THE UNIVERSITY OF CHICAGO

CHARACTERIZATION OF PROTEIN FOLDING INTERMEDIATES
FOR DELINEATION OF FOLDING PATHWAYS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE Degree OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY

ZHONGZHOU ZHENG

CHICAGO, ILLINOIS

JUNE 2010
TO MY WIFE, MI, AND MY PARENTS
Table of Contents

LIST OF FIGURES .......................................................................................................... vii
LIST OF TABLES ........................................................................................................... ix
LIST OF ABBREVIATIONS ............................................................................................ x
ACKNOWLEDGEMENTS ................................................................................................. xi
ABSTRACT ..................................................................................................................... xii

Chapter
1 Introduction .................................................................................................................... 1
  1.1 Why study protein folding ...................................................................................... 1
  1.2 Time scales and energies ....................................................................................... 3
    1.2.1 How fast should a protein fold ........................................................................ 3
    1.2.2 What drives proteins to fold ............................................................................ 4
      1.2.2.1 The hydrophobic effect ............................................................................ 4
      1.2.2.2 Hydrogen bonds ..................................................................................... 5
      1.2.2.3 Electrostatic interactions ......................................................................... 5
      1.2.2.4 van der Waals force ............................................................................... 6
      1.2.2.5 Conformational entropy .......................................................................... 6
  1.3 Folding mechanism, pathway and intermediate ..................................................... 6
    1.3.1 Early models .................................................................................................... 7
    1.3.2 Folding intermediates ..................................................................................... 9
    1.3.3 Native-state hydrogen exchange and sequential foldon-formation
        Mechanism ......................................................................................................... 11
1.3.4 Early folding events: funneled landscape or sequential pathway ..... 15

1.4 Scope of this thesis ........................................................................................................ 18

2 Protein Vivisection Reveals Elusive Folding Intermediates ............................................. 19

2.1 Abstract ............................................................................................................................ 19

2.2 Introduction ..................................................................................................................... 19

2.3 Results ................................................................................................................................ 23

2.3.1 Rationally populating intermediates ...................................................................... 23

2.3.2 Thermodynamic and kinetic properties .................................................................. 25

2.3.3 NMR analysis ............................................................................................................ 29

2.3.4 Hydrogen exchange .................................................................................................. 32

2.3.5 Small-angle X-ray scattering ............................................................................... 33

2.3.6 Denaturant dependence of HX .............................................................................. 39

2.3.7 Other charged substitutions ................................................................................... 41

2.4 Materials and methods .................................................................................................. 45

2.4.1 Proteins ..................................................................................................................... 45

2.4.2 Equilibrium and kinetic measurements ................................................................ 45

2.4.3 NMR spectroscopy and HX chemistry ................................................................. 45

2.4.4 Small-angle X-ray scattering ............................................................................... 46

2.5 Discussion ......................................................................................................................... 47

2.6 Conclusions ...................................................................................................................... 51

3 Early Ubiquitin Folding Intermediates and Pathway ......................................................... 52

3.1 Abstract ........................................................................................................................... 52

3.2 Introduction ..................................................................................................................... 52

3.3 Results ............................................................................................................................. 56
3.3.1 Engineering ubiquitin fragment 1 to 36 to mimic the hairpin-helix motif ................................................................. 56

3.3.2 Peptide complementation for characterizing hydrogen bonds in the hairpin-helix motif ........................................... 60

3.3.3 A stable β1β2 hairpin-α-helix-β3 intermediate suggested by NSHX ........................................................................... 63

3.3.3.1 CPMG relaxation dispersion NMR spectroscopy ............ 65

3.3.3.2 Trans- and cis-reconstruction of the β1β2 hairpin-α-helix-β3 early intermediate ....................................................... 65

3.4 Materials and methods ...................................................................................................................................................... 68

3.4.1 Proteins and peptides ..................................................................................................................................................... 68

3.4.2 Equilibrium and kinetic measurements .......................................................................................................................... 69

3.4.3 NMR spectroscopy ......................................................................................................................................................... 69

3.5 Summary ............................................................................................................................................................................ 69

4 Conclusions and Future Directions .................................................................................................................................. 72

4.1 Conclusions ........................................................................................................................................................................ 72

4.2 Future directions ................................................................................................................................................................. 73

References Cited ....................................................................................................................................................................... 77
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Reaction Coordinate of protein folding reaction</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Schematic view of early proposals about protein folding mechanism</td>
<td>8</td>
</tr>
<tr>
<td>1.3</td>
<td>Principle of native-state hydrogen exchange</td>
<td>13</td>
</tr>
<tr>
<td>1.4</td>
<td>Sequential pathway Versus funneled energy landscape view of protein folding</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>NSHX can fail to detect subglobal openings</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Post-TS event in ubiquitin folding</td>
<td>24</td>
</tr>
<tr>
<td>2.3</td>
<td>Equilibrium and kinetic folding behavior of UbL50E</td>
<td>26</td>
</tr>
<tr>
<td>2.4</td>
<td>NMR analysis of pWTUb and UbL50E</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>pH titration of UbL50E from pH 4 to pH 7.8</td>
<td>31</td>
</tr>
<tr>
<td>2.6</td>
<td>Native-state hydrogen exchange on UbL50E I state at pH (pD_{\text{read}}) 7.5</td>
<td>34</td>
</tr>
<tr>
<td>2.7</td>
<td>Isotherms of NSHX on UbL50E I state at (pD_{\text{read}}) 7.5</td>
<td>35</td>
</tr>
<tr>
<td>2.8</td>
<td>Dimensions of pWTUb and UbL50E intermediate measured using SAXS</td>
<td>38</td>
</tr>
<tr>
<td>2.9</td>
<td>Dependence of UbL50E stability on (D_2O) level</td>
<td>40</td>
</tr>
<tr>
<td>2.10</td>
<td>Post-TS free energy surface of ubiquitin folding</td>
<td>42</td>
</tr>
<tr>
<td>2.11</td>
<td>Charge-burial partial unfolding strategy applied at other positions</td>
<td>43</td>
</tr>
<tr>
<td>2.12</td>
<td>Interpretation of possible chevrons generated by the charge-burial strategy</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Proposed ubiquitin early folding pathway</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>CD characterization of (\alpha)-helix formation driven by Ni(^{2+}) ion in Ub*(1-36)</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>NMR characterization of the hairpin-helix motif driven by Ni(^{2+}) ion</td>
<td>59</td>
</tr>
<tr>
<td>3.4</td>
<td>Coupled binding and folding of Ub*(1-36) with Ub(37-76)</td>
<td>61</td>
</tr>
<tr>
<td>3.5</td>
<td>Possible explanations for NSHX behavior of ubiquitin group B hydrogen bonds</td>
<td>64</td>
</tr>
</tbody>
</table>
3.6  CPMG relaxation dispersion measurements on UbL50E I state

3.7  Trans- and cis-reconstruction of the β1β2 hairpin-α-helix-β3 intermediate

3.8  Reconstruction of energy landscape for major ubiquitin early folding pathway

4.1  Reshaped folding energy landscape of ubiquitin early folding pathway by the
    T9D mutation and Ni^{2+}:bi-Histidine binding

4.2  Distinguishing possible causes for NSHX results of ubiquitin group B
    hydrogen bonds
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Kinetic parameters obtained from chevron analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.2</td>
<td>Native-state hydrogen exchange on UbL50E intermediate, $pD_{read}$ 7.5</td>
<td>36</td>
</tr>
</tbody>
</table>
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>biHis</td>
<td>Designed two histidine site for divalent metal ion binding</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CPMG</td>
<td>Carr-Purcell-Meiboom-Gill</td>
</tr>
<tr>
<td>GdmCl</td>
<td>Guanidinium chloride</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-(2-hydroxyethyl)piperazine-N'-(2ethanesulfonic acid)</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear Single Quantum Coherence</td>
</tr>
<tr>
<td>Im7</td>
<td>Colicin immunity protein 7</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NSHX</td>
<td>Native-state hydrogen exchange</td>
</tr>
<tr>
<td>PUF</td>
<td>Partially unfolded form</td>
</tr>
<tr>
<td>pWTUb</td>
<td>Pseudo wild type ubiquitin with F45W/H68N mutation</td>
</tr>
<tr>
<td>RD</td>
<td>Relaxation dispersion</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small-angle X-ray scattering</td>
</tr>
<tr>
<td>TMAO</td>
<td>Trimethylamine N-oxide</td>
</tr>
<tr>
<td>TS</td>
<td>Transition state</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
Acknowledgements

Being able to pursue a Ph.D. at the University of Chicago is definitely one of the biggest fortunes in my life. I benefited a lot from the incredibly stimulating academic environment with so many brilliant and dedicated scientists around. I am thankful to those persons including Tony Kossiakoff (chair) and Lisa Anderson who have been continuing to make the department even more vibrant. I would like to thank my advisor Tobin Sosnick who guides me through this journey with his scientific expertise and carries me through those frustrating moments with his enthusiasm and never-say-die attitude. I am grateful to all the Sosnick lab pals for making such an atmosphere that makes research really enjoyable. Meetings with my thesis committee, Stephen Kent (chair), Shohei Koide, and Robert Keenan afforded thoughts provoking discussions and valuable perspective on my project. With the generous help of Stephen Kent and his former and present lab staff, Brad Pentelute, Kalyaneswar Mandan, Ryo Okamoto, Fangkun Deng, and others, I was able to learn chemical synthesis of peptide and protein, which has not only been essential for my thesis research but become precious asset for my future career as well. I would like to thank Josh Kurutz, former director of the BioNMR facility, and Shohei Koide for assistance with NMR. It was Josh’s day in and day out efforts that kept the NMR facility running smoothly, of which we are reminded after his leave. Special thanks to Francesca Massi, now assistant professor at University of Massachusetts, who brought relaxation dispersion technique to the University of Chicago community and I happened to be the first to benefit from it. Finally, I wish to thank my wife for virtually giving up everything in her life to support me and my career. I can never pay off what I have owed her in my life.
Abstract

A solution to the protein folding puzzle -- how amino acid sequence determines protein structure – will have huge impact in a variety of research areas ranging from protein design and production to aggregation-related diseases and disorders. The characterization of protein folding intermediates likely holds a crucial key to understanding protein folding mechanism. In spite of the countless folding intermediates which have been successfully studied, two major issues remain to be addressed: 1. The need of a general method to find and trap folding intermediates. 2. Lack of knowledge about protein folding at the early stages of the folding reaction.

Using ubiquitin as model protein, I developed a novel “Protein Vivisection” strategy that involves replacing buried hydrophobic residues with charged ones to partially unfold a protein. Using this strategy, I identified two folding intermediates which had been unobservable using previously available methods. Characterization of these intermediates led to full delineation of ubiquitin post-transition state pathway.

I then moved on to investigate the pre-transition state events of the ubiquitin folding reaction. Peptide fragments of ubiquitin were used as mimics of early folding intermediates to demonstrate 1. Secondary structure elements formed at early stage are stabilized by native-like tertiary contacts. 2. Marginally stable early folding intermediate may be identified by native-state hydrogen exchange data. Together, these results are consistent with the view that at early stages of folding, formation of secondary structure and native tertiary contact is concerted and there is a preferred pathway for folding to proceed.
1 Introduction

1.1 Why study protein folding?

Proteins are the most important functional biomolecules. Half of human body’s dry weight is made up of protein (Harper et al., 1977). Proteins serve a wide range of functions including catalyzing reactions, mediating signal transduction, regulating gene expression, immune system recognition and response, providing mechanical scaffold for cells, and many others. Newly synthesized proteins from ribosome begin as a string of amino acids with a random coil-like conformation. It is not until the random coil folds into a unique and well-defined 3-dimensional structure that the protein becomes functional. People have been working hard to understand how this unfolded to folded state transition occurs (Figure 1.1).

The first milestone emerged when Anfinsen demonstrated that simply removing denaturant is sufficient for a fully denatured protein to refold (Anfinsen, 1973). This work demonstrated that all the driving force and guiding information for folding is embedded in the amino acid sequence (primary structure) and no external templating is required. Up until this point, however, it is still not well understood how the primary structure translates into a distinct 3-D structure. This remains as the last piece of unsolved puzzle in the “central dogma” of molecular biology.

Delineating folding mechanisms has tremendous implications to human health and biological function. Folding is involved in various biological activities including protein translocation (Krantz et al., 2005), signal transduction (Young et al., 2001), vesicle fusion (Sutton et al., 1998), and many others. Diseases caused by protein misfolding are ubiquitously
A protein is synthesized as an unstructured, random coil-like polypeptide chain. Such state is called the “unfolded state”. The unfolded state will spontaneously fold into a “native state” with distinct 3-D structure because the native state has lower free energy so that such transition is energetically favorable. During the folding reaction, the chain must climb up the reaction surface until it reaches the highest point called the “transition state”, as indicated by “‡”. After the transition state folding goes downhill to reach the native state. All the information needed to guide the folding reaction along the reaction surface is encoded in the amino acid sequence of the polypeptide chain.

*Figure 1.1 Reaction Coordinate of protein folding reaction.*
found as in cystic fibrosis, scurvy, phenylketonuria, Marfan Syndrome, and Alzheimer’s/Parkison’s Disease. (Thomas et al., 1995; Dobson, 2001; Hashimoto et al., 2003). Sometimes even the “sequence determines folding” rule gets challenged as in the transmissible spongiform encephalopathy (prion diseases) where the correct amino acid sequence can lead to an infectious abnormal fold (Pan et al., 1993; Priola et al., 2003). In vitro, folding errors can obstruct laboratory studies and industrial protein production (Kim et al., 2009). Last but not the least, the principles and methods used in protein folding are applicable to protein design (Strickland et al., 2008) and protein-protein association (Horn et al., 2009).

