Electronic Thesis and Dissertation Repository

4-17-2012 12:00 AM

Structural Characterization of Protein Folding Intermediates by Oxidative Labeling and Mass Spectrometry

Bradley B. Stocks, *The University of Western Ontario*

Supervisor: Lars Konermann, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry
© Bradley B. Stocks 2012

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biochemistry, Biophysics, and Structural Biology Commons

STRUCTURAL CHARACTERIZATION OF PROTEIN FOLDING INTERMEDIATES BY OXIDATIVE LABELING AND MASS SPECTROMETRY

(Spine title: Protein Folding Studied by Oxidative Labeling and ESI-MS)

(Thesis format: Integrated-Article)

by

Bradley B. <u>Stocks</u>

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Biochemistry
The University of Western Ontario
London, Ontario, Canada

© Bradley B. Stocks 2012

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor   Examiners

_____   _____
Dr. Lars Konermann   Dr. James Choy

                                                              _____
Supervisory Committee   Dr. Gilles Lajoie

_____
                                                              _____
                                                              Dr. Mark Bernards

_____   _____
                                                              Dr. Gerold Schmitt-Ulms

The thesis by

Bradley B. <u>Stocks</u>

entitled:

Structural Characterization of Protein Folding Intermediates by Oxidative Labeling and Mass Spectrometry

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date_____   _____
                                                              Chair of the Thesis Examination Board

Abstract

A key challenge associated with protein folding studies is the characterization of short-lived intermediates that become populated en route to the native state. In this work, a covalent labeling method was developed that provides insights into the structures of these transient species. Hydroxyl radical (·OH) reacts with solvent-exposed side chains, whereas buried residues are protected. Mass spectrometry is used for monitoring the locations and the extent of labeling. Pulsed ·OH labeling of proteins at selected time points during folding results in high temporal and spatial resolution when compared to existing other labeling methods.

This novel technique was validated by studying the kinetic unfolding and refolding of holomyoglobin (hMb) and cytochrome *c* (cyt *c*), respectively. The noncovalent prosthetic heme group in hMb was shown to drastically affect the unfolding pathway. Cyt *c* refolding was found to fold in a stepwise manner. The population of a misfolded cyt *c* intermediate was also detected. Results in both cases were in accord with published data.

Many cellular proteins exist as oligomers. Pulsed ·OH labeling method was therefore extended to monitor the folding and assembly of a 22 kDa homodimeric protein, S100A11. Prior to this study very little information regarding the folding mechanism of this protein was available. ·OH labeling reveals that disruption of the native dimer is followed by the formation of non-native hydrophobic contacts within the denatured monomers. The folding/binding pathway was shown to progress through monomeric and dimeric intermediates.

In the final section of this study we applied ·OH labeling to a large monomeric protein that folds to a metastable state. The folding pathway of the 44 kDa protease inhibitor, α_1-antitrypsin, was characterized and compared with complementary data from hydrogen/deuterium exchange studies. Our results show that the formation of early tertiary contacts and specific hydrogen bonds guide the protein towards its active, metastable structure. Structural correlation is also seen between a late kinetic species and a previously characterized equilibrium intermediate of a pathogenic mutant.

Overall, the results presented highlight the ability of the technique developed in this work to provide in-depth information about the mechanisms of protein folding.

Statement of Co-Authorship

The work in Chapters 2, 3 and 4 were published in the following articles respectively:

Bradley B. Stocks and Lars Konermann (2009). Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry. *Anal. Chem.* 81: 20-27. Reproduced with permission. © 2009 American Chemical Society

Bradley B. Stocks and Lars Konermann (2010). Time-Dependent Changes in Side Chain Solvent Accessibility during Cytochrome *c* Folding Probed By Pulsed Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* 398: 362-373. Reproduced with permission. © 2010 Elsevier Ltd.

Bradley B. Stocks, Atoosa Rezvanpour, Gary S. Shaw, and Lars Konermann (2011). Temporal Development of Protein Structure during S100A11 Folding and Dimerization Probed by Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* in press. Reproduced with permission. © 2011 Elsevier Ltd.

The work in Chapter 5 has been incorporated into the following article:

Bradley B. Stocks, Patrick L. Wintrode, and Lars Konermann (2012). Folding Mechanism of α_1-Antitrypsin to its Metastable State: Insights from Pulsed Oxidative Labeling and Mass Spectrometry. *In preparation.*

The original draft for each of the above articles was prepared by the author. Subsequent revisions were by the author and Dr. Lars Konermann. Protein purification for the works in Chapters 4 and 5 was completed by the labs of Dr. Gary Shaw and Dr. Patrick Wintrode, respectively. HDX measurements used to generate Figure 5.6 were collected by the lab of Dr. Patrick Wintrode. All other experimental work and data analysis was performed by the author under the supervision of Dr. Konermann.

Not knowing where you're going is the best way to get someplace you've never been

-   *J. Peterman 'Seinfeld'*

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisor, Dr. Lars Konermann. He has somehow taken my interest in science and transformed it into a passion for research. And he has done so all the while making it seem as though it were my own idea. He has provided me with outstanding leadership while encouraging me to think freely. Returning my lab keys will not be an easy task.

Thank you to my departmental committee members Drs. James Choy and Stan Dunn. They have provided much useful feedback regarding my research over the past 5 years. Their genuine interest in my work and progress has been greatly appreciated.

I would not have been successful without the help of a cast of truly outstanding labmates. Brian, Mark, Peter, Yuhong, and Xin were there to help when I first joined the lab. The second wave included Elias, Jingxi, Jenna, and Lucy. Significant thanks goes to Siavash for the countless hours of conversations regarding oxidative labeling, ion mobility, and Persian culture. I leave the lab in good hands.

A final thanks to my crew from Biochemistry. Whether things were going well in the lab or nothing was working, I could always count on Lazar, Jake, Chris, Erin, Piya, Elnaz, and Joey to be up for pints at the Grad Club.

Table of Contents

List of Figures

Chapter 4

Chapter 5

List of Appendices

List of Symbols and Abbreviations

·OH – hydroxyl radical

A_{695} – absorbance at 695 nm

aMb – apomyoglobin

ANS – 1-anilino-8-naphthalene sulfonate

A_{tot} – total peak area

A_u – area of unmodified peak

CD – circular dichroism

CIDNP – chemically induced dynamic nuclear polarization

CRM – charged residue model

cyt c – cytochrome c

Da – Dalton

D_I – dimeric protein folding intermediate

D_N – native protein dimer

ESI – Electrospray Ionization

FRET - Förster resonance energy transfer

F_u – fraction unmodified

fwhm – full width, half-maximum

GdnHCl – guanidinium hydrochloride

H_2O_2 – hydrogen peroxide

HDX – hydrogen/deuterium exchange

hMb – holomyoglobin

IEM – ion evaporation model

KrF – krypton fluoride

MALDI – Matrix-Assisted Laser Desorption/Ionization

MD – molecular dynamics

M_F – folded protein monomer

M_I – monomeric protein folding intermediate

MS – Mass Spectrometry

MS/MS – tandem mass spectrometry

M_U – unfolded protein monomer

NMR – Nuclear Magnetic Resonance

NOL – normalized oxidation level

R – unoxidized peptide relative signal

Serpin – serine protease inhibitor

UPLC – ultra performance liquid chromatography

UV-Vis – ultraviolet-visible

Xaa – variable amino acid

$\alpha(t)$ – solvent accessibility as a function of folding time

$\alpha_1 AT$ – α_1-antitrypsin

α_{rel} – relative solvent accessibility

λ_{em} – emission wavelength

λ_{max} – wavelength of maximal emission

τ – duration of labeling pulse

Chapter 1 – Introduction

1.1 Protein Structure and Folding

While DNA encodes all of the information necessary for life, proteins are the workhorse molecules of the cell. From cytoskeletal support and energy conversion to transport and catalysis, the list of cellular tasks performed by proteins is seemingly interminable. It is widely accepted that structure and function are intimately connected. Protein functionality is achieved by utilizing the 20 naturally occurring amino acids which represent a relatively small set of structural building blocks. All amino acids (except for proline) have the same basic structure

