophobic core of ILBP are also warranted based on the results presented here. One possible suggestion for a continuation of these studies is the construction of mutations similar to those developed in Dalession (2005) in which the hydrophobic core of ILBP is substituted with additional, $^{19}$F-labeled, residues. The single-site labeling system presented here could be effectively used in this context to provide information of the behavior of additional hydrophobic residues in the core, as well as give an indication of changes caused in the native hydrophobic residues as a result of these mutations. The latter would be extremely informative, as one potential interpretation of the initial studies is that the additional hydrophobic residues significantly altered the behavior of the hydrophobic core (Dalessio 2006).
Finally, several different models attempting to explain the unusual behavior of the phenylalanine residues in ILBP have been presented here. Each of these models could be further explored and developed utilizing the methods described here, further mutational analysis, as well as experiments involving construction of short peptides mimicking the regions around ILBP aromatic residues. $^{19}$F-NMR has been shown to be an important tool in the analysis of protein structure, especially in the field of protein folding, for which very few other methods are as informative. Although many questions remain as to the exact nature of $\beta$-barrel protein folding, the results presented here demonstrate that there is certainly a role for hydrophobic collapse early in the folding pathway, and that differences in the nature of the hydrophobic core regions between related proteins have profound effects on the subtleties of the interactions caused by this collapse.
References:


Bezsonova, I., D.M. Korzhnev, R.S. Prosser, J.D. Forman-Key, and L.E. Kay. 2006. Hydration and packing along the folding pathway of SH3 domains by pressure-dependent NMR. *Biochemistry*. accelerated publication


Vita

Heather K. Basehore

EDUCATION

**Ph.D.** in Integrative Biosciences
Molecular Medicine Option
The Pennsylvania State University College of Medicine, Hershey, PA

**B.S.** in Chemistry, Biochemistry emphasis
Mansfield State University of Pennsylvania, Mansfield, PA

FELLOWSHIPS AND AWARDS


Senior Award: American Institute of Chemists. 2001

Certificate: Academic Achievement in Chemistry, Mansfield State University of Pennsylvania. 2001

PUBLICATIONS