1.2 Times scales and energies

Given the sub-second time scale of most biological reactions, protein folding must occur at a rate fast enough to allow a protein to carry out its biological function. Also to maintain the native state stable over a reasonably long period of time, there must be physical processes that contribute free energy to stabilize native state. Here we take a closer look at these two issues.

1.2.1 How fast should a protein fold?

For a protein to fold into native conformation, there are intrinsic steps that the protein must perform. They include secondary structure formation, long-range tertiary contacts, desolvation of hydrophobic groups and hydrogen bond donor/acceptor, side chain packing, and others. The fact that the fastest folding proteins fold in microsecond time scale (Kubelka, et al., 2004) indicates that the intrinsic steps are microsecond or even faster processes. A collection of
in vitro studies demonstrates that duration of the folding process can vary by 10 orders of magnitude depending on the protein. While milliseconds time scales are most frequently found for sub 100 amino acid proteins, it may take minutes or even hours for some of the very large, multi-domain proteins to finish folding in vitro (Kim, et al., 1990). Even for the slowest folders, however, the amount of time needed is negligible compared to the astronomical time it will take for the polypeptide chain to sample all possible conformations to find the lowest energy native structure. This apparent contradiction between actual folding times and a simple estimate of the time it would take for a random search, known as the Levinthal paradox (Levinthal, 1968), points out that protein folding could not be accomplished in the “trial and error” fashion. There likely is a preferred pathway where localized interactions and structures favoring the native conformation sequentially form to guide and speed up the folding reaction to the finish point (Rooman et al., 2002).

1.2.2 What drives protein to fold?

Several energetic terms are involved in the folding reaction, some working in favor of folding while others against it. The net result is usually a relatively small gain of free energy stabilizing the folded state of protein.

1.2.2.1 The hydrophobic effect

For hydrophobic residues exposed in the water solution, water molecules will form a layer of highly organized, ice-like structure on the surface of the hydrocarbon side chain. Such structure is often referred to as “clathrate”, which is energetically costly to form due to the reduction in solvent entropy (see Horvath et al., 1976 and references therein). When hydrophobic
residues cluster to reduce water-exposed non-polar surface area, there is a net decrease of the
amount of clathrate structure and a corresponding increase in entropy, thus providing driving a
force for the clustering of hydrophobic residues. The hydrophobic effect is probably the most
stabilizing factor in protein stability. Quantification of the energetic consequences of
hydrophobic effect was initiated by Kauzmann with his solvent transfer model (Kauzmann, 1959)
and followed by numerous other studies (Lee and Richards, 1971; Hermann, 1972, Chothia,
1974; Reynolds, 1974; Tanford, 1980).

1.2.2.2 Hydrogen bonds

A hydrogen bond is the attractive interaction between a hydrogen atom with an
electronegative atom. For proteins, hydrogen bonds between backbone amide hydrogen and
backbone carbonyl oxygen are the most important ones involved in native state stabilization.
However, it is controversial as to whether hydrogen bonds make significant net energetic
contribution to drive folding because unfolded protein also forms hydrogen bonds albeit with
water (Dill, 1990; Yang and Honig, 1995; Pace et al., 1996; Kentsis and Sosnick, 1998; Krantz
et al., 2000). It remains a possibility that energy of hydrogen bonds from unfolded state and
folded state are nearly equal, resulting in a near zero net gain.

1.2.2.3 Electrostatic interactions

The attractive interactions between oppositely charged polar-groups on protein surface
are often accompanied by entropic penalties for immobilizing the charged pair. Thus the net gain
of free energy is less than the calculated value from Coulomb’s Law. Due to the unfavorable
repelling interactions between like charges, efforts for engineering stabilizing charge-charge pair
on protein surface can be challenging and require consideration of the entire surface charge distribution (Schweiker and Makhatadze, 2009).

1.2.2.4 van der Waals force

van der Waals forces are caused by correlations in the fluctuating polarizations of nearby particles. The Lennard-Jones potential is often used as an approximate model for van der Waal force as a function of distance. Since both unfolded and folded states contain a large number of van der Waals interactions, the net contribution for folding may be near zero.

1.2.2.5 Conformational entropy

Conformational entropy is the major energy term that opposes folding of protein. The unfolded state of protein provides a myriad of conformations that the unstructured polypeptide chain can sample. Upon folding such freedom is gradually reduced, which leads to an uphill climb of the folding energy landscape. As the folding reaction proceeds, the free energy eventually decreases as the loss of entropy starts to be compensated by the favorable hydrophobic and other interactions. Therefore, after the highest point of the free energy surface (transition state) is passed, folding undergoes a downhill process to reach the low energy native state.

1.3 Folding mechanism, pathway and intermediate

As mentioned in section 1.2.1, a random, unbiased search of all possible conformations to find the native one is not a practical strategy for proteins to fold. Proteins must
follow a pathway that guides non-random search. The question arises: is there a single mechanism that governs the folding behavior of all proteins?

1.3.1 Early models

One of the very first models, nucleation-growth model, proposed that the polypeptide chain of the protein form some local secondary structure element that serves as a nucleus (Wetlaufer, 1973). Then the nucleus grows rapidly by adding peptide segments that are close to the nucleus in amino acid sequence. The nucleation-growth model dropped out of favor because it argues against existence of folding intermediates that had been supported by substantial amount of experimental evidence (Kim and Baldwin, 1982; Ptitsyn, 1987).

After the nucleation-growth model, two alternative models became popular. The first one is the framework model (and the related diffusion collision model) in which secondary structure formation precedes tertiary interactions (Figure 1.2). This model gained support primarily from studies on relatively stable helical fragment of the protein. It was found that the peptide comprising the first 13 residues of ribonuclease A favors helix formation under low temperature and high ionic strength (Brown and Klee, 1971), while engineered salt bridge can drive C-peptide of RNase A into helix formation (Bierzynski et al., 1982; Shoemaker et al., 1985). However, 90% of the isolated peptides are disordered even though they form regular secondary structures in proteins (Epand and Scheraga, 1968). The generality of framework model is questionable.

The second model, hydrophobic collapse model (Figure 1.2), argues that since the hydrophobic effect is the major driving force for protein folding, formation of a hydrophobic
Figure 1.2 Schematic view of early proposals about protein folding mechanism. Framework model requires stable secondary structure formation of isolated peptide fragment of the protein. Hydrophobic collapse model involves an initial collapse of hydrophobic residues to expel water and provide driving force for folding.
residue cluster that repels water must be the first step of folding. The problem with this model is that such quick collapse must involve a lot of non-specific hydrophobic interactions. It is hard to imagine that the excessive non-native contact will not pose hindrance to the reorganization of polypeptide chain to make native contacts. In addition, random collapse results in partial desolvation of backbone that is unlikely to be fully compensated for by native hydrogen bonding.

1.3.2 Folding intermediates

Studying protein folding intermediates has been playing a crucial role in search for a unifying mechanism of folding reaction. Existence of protein folding intermediate was first proposed by Paul Flory in a seminar in 1968 in which Levinthal’a paradox was brought out (Baldwin, 1994). However, early on scientists studying folding were often confronted with a skeptical wider community that wondered whether it was possible to detect folding intermediates. The reason for this view was the deeply rooted “two-state (U ↔ N)” model of folding (Lumry et al., 1966). The situation started to change from early 80’s when an increasing amount of evidence accumulated pointing to the existence of folding intermediates. Urea-gradient electrophoresis identified compact kinetic intermediates near 0 °C (Creighton and Pain, 1980). Fast circular dichroism (CD) measurements revealed the presence of early folding intermediates in kinetic studies (Labhardt, 1984; Kuwajima et al., 1987). Fragments of small proteins were found to form weak but folded structures, suggesting that the entire polypeptide chain is not needed for folding to occur (Dyson et al., 1988a,b; Oas and Kim, 1988; Staley and Kim, 1990). Hydrogen exchange pulse labeling studies identified transient structural intermediates of cytochrome c and ribonuclease A (Roder et al., 1988; Udgaonkar and Baldwin, 1988). Probably the most convincing piece of evidence came from studies on disulfide bond formation in folding
of bovine pancreatic trypsin inhibitor (BPTI) (Creighton 1978; 1985). Native BPTI has three
disulfide bonds (referenced as cysteine residue numbers enclosed in brackets): [30-51], [5-55]
and [14-38] and BPTI with partial formation of native disulfide bonds were trapped with thiol-
blocking agents (Creighton, 1977). Because the process of disulfide reoxidation is coupled to the
refolding of BPTI, these partial disulfide bonds intermediates were argued to be authentic folding
intermediates (Creighton, 1983; Creighton and Goldenberg, 1984).

A so-called ―molten globule‖ model started to emerge in the 60’s and dominated the
field of protein folding in early 90’s (Tanford, 1968; Wong and Hamlin, 1974; Kuwajima et
al., 1976; Nozaka et al., 1978; Feng et al., 1991; 1994; James et al., 1992; Mayo et al., 1992;
Palleros et al., 1992; Philo et al., 1993; Narhi et al., 1993; Hua et al., 1993). Although the exact
nature of molten globule is never strictly defined, a useful description will be an intermediate
state where native secondary structures are formed and no fixed tertiary contacts are made
(Dobson, 1992; Ptitsyn, 1995a). It was suggested that the molten globule state is a general kinetic
folding intermediate state and the conversion from molten globule to native state is the rate-
limiting step of folding reaction (Ptitsyn, 1995b).

Two major discoveries started to question molten globule as an essential kinetic
intermediate of folding. 1. ―Two-state‖ model was rejuvenated by Jackson and Fersht who
elegantly demonstrated that chymotrypsin inhibitor II from barley (CI2) folded in a two-state
manner (Jackson and Fersht, 1991). Sosnick and Englander further suggested that 2-state folding
behavior is a general rule (> 90%) for small, single-domain proteins and kinetic intermediates
observed are misfolded species for which error correction and reorganization can be slow
(Sosnick et al., 1994; 1996; 1997; Krantz and Sosnick, 2000a; Krantz et al., 2002a). 2. The
molten globule to native transition of cytochrome c was shown to be very fast (< 1 msec) and thus cannot be the rate-limiting step of folding (Sosnick et al., 1994; 1996).

It is noteworthy that two-state folding behavior does not rule out the presence of intermediate states along the folding route. It simply means folding intermediates do not sufficiently accumulate to be observed by kinetic measurements. Therefore in two state folding, any stable folding intermediate (relative to unfolded state) must be post-transition (late) intermediate that only exists on native side of the folding energy barrier.

1.3.3 Native-state hydrogen exchange and sequential foldon-formation mechanism

How do we identify stable late folding intermediates? Serendipitously found mutations can perturb the folding energy landscape to trap folding intermediates in equilibrium (Whittaker et al., 2007; Feige et al., 2008; Connell et al., 2009). The more recent advent of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion (RD) NMR spectroscopy correlates accelerated transverse relaxation rate (R₂) with existence of sparsely populated states in equilibrium with the major state (Palmer III et al., 2001). In folding, the equilibrium between folding intermediates and native state of protein allows CPMG RD to identify folding intermediates with thermodynamic, kinetic and structural information (Korzhnev et al., 2004; Grey et al., 2006; Schanda et al., 2008; O’Connell et al., 2009).

Among all the methods, native-state hydrogen exchange (NSHX) stands out as the most powerful and systematic one (Bai, 2006). Hydrogens on the polar groups of proteins engage in continuous exchange with hydrogens of solvent unless they are structurally protected (buried or involved in hydrogen bond). Linderstrom-Lang and his colleagues worked out a two-state
model in which hydrogens are either non-exchangeable in protected state (closed) or susceptible to exchange in some transiently open form (open) (Scheme 1.1) (Hvidt and Linderstrom-Lang, 1955).

\[
\begin{align*}
\text{Closed state} & \quad \xrightarrow{k_{\text{open}}} \quad \text{Open state} & \quad \xrightarrow{k_{\text{int}}} \quad \text{exchange} \\
& \quad \quad _{k_{\text{close}}} \quad & \quad \\
\end{align*}
\]

(Scheme 1.1)

Here \(k_{\text{open}}\) and \(k_{\text{close}}\) are opening and closing rate of the protecting structure. The chemical exchange rate of freely available, unprotected amide hydrogens, \(k_{\text{int}}\), depends on a variety of conditions (pH, temperature, etc.) and can be obtained by straightforward calculation (Bai \textit{et al.}, 1993; Connelly \textit{et al.}, 1993). Most of the hydrogen exchange literature assumes stable structure \((k_{\text{open}} \ll k_{\text{close}})\) and fast equilibrium between open and close state \((k_{\text{int}} \ll k_{\text{close}}, \text{so-called “EX2 limit”})\). Thus, the observed hydrogen exchange rate reduces to the fraction of time the protein is in the excited state multiplied by the intrinsic rate of exchange, \(k_{\text{obs}} = k_{\text{open}}k_{\text{int}}/k_{\text{close}} = K_{eq}k_{\text{int}}\) (Equation 1.1). The free energy of the open state is calculated according to \(\Delta G_{\text{HX}} = -RT \ln (K_{eq}) = -RT \ln (k_{\text{obs}}/k_{\text{int}})\) (Equation 1.2), where \(RT\) is the product of the gas constant and temperature.