$$H_2N \text{---} CH \text{---} COOH$$
$$|$$
$$R$$
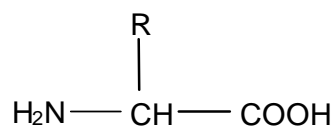
where R is the side chain. Proline has a secondary amino nitrogen due to bonding with its R-group and the resulting conformational strain is responsible for a great deal of interesting structural effects (1). While all proteins are polymers of amino acids, most must adopt a unique three dimensional structure in order to carry out their specific functions (2). It still remains unclear how this protein structure is encoded within a linear sequence (3). Four levels of protein structure can be distinguished.

Primary structure refers to the linear sequence of amino acids. This sequence is encoded within DNA and it is assembled on ribosomes (4). With 20 naturally occurring amino acids, an astronomical number of potential sequences exist. Only a tiny fraction of

these sequences are encoded. Great diversity of function and regulation is obtained through protein post-translational modifications (5). *Secondary structure* refers to the organization of local contacts mediated by hydrogen bonds along the peptide backbone. α-helices, β-sheets, and β-turns comprise the dominant secondary structural elements (4). The three-dimensional orientation of such elements within a single protein chain represents its *tertiary structure.* For many proteins the structural hierarchy stops there. However, a considerable number of proteins exist as dimers, trimers, or larger oligomers (6). Such association of distinct polypeptide chains, each with its own tertiary structure, is referred to as *quaternary structure.*

1.1.1 Native Protein Structure

The native structure of a protein is only marginally stable compared to the unfolded state. There is a large entropic penalty associated with the transition from a disordered chain to one of highly defined structure, as well as an enthalpic penalty for disrupting many interactions with the solvent. The protein must overcome these unfavourable contributions through intramolecular contacts. Some important examples are discussed below:

Hydrophobic interactions. Globular proteins exist in an aqueous environment however, many amino acid side chains are hydrophobic. The clustering of said residues in the protein interior results in a stabilizing contribution that is proportional to the buried side-chain surface area (7). This leads to the formation of a hydrophobic core and leaves the

hydrophilic residues mostly on the surface where they interact with the solvent. Recent experiments have demonstrated that for small-to-medium sized proteins, the hydrophobic effect contributes approximately 60% to the overall stability (8).

Hydrogen bonding. Hydrogen bonds require both a donor group and an acceptor group. Donors have a hydrogen atom bound to an electronegative heteroatom, eg. -NH, -OH, -SH, and acceptors have lone electron pairs, eg. C=O. The peptide backbone, as well as side chains, of an unfolded protein makes numerous hydrogen bonds with the solvent. Upon folding, loss of those contacts must be compensated for with intramolecular hydrogen bonds, most of which are found in α-helices and β-sheets. It has been estimated that the average hydrogen bond stabilizes a protein by ~2-5 kJ mol^{-1} (9, 10), however this bond energy is context sensitive (11).

Salt bridges. The positive charge on the side chains of Arg and Lys, as well as the negative charge on those of Asp and Glu, often result in their localization on the protein surface. Electrostatic pairing of these side chains is stabilizing. Occasionally such salt bridges can also occur within the protein core. Charge-charge interactions on the surface are weaker due to solvent screening. While perhaps not as important as other intramolecular forces, it has been shown that optimizing the charged interactions on the protein surface can increase stability over the wild-type molecule (12).

Cation-π interactions. While the positively-charged side chains of Arg and Lys can be involved in salt bridges, they are also often found interacting with the π-electrons of the aromatic amino acids Phe, Tyr, and Trp. A survey of the structures in the Protein Data Bank revealed ~25% of all Trp residues involved in a significant cation-π interaction (13). A bias for Arg-Tyr interactions has been shown to exist at protein-protein interfaces, with the stabilizing electrostatic energy estimated to be ~10 kJ mol^{-1} (14). A recent study found a specific cation-π interaction to be vital for the regulation of integrin affinity and function (15). Similar interactions occur between aromatic residues. The hydrogens on the aromatic rings are slightly positive. This allows the aromatic side chains to interact favourably in an edge-to-face orientation, with the hydrogens of one in close proximity to the π-electrons of another. Face-to-face side chain stacking, akin to base stacking in nucleic acids, is also a possibility albeit with a slight offset to avoid electrostatic clash between the electron clouds.

Disulfide bonds. The thiol side chains of cysteine residues can be oxidized to form a disulfide (S-S) bond. These bonds can be *intra*chain as in lysozyme, or *inter*chain as for insulin. Disulfides serve to stabilize the native state by reducing the conformational entropy of the unfolded state. Also, residue interactions required in the native state can be promoted by disulfide linkages (16).

It has been over 50 years since the first X-ray structure of myoglobin was determined (17). Currently the Protein Data Bank contains nearly 80000 protein structures, determined mainly by two methods. Approximately 7 out of every 8 protein

structures have been solved by X-ray crystallography. Solution NMR has been used for ~12% of structures. Crystallography is preferred for large proteins, and NMR is required for those with highly dynamic regions. Static X-ray structures (Figure 1.1A) can be helpful for visualizing binding interfaces, enzyme active sites, etc. but they are only informative regarding native proteins. Partially folded states are often too flexible for a detailed structural elucidation by X-ray methods. NMR can characterize these disordered proteins, and it is capable of probing dynamic aspects of native proteins (Figure 1.1B).

1.1.2 Protein Folding

Anfinsen's experiments on ribonuclease showed that a denatured protein will spontaneously refold to its correct native structure when placed in a suitable solvent environment. Hence, the sequence of amino acids contains all of the information needed to rearrange from a disordered polypeptide chain into a highly ordered native state (18). This observation begs the question: How does a protein "know" its correct secondary, tertiary and quaternary structure based only on primary structure? Much progress has been made towards answering this question in the decades since Anfinsen's discovery through experiment and simulation, although many details remain elusive.