In protein backbone amide hydrogens are involved in hydrogen-bond network that stabilizes native state of the protein, so their hydrogen exchange behaviors are the most studied. For a given hydrogen-bonded backbone amide hydrogen, there are usually three types of openings through which hydrogen exchange with water can occur: globally unfolded state, partially unfolded form (PUF) and local structure fraying (Figure 1.3). The relative free energy levels of the states (i.e. their relative populations) decide through which state hydrogen exchange primarily occurs. When small amount of denaturant is added to the protein, free energy levels of
Figure 1.3 Principle of native-state hydrogen exchange. 
a. In the native state, a residue’s (blue dot) hydrogen-bonded backbone amide is protected from exchanging with solvent water (D₂O). Native state is sampling in equilibrium three types of exchange-capable, high energy states (globally unfolded, PUF, local structure fraying). HX occurs primarily through the lowest energy state of the three, at a rate determined by free energy difference between native state and the lowest energy, exchange competent state ($\Delta G_{HX}$), along with the intrinsic exchange rate ($k_{int}$). 
b. A typical HX isotherm. Each dashed line indicates a stage in which HX occurs through a primary exchange competent state. Under native condition with no denaturant, local structure fraying is usually the lowest energy state for HX to occur. Because of the differential effect of denaturant on different states (globally unfolded > PUF > local fraying ≈ 0), the isotherm ($\Delta G_{HX}$ plotted against denaturant) starts with a denaturant independent phase with zero slope, indicating exchange with a local fraying event. At some point, a PUF surpasses locally frayed state to become the lowest energy, exchange competent state and the slope of the isotherm decreases. Due to the globally unfolded state’s large amount of surface exposure, it dominates HX, resulting in an isotherm phase with greatest slope. The intercept of extrapolation of each phase to y axis gives free energy level of each state relative to native state. The PUF phase of isotherm is the most information-rich one. A group of backbone amide groups (a stretch of $\alpha$-helix or $\beta$-hairpin, etc.) displaying a same HX behavior (same slope and same y axis intercept) indicates presence of a folding intermediate with a cooperatively-unfolded fragment (foldon).
the states will be perturbed differentially according to their specific levels of response to denaturant. Incremental addition of denaturant with measured $\Delta G_{HX}$ will produce a so-called “isotherm” on which each of the three states can usually be distinguished by its denaturant dependence.

In this manner, folding intermediates that exist as infinitesimally populated high energy PUFs can be detected under native conditions, hence the name “native-state hydrogen exchange” (Bai et al., 1995; Hollien and Marqusee, 1999; Chu et al., 2002; Kato et al., 2007b). The openings of PUFs represent cooperative unfolding of individual or groups of secondary structure elements (foldons) corresponding to events on the unfolding pathway leading back up to the transition state. When viewed from the folding direction, these events represent sequential buildup of foldons after the rate-limiting transition state where unfolded regions dock onto pre-existing structure. Since most secondary structures are not stable in isolation (Epand and Scheraga, 1968), the newly formed foldons are presumably stabilized by making tertiary contacts (e.g. hydrophobic burial) with the rest of the protein. This view is supported by kinetic H/D amide isotope effect demonstrating commensurate hydrogen bond formation and surface burial (Krantz et al., 2000b; 2002b). Using protein engineering to delete a foldon, Bai and colleagues managed to trap and determine high-resolution NMR structures of the PUFs revealed by NSHX (Takei et al., 2002; Feng et al., 2005; Kato et al., 2007a; Zhou et al, 2008). The structures suggested that the folded parts of folding intermediates have native backbone conformation and some non-native side chain packing.
1.3.4 Early folding events: funneled landscape or sequential pathway?

Compared to the stable late folding intermediates, unstable early folding intermediates before the transition state are unobservable with any currently available method including NSHX and CPMG RD. Consequently, little is currently known about how the unstructured polypeptide chain reaches transition state. Theorists proposed a funnel-shaped folding energy landscape which suggests that folding reaction can proceed through a multitude of distinct routes to reach the lowest energy native state (Bryngelson et al., 1995; Dill and Chan, 1997; Wolynes, 1997; Shakhnovich, 1998; Socci et al., 1998) (Figure 1.4). Although recently the funnel picture has adopted a more “restricted choice of road” view for the late stage of folding, substantial pathway heterogeneity remains as a major theme for pre-transition state region of the landscape. Although Boltzmann distribution suggests there is nothing fundamentally incorrect about the funneled energy landscape theory, this “anything is possible” view fails to provide clues as to the guidance folding reaction follows in search for the native state. Therefore, contributions to unraveling folding mechanism may be very limited by this funneled energy landscape model.

On the other hand, based on the aforementioned kinetic H/D amide isotope effect (Krantz et al., 2000b; 2002b) and the presence of native secondary structure elements in protein folding transition state defined by ψ-analysis (Krantz et al., 2004; Sosnick et al., 2006), Sosnick and colleagues set out trying to extend the mechanism of sequential foldon stabilization, which governs the late stage of folding as revealed by NSHX, to folding events before transition state (Figure 1.4). According to their view (Sosnick et al., 2008), the structural freedom and flexibility of the unfolded polypeptide chain at the starting point of folding allows search through a myriad
Figure 1.4 Sequential pathway Versus funneled energy landscape view of protein folding.
Sequential and step-wise foldon stabilization by tertiary contact as the mechanism for the entire folding reaction was brought out by Sosnick and colleagues (Krantz et al., 2004; Sosnick et al., 2006; Sosnick, 2008) based on NSHX results (Bai, 2006), kinetic isotope effects (Krantz et al., 2000b; 2002b), and ψ-analysis (Krantz et al., 2004; Sosnick et al., 2006). Funneled energy landscape view suggests a multitude of heterogeneous routes that folding reaction can follow, especially at early stage of folding.

Funneled energy landscape graphic adapted from Ken Dill (http://www.dillgroup.ucsf.edu)
of conformations (e.g., local secondary structures), most of which are unproductive and rapidly fall apart. Yet when two or more pieces of natively hydrogen bonded foldons are present and make native tertiary contacts to bury surface area, hydrogen bonds of the participating foldons are stabilized and this conformation serves as a more accessible foundation for subsequent addition of new foldons. Although propagation of such foldon addition may not follow a hierarchical order as pronounced as that of late folding pathway and some bifurcations may occur if two foldons can be added with comparable energy, the energetically uphill early folding events are overall linear with no more than a few kinetically preferred pathways if not a single one.

Using this reasoning along with experimental results, complete folding pathway of ubiquitin has been proposed (Krantz et al., 2004; Sosnick et al., 2006) (see section 3.2 for more details). Each addition of foldon occurs with commensurate level of hydrogen bond formation and surface burial, which is suggested by the aforementioned kinetic isotope effect (Krantz et al., 2000b; 2002b). Uversky and Fink’s analysis on 41 native and partially unfolded states, which established a good correlation between decrease of hydrodynamic volume and increase in secondary structure during folding, also supports this view (Uversky and Fink, 2002). Apparently sequential pathway view clashes against the general funneled landscape view, especially regarding the early stages of folding. Results from experimental studies on early folding events are greatly anticipated to allow testing of proposals.
1.4 **Scope of this thesis**

As a firm believer of the sequential foldon stabilization mechanism of folding, I devote this thesis to the following two major aspects that require effort.

First, although in principle NSHX should be able to detect any stable folding intermediate at the native side of barrier, in practice such stable intermediates can become unobservable under some circumstances (see section 2.2 for detailed discussion). Considering the generally superior sensitivity of NSHX for identifying folding intermediates compared to other methods, there still exist cases where identification and characterization of folding intermediates is not possible. Chapter 2 reports development of an effective and generally applicable strategy ("protein vivisection") for trapping late folding intermediates, which can then be interrogated with NSHX. The idea behind the strategy – foldon unfolding by disrupting hydrophobic burial – comes from concomitant secondary and tertiary structure formation, which is first suggested by kinetic isotope effect (Krantz *et al.*, 2000b; 2002b) and then directly proven by the success of vivisection strategy in reverse.

Second, early protein folding events still remain almost untouched by experimental approaches due to the unstable nature of early folding intermediates. In chapter 3 I share my successful and unsuccessful experiences in characterizing early folding intermediates. The results support the proposed foldon stabilization by native tertiary interactions at early stage of folding. The dynamic nature of early folding intermediates proves to be the greatest challenge for characterization and points to a method that is most promising for solving the problem.
2 Protein Vivisection Reveals Elusive Folding Intermediates

2.1 Abstract

Although most folding intermediates escape detection, their characterization is crucial to the elucidation of folding mechanisms. Here, we outline a powerful strategy to populate partially unfolded intermediates: A buried aliphatic residue is substituted with a charged residue (e.g., Leu→Glu−) to destabilize and unfold a specific region of the protein. We applied this strategy to ubiquitin, reversibly trapping a folding intermediate in which the β5-strand is unfolded. The intermediate refolds to a native-like structure upon charge neutralization under mildly acidic conditions. Characterization of the trapped intermediate using NMR and hydrogen exchange methods identifies a second folding intermediate and reveals the order and free energies of the two major folding events on the native side of the rate limiting step. This general strategy may be combined with other methods and have broad applications in the study of protein folding and other reactions that require trapping of high-energy states.

2.2 Introduction

Multiple points on the free energy surface must be structurally and thermodynamically characterized to fully delineate protein folding mechanisms. However, this task is extremely challenging, in particular for proteins that fold in an apparent two-state manner. Folding intermediates tend to be transient and weakly populated in both kinetic and equilibrium studies. Two major strategies are used to address this challenge. The most straightforward and
information-rich method is direct trapping followed by structural characterization. However, trapping often involves the use of unnatural conditions (e.g., low pH, cosolvents) (Pan and Briggs, 1992; Jourdan and Searle, 2001), or relies on protein-specific mutagenesis (Eliezer et al., 1998; Korzhnev et al., 2004; Valle-Belisle and Michinick, 2007; Whittaker et al., 2007; Rea et al., 2008; Connell et al., 2009).

The other major strategy involves the detection of sparsely populated intermediates—for example, using NMR relaxation dispersion (RD) (Korzhnev et al., 2004; Neudecker et al., 2009), native-state hydrogen exchange (NSHX) (Bai et al., 1995; Englander et al., 1997), chemical modification (Silverman and Harbury, 2002; Sridevi and Udgaonkar, 2002), or proteolysis (Wang and Kallenbach, 1998; Park and Marqusee, 2004). RD methods can characterize excited states that are populated as little as 0.5% and provide interconversion rates, stability, and the chemical shifts of the excited state (Neudecker et al., 2009). RD methods have characterized intermediates that reside either before or after the rate-limiting transition state (TS) for a number of proteins (Grey et al., 2006; Tang et al., 2006; Korzhnev and Kay, 2008).

NSHX can detect partially unfolded intermediates by examining the denaturant dependence of the HX rates. The regions of the protein that unfold cooperatively in subglobal openings are termed foldons. Foldons are typically associated with elements of secondary structure. Trapped intermediates have been created by selectively destabilizing or deleting a foldon (Chamberlain et al., 1999; Feng et al., 2004; Bai et al., 2007; Kato et al., 2007; Whittaker et al., 2007). The resulting intermediates generally are on-pathway, located before or after the rate-limiting step in two-state or multi-state folding, respectively.

RD and NSHX methods are powerful but have certain limitations. RD requires appropriate interconversion rates and chemical shift differences, and the intermediate must
comprise at least ~0.5% of the total population. NSHX requires that the hydrogen bonds broken in the subglobal opening are not hidden by faster exchange processes, including smaller-scale (local) openings (Figure 2.1). In addition, the free energy and the size of the subglobal opening must be such that its HX “isotherm” is distinct from those of global and local unfolding processes. These requirements may explain why NSHX studies on mammalian ubiquitin (Ub) did not reveal any subglobal openings (Sidhu and Robertson, 2004).

These issues call for new strategies for trapping folding intermediates. In this study, we explored a strategy that involves reversibly destabilizing a foldon. The destabilization is accomplished by replacing a buried aliphatic residue with a nearly isosteric but charged acidic residue (e.g., Leu→Glu⁻) (Harper et al., 2004; Isom et al., 2008). At neutral pH, the energetic penalty associated with the burial of the acidic group destabilizes the foldon, thereby creating an equilibrium intermediate amenable for detailed characterization. We applied this strategy to Ub to reversibly populate an intermediate that had escaped detection by NSHX. Kinetic folding experiments indicate that it is a late folding intermediate on the native side of the rate-limiting barrier. The application of NSHX to the trapped folding intermediate identifies a second unfolding intermediate. These results reveal a hierarchical unfolding pathway beginning with the loss of the β5-strand followed by the disruption of a small 3\textsubscript{10}-helix.
Figure 2.1 NSHX can fail to detect subglobal openings. a. Multiple subglobal unfolding events (S1, S2, green, red) can exist but none of them are detected because their isotherms are masked by the faster local unfolding events (grey) at low denaturant, and by global unfolding events at high denaturant concentration (U, blue). The S1 subglobal opening can be trapped (N-S1) and characterized using the charge-burial strategy (right panel). NSHX can be applied on this trapped PUF to identify subglobal opening S2. Local unfolding events are assumed to be at a fixed energy above the ground state of the system. b. If the subglobal opening involves little exposure of surface area, the subglobal unfolding cannot easily be distinguished from a local unfolding event. In this case the structural identity of the subglobal opening can be identified by populating the intermediate using the charge-burial strategy and then identifying which hydrogen bonds are lost using HX at high pH and NMR readout on the N* state where the amide resonances can be assigned.
2.3 Results

2.3.1 Rationally populating intermediates

We have previously proposed a detailed folding pathway for Ub using a combination of results including HX (Pan and Briggs, 1992) and ψ-analysis, our bihistidine metal binding method that identifies inter-residue contacts in the TS ensemble (TSE) (Krantz et al., 2004; Sosnick et al., 2006; Baxa et al., 2009). Each major pathway step both on the way to and from the TSE is the addition or consolidation of regions of secondary structure. These events occur with a commensurate level of hydrogen-bond formation and surface burial, as indicated by our kinetic isotope effect studies (Krantz et al., 2000b; 2002b). The only secondary structures that form after the TS are the small 3_{10}-helix and the β5-strand (Figure 2.2). This identification suggests the presence of one or two post-TS, kinetically silent intermediates lacking one or both of these structural elements. Such intermediates are likely to be unfolding intermediates whose thermodynamics may be characterized using HX. An earlier HX study (Pan and Briggs, 1992) indicated that hydrogen bonds between β4 and β5 generally are ~1 kcal mol⁻¹ less stable than bonds in the 3_{10}-helix. This difference suggests that the β5-strand unfolds prior to the helix on the unfolding pathway (N → N^{β5} → N^{β5,3_{10}} → TS, Figure 2.1).