In 1969 Cyrus Levinthal posited that it should take many times the age of the universe for a polypeptide chain to find its native conformation based on a random search (19). Proteins are continually synthesized in living organisms, and most fold within a few seconds. This mismatch of timescales became known as the "Levinthal Paradox", and it provided the basis for what is now known as the "protein folding problem".
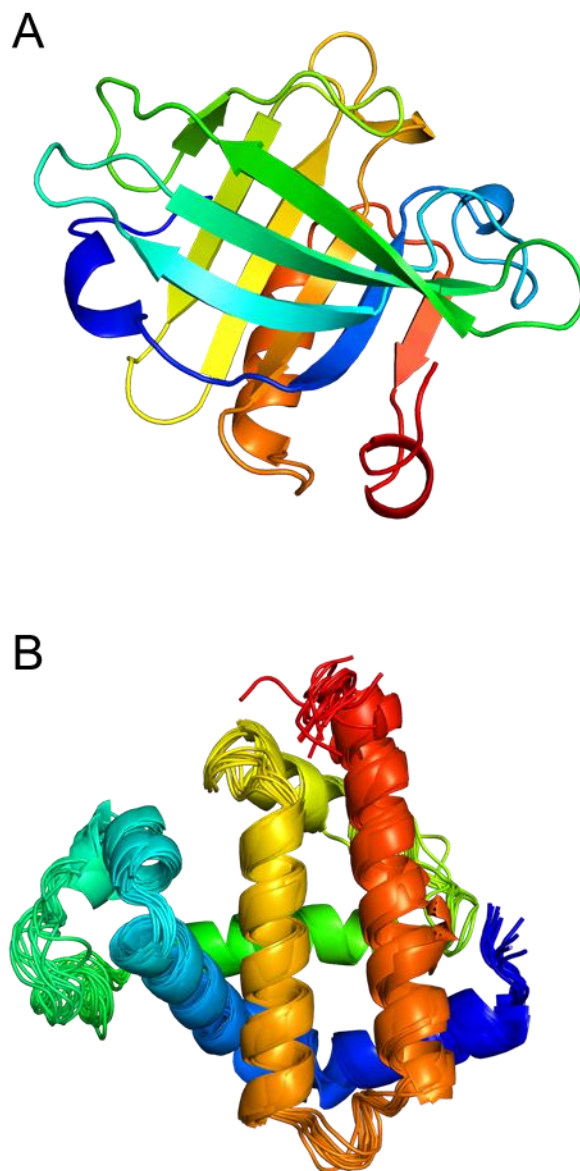
Figure 1.1. 3D protein structures. (A) X-ray crystal structure of the bovine β-lactoglobulin, a protein with a high β-sheet content (PDB ID: 1BEB). (B) Ensemble of NMR structures for sperm whale myoglobin (PDB ID: 1MYF).

Levinthal assumed that each partially folded conformer was approximately degenerate in energy with the unfolded state. The native conformation corresponds to a deep energy well which can only be attained by an extensive random search (Figure 1.2A). Levinthal's proposed solution involved the presence of defined pathways along which proteins can fold on a biologically relevant timescale (20). Many different mechanisms have been proposed to explain the properties of these putative pathways.

1.1.3 Protein Folding Mechanisms

The *framework* model suggests that local secondary structural elements can form independently (21). These structural elements would then diffuse and collide, ultimately coalescing into native tertiary structure. At the other end of the spectrum is the *hydrophobic collapse* model. This proposal asserts that folding is initiated with an extensive collapse of hydrophobic residues, followed by formation of secondary structure. The *nucleation* model took up residence somewhere in the middle and stated that partial native secondary structure would serve as a folding nucleus from which further structure would propagate (22).

All three of the proposed models implied the population of semi-folded structures (intermediates) along the pathway. Jackson and Fersht, however, reported that folding of chymotrypsin inhibitor 2 occurred in a two-state manner, with no detectable intermediates (23); a mechanism termed *nucleation-condensation*. Subsequent ϕ-value analysis, which probes the effects of amino acid substitutions on transition state stability,

A

B
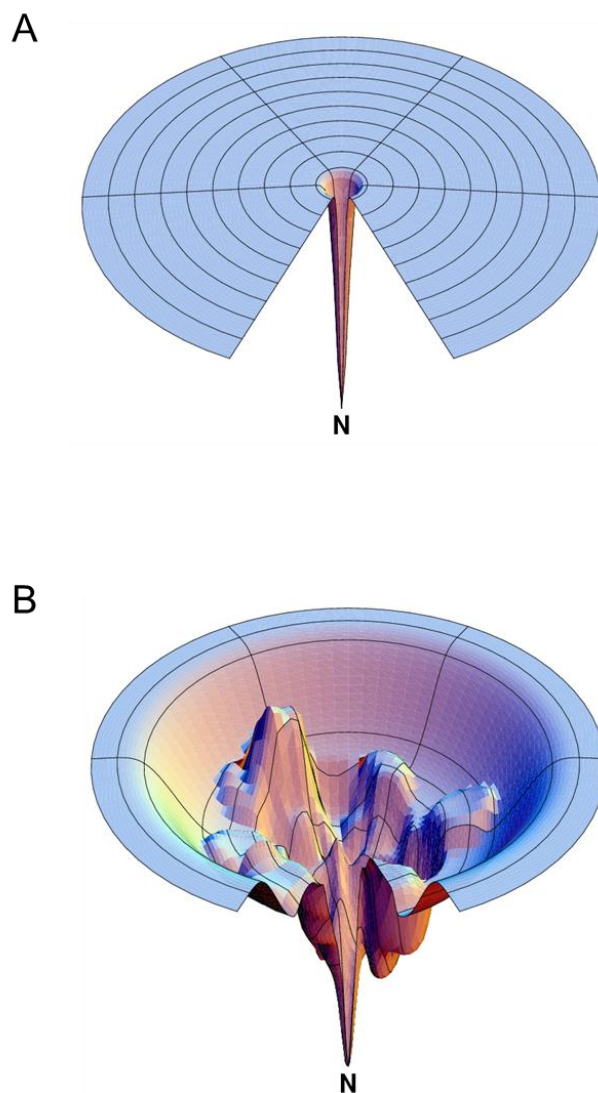
Figure 1.2. Protein folding landscapes. (A) Levinthal folding funnel involves a random conformational search for the native state. (B) A rugged free energy landscape that funnels partially folded conformations toward the native state (24).

confirmed the two-state folding. Recently Fersht and coworkers have proposed a unifying mechanism (25, 26), that consists of three basic steps: 1) formation of a nucleus with native-like topology, 2) polypeptide chain collapse, and 3) achievement of the native conformation through structural consolidation. Under the posited unifying mechanism proteins may appear to fold via the nucleation-condensation or framework models depending on the stability of their relative folding intermediates and the height of the corresponding transition barriers (27).

The current generally accepted view of protein folding extends Levinthal's pathway notion by incorporating the presence of many possible routes to the native state (28, 29). Support for this view came from studies on ultrafast folding proteins, and the finding that the folding rate is proportional to the number of folding routes (30). The presence of many possible pathways leading to the native structure forms the basis of the *folding funnel* model (24, 31). This model envisions folding as a conformational search on a rugged free energy landscape which is biased towards the native state (Figure 1.2B). The population of folding intermediates, or lack thereof, can then be explained by the degree of smoothness of the landscape.

The energetic bias of the funnel explains, at least partially, how proteins can adopt complex structures on a biologically relevant timescale. Another factor that affects the folding rate is the amount of native-like structure present in the unfolded state. The denatured state ensemble (DSE) of a protein can be highly heterogeneous. However, many recent accounts suggest the retention of some residual structure (32-34). This would alleviate the difficulties envisioned by Levinthal by limiting the conformations populated in the DSE to a small fraction of all possible conformations.

1.1.4 Folding Intermediates

Determining the mechanism by which a protein attains its unique native state requires the structural characterization of the entire folding pathway. As mentioned above, ϕ-value analysis has been useful for characterizing transition states. However, it is critical to also assess the partially folded states that become populated en route to the native state (35). The formation of intermediates for multidomain proteins is well established (36) but the case for small proteins remains unclear (37).

Many intermediates are formed via productive folding events (38). In addition, off-pathway species can become populated, which must then unfold before joining the pathway to the native structure (39). The funnel model suggests that many parallel routes are available to reach the native state, and that intermediate species may reflect kinetic traps (Fig. 1.1B) (31). Wildegger and Kiefhaber have utilized double-jump folding experiments to show that lysozyme refolding can proceed directly to the native state, or via such a kinetic trap (40).

Many protein folding transitions are highly cooperative, implying that intermediate species cannot be detected. Some proteins do populate equilibrium intermediates that are similar to their kinetic counterparts (41), but this does not appear to be universal. It is postulated that intermediates are always present (42). However, due to their transient nature time-resolved experiments are required for their characterization. Detection of folding intermediates has become a fairly common practice, but their detailed structural characterization continues to present a challenge.

1.1.5 Misfolding

Elucidating protein folding mechanisms remains one of the most ardently pursued goals in biology. Much of the interest in solving the protein folding problem stems from the fact that many diseases result from *mis*folding (43). Some misfolding diseases, such as cystic fibrosis, stem from a loss-of-function mutation which results in an altered protein structure. The other main class of folding disorders involves protein aggregation. Diseases of this type, *i.e.* Alzheimer's, Parkinson's, type 2 diabetes, and prion diseases, have elicited intense research efforts from the medical community due to their prevalence in the population and the resulting economic pressure (44). Misfolding intermediates can be prone to aggregation and are potentially suitable drug targets. However, akin to their correctly folded counterparts, the structural characterization of those species remains a challenge. A recent combination of experimental and computational approaches allowed for a detailed description of a misfolded intermediate in a small protein domain (45). A thorough understanding of protein folding mechanisms will facilitate the development of targeted therapeutics (46-48).

1.1.6 Structure Prediction and De Novo Protein Design

The ultimate goals of protein folding research are related but distinct: 1) predicting the tertiary fold of a protein solely from its amino acid sequence, and 2) synthesizing a protein to carry out a desired function. While some success has been achieved in predicting folding rates from sequences (49), structure prediction is the crucial objective. Good agreement between computational modeling and X-ray or NMR structures can be achieved for relatively small proteins but large systems represent a

challenge (50). Understanding amino acid positioning within a protein active site is straightforward, but elucidating how the remainder of the sequence interacts to achieve such positioning remains a mystery. Recently, Koder *et al.* designed a heme-binding oxygen transport protein that exhibited many similar characteristics to natural proteins, although the engineered molecules exhibited a higher affinity for oxygen than for carbon monoxide (51). Successes in the design of proteins with improved or novel functions will likely become more common as new aspects of protein folding are uncovered.

1.2 Studying Protein Folding

Protein folding is an extremely rapid process, in most cases complete after only a few seconds, with some ultra-fast domains requiring only a few *microseconds* (52). This short timescale can complicate measurements, and thus equilibrium studies are often used as a substitute. Much information about protein folding pathways has been garnered from equilibrium experiments, which monitor protein structure as a function of denaturant; frequently used denaturants include pH, chaotropic agents, and temperature. Many salient details regarding *in vivo* protein folding can be uncovered through *in vitro* kinetic experiments. These studies begin with a denatured protein which is exposed to a rapid change in environment to trigger folding. Conventional triggers include rapid mixing steps, most commonly stopped-flow, to introduce a pH jump or denaturant dilution (53). Photochemical and temperature-jump methods have also been described (54). In addition to a folding trigger, kinetic studies also require a suitable detection method, some examples of which will be discussed further.

1.2.1 Optical Methods

Absorption spectroscopy. While UV-Vis absorption spectroscopy is most commonly used for protein concentration measurements, it can also be used to monitor time-resolved conformational changes. Proteins themselves do not absorb much light in the UV/visible range so this is typically only used for those with external chromophores incorporated into their structure. Changes in absorption result from alterations to the environment of a chromophore. Two intermediates in the folding of cytochrome *c* after a pH jump were detected with absorption spectroscopy, providing evidence for the proposal of a four-state mechanism (53).

Circular dichroism. CD refers to the differential absorption of left- and right-handed circularly polarized light within a chiral molecule. CD is most often utilized to investigate the protein backbone (55). Characteristic CD profiles are obtained for α-helices, β-sheets, and random coils. Thus, the secondary structure of a protein can be interrogated in a straightforward manner.

Fluorescence spectroscopy. Aromatic amino acid residues are intrinsically fluorescent, although usually only tryptophan fluorescence is measured. The emission properties of the Trp side chain are highly sensitive to the solvent environment. The λ_{max} in water is around 355 nm, while in a hydrophobic environment (such as the protein core) it is around 335 nm (56). Interpretation of fluorescence data is greatly simplified in proteins which contain only a single Trp residue. Data analysis for proteins with multiple Trp

residues can be complicated but single Trp mutants can be used to deconvolute the fluorescence signal and localize the structural changes taking place during folding (57).

Förster resonance energy transfer. FRET involves energy transfer between a donor and an acceptor. The emission spectrum of the donor must partially overlap with the absorption spectrum of the acceptor. After excitation of the donor, fluorescence intensity measured from the acceptor reports on their relative distance. FRET is often termed a 'molecular ruler' for this reason. Care must be taken when choosing the fluorophores for protein studies, as large bulky dyes can sometimes interfere with folding. The incorporation of fluorescent unnatural amino acids represents an interesting alternative (58). FRET studies are particularly intriguing as they offer the possibility of monitoring the dynamics of single protein molecules (59).

1.2.2 Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy can provide a comprehensive portrayal of protein dynamics at the atomic level. NMR active nuclei, typically ^{15}N, ^{13}C and ^{19}F, can be incorporated into protein sequences to act as molecular reporters, along with naturally occurring ^{1}H, through the use of myriad dynamics techniques (60). 1D ^{19}F NMR combined with stopped-flow mixing allowed the folding of *E. coli* DFTR, with engineered fluorinated tryptophans, to be followed on the order of seconds (61), however 2D experiments are more common. Relaxation dispersion measurements have been used to uncover unfolding/refolding events in aMb (62) and a transient folding intermediate in a small protein domain (63). A recently described method involving *in situ* rapid mixing and

photo-CIDNP spectroscopy allows for structural transitions on the order of tens of milliseconds to be characterized (64). This method, which measures the cross-polarization of amino acid side chains from an activated aromatic residue, has revealed a preformed hydrophobic cluster in the unfolded state of a small Trp-cage protein (65). Because ^{1}H nuclei result in an NMR signal but ^{2}H nuclei do not, NMR is a suitable detection method for hydrogen/deuterium exchange experiments (see section *1.4.1*) and has been used to characterize folding intermediates populated as early as ~0.5 milliseconds (66).

1.2.3 Computer Simulations

While experimental methods routinely interrogate folding processes on the (sub)millisecond timescale, they struggle to access faster regimes. Computer simulations historically have had the opposite problem; accession of biologically relevant timescales requires an enormous amount of computing power and time. This is troublesome because experimental validation of computational outputs is required. Since the advent of molecular dynamics (MD) in the late 1970s (67) much improvement has been made in the time resolution of simulated folding studies (68). Recent work from Shaw and coworkers has caused a paradigm shift by extending MD timescales by over two orders of magnitude (69). A new supercomputer, ANTON, has allowed for a greater overlap between simulation and experiment. Folding trajectories for many small protein domains have been simulated for upwards of 1.1 *milliseconds*, producing structures that are in excellent agreement with their experimentally determined counterparts (34).