In order to investigate the sequential order of unfolding events, we set out to trap one or both of these intermediates. The buried leucine on β5 is replaced with a glutamic acid (UbL50E). The energetic penalty for burying the charged Glu⁻, or, more precisely, the energy required to shift acid's pKa so that it remains neutral at a pH above its intrinsic pKa (Isom et al., 2008), is anticipated to unfold the β-strand, thereby populating the N^{−β5} conformation. This
Figure 2.2 Post-TS event in ubiquitin folding. The present study indicates that the folding of the 3$_{10}$ helix and β5 strand preferentially occur along the upper pathway. Mutations employed in the present study strategy are highlighted (V5, yellow; L43, blue; L50, magenta; L56, orange; L67, green), except I30E on the interior face of the helix.
unfolding event would be analogous to the disruption of the LOV2 protein's signaling helix upon an I→E substitution (Harper et al., 2004) and the partial unfolding of a helix in SNase (Karp et al., 2007), but it is possibly different from other results with this protein as nearly all substitutions are well tolerated (Isom et al., 2008). For UbL50E at pH below glutamic acid's intrinsic pKa, the nearly isosteric Glu$^0$ may adequately mimic the Leu, allowing the β-strand to refold and generate the native-like N* conformation. If successful, this strategy provides a system whereby pH can be used to reversibly convert Ub from a native-like state to a late folding intermediate, which can be characterized using equilibrium methods.

2.3.2 Thermodynamic and kinetic properties

Structural, kinetic, and thermodynamic measurements are performed on UbL50E to investigate the ability of the charge-burial strategy to trap a partially unfolded state. A pH titration of UbL50E at 4 °C, monitoring near-UV circular dichroism (CD), indicates the presence of a minor unfolding transition with a midpoint near pH 6.0 (Figure 2.3a; a similar transition is observed by NMR following the chemical shift of amide protons, see below). At pH 4.0, the protein has approximately the wildtype $[\theta]_{230\text{ nm}}$ value, while the high-pH species has lost structure ($\Delta[\theta]_{230\text{ nm}}=2700$ deg cm$^2$ dmol$^{-1}$). The observed pH transition midpoint is 1.7 units above the intrinsic pK$_a^{\text{Glu}}$ because β5's folding energy helps stabilize the folding transition by 2 kcal mol$^{-1}$. This amount is 4 kcal mol$^{-1}$ less than the free energy for opening β5 in the F45W pseudowild-type Ub (pWTUb; see HX results below), presumably because the Glu$^0$ does not fully mimic the Leu (Karp et al., 2007) in the low-pH N* species.

The folding rates of UbL50E are measured as a function of denaturant concentration using chevron analysis and stopped-flow methods following F45W fluorescence (Figure 2.3) to
Figure 2.3 Equilibrium and kinetic folding behavior of UbL50E. a. Upon acidification, the protein shifts from an intermediate to a native-like state. b. At both high and low pH, the chevron plot of UbL50E and pWTUb have nearly overlapping folding arms but distinct unfolding arms. At high pH, the UbL50E unfolding arm is shallower, suggesting disruption of β5 strand in the intermediate state. Illustrative folding energy landscapes are shown at right. Chevron values for the pWTUb have been shifted by the amount indicated by the arrows to account for the addition of 2 or 2.3 M guanidine hydrochloride required to unfold the pWTUb.
further characterize the low- and high-pH conformations. At pH 7.8, the folding rates of UbL50E and pWTUb are statistically identical (ΔΔG‡ = −0.25±0.32 kcal mol⁻¹) in spite of the considerable destabilization imparted by the substitution of the acidic residue (ΔΔGeqL50E of ~6.4±0.4 kcal mol⁻¹). At pH 4.0, the folding rate of UbL50E is mildly slower than that of pWTUb. In energetic terms, however, the difference ((ΔΔG‡ = 1.1±0.2 kcal mol⁻¹) remains much smaller than the destabilization (ΔΔGeqL50E = 6.8±0.3 kcal mol⁻¹). The resulting mutational ϕL50E values are 0.16±0.03 and 0.04±0.05 at pH 4.0 and pH 7.8, respectively. These low values are consistent with our identification of β5 as a structure formed after the TS along the folding pathways of pWTUb and UbL50E.

At both low and high pH, the folding rates of the two proteins have the same urea dependence (mᵢ-value; Table 2.1). This correspondence indicates that a comparable amount of urea-sensitive surface area is buried in their TSs (relative to their denatured states). This result, along with the similarity of their folding rates and the lack of β5 in their TS, suggests that the two proteins have very similar TS structures at both pH conditions.

Although the folding rate of UbL50E is largely unchanged, its unfolding rate at both pH 4.0 and pH 7.8 is much faster than the unfolding rate of pWTUb. This acceleration is due to the stabilizing effect of burying the acidic group in UbL50E. In addition at higher pH, the slope of the chevron unfolding arm (mu-value) for UbL50E is shallower by ~15% of the equilibrium mo-value as compared with pWTUb. This decrease in slope for UbL50E indicates that the starting state of its unfolding reaction has lost some structure, presumably the destabilized β5-
Table 2.1 Kinetic parameters obtained from chevron analysis.

<table>
<thead>
<tr>
<th>Protein (Figure)</th>
<th>Solvent condition</th>
<th>$m_r$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$m_u$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$\Delta G_f^\circ (0)$ (kcal mol$^{-1}$)</th>
<th>$\Delta G^\circ$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWTUbc (Figure 2.3)</td>
<td>N-state pH 7.8</td>
<td>0.60 ± 0.04</td>
<td>0.35 ± 0.02</td>
<td>3.97 ± 0.24</td>
<td>-8.64 ± 0.35</td>
</tr>
<tr>
<td>pWTUbc (Figure 2.3)</td>
<td>N-state pH 4.0</td>
<td>0.64 ± 0.03</td>
<td>0.36 ± 0.02</td>
<td>4.62 ± 0.21</td>
<td>-9.18 ± 0.28</td>
</tr>
<tr>
<td>L50E (Figure 2.3)</td>
<td>I-state pH 7.8</td>
<td>1.08 ± 0.15</td>
<td>0.14 ± 0.02</td>
<td>4.22 ± 0.21</td>
<td>-2.29 ± 0.17</td>
</tr>
<tr>
<td>L50E (Figure 2.3)</td>
<td>N*-state pH 4.0</td>
<td>0.97 ± 0.05</td>
<td>0.32 ± 0.01</td>
<td>3.47 ± 0.06</td>
<td>-2.43 ± 0.04</td>
</tr>
<tr>
<td>L50E (Figure 2.9, HX condition)</td>
<td>I-state pD$_{read}$ 7.5</td>
<td>0.71 ± 0.02</td>
<td>0.18 ± 0.01</td>
<td>4.18 ± 0.05</td>
<td>-2.98 ± 0.07</td>
</tr>
<tr>
<td>L67E (Figure 2.11)</td>
<td>I-state pH 7.8</td>
<td>0.63 ± 0.02</td>
<td>0.21 ± 0.01</td>
<td>4.31 ± 0.04</td>
<td>-4.00 ± 0.07</td>
</tr>
<tr>
<td>L67E (Figure 2.11)</td>
<td>N*-state pH 4.0</td>
<td>0.68 ± 0.02</td>
<td>0.28 ± 0.01</td>
<td>4.10 ± 0.08</td>
<td>-4.39 ± 0.07</td>
</tr>
</tbody>
</table>
strand. This result is suggestive that high-pH conformation of UbL50E is a folding intermediate on the native side of the major folding barrier.

2.3.2 NMR analysis

A series of $^1$H–$^{15}$N heteronuclear single quantum coherence (HSQC) spectra were acquired from pH 4 to pH 7.8 (Figures 2.4 and 2.5) to characterize the low- and high-pH forms of UbL50E in detail. The spectra of the high- and low-pH states contain well-dispersed sharp peaks, indicating that both states contain well structured regions. At pH 4.0 (Figure 2.4) and slightly higher pH, most of the amide resonances overlapped or remained close to the pWTUb's resonances, suggesting that the UbL50E with Glu$^0$ retains a very native-like structure.

Resonances were assigned using triple resonance HNCACB and CBCACONH measurements in conjunction with the pWTUb assignments (F. Massi, unpublished data). To obtain a concentrated protein sample for the three-dimensional measurements, we used a solubility-enhancing buffer (Golovanov et al., 2004) (15 mM sodium phosphate, 225 mM sodium chloride, 50 mM glutamic acid, and 50 mM arginine). The enhancement is best near neutral pH, but the NMR spectra gradually deteriorate from pH 5.0 to pH 7.8. Accordingly, assignments were conducted at pH 5.3, where all NH resonances are assignable except for residues 48–51 on the $\beta$5-strand. These four resonances were not observed, presumably due to exchange broadening (pH 5.3 is in the middle of the pH-induced folding transition). At pH 4.0, four additional peaks appeared in the region appropriate for amide NH resonances. Because the other amide resonances are accounted for, we associated the new resonances with the four missing $\beta$5 resonances (Figure 2.4). The assignment of peaks at other solvent conditions was accomplished by tracking peak movement during the pH titration (Figure 2.5a and b).
Figure 2.4 NMR analysis of pWTUb and UbL50E. Overlay of $^1$H-$^{15}$N HSQC spectra of pWTUb and UbL50E at pH 4 (upper panel). The extensive peak overlap indicates that the two proteins have similar conformations. At pH 7.8, UbL50E has sharp and well-dispersed peaks, although with fewer resonances than at pH 4 (lower panel). The four peaks (blue circles) observable at pH 4 but not pH 5, are attributed to the four unassigned residues on $\beta5$. 
Figure 2.5 pH titration of UbL50E from pH 4 to pH 7.8. a. Representative resonance which shifts and then disappears due to exchange broadening. b. In the pH 6.2 to 7.8 range, some resonances move while others do not (right panel). c. Resonances classified as non-moving if their chemical shift remained unchanged from pH 6.2 to 7.8 (red box). d. The three classes of behavior of UbL50E peaks are mapped onto the structure: unchanged (blue), changed (magenta), and chemical shift changed with peak disappearing at neutral pH due to exchange broadening (orange).
In the pH titration, all the NH peaks shift, and some start to disappear near pH 6.0. The shifts reflect both the folding transition and the less interesting ionization of the surface-exposed Glu and Asp groups. Fortunately, their ionization equilibria are ~2 pH units below the midpoint of the CD-monitored folding transition. Consequently, the surface Glu and Asp side chains are nearly fully ionized by pH 6.0, and hence the remaining chemical shift movement at higher pH levels reflects the partial unfolding transition.

Across the pH 6.2–pH 7.8 range, 48%, 28%, and 24% of the observable NH resonances remained unchanged (\(|\Delta \delta_{HN}| < 0.05 \text{ ppm} \); \(|\Delta \delta_{N}| < 0.15 \text{ ppm} \)), shifted, and shifted and then disappeared, respectively (Figure 2.5). These three classes are interpreted as minimal or significant backbone conformational change, which may also be accompanied by extreme exchange broadening due to millisecond timescale dynamics. The residues composing each class map onto the native structure of ubiquitin’s according to their distance from the β5-strand (Figure 2.5d) and ψ-values (Krantz et al., 2004). The resonances for the β1–β2 hairpin and carboxy-terminus of the α-helix, the most distal regions, are the most invariant, while the resonances for β4, 3_{10}-helix, located adjacent to β5, are the most perturbed. The exchanged broadened peaks do not reappear as sharp peaks with random-coil chemical shifts. Hence, the β5-strand is not behaving as a completely unfolded loop. Presumably, the loop interacts dynamically with the rest of the protein.

2.3.4 Hydrogen exchange

HX measurements are conducted on the intermediate form of UbL50E to characterize its structure and thermodynamics, and potentially identify other partially unfolded states. Under our experimental conditions (pD_{read} 7.5, 4 °C), intrinsic exchange rates are from seconds to
minutes, which are faster than standard HSQC acquisition times. Accordingly, a pulse-labeling strategy is employed to measure HX rates. HX is initiated with a 1:4 dilution with D$_2$O to a pD$_{\text{read}}$ of 7.5. After a specified time, H-to-D exchange is quenched by reducing the pD$_{\text{read}}$ to 2.9 with DCl. In addition to quenching HX, acidification also generates the N* state where the exchange-broadened resonances in the intermediate state become sharp and assignable. Hence, the HX rates for otherwise NMR invisible residues in the intermediate can be determined using this protocol. HX likely occurs via the EX2 mechanism as $k_{\text{int}}$ is much slower than global refolding rates. Hence, the stability of the hydrogen bonds can be determined for $\sim$80% of the intramolecular hydrogen bonds observable in the crystal structure.