1.3 Mass Spectrometry

MS measures the mass-to-charge ratio (m/z) of gas phase ions. MS was introduced nearly a century ago. Unfortunately, traditional ionization sources were too harsh for the transfer of large, intact biomolecules into the gas phase. This feat was first accomplished nearly 25 years ago with the advent of two 'soft' ionization techniques. The 2002 Nobel Prize in Chemistry was awarded for techniques to identify and structurally characterize biomolecules of which Koichi Tanaka and John Fenn shared one half for their development of MALDI and ESI techniques, respectively. Over the past 25 years MS has become a widely used technique for the analysis of biological molecules (70). A typical mass spectrometer comprises an ion source, a mass analyzer, and a detector.

1.3.1 Ionization Techniques

1.3.1.1 Matrix-Assisted Laser Desorption/Ionization

MALDI utilizes a pulsed laser to desorb biomolecular analytes from a solid matrix. Organic molecules in the matrix absorb UV photons and desorb from the surface taking proteins with them. Charge is then transferred from the matrix molecules to the gaseous proteins. Because MALDI is a pulsed ionization method and produces very low charge states, it is most often interfaced with a time-of-flight analyzer (71). Tanaka *et al.* were able to detect singly charged proteins upwards of 25 kDa along with multimers up to 100 kDa, albeit with greatly reduced sensitivity (72). Around the same time, Karas and Hillenkamp showed that singly charged molecular ions for proteins between ~15-70 kDa

could be detected (73). MALDI is still routinely used for protein identification, and it has also found applications for tissue imaging (74).

1.3.1.2 Electrospray Ionization

ESI was first conceived by Dole *et al*. over 40 years ago (75). An explosion in the usage of ESI-MS occurred after Fenn and coworkers demonstrated its utility for studying biological macromolecules (76). ESI introduces analyte ions into the MS directly from solution and is thus commonly interfaced with liquid chromatography. Analyte solution is pumped through a spray needle which is held at a high potential, ~3-5 kV. In positive ion mode, enrichment of positive charges occurs at the needle tip resulting in a Taylor cone (Figure 1.3). With the help of a heated, concentric nebulizing gas flow, a mist of solvent droplets is emitted toward the MS. The droplets shrink through solvent evaporation until they reach the Rayleigh limit (77) at which point Coulombic fission occurs. This process of evaporation and fission is repeated until naked gas-phase ions are produced (78). Two mechanisms have been proposed for this final step.

Ion evaporation model. The IEM was first put forth by Iribarne and Thompson to describe how small molecules became ionized from droplets close to the Rayleigh limit (79). In this model, solvated ions are ejected from the droplet surface. Computer simulations have shown this to be the likely ionization mechanism for small molecules and salt clusters (80, 81). Ahadi and Konermann recently utilized a simple MD model to show that the ionization of unfolded protein chains likely proceeds via the IEM as well (82).
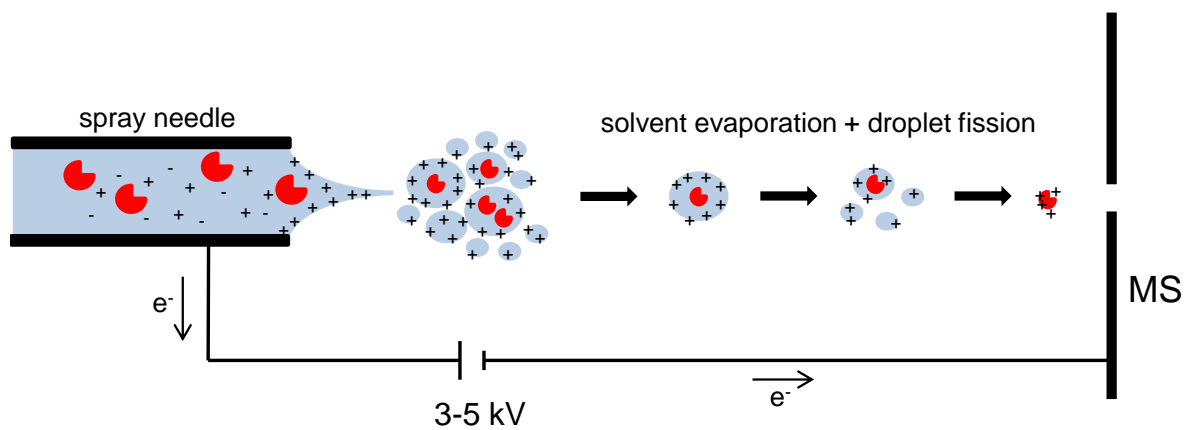
Figure 1.3. ESI Process. Simplified schematic representation of the electrospray ionization technique. Red objects represent analyte molecules.

Charged residue model. The CRM posits that gas phase ions are produced via complete solvent evaporation. Folded globular proteins are thought to become ionized by this mechanism, as the maximum ESI-MS charge state for a given protein shows good agreement with the Rayleigh charge for a water droplet of the same radius (83). Recent MD simulations have provided further evidence supporting the notion that ionization of folded proteins proceeds via the CRM (82).

1.3.2 Mass Analyzers

Once gas-phase ions have been produced and introduced into the MS, they must be separated and detected based on their m/z. Various mass analyzers are available, including quadrupole ion traps and ion cyclotron resonance cells. The two most common are linear quadrupoles and time-of-flight analyzers.

1.3.2.1 Quadrupole Mass Analyzer

A quadrupole is a set of 4 parallel cylindrical rods. An RF voltage and a superimposed DC voltage are applied to each opposite rod pair (84). For a given voltage setting only ions of a certain m/z will be transmitted through the quadrupole. All other ions will collide with the rods due to their unstable trajectories. Sequential transmittance of all ions in a given m/z range is accomplished by changing the voltages in small steps. Quadrupoles exhibit limited resolution and mass range. Therefore, they are often coupled to other mass analyzers, such as a time-of-flight unit. In such a situation, the quadrupole can be operated in RF-only mode (with no DC voltage). In this case the quadrupole acts as an *ion guide* that allows transmittance of most m/z simultaneously.

A benefit of linking two analyzers in tandem is the possibility to carry out MS/MS, or tandem-MS, measurements. In these experiments, the first quadrupole voltages are 'parked' to allow transmittance of ions of a certain m/z. These selected ions are then fragmented in a collision cell, and the resulting fragments are detected in the second mass analyzer. Quadrupole time-of-fight (Q-TOF) mass spectrometers can easily be interfaced with liquid chromatography systems and have thus become the workhorses of the proteomics and pharmaceutical industries.

1.3.2.2 Time-of-Flight Mass Analyzer

Ions are separated in a TOF analyzer based on their m/z. An electric field is applied to a packet of ions and this potential energy is transformed into kinetic energy as the ions are accelerated through the TOF tube.

$$(ze)\Delta U = \frac{1}{2}mv^2 \tag{1-1}$$

where *ze* is the ion charge and ΔU is the applied potential gradient. Rearrangement of Eq. 1-1 yields

$$v = \sqrt{\frac{2ze\Delta U}{m}} \tag{1-2}$$

The flight time on an ion is hence given by

$$t = \frac{L}{v} = \frac{L}{\sqrt{2e\Delta U}} \cdot \left(\frac{m}{z}\right)^{1/2} \tag{1-3}$$

where L is the path length. A time-to-digital converter is used to transform the measured drift times into a mass spectrum.

The applied electric voltage in the TOF may not affect all ions of a given m/z equally, resulting in a distribution of kinetic energies and thus flight times (85). This is seen as peak broadening that decreases resolution in the mass spectrum. The issue can be addressed by incorporating a reflectron (Fig. 1.3). This device acts as an ion mirror and consists of a stack of rings with increasing potential. Ions with larger velocities will penetrate deeper into the reflectron, while slower ones will follow a shallower trajectory. This change in path length compensates for slight differences in kinetic energy and allows all ions of a given m/z to be detected simultaneously, improving peak resolution by a factor of ~5.

1.3.3 Protein Folding Studied by Electrospray Mass Spectrometry

The protein charge state distribution seen in an ESI mass spectrum has been shown to be dependent on its solution phase conformation (86) and this feature has been exploited in protein folding studies (87). While equilibrium studies have been performed in this way (88), the fast detection of MS provides the opportunity to investigate kinetic folding processes (89) and even enzymatic reactions (90). Such time-resolved experiments utilize solution-phase mixing interfaced directly with a mass spectrometer (91, 92). Additional structural information can be obtained through the incorporation of a conformationally sensitive labeling step prior to MS injection. Various such labeling strategies will be discussed in the following section.
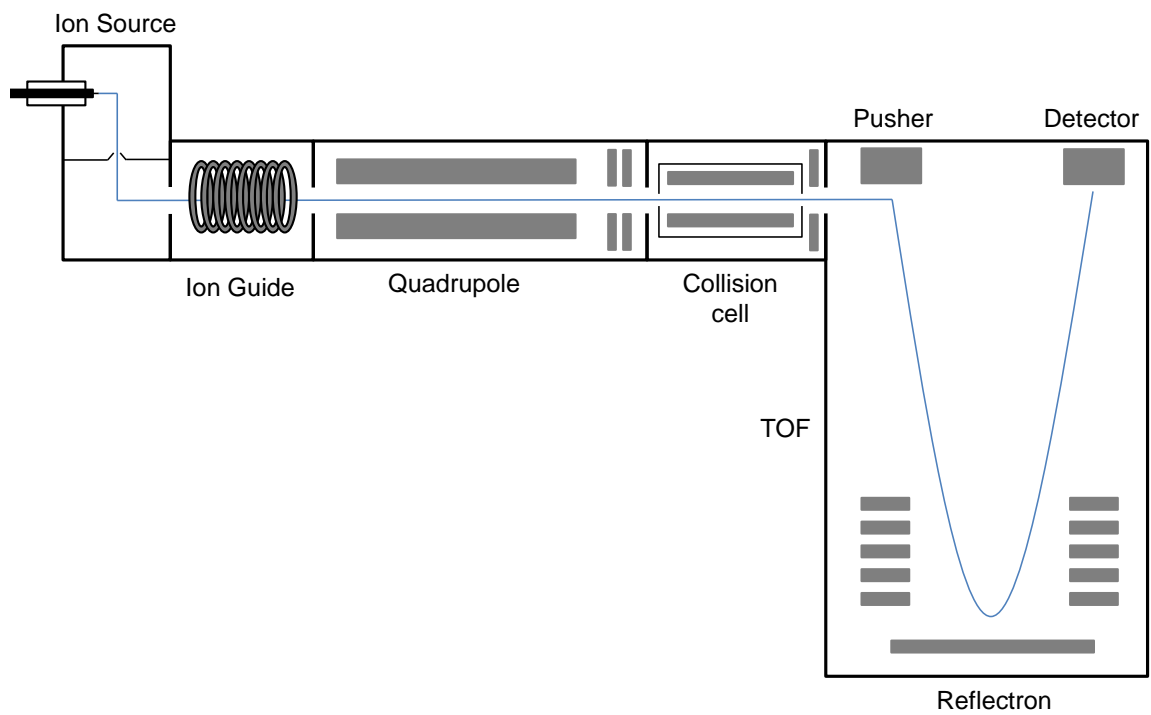
Figure 1.4. Schematic of a Q-TOF. Simple representation of the major components of a quadrupole time-of-flight mass spectrometer. The blue line indicates the path travelled by ions.

1.4 Protein Labeling

Most optical techniques only report on global structural features of proteins. To garner detailed mechanistic insights into protein structure, dynamics, and folding, spatially resolved information is required. There are two general methods that can be coupled to MS for such a purpose: hydrogen/deuterium exchange (HDX) and covalent labeling. In broad terms, HDX monitors hydrogen bonding patterns in a polypeptide backbone and covalent labeling probes the solvent accessibility of amino acid side chains. Thus, results obtained from these two techniques are complementary and can provide a comprehensive picture of the protein of interest (93).

1.4.1 Hydrogen/Deuterium Exchange

The exchange of hydrogens attached to electronegative heteroatoms in proteins with deuterium from the solvent was first described by Linderstrøm-Lang in the early 1950s (94). The kinetics of HDX, and their dependence on temperature and pH have since been characterized (95) as well as the effects of neighboring amino acid residues (96). The main principle of HDX is that exposed hydrogens will exchange with solvent quite rapidly while those involved in hydrogen bonds, as in the case of α-helices and β-sheets, will exchange much more slowly. Significant mechanistic understanding of HDX has come from NMR experiments conducted by Englander and coworkers (97). The nuclear spin of hydrogen (^{1}H) gives rise to an NMR signal while that of deuterium (^{2}H) does not, making NMR a preferred readout for hydrogen exchange experiments. Many significant discoveries have been made with this method (98). Hierarchical folding

pathways for model proteins such as cytochrome *c* (99) and apomyoglobin (66) have been characterized.

In addition to altering nuclear spin, the exchange of hydrogen for deuterium also increases the mass of an analyte. This effect can be exploited for MS experiments (100). The classic "bottom-up" approach was first described by Zhang and Smith nearly 20 years ago (101) and it involves monitoring the mass increase of proteolytic peptides, as a function of deuteration time. Bottom-up HDX-MS has been utilized extensively ever since (102), including the study of changes in receptor dynamics after drug binding protein-interactions (103), the thermodynamic stability of proteins (104), and the cyclic motions of molecular machines (105). Pulsed labeling methods have been described whereby the hydrogen bonding status of folding intermediates can be reported, often with single-residue resolution (106), however structural data should be interpreted with caution due to the basic pH required for labeling (107). Despite being widely used, HDX-MS has two main drawbacks that must be addressed: 1) deuterium scrambling and 2) back-exchange.

Akin to the mechanism of peptide fragmentation (108), deuterium scrambling is the process by which protons and deuterons can migrate along the backbone of a collisionally activated peptide, thus reducing the measured spatial resolution (109, 110). Radical-based fragmentation techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) have been described as alternatives to classical collision techniques. They have been shown to be less sensitive to neighboring amino acids, resulting in a higher sequence coverage (111) as well as having the capability to maintain the solution deuteration pattern of small peptides (112). Subsequent extension to

intact proteins have shown good agreement with NMR measurements with near single amino acid resolution (113, 114). These "top-down" approaches are conducted by fragmentation of intact proteins, and hence they do not require proteolytic digestion. A combination of proteolytic digestion and electron-based fragmentation has been termed "middle-down". This approach is capable of measuring accurate deuteration levels with amino acid resolution as in "top-down" experiments but with simplified data analysis due to peptide, rather than protein, precursor ions (115).

Back exchange is another issue that must be considered. In experiments involving proteolytic digestion, sample clean-up and peptide separation is performed via liquid chromatography (LC) prior to MS analysis. LC is performed with a gradient of aqueous and organic solvents which implies that deuterated peptides are in contact with protiated water for much of the analysis. This allows deuterium to exchange with hydrogens from the LC solvent, thereby diminishing the amount of structural information. The inclusion of fully-labeled control samples can correct back-exchange to some extent.