In the intermediate, significant HX protection is observed for all of the native secondary structures except for the single measured hydrogen bond between strands $\beta$4 and $\beta$5 (NH$_{W45}^\alpha$→O=C$_{K48}^\alpha$). For this bond, $k_{\text{HX}}/k_{\text{int}}=0.3\pm0.3$ s$^{-1}$/0.3 s$^{-1}$. Hence, the protection factor essentially is unity. The four unassigned peaks associated with $\beta$5 also have a $k_{\text{HX}}/k_{\text{int}}$ of $\sim$1 (observed HX is complete within 3 s; $k_{\text{int}}$ of $\sim$0.5–3 s$^{-1}$). Hence, the hydrogen bonds between $\beta$4 and $\beta$5 are broken most of the time. For the remaining hydrogen bonds, $\Delta G_{\text{HX}}$ is between 1.7 and 3.5 kcal mol$^{-1}$ (Figure 2.6; 2.7; Table 2.2).

2.3.5 Small-angle X-ray scattering

We compared the global dimensions of pWTUb with the UbL50E intermediate using small-angle X-ray scattering (SAXS) (Figure 2.8). At pH 7.8, the measured radii of gyration ($R_g$'s) are 13.2±0.2 and 13.8±0.1 Å for the two proteins, respectively. This difference matches the 0.5-Å difference between $R_g$ values of the native state and the model for the N$^{-}\beta$5 intermediate but is less than the 1.6-Å difference for the N$^{-}\beta$5$_{10}$ intermediate (Figure 2.2).
Figure 2.6 Native-state hydrogen exchange on UbL50E I state at pHread 7.5- I. a. Hydrogen bond network with direction of bond (NH→O=C) and $K_{eq}$ noted. Arrows are color-coded according to their stability (red = global, yellow= $3_{10}$ opening, orange = other, white = absent) while boxes are color-coded according to the type of chemical shift movement during the pH titration (adapted from Figure 2.5). b. Representative NSHX denaturant dependent isotherms. Isotherms for protons on the $3_{10}$ helix have lower $\Delta G_{HX}$ and m values compared to the rest of the isotherms. c. Histogram of $\Delta G_{HX}$ and m-values. The positions of the strands (blue), $\alpha$-helix (red), and $3_{10}$-helix (green) are noted at the top.
Figure 2.7 Isotherms of NSHX on UbL50E I state at pD_{read} 7.5. Denaturant dependent isotherms for all measured residues classified according to exchange mechanism (upper left: global, upper right: near-global I, lower left: near-global II, lower right: subglobal 3_{10}-helix).
### Table 2.2 Native-state hydrogen exchange on UbL50E intermediate, pD\textsubscript{read} 7.5

<table>
<thead>
<tr>
<th>2\textsuperscript{nd} structure of H-bond donor</th>
<th>H-bond donor (-NH)</th>
<th>H-bond acceptor (-C=O)</th>
<th>ΔG\textsubscript{HX} (kcal mol\textsuperscript{-1})</th>
<th>m (kcal mol\textsuperscript{-1} M\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>β2 strand</td>
<td>I3</td>
<td>L15</td>
<td>2.90±0.05</td>
<td>1.03±0.09</td>
</tr>
<tr>
<td></td>
<td>F4</td>
<td>S65</td>
<td>3.16±0.04</td>
<td>1.11±0.07</td>
</tr>
<tr>
<td></td>
<td>V5</td>
<td>I13</td>
<td>3.10±0.01</td>
<td>1.12±0.03</td>
</tr>
<tr>
<td></td>
<td>K6</td>
<td>L67</td>
<td>3.01±0.01</td>
<td>1.00±0.02</td>
</tr>
<tr>
<td></td>
<td>T7</td>
<td>K11</td>
<td>3.20±0.09</td>
<td>1.16±0.16</td>
</tr>
<tr>
<td>β1 strand</td>
<td>I13</td>
<td>V5</td>
<td>3.06±0.03</td>
<td>1.04±0.06</td>
</tr>
<tr>
<td></td>
<td>L15</td>
<td>I3</td>
<td>3.38±0.07</td>
<td>1.16±0.06</td>
</tr>
<tr>
<td></td>
<td>V17</td>
<td>M1</td>
<td>3.48±0.03</td>
<td>1.16±0.06</td>
</tr>
<tr>
<td>α-helix</td>
<td>V26</td>
<td>T22</td>
<td>3.21±0.06</td>
<td>1.00±0.15</td>
</tr>
<tr>
<td></td>
<td>K27</td>
<td>I23</td>
<td>3.31±0.12</td>
<td>1.17±0.19</td>
</tr>
<tr>
<td></td>
<td>A28</td>
<td>E24</td>
<td>3.55±0.04</td>
<td>1.18±0.07</td>
</tr>
<tr>
<td></td>
<td>K29</td>
<td>N25</td>
<td>3.30±0.06</td>
<td>0.95±0.09</td>
</tr>
<tr>
<td></td>
<td>Q31</td>
<td>K27</td>
<td>2.68±0.13</td>
<td>0.75±0.20</td>
</tr>
<tr>
<td></td>
<td>E34</td>
<td>I30</td>
<td>2.71±0.03</td>
<td>0.77±0.05</td>
</tr>
<tr>
<td>β4 strand</td>
<td>R42</td>
<td>V70</td>
<td>2.87±0.04</td>
<td>0.93±0.08</td>
</tr>
<tr>
<td></td>
<td>I44</td>
<td>N68</td>
<td>2.36±0.05</td>
<td>0.60±0.13</td>
</tr>
<tr>
<td></td>
<td>W45</td>
<td>K48</td>
<td>~0\textsuperscript{1}</td>
<td>ND\textsuperscript{2}</td>
</tr>
<tr>
<td>3\textsubscript{10} helix</td>
<td>L56</td>
<td>D21</td>
<td>1.71±0.11</td>
<td>0.47±0.29</td>
</tr>
<tr>
<td></td>
<td>S57</td>
<td>P19</td>
<td>1.93±0.03</td>
<td>0.42±0.07</td>
</tr>
<tr>
<td></td>
<td>D58</td>
<td>T55</td>
<td>1.86±0.03</td>
<td>0.32±0.07</td>
</tr>
<tr>
<td></td>
<td>Y59</td>
<td>L56</td>
<td>1.73±0.01</td>
<td>0.23±0.02</td>
</tr>
<tr>
<td>β3 strand</td>
<td>L67</td>
<td>F4</td>
<td>2.14±0.16</td>
<td>0.73±0.25</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>N68</td>
<td>I44</td>
<td>2.38±0.05</td>
<td>0.51±0.06</td>
<td></td>
</tr>
<tr>
<td>L69</td>
<td>K6</td>
<td>2.51±0.03</td>
<td>0.57±0.08</td>
<td></td>
</tr>
<tr>
<td>V70</td>
<td>R42</td>
<td>2.42±0.04</td>
<td>0.82±0.08</td>
<td></td>
</tr>
<tr>
<td>R72</td>
<td>Q40</td>
<td>2.82±0.07</td>
<td>0.87±0.12</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2.2 (continued)*

1. $k_{HX}/k_{int} \sim 1$

2. Not determined
Figure 2.8 Dimensions of pWTUb and UbL50E intermediate measured using SAXS. $R_g$ values are obtained from a Guinier analysis performed over the data range highlighted. Protein concentration is ~2 mg/ml. Solvent condition was the same as used for NMR assignments except at pH 7.8, 0% D$_2$O.
In summary, a combination of results — the loss of secondary structure (CD, pH titration) that is formed on the native side of the TS (shallower unfolding chevron arm), the dynamic behavior of the β5-strand relative to other regions (HSQC assignments, line broadening), the loss of hydrogen bonds only between β4 and β5 (HX), and the 0.5-Å increase in $R_g$ (SAXS) — demonstrates that the L50E mutant selectively disrupts the β5-strand at neutral pH, thereby trapping the protein in the N$^{-}\beta_5$ conformation.

2.3.6 Denaturant dependence of HX

I next investigated the denaturant dependence of HX on the N$^{-}\beta_5$ state of UbL50E to identify the nature of the openings leading to exchange for each amide proton (Figure 2.6; 2.7; Table 2.2). From the linear decrease in stability as a function of added urea, $\Delta G_{\text{HX}}([\text{denaturant}])=\Delta G_{\text{HX}}(0)+m[\text{denaturant}]$, the surface exposure of the structural fluctuation leading to exchange is determined. Nearly half of the amide protons exchange with the same $\Delta G_{\text{HX}}$ and m-value as measured for global unfolding, 3 kcal mol$^{-1}$ and 1 kcal mol$^{-1}$M$^{-1}$, respectively, determined by equilibrium denaturation and kinetic measurements (Table 2.1; Figure 2.9). These amide protons are exchanging from the globally unfolded state and are located in the β1–β2 hairpin and the N-terminal half of the α-helix, as well as two long-range bonds between strands β2 and β3 (NH$^{K6}_4\rightarrow O=C^{L67}_3$ and NH$^{F4}_4\rightarrow O=C^{S65}_3$).

The hydrogen bonds on the 3_10-helix stand out with their uniformly low stability ($\Delta G_{\text{HX}}=1.8–2.0$ kcal mol$^{-1}$) and mild response to denaturant (m=0.25–0.50 kcal mol$^{-1}$ M$^{-1}$). Their uniformity indicates that the small helix is a second foldon that unfolds in a concerted event, after the loss of the β5-strand. The similarity of the m-value for the 3_10-helix and the m$_{u}$-
Figure 2.9 Dependence of UbL50E stability on $D_2O$ level. Comparison of chevrons in $H_2O$, pH 7.8 (15 mM phosphate, 225 mM NaCl) and 80% $D_2O$, $pD_{\text{read}}$ 7.5 used in the HX studies (extra 10 mM Glu, 10 mM Arg, 80% $D_2O$). HX buffer stabilizes UbL50E I by ~ 1 kcal mol$^{-1}$. The $K_{eq}$ determined from the kinetic and equilibrium data agrees with $DG_{HX}$ of global unfolding from HX study under the same solvent condition.
value from the chevron analysis indicates that the amount of surface burial exposed in the unfolding of the foldon is nearly the same as that exposed in the TSE.

The hydrogen bonds between $\beta$3 and $\beta$4 and in the carboxy-terminal portion of the $\alpha$-helix have slightly lower $\Delta G_{\text{HX}}$ and m-values compared with the globally exchanging hydrogen bonds. Although this difference is near our experimental error, it is observed uniformly across these regions. Potentially, the mild decrease is due to a slight amount of mixing between a global exchange process and a smaller-scale event. Alternatively, the unfolded state may contain some residual structure in the $\beta$1–$\beta$3 strands and $\alpha$-helix, which is marginally stable only under aqueous conditions (see section 3.3.3 for further discussions).

A free energy surface can be constructed for pWTUb and the UbL50E using the kinetic, NMR and NSHX data (Figure 2.10). Structurally, the unfolding pathway goes through two PUFs, $N \leftrightarrow N^{\beta 5} \leftrightarrow N^{\beta 5,3_{10}} \leftrightarrow \text{TS}$. At pH 7.8, UbL50E, $\Delta G_N$ is 0.9 kcal mol$^{-1}$ less stable than $\Delta G_{N^{\beta 5,3_{10}}}$, according to calculations based on apparent pKa from pH-induced folding titration monitored by CD (Figure 2.3). Hence, the ladder of free energies for UbL50E at pH 4.0 and pWTUb at both pH 4.0 and pH 7.8 is $\Delta G_N < \Delta G_{N^{\beta 5}} < \Delta G_{N^{\beta 5,3_{10}}}$, whereas for UbL50E at pH 7.8, the ladder is $\Delta G_{N^{\beta 5}} < \Delta G_{N^{\beta 5,3_{10}}} < \Delta G_N$.

2.3.7 Other charged substitutions

We individually introduced five other buried aliphatic-to-charge mutations into $\beta$1, the $\alpha$-helix, $\beta$4, $3_{10}$-helix, and $\beta$3 (V5D, I30E, L43E, L56E, and L67E; Figures 2.2 and 2.11). At neutral pH, the L56E substitution on the $3_{10}$-helix is anticipated by folding hierarchy to unfold both this helix and the $\beta$5-strand, thereby populating the $N^{\beta 5,3_{10}}$ state. This species is unstable
Figure 2.10 Post-TS free energy surface of ubiquitin folding. Values for $N^{-\beta_5}$ were derived from kinetic data for UbL50E and pWTUb, while values for $N^{-\beta_5, -3_{10}}$ were obtained from NSHX studies on the UbL50E intermediate. The energy of the UbL50E at pH 7.8 with the $\beta_5$-strand folded was obtained from the apparent pKa of the CD-monitored pH-induced unfolding transition (pH 5.8; see Figure 2.3).
Figure 2.11 Charge-burial partial unfolding strategy applied at other positions. a. NMR $^1$H–$^{15}$N HSQC spectra. b. Corresponding chevron plots of UbL67E at low and high pH compared with pWTUb.
and prone to precipitate with a virtually blank HSQC spectrum presumably due to extreme line
broadening. Nevertheless, this species refolds at pH 4.0 with an NMR spectrum having similar
dispersion as the native protein. The potential loss of the 3_{10}-helix could be tested with HX
labeling at neutral pH followed by NMR analysis at low pH, as done for UbL50E. Irrespective of
the exact structural content, the NMR spectrum of the L56E species lacks sharp and dispersive
peaks at neutral pH, suggesting that the protein becomes more dynamic on the unfolding
pathway back toward the TS.