1.4.2 Covalent Labeling

An entirely different MS-based structural approach follows a strategy where the target protein is modified with covalent labels. This provides the opportunity for extensive downstream sample handling such as dialysis, lyophilization, and chromatography. The amount of labeling achieved at any given site is modulated by both the intrinsic reactivity of the amino acid with the label and the solvent accessibility of the side chain. Figure 1.5A depicts the typical workflow in a covalent labeling experiment. How labeling can report on various protein conformations is shown in Figure 1.5B.

Extensive labeling is seen for unfolded proteins, whereas the decreased solvent accessibility of the folded state results in fewer modifications. Side chain protection at protein-protein binding interfaces results in protection as well. Covalent labels fall into two general categories, those that react with specific amino acids and those that exhibit more promiscuous reactivity.

1.4.2.1 Specific Labeling

Specific labeling reports directly on solvent accessibility since all labeled sites will have identical reactivities. Common targets are hydrophilic residues on the protein surface such as Lys, Arg, Asp, Glu, and Cys. Predictable mass increases at labeled residues can be easily tracked by MS. Many different labeling agents targeting various amino acids have been described in the literature (116) but discussion will be limited to a select few.

Lysine modification and cross-linking. The positively charged side chain of lysine provides an ideal modification target due to its propensity to reside on protein surfaces. Acetylation of exposed lysines with MS detection dates back 20 years (117). Most often lysine modification is utilized to map protein binding interfaces such as HIV-1 reverse transcriptase contacts within the viral RNA:tRNA complex (118). While many lysines are solvent exposed, some are globally protected within the protein core. Fitzgerald and coworkers monitored the modification of buried lysines as a function of denaturant concentration to investigate the thermodynamic stability of proteins in a multicomponent mixture (119). The surface affinity and relatively high abundance of lysines has also led

A

B

unfolded momomer   folded momomer   folded dimer
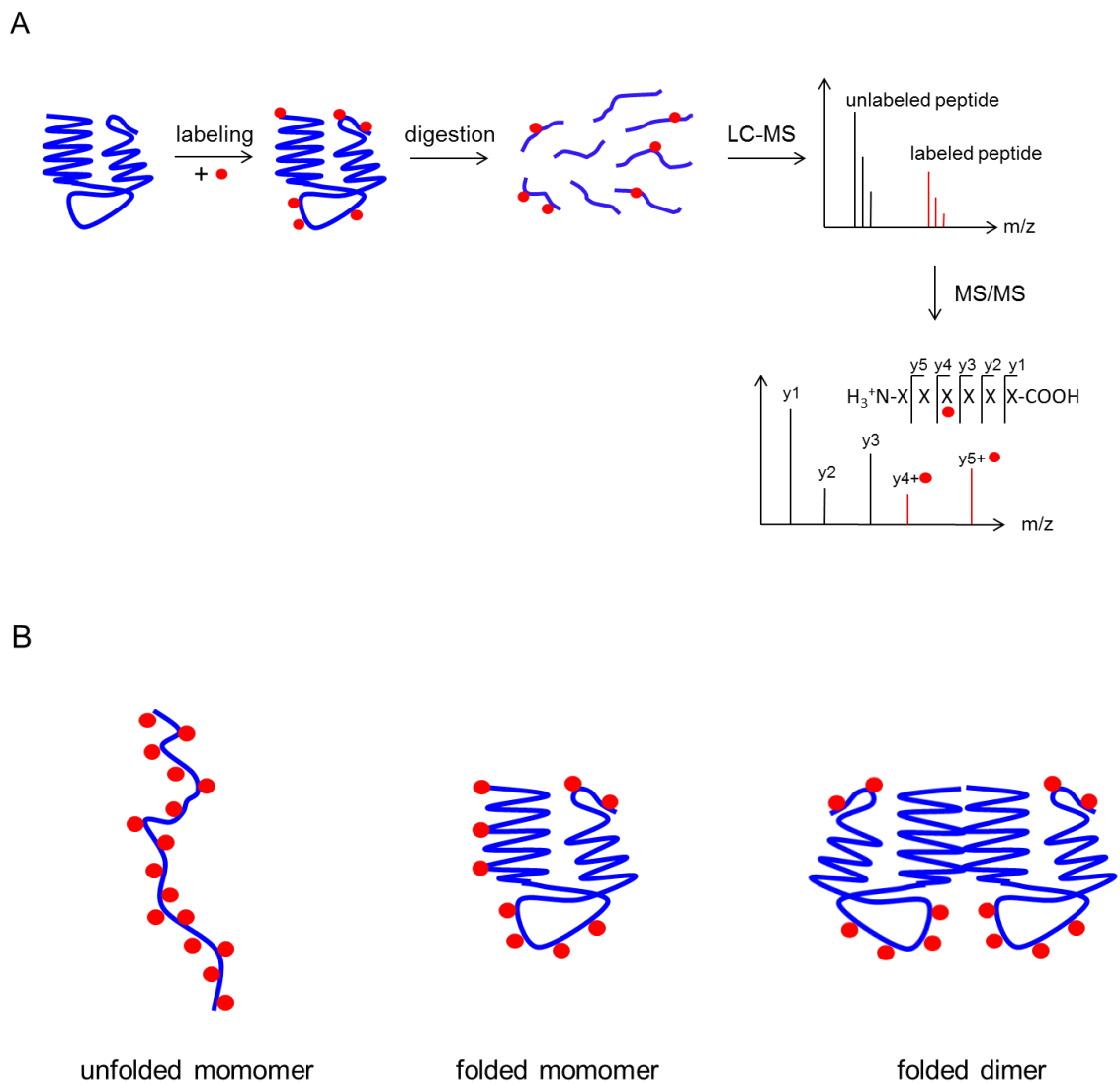
Figure 1.5. Covalent labeling. (A) Typical workflow of a covalent labeling experiment. Proteins are labeled in solution and digested. Resultant peptides are separated and detected with LC-MS. Label site determination is achieved through MS/MS experiments. (B) Depiction of how the extent of labeling reports on the degree of side chain burial.

to their being used in crosslinking experiments. These studies use bifunctional molecules of varying lengths to determine distances between lysines on a protein or within interacting proteins (120). Crosslinking data recently provided distance constraints for computational modeling that determined the subunit arrangement of a eukaryotic 16-mer chaperonin (121).

Carboxylate labeling. Similar to lysine, aspartic and glutamic acid are abundant in proteins and most often found on the surface. Mapping protein solvent accessible surface area via modification of acidic residues has forged ahead through the efforts of Gross and coworkers. The original study used glycine ethyl ester and MS to determine the orientation of the FMO antenna protein on a bacterial membrane (122). Recently the method has been extended to monitor differences in solvent accessibility of multiple conformational states of calmodulin (123).

Thiol labeling. Cysteine residues not involved in disulfide bridges contain a free –SH group that is available for modification. Such free Cys residues are rare in most proteins. Genetic engineering, however, allows the introduction of additional Cys residues. Due to the small, polar nature of the side chain these mutations can be tolerated to a far greater extent than Lys, Glu, or Asp (124). While the solvent accessibility of Cys residues as a function of denaturant has led to structural characterization of protein folding intermediates under equilibrium conditions (125), thiol labeling has predominantly been used in kinetic experiments. While many amino acids require extended incubation with a labeling reagent, the high reactivity of the -SH allows labeling to be completed in a few

milliseconds. As a result, the temporal changes in Cys labeling have provided structural information on the folding pathways of proteins such as barnase and monellin (126-128).