The L67E and L43E substitutions are located on β3 and β4, respectively. These two
strands are present in the TS according to ψ-analysis. Both mutants have well-dispersed HSQC
spectra at both pH 4.0 and pH 8.0. The L67E species has a native number of peaks, while L43E
has slightly less at pH 8.0. The kinetic folding properties of the L67E mutant are similar to those
of L50E, in particular possessing a shallower chevron unfolding arm slope at pH 8.0 as
compared with pWTUb (Figure 2.11b; Table 2.1). However, the ~10% reduction of the mw-value
for L67E is insufficient to account for disruption of the entire β3-strand, which makes multiple
interactions with β2 and β4 in the TS. We infer that the L67E mutation induces some local
structural fraying of β3 at high pH that exposes some buried surface area, rather than the loss of
the entire strand.

The V5D and I30E mutations are located on the amino-terminal hairpin and the α-
helix, respectively. These two secondary structures form prior to the TS in the folding pathway,
and the two substituted positions are located in the core of the protein. The HSQC spectrum of
V5D exhibits extensive peak broadening at pH 8.5, whereas I30E’s spectrum remains dispersive
with extra sets of peaks. At pH 4.0, the V5D spectrum has less dispersion than I30E, which is
similar to pWTUb. The V5D substitution is disruptive, indicating that charge burial can identify regions of the protein that cannot be unfolded without disrupting the rest of the protein. Such regions are likely to form early in the folding pathway.

2.4 Materials and Methods

2.4.1 Proteins

UbL50E and other mutants, based on a pWTUb (with F45W and H68N), were prepared according to Krantz et al (Krantz et al., 2004). Proteins in inclusion bodies were solubilized in 8 M urea buffer before HPLC purification.

2.4.2 Equilibrium and kinetic measurements

CD experiments were performed with a Jasco 715 spectropolarimeter with a path length of 1 cm in 15 mM sodium phosphate and 225 mM sodium chloride. Rapidmixing fluorescence experiments used a Biologic SFM-400 stopped-flow apparatus connected via a fiber optic cable to a PTI A101 arc lamp. Fluorescence spectroscopy used excitation and emission wavelengths of 280–290 nm and 300–400 nm, respectively.

2.4.3 NMR spectroscopy and HX chemistry

All the NMR experiments were run on a 600-MHz Varian Unity Inova spectrometer equipped with a cryoprobe. HNCACB and CBCACONH experiments were run on $^{15}$N,$^{13}$C-labeled UbL50E at 0.5 mM concentration in 15 mM sodium phosphate, 225 mM sodium chloride,
50 mM Glu, 50 mM Arg, pH 5.3, and 5% D2O to obtain resonance assignments. The two-dimensional $^{15}$N HSQC spectra were taken with $^{15}$N-labeled UbL50E at 0.1 mM protein concentration in the same buffer with added 10 mM Glu and Arg. The protein at 0.5 mM, pH 6.6, was diluted 1:4 to a final condition of 15 mM sodium phosphate, 225 mM sodium chloride, 10 mM Glu, 10 mM Arg, 80% D2O, and pD$_{\text{read}}$ 7.5 to initiate HX. Exchange was quenched by the addition of 3.5 $\mu$L of 6M DCl to 500 $\mu$L of protein solution, which dropped pD$_{\text{read}}$ to 2.9. The HX rate at the quenching condition ($k_q$) was also measured for a fully protonated sample and fit with the following equation in order to correct for back-exchange during the NMR measurement:

$$V_T = 0.8 \times V_0 \times \left( \frac{(1 - \exp(-k_q \times L))}{k_q \times L} \right) \times \exp(-k_q \times T) + 0.2 \times V_0$$ (equation 2.1)

where $T$ is the HSQC starting time, $L$ is the length of the HSQC measurement, $V_T$ is the NMR peak volume at $T$, and $k_q$ and $V_0$ are fitting parameters. The HX results were fit with the equation:

$$V_t = 0.8 \times V_0 \times \left( \frac{(1 - \exp(-k_q \times L))}{k_q \times L} \right) \times \exp(-k_{\text{HX}} \times t) + 0.2 \times V_0$$ (equation 2.2)

where $t$ is the length of the labeling period at pD$_{\text{read}}$ 7.5, $V_t$ is the peak volume at time $t$, and $k_{\text{HX}}$ and $V_0$ are fitting parameters.

2.4.4 Small-angle X-ray scattering

Data were collected at the BioCAT beamline at the Advanced Photon Source, Argonne National Laboratory. Samples were flowed through a 1.5-mm capillary at a rate of 2 $\mu$L/s, and 15 exposures of $\sim 1^s$ duration were collected at room temperature. The exposures were averaged, buffer blanks were subtracted, and Guinier analysis was performed using IGOR Pro.
2.5 Discussion

Rather than relying on serendipitously found mutations, our charge-burial strategy is a rational, active, and generally applicable approach for reversibly trapping partially unfolded states and delineating folding pathways. For a protein where little is known about its folding behavior, the charge-burial strategy can be extremely revealing. Leu/Ile→Glu or Val→Asp substitutions can be introduced into each potential foldon. The acquisition of NMR spectra under acidic and neutral conditions can identify which parts of the protein can be selectively disrupted without unfolding the entire protein. Once a potential folding intermediate is found through such substitutions, it can be further interrogated with HX, chevron analysis (Figure 2.12), and, possibly, structure determination. The intermediate can also be used as the new ground state to perform additional NSHX and RD measurements to “walk” up the energy surface, identify other folding intermediates, and test tentative conclusions from the preliminary scan using charged residues.

By introducing the L50E substitution into ubiquitin, we trapped the N−β5 species, a late folding intermediate, and characterized its structure, thermodynamics, and dynamics using CD, NMR, SAXS, and chevron analysis. This trapped species has very native-like regions on the opposite side of the protein (no shift in the sharp $^{15}\text{N}–^{1}\text{H}$ resonances). The disrupted β5-strand and neighboring regions, however, undergo extreme exchange broadening due to millisecond timescale dynamics, although the backbone hydrogen bonds remain in the $3_{10}$-helix and between strands β3 and β4.

I performed NSHX measurements on the intermediate, taking advantage of the well-behaved NMR spectrum of the low-pH native-like species. These measurements identified an
Figure 2.12 Interpretation of possible chevrons generated by the charge-burial strategy. For an aliphatic-to-charged substitution, the chevron can have a similar folding arm but an altered unfolding arm at neutral pH. The change in slope can be used to estimate the amount of structure lost due to the substitution. If the decrease of slope is consistent with the estimate (red), it is suggestive of the creation of a folding intermediate lacking structure near the site of the substitution that otherwise is formed after the TS in the folding of the wild-type protein. If the decrease of slope is smaller than the estimate (not shown), it may be indicative of a locally frayed conformation rather than a genuine folding intermediate. Alternatively, the decrease may be consistent with the loss of an additional foldon (blue) if folding is hierarchical, being dependent on the perturbed foldon. This possibility can be tested by further characterization (e.g., with NSHX). Changes in the folding arm indicate that the TS is destabilized by the substitution. When the slope of the folding is decreased, some structure has been lost in both the TS and the ground state (cyan). More complicated scenarios are also possible.
additional unfolding intermediate lacking the $3_{10}$-helix. As far as I am aware, this is the first report of a folding intermediate identified through NSHX applied to another intermediate.

Even for an equilibrium study that identifies an extensive ladder of unfolding intermediates, the kinetic connectivity can only be inferred. In the case of ubiquitin, we have additional kinetic information. The only secondary structures that form between the TS native state are the $\beta 5$ strand and the $3_{10}$-helix (Krantz et al., 2004). Hence, the unfolding of either of these elements is highly likely to be the first step along the unfolding pathway. Our work demonstrates that a species lacking only the $\beta 5$-strand lies at a lower free energy than the state lacking the $3_{10}$-helix and that the $\beta 5$ strand can be individually disrupted. Furthermore, the stability of the $3_{10}$-helix responds to the L50E substitution to the same degree as does the $\beta 5$ strand, thereby demonstrating that a structural hierarchy exists between these two foldons. We also found that the unfolding rate starting from the $N^{-\beta 5}$ species is faster than that from $N$ yet the pathways go through the same TS, and the trapped species is in between the native state and TS when considering urea-sensitive surface burial (m-value). All these results point to the early stages of the kinetic unfolding pathway going through two intermediates, $N \rightarrow N^{-\beta 5} \rightarrow N^{-\beta 5, 3_{10}} \rightarrow TS$.

When NSHX was performed with wild-type ubiquitin, neither of these two intermediates was readily detectable (Sidhu and Robertson, 2004). Similarly, RD methods on wild-type protein would not have detected these intermediates because their populations are too low (<0.002%). However, RD measurements could be applied to the trapped L50E intermediate. Such measurements should detect the $N^{-\beta 5, 3_{10}}$ species as it now comprises 3% of the total population. This potential application illustrates the possibilities of combining the charge-burial
strategy with other methods to enable structural and thermodynamic characterization of folding intermediates.

Other protein engineering strategies have been used to trap intermediates. However, these methods often lack the degree of control possible with the charge-burial strategy, and they typically require advanced knowledge of the structural content of the intermediate. For example, Bai and coworkers used HX in conjunction with truncation and glycine substitutions to create analogs of intermediates of *Thermus thermophilus* RNase H (Zhou *et al*., 2008) and T4 lysozyme (Kato *et al*., 2007). They conducted NSHX on the T4 lysozyme analog but found no additional intermediate (Kato *et al*., 2007). Based on HX results on Escherichia coli RNase H, Marqusee and coworkers generated a stable mini-core (Chamberlain *et al*., 1999) and a fractionally populated intermediate using a single I25A substitution (Connell *et al*., 2009). Other late intermediates have been created using glycine substitutions in cytb562 (Feng *et al*., 2004) and IM9 (Whittaker *et al*., 2007).

The NMR characterization of the IM9 intermediate indicated that it has a fluid core with fluctuations on the NMR timescale due to the repacking of the remaining three helices. This repacking is similar to that of the remaining helices in the cyt-b562 intermediate (Feng *et al*., 2004). However, this intermediate and the *Thermus thermophilus* RNase H intermediate (Zhou *et al*., 2008) remain very native-like according to NMR line widths, as do the regions distal to the perturbation in the UbL50E intermediate. In contrast, the E. coli I25A intermediate has more dynamic behavior (Connell *et al*., 2009). We suggest that an I25D charged substitution would fully populate the *E. coli* intermediate and help address the apparent contradiction between the RNase H studies of the groups of Bai and Marqusee. This example further highlights the utility of the charge-burial strategy.
2.6 Conclusions

The buried aliphatic-to-charged substitution strategy is a general method for trapping and characterizing folding intermediates under equilibrium conditions. In the case of ubiquitin, the intermediate's stability relative to a native-like species is controllable, with acidification generating a well-dispersed NMR spectrum under conditions where HX is conveniently quenched. As a result, NSHX measurements can be performed on intermediates with poor NMR spectra. The strategy is particularly powerful in combination with other methods such as NSHX and RD. Although we applied the strategy with foreknowledge of ubiquitin's folding behavior, we envision that even without this foreknowledge, a series of charged substitutions throughout a protein can help identify foldons and their relationship to folding pathways in other proteins.
3 Early Ubiquitin Folding Intermediates and Pathway

3.1 Abstract

In this chapter of the thesis, I detail my search to find experimental evidence to address two major questions concerning early steps in protein folding pathways: 1. Are there native-like, mutually stabilizing tertiary interactions between foldons? 2. Do native-like hydrogen bond patterns exist in early folding intermediates? Studies suggest that there are stabilizing native-like tertiary interactions between $\beta_1\beta_2$ hairpin and $\alpha$-helix, which are believed to be the first and second foldon to form in the proposed ubiquitin early folding pathway (Krantz et al., 2004). Previous NSHX studies (Sidhu and Robertson, 2004; Zheng and Sosnick, 2010) suggest potential existence of a stable early intermediate with native-like hydrogen bond pattern. The results suggest performing NSHX on native state ubiquitin to characterize hydrogen bonds of stabilized early intermediates.

3.2 Introduction

As discussed in section 1.3.4, one major controversy in protein folding is whether the early, pre-transition state events follow a well-defined pathway, potentially with some minor parallel routes, or are there a multitude of distinct pathways in a funneled energy landscape. The funneled landscape view is primarily supported by theorists based on computational results (often with highly simplified models) while the pathway view is favored by experimentalists
(who struggle to characterize early folding events). If a series of sequentially stabilized early folding intermediates can be characterized, the pathway view will be strongly supported. However the two-state folding behavior of small single-domain proteins mandates that any pre-transition state intermediate cannot be more stable than the unfolded state (or else folding would be multi-state) (Sosnick et al., 1996). As a result, any potential intermediate is minimally populated and likely to be observed even by single molecule or ultra-fast methods. Currently, all the available methods, including our new protein vivisection strategy described in the previous chapter, can only be used to trap or detect folding intermediates that are more stable than the unfolded state. Our inability to characterize early folding intermediate leaves a huge gap in our knowledge of protein folding behavior, and leaves the pathway versus funnel controversy unresolved.

For ubiquitin folding reaction, transition state and post-transition state pathway has been defined by $\psi$-analysis (Krantz et al., 2004) and protein vivisection (Zheng and Sosnick, 2010), respectively. Little is known about how folding reaction reaches transition state from the unfolded state. A major pathway was proposed with the sequential formation of foldons, in a similar as late folding events (Figure 3.1) (Krantz et al., 2004). The $\beta_1\beta_2$ hairpin is thought to be the first structure element to form because the isolated hairpin populates the native geometry as in the full-length protein (<20%) under non-denaturing conditions (Zerella et al., 1999). Such native geometry can be stabilized by a single mutation (T9D) that introduces a salt-bridge across the hairpin (D9 and K11) because NMR structural characterization of the mutant peptide indicates a hairpin with native-like register (Zerella et al., 2000). The second piece of structure to form after the $\beta_1\beta_2$ hairpin probably is the $\alpha$-helix region which has very low intrinsic helicity.
Figure 3.1 Proposed ubiquitin early folding pathway. Foldon formation events connected by filled arrows are considered the major folding pathway as it conforms to the idea of commensurate formation of hydrogen bonds and hydrophobic burial. Existence of minor parallel pathways indicated by dashed arrows remains a theoretical possibility.
(<3%) (Munoz and Serrano, 1997). The hairpin and helix form tertiary contact through hydrophobic interactions to stabilize the otherwise unstable helix, which is consistent with the kinetic isotope data indicating that hydrophobic burial is commensurate with hydrogen bond formation in the transition state (Krantz et al., 2000b; 2002b). The hairpin-helix motif is subsequently joined by β3 strand and then the more distal β4 strand before reaching the transition state.

A possible parallel route could be that β3 strand joins β1β2 hairpin before the helix formation. However the formation of this three-stranded structure would require that the amino and carboxyl termini form a parallel β-sheet and close a large, ~35 residue loop. Furthermore, the sheets would bury less hydrophobic surface than the helix-hairpin nucleus. Hence, this route is less probable than the major route, whereby the helix associates with β1β2 hairpin before joining of β3 strand.

In order to test the proposed pathway, two major questions to be answered experimentally are: 1. Do foldons stabilize themselves by making native-like tertiary contact, as is suggested by kinetic isotope effect results? 2. Are there early intermediates with a native hydrogen bonding pattern, as is predicted by the step-wise foldon addition mechanism? A recent study on metal binding of engineered bi-Histidine sites of ubiquitin suggested the presence of high-energy early folding intermediates and concluded that the highest free energy level of these intermediates must be within 4 kcal mol\(^{-1}\) relative to the unfolded state (Bosco et al., 2009). Because of this relatively low upper bound and the various methods available for structure stabilization (e.g., T9D mutation, bi-Histidine site, osmolytes including trimethylamine N-oxide
(TMAO) and sarcosine), chances seem good that we can engineer at least some of the early intermediates to be more stable than unfolded state. In this chapter of the thesis, I present my attempts to stabilize and study ubiquitin early folding intermediates in an effort to answer the questions of concern. Although the several methods and ideas that I have tried have failed to generate definitive conclusions, I was able to gather information favoring the proposed ubiquitin early folding pathway and mechanism. The results also suggest other directions to pursue for more decisive evidence.

3.3 Results

3.3.1 Engineering ubiquitin fragment 1 to 36 to mimic the hairpin-helix motif

The ubiquitin fragment containing residues 1 to 36, Ub(1-36), contains structure elements β1β2 hairpin, Ub(1-17) and α-helix (residue 23-33) in the native structure. To stabilize the predicted β1β2 hairpin-α-helix intermediate, I incorporated the T9D mutation that has been reported to stabilize β1β2 hairpin (Zerella et al., 2000). A bi-Histidine pair (A28H, D32H) was engineered on α-helix to stabilize the helix in the presence of divalent metal ion (Krantz et al., 2004). The engineered fragment, Ub*(1-36), was subjected to far-UV CD characterizations in the absence and presence of Ni^{2+} ion to observe the effects of α-helix stabilization (Figure 3.2a). The shape of the difference CD spectrum (CD spectrum of Ub*(1-36) in the presence of Ni^{2+} - in the absence of Ni^{2+}) indicates mixture of signals from both α-helix
Figure 3.2 CD characterization of $\alpha$-helix formation driven by Ni$^{2+}$ ion in Ub*(1-36). a. The difference CD spectrum (green triangle), generated by subtracting CD spectrum of Ub*(1-36) in the absence (black rectangle) of Ni$^{2+}$ ion from that in the presence (red circle) of Ni$^{2+}$ ion, indicates that Ni$^{2+}$ ion primarily drives $\alpha$-helix formation based on similarity to known CD spectra of $\alpha$ helix, $\beta$ sheet and random coil (inset adapted from http://www.ap-lab.com/circular_dichroism.htm). b. In an $\alpha$-helix unfolding $\leftrightarrow$ folding equilibrium, plotting of folded helix population increase (%) by 0.9 kcal mol$^{-1}$ helix stabilizing energy against initial $\Delta G$ shows that an initial $\Delta G$ of $\sim$ 0.4 kcal mol$^{-1}$ gives maximum helix population increase ($\sim$ 40%).
and β sheet, although α-helix signal is probably the major contributor since the lowest ellipticity point is at ~ 222 nm wavelength. This fact, together with the roughly 3:1 ratio of absolute ellipticity at 222 nm ([θ]_{222nm}) from α-helix and β sheet, supports the assumption that the ellipticity of the difference CD spectrum at 222 nm primarily came from α-helix formation driven by the bi-Histidine site and Ni^{2+} ion. Calculations based on [θ]_{222nm} of pure α-helix and the difference spectrum normalized against the length of α-helix in Ub*(1-36) fragment suggest that the 0.9 kcal mol^{-1} stabilization energy provided by Ni^{2+}:bi-Histidine binding (Krantz et al., 2004) drove ~40% α-helix formation of Ub*(1-36). In a hypothetical 2 state equilibrium of helix formation, the increase in helix formation driven by the Ni^{2+} ion binding is determined by the intrinsic helix stability, \( \Delta P\% = \frac{1}{\exp((\Delta G-0.9)/0.55)+1} - \frac{1}{\exp(\Delta G/0.55)+1} \)*100 (Equation 3.1), where \( \Delta P\% \) is the percentage increase of helix formation and \( \Delta G \) is the intrinsic helix stability. A plot of \( \Delta P\% \) against \( \Delta G \) (Figure 3.2b) indicates that the ~40% increase in helix formation corresponds to an intrinsic helix stability (without Ni^{2+} binding stabilization) of ~ 0.4 kcal mol^{-1} (favoring helix unfolded) in the context of Ub*(1-36) where β1β2 hairpin is present. However the < 3% intrinsic helicity of the isolated ubiquitin α-helix alone (Munoz and Serrano, 1997) translates into an unfavorable free energy difference of α-helix formation greater than 1.9 kcal mol^{-1}. The > 1.5 kcal mol^{-1} discrepancy probably reflects the free energy gain from tertiary contacts between β1β2 hairpin and α-helix. Ni^{2+} ion stabilization results in ~71% of α-helix formed in the hairpin-helix context.

\(^{15}\)N labeled Ub*(1-36) was subjected to NMR characterization in the absence and presence of Ni^{2+} ion (Figure 3.3a). Ni^{2+} ion led to universal peak broadening of the collapsed
Figure 3.3 NMR characterization of the hairpin-helix motif driven by Ni$^{2+}$ ion. a. Spectrum on the left is the overlay of HSQC spectra of Ub*(1-36) in the absence (black) and presence (red) of Ni$^{2+}$ ion. A slice of 1D HN spectrum at the dashed line shows the peak broadening upon addition of Ni$^{2+}$ ion. By comparing to the HSQC spectrum of full length ubiquitin (right spectrum), the resonance signal of Ub*(1-36) that appears in the presence of Ni$^{2+}$ ion (green dashed circle) likely represents residue S20. b. Conformation of Ub*(1-36) in the context of full-length, folded ubiquitin structure. Residue S20 located in the loop connecting hairpin and helix is highlighted. Backbone angle of S20 is likely to be the primary determinant of its HSQC chemical shift. Native chemical shift of S20 suggests formation of native-like orientation of hairpin-helix motif driven by Ni$^{2+}$ ion.
HSQC spectrum of Ub*(1-36). The most likely explanation for this broadening is the micro-heterogeneity of the tertiary contacts between β1β2 hairpin and α-helix that result in multiple micro-states. Each of the micro-state has a slightly different orientation between hairpin and helix, therefore chemical shift of each micro-state is slightly different. Intermediate NMR time-scale exchange among those micro-states causes the peak broadening. The most interesting change of the Ub*(1-36) HSQC spectrum after addition of Ni^{2+} ion is the appearance of a resonance signal at the top-right corner of Figure 3.3a. Compared to the HSQC spectrum of full length ubiquitin (with T9D and bi-Histidine site), the resonance signal likely represents residue S20 located in the loop connecting hairpin and helix because of its almost identical chemical shift in the sparsely populated region of the spectra. The observation that Ni^{2+} ion-driven hairpin-helix motif formation gives rise to native-like S20 resonance signal suggests that both backbone angle of residue S20 and the overall orientation of hairpin relative to helix are native-like. Consequently the tertiary interaction between hairpin and helix is likely native-like as well.

3.3.2 Peptide complementation for characterizing hydrogen bonds in the hairpin-helix motif

To test whether the hairpin-helix motif, which is stabilized by Ni^{2+} binding, is in a natively hydrogen-bonded conformation, I used hydrogen exchange method. However as shown in figure 3.3, the HSQC peaks of Ub*(1-36) are too collapsed and/or broadened to conduct direct hydrogen exchange measurement. In order to circumvent this problem, a peptide complementation strategy was attempted. By mixing ^15N-labeled Ub*(1-36) with unlabeled ubiquitin fragment Ub(37-76), the two fragments bind and fold simultaneously to form the native conformation of ubiquitin (Figure 3.4a). The resulting HSQC spectrum contains well dispersed, sharp peaks for residues in Ub*(1-36) in the complex. There still exists a small portion of Ub*(1-
Figure 3.4 Coupled binding and folding of Ub*(1-36) with Ub(37-76). a. HSQC spectrum of $^{15}$N-labeled Ub*(1-36) mixed with unlabeled Ub(37-76). The degree of HSQC peak dispersion suggests that the two fragments bind and fold into a native structure of ubiquitin. More importantly, the HSQC spectrum is suitable for hydrogen exchange measurements. b. Binding on-rate of Ub*(1-36) and Ub(37-76) fragment measured to be $5.5 \times 10^4$ sec$^{-1}$ M$^{-1}$ by stopped-flow apparatus (as described in Zheng and Sosnick, 2010).
36) in the free form as mandated by the dissociation constant. Measurements of hydrogen exchange rate for the resonances of Ub*(1-36) in the complex should provide a readout of the hydrogen exchange rate of free Ub*(1-36) according to equation $k_{\text{obs}} = k_{\text{free}} \times [\text{Ub}^*(1-36)]\%$ (Equation 3.2), where $k_{\text{obs}}$ is the measured exchange rate of Ub*(1-36) in the complex, $k_{\text{free}}$ is the exchange rate of free Ub*(1-36) and $[\text{Ub}^*(1-36)]\%$ is the fraction of free Ub*(1-36) that can be calculated from dissociation constant and concentration of Ub(37-76).

However, one important condition must be satisfied for equation 3.2 to be valid: the interconversion rate of Ub*(1-36) between free form and complimented form must be much faster than intrinsic exchange rate so that hydrogen exchange falls into the so-called EX2 regime. In this regime, hydrogen exchange reports on free energies of a fully equilibrated system, rather than a subset of states or their interconversion rates. Upon the coupled binding and folding of the fragments, an exposed tryptophan residue (W45) on Ub(37-76) becomes buried and induces a fluorescence signal change, analogous to that observed in the folding of the intact protein. This signal allows for direct measurement of the binding on-rate using a Biologic SFM4 stopped flow apparatus. To measure the binding rate, 10x Ub*(1-36) was mixed with 1x Ub(37-76) and the time course of fluorescence signal change is recorded and fit to obtain a binding on-rate of $5.5 \times 10^4 \text{ sec}^{-1} \text{ M}^{-1}$ (Figure 3.4b). This on-rate is such that under the hydrogen exchange conditions, it is much slower than the intrinsic hydrogen exchange rate ($2.5 \text{ sec}^{-1}$ versus $\sim40$-$50 \text{ sec}^{-1}$). Hence the necessary conditions for an EX2 measurement are not satisfied. As a result, my complementation strategy is unsuitable for measuring the hydrogen bond stability in free Ub*(1-36).
3.3.3 A stable β1β2 hairpin-α-helix-β3 intermediate suggested by NSHX

A close scrutiny of two hydrogen bond groupings (group A: β1β2 hairpin + N-terminal of α-helix + F4-S65 and K6-L67 between β2 and β3 strand; group B: between β3 and β4 strand + C-terminal of α-helix + L69-K6 and L67-F4 between β2 and β3 strand) from the NSHX results on UbL50E I state suggests an interesting possibility. The hydrogen bonds from both groups have large ΔG_{HX} and denaturant m-values suggesting hydrogen exchange occurs via global unfolding. However, the ΔG_{HX} and m-values of group B hydrogen bonds are slightly smaller than those of group A hydrogen bonds. NSHX results on wild type ubiquitin display the same difference between the exact two groups of hydrogen bonds (Sidhu and Robertson, 2004). The size of error bars suggests that this difference is statistically significant.

The intuitive interpretation of the NSHX results regarding the slight difference between group A and group B can lead to a significant hypothesis: an early, marginally-stable folding intermediate exists with the β1β2 hairpin, N-terminal of α-helix and β3 strand formed. This intermediate is more stable than the unfolded state, although only by ~ 0.5 kcal mol\(^{-1}\) (Figure 3.5a). An alternative explanation is that group B hydrogen bonds fray in the ground state of the protein under the conditions where NSHX is conducted (either wild type ubiquitin or UbL50E I state at pH 8) (Figure 3.5b). In the second scenario, the free energy level of the locally frayed state fortuitously happens to be very close to that of the globally unfolded state so that the very short, flat (horizontal) local fraying phase of the NSHX isotherm mingles with the globally unfolded phase (Figure 3.5c). In practice, the starting portion of the isotherm may be interpreted as a phase with slightly shallower slope (m-value) and lowered intercept with the
Figure 3.5 Possible explanations for NSHX behavior of ubiquitin group B hydrogen bonds. Two scenarios are most plausible: a. Presence of an early folding intermediate (I) composed of β1β2 hairpin-α-helix-β3 through which hydrogen exchange can occur for group B backbone amides. b. Group B backbone amides can exchange through local (small-scale) opening (L). The free energy and m-values for the various states (U, N, I, L) in the depicted energy landscape are arbitrarily chosen but very closely resemble those for wild type ubiquitin. c. Predicted NSHX isotherm for group B backbone amide based for each scenario. Fitting either isotherm results in slightly lower ΔGHX and m-values than those of global exchange (blue dashed line).
vertical axis ($\Delta \text{G}_{\text{HX}}$), thus leading to a false conclusion of an early intermediate.

In the following sections, I describe my attempts to identify the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ early intermediate.

### 3.3.3.1 CPMG relaxation dispersion NMR spectroscopy

Because the I state of UbL50E trapped by protein vivisection strategy is 2.3 kcal mol$^{-1}$ more stable than the unfolded state (U comprises 0.5% of population), a potential folding intermediate with similar free energy to unfolded state may be detected by CPMG relaxation dispersion performed on UbL50E I state. However, no decrease of $R_2$ was observed with increase of refocusing frequency $1/\tau_\text{cp}$ (Figure 3.6). This result suggests that even if the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ early intermediate does exist, it samples a variety of conformational microstates, similar to that observed for the hairpin-helix motif (section 3.3.1). The interconversion between micro-states results in loss of magnetization coherence that CPMG refocusing pulses are not able to bring back. Therefore CPMG relaxation dispersion may not be a suitable tool for studying early folding intermediates which sample microstates on the submillisecond timescale.

### 3.3.3.2 Trans- and cis-reconstruction of the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ early intermediate

Trans-reconstruction was attempted by mixing Ub*(1-36) with 10 fold excess of the peptide fragment corresponding to $\beta_3$ strand. If the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ early intermediate forms, there will be a random coil to $\beta$ sheet transition for the $\beta_3$ strand fragment. However, CD measurements showed no sign of complementation and $\beta$ sheet formation (Figure 3.7a). In retrospect this result is not surprising because in the real folding reaction, the hairpin-helix motif
Figure 3.6 CPMG relaxation dispersion measurements on Ubl50E I state. Experiments were run at pH 7.8, 4 °C. $R_2$ values of two representative residues (V17 on β1β2 hairpin and V26 on α-helix) remain constant in response to increase of refocusing pulse frequency $1/\tau_{cp}$. 
Figure 3.7 Trans- and cis-reconstruction of the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ intermediate. a. CD signal of Ub*(1-36) and 10 fold $\beta_3$ strand fragment was recorded separately and added mathematically to constitute the background spectrum (black rectangle). The two fragments were mixed and CD signal of the mixture was recorded (red rectangle). Because the signal almost overlaps with background, there is no obvious coil to $\beta$ sheet transition for the $\beta_3$ strand fragment and trans-reconstruction is unsuccessful. b. Both HSQC spectra of the Ub*(1-36+GS26+$\beta_3$) construct in the absence of presence of Ni$^{2+}$ ion show peak broadening, which may arise from the desired native tertiary interactions or from less interesting origins such as non-specific aggregation.
is linked with β3 strand by a stretch of unfolded polypeptide chain. Therefore the effective concentration of β3 strand around the hairpin-helix motif is much higher than that used in the trans-reconstruction. To better mimic the folding situation and enhance the chances of reconstruction, a flexible, 26 amino acids-long glycine-serine linker was added to connect Ub*(1-36) with β3 strand fragment. HSQC spectrum of this construct (Ub*(1-36+GS26+β3)) showed similar peak broadening as the HSQC spectrum of Ub*(1-36) with Ni²⁺ ion (Figure 3.7b). Addition of Ni²⁺ ion to Ub*(1-36+GS26+β3) resulted in even more extensive peak broadening. By comparison with the more random-coil like HSQC behavior of Ub*(1-36) (collapsed sharp peaks), the results suggest that Ub*(1-36+GS26+β3) in the absence of Ni²⁺ ion exhibits signs of tertiary interactions. It is not clear, however, at this point whether such peak broadening truly reflects the native-like tertiary contact of β1β2 hairpin-α-helix-β3 intermediate, or it is simply caused by non-specific aggregation or other sources.

### 3.4 Materials and Methods

#### 3.4.1 Proteins and peptides

Ub*(1-36) peptide (¹⁵N labeled or unlabeled) was expressed in *E. coli* BL21(DE3) strain and purified with the same protocol developed to purify wild type ubiquitin according to Krantz *et al.* (Krantz *et al.*, 2004). Ubiquitin β3 strand fragment and fragment 37_76 was synthesized using tBoc chemistry. Ub*(1-36+GS26+β3) construct was expressed and purified from inclusion body as described by Zheng and Sosnick (Zheng and Sosnick, 2010).
3.4.2 Equilibrium and kinetic measurements

CD experiments were performed with a Jasco 715 spectropolarimeter with a pathlength of 1 cm in 15 mM sodium phosphate, 225 sodium chloride. Rapid mixing fluorescence experiments used a Biologic SFM-400 stopped-flow apparatus connected via a fiber optic cable to a PTI A101 arc lamp. Fluorescence spectroscopy used excitation and emission wavelengths of 280-290 nm and 300-400 nm, respectively.

3.4.3 NMR spectroscopy

All the NMR experiments were run on a Varian Unity Inova 600 MHz spectrometer equipped with cryo-probe. If not otherwise specified, all the samples were prepared with a buffer containing 15 mM HEPES, 225 mM NaCl, pH 7.8. CPMG relaxation dispersion pulse sequence was a kind gift from Francesca Massi at the University of Massachusetts. The pulse sequence was initially developed by Patrick Loria at Columbia University.

3.5 Summary

As mentioned in the introduction of this chapter, two major questions to be answered in investigating early protein folding intermediates and pathways are: 1. Are there native-like, mutually stabilizing tertiary interactions between foldons? 2. Do native-like hydrogen bond patterns exist in early folding intermediates? My studies on Ni$^{2+}$ ion-driven hairpin-helix motif formation suggest a positive answer to the first question. I demonstrated that the greater than 1.5 kcal mol$^{-1}$ stabilizing energy for the hairpin-helix motif (0.4 kcal mol$^{-1}$ and >1.9 kcal mol$^{-1}$ helix
stability in the hairpin-helix and isolated helix context, respectively) probably comes from native-like tertiary interactions (NMR peak broadening and likely appearance of the native S20 HSQC resonance). Native-state hydrogen exchange results on native ubiquitin as well as N$^{\beta5}$ intermediate point to the presence of a marginally-stable, natively hydrogen-bonded $\beta1\beta2$ hairpin-$\alpha$-helix-$\beta3$ early intermediate. The information so far available allows preliminary reconstruction of the energy landscape for the major early folding pathway of ubiquitin (Figure 3.8). However, dynamic nature of the tertiary interactions of early intermediates lead to poor quality of NMR spectra (extensive peak broadening) for both hairpin-helix motif (Ub*(1-36) with Ni$^{2+}$ ion) and reconstructed $\beta1\beta2$ hairpin-$\alpha$-helix-$\beta3$ intermediate (Ub*(1-36+GS26+\beta3), even if they are more stable than the unfolded state. Therefore direct characterization of early intermediate hydrogen bonds by NMR is extremely difficult. At this point, with all the successful and failed efforts associated with studies on early intermediate, one method stands out as the most promising one to enable complete delineation of early folding pathway – early intermediate stabilization followed by detection using NSHX. Details about experimental plan will be discussed in the next chapter.
Figure 3.8 Reconstruction of energy landscape for major ubiquitin early folding pathway. The greater than 0.7 kcal mol\(^{-1}\) free energy difference between unfolded state and β1β2 hairpin is from the sub 20% population of hairpin that the isolated peptide contains in non-denaturing conditions (Zerella et al., 1999). Free energy difference between hairpin and hairpin-helix is from section 3.3.1. Potential presence of a stable β1β2 hairpin-α-helix-β3 intermediate is from NSHX results on both wild type ubiquitin and UbL50E I.
4 Conclusions and Future Directions

4.1 Conclusions

A folding pathway containing the step-wise formation of foldons stabilized by tertiary interactions (e.g., hydrophobic burial), and the presence of natively hydrogen bonded folding intermediates are two major affirmations of my view of protein folding mechanism. In Chapter 2, successful trapping of ubiquitin folding intermediate $N^{\beta 5}$, which had eluded other detection methods including NSHX, by our charge-burial strategy provides direct evidence for coupled secondary and tertiary structure formation. Characterization of the $N^{\beta 5}$ intermediate revealed that the folded part contained an extensive hydrogen bond pattern. The unfolded part made transient interactions with other parts of the protein instead of being fully unfolded. The vivisection strategy can be even more powerful if it is combined with other methods such as NSHX or CPMG RD NMR spectroscopy. A second ubiquitin folding intermediate $N^{\beta 5-3\alpha}$ was uncovered by performing NSHX on $N^{\beta 5}$ intermediate, which led to total delineation of ubiquitin post-transition state folding pathway. This study was the first application of NSHX to a trapped intermediate which led to the identification of another intermediate. This general strategy may be combined with other methods and have broad applications in the study of protein folding and other reactions that require trapping of high energy states.

In Chapter 3, I tested whether the two major affirmations mentioned above about protein folding mechanism also govern the early, pre-transition state folding events. Studies on
ubiquitin β1β2 hairpin-α-helix motif suggest that there are native tertiary interactions between
hairpin and helix that provide stabilizing free energy for secondary structures. However the
dynamic nature of early folding intermediates, even if they are stable, makes NMR
characterizations extremely challenging due to extreme peak broadening. As discussed in the
next section, I believe that the application of NSHX to infer hydrogen bond pattern of stabilized
early intermediates appears to be the most likely method to succeed.

4.2 Future Directions

Results from Chapter 3 suggest that introduction of mutation T9D and bi-Hitidine site
(A28H, D32H) in the presence of Ni²⁺ ion into wild type ubiquitin reshape the folding energy
landscape of early folding pathway (Figure 4.1). β1β2 hairpin and hairpin-helix intermediate
should be rendered more stable than unfolded state so that I can apply NSHX to detect these
intermediate and measure the surface area exposure of each intermediate.

As discussed in section 3.3.3, the NSHX results indicate that a group of hydrogen
bonds (group B) has slightly lower ΔG_{HX} and m-values than those for hydrogen bonds that
exchange through global unfolding. This can be due to presence of either an early folding
intermediate (β1β2 hairpin-α-helix-β3) or fraying of local structure. In order to distinguish
between the two scenarios, I plan to introduce T9D mutation to wild type ubiquitin and perform
NSHX on the T9D mutant protein in the presence of TMAO that stabilizes hydrogen bonds. The
combination of T9D mutation and TMAO is estimated to provide ~1 kcal mol⁻¹ stabilization to
Figure 4.1 Reshaped folding energy landscape of ubiquitin early folding pathway by the T9D mutation and Ni\textsuperscript{2+}:bi-Histidine binding. T9D mutation will make $\beta_1\beta_2$ hairpin more stable than U state by an undetermined value (Zerella et al., 2000). Results from section 3.3.1 suggest that Ni\textsuperscript{2+} binding to the bi-Histidine site (A28H, D32H) will favor $\alpha$-helix formation by $\sim$0.5 kcal mol\textsuperscript{-1} in the presence of $\beta_1\beta_2$ hairpin. Thus, both $\beta_1\beta_2$ hairpin and hairpin-helix intermediates should become more stable than the U state and can be detected by performing NSHX. It is unclear if the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ and $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$-$\beta_4$ intermediate can also be detected.
either the $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ intermediate or the locally frayed conformation. The change of NSHX isotherm of group B hydrogen bonds as a result of the 1 kcal mol$^{-1}$ stabilization will be different for the early intermediate scenario and the local fraying scenario (Figure 4.2). In the former situation, the more linear-like isotherm (Figure 3.5) will remain linear after the 1 kcal mol$^{-1}$ stabilization. In the latter situation, the previously omitted small curvature representing the transition from flat local fraying phase to global exchange phase will be enhanced as the flat local fraying phase gets extended out by the 1 kcal mol$^{-1}$ stabilization. Actually both the scenarios may exist for group B hydrogen bonds and they can be easily distinguished to identify the potential $\beta_1\beta_2$ hairpin-$\alpha$-helix-$\beta_3$ intermediate.
Figure 4.2. Distinguishing possible causes for NSHX results of ubiquitin group B hydrogen bonds. Starting energy landscapes remain the same as those in Figure 3.5. The 1 kcal mol\(^{-1}\) stabilizing energy by T9D mutation and TMAO will reshape the energy landscapes (red dashed well) and lead to different changes of NSHX results for group B hydrogen bonds in the two scenarios. If the \(\beta_1\beta_2\) hairpin-\(\alpha\)-helix-\(\beta_3\) intermediate exists (top), the linear-like isotherm will remain linear-like after stabilization. If local structure fraying exists (bottom), the flat local fraying phase will be extended and the curvature of local-to-global transition should become more significant and easily distinguished. Blue dashed line indicates hydrogen exchange through global unfolding.
References Cited