1.4.2.2 Non-Specific Labeling

While residue-specific labeling can simplify data analysis the spatial resolution often suffers. Increased sequence coverage can be attained through the combination of multiple labeling reagents (129). A conceptually simpler method is the use of semi- or non-specific labels. Vachet and coworkers have shown that diethylpyrocarbonate (DEPC) can modify Ser, Thr, His, and Tyr residues resulting in *ca*. 25% coverage of an average protein. Labeling times as short as a few minutes provide significant modification levels with minimal perturbation to protein structure (130). Jumper and Schriemer described a method for interrogating protein topography via production of carbene labels by photolysis of a diazarine-modified precursor (131). This carbene method provided good temporal resolution, as well as the potential to probe every amino acid in a protein sequence. By far however, the preferred non-specific labeling method for biomolecular structure analysis has been the production of hydroxyl radicals (·OH) (132).

1.4.3 Hydroxyl Radical Labeling

In 1978 Galas and Schmitz introduced the term *footprinting* to describe experiments that sought to determine regions of nucleic acids that were protected from nucleases by proteins (133); the extension to proteolytic cleavage of proteins came much later (134). This term persisted when researchers began to use hydroxyl radicals rather

than enzymes to cleave nucleic acids (135). Hydroxyl radical footprinting continues to be used to study dynamics of nucleic acids in both equilibrium and time-resolved experiments (136, 137),

The pioneering work of Chance and coworkers showed that ·OH could modify solvent exposed amino acid side chains of a protein without backbone cleavage (138). The term "footprinting" is now commonly used for hydroxyl radical labeling of protein surfaces (139), binding interfaces (140), and folding mechanisms (141). While hydroxyl radicals have the potential to modify any of the 20 amino acids, only ~50% are typically seen in a protein labeling experiment. The most reactive residues tend to be the sulfur-containing side chains (Cys and Met), followed by the aromatics (Trp, Tyr, Phe). Other commonly oxidized side chains are those of His, Leu, Ile, Val, Pro, Glu, and Arg. The oxidation of these side chains can progress through complex mechanisms which have been well described (142, 143). The most frequently observed oxidation products involve the incorporation of an oxygen atom, resulting in a +16 Da mass shift, although other products are possible. Some typical oxidation products are shown below.
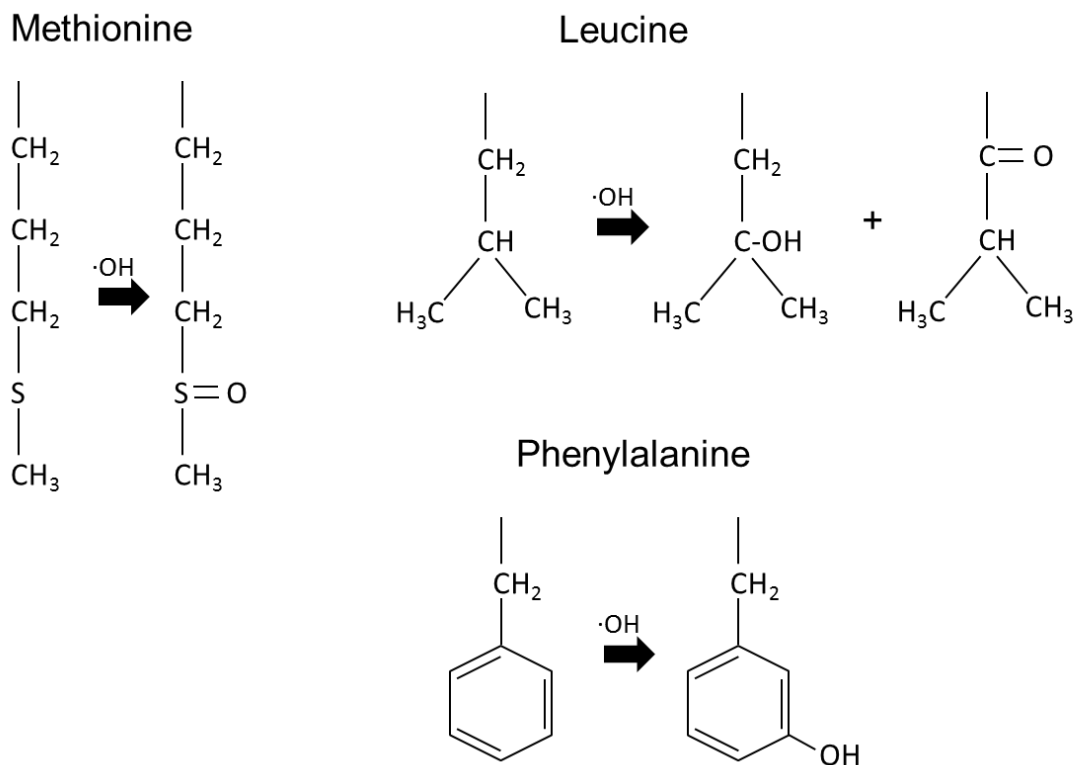
There are several radical generation approaches for protein labeling studies, such as Fenton chemistry (144), electrochemistry (145), and corona discharge methods (146, 147). Labeling has been shown to be rather insensitive to the method used, however only two have really gained a foothold in the field.

1.4.3.1 Radiolysis of Water

The generation of ·OH directly from water via synchrotron radiation for protein structural studies was pioneered by Chance when he utilized the amount of labeling of aMb as a function of denaturant to probe its unfolding pathway (148). A series of articles by Xu and Chance outlined much of the chemistry involved in ·OH labeling (149-151). Also, those authors determined a reactivity series for residue within proteins that was in

accord with published rate constants for the free amino acids (152). ·OH labeling via millisecond X-ray radiolysis of water has been used by Chance and coworkers to gain insights into enzyme mechanisms (153) and the structure/function relationship of integral membrane proteins (154). Since access to synchrotron light sources is limited, the Chance group has maintained a near monopoly in this field.

A water radiolysis method with broader accessibility uses γ-rays to generate ·OH. Because of the relatively long labeling time required (minutes), γ-ray radiolysis is well suited for equilibrium experiments and has been exploited mainly in differential studies. Experiments of this type seek to gain topological information on discrete conformational states of a protein. Examples include the calcium induced conformational change of calmodulin (155), the effect of heme removal from myoglobin (156), and the prepore to pore transition of the protective antigen from *Bacillus anthracis* (157).

To overcome the deficiencies of limited synchrotron access and long labeling times in γ-ray studies, Sharp and coworkers recently used a pulsed electron gun to generate ·OH from water on the submicrosecond timescale (158). This proof-of-principle experiment monitored the solvent accessibility of residues within model proteins ubiquitin and β-lactoglobulin. The extremely fast labeling time achieved with this method is highly desirable for the study of protein folding in a time-resolved manner, however this avenue remains unexplored.

1.4.3.2 Photolysis of Hydrogen Peroxide

Early studies mapped protein surfaces using concentrated (~15% v/v) H_2O_2 and prolonged exposure to low intensity UV light (159). While this method produces modified proteins, spurious 'background' oxidation can occur from the presence of the peroxide. More recently, a photolytic approach has evolved for generating radicals from dilute (~0.1% v/v) solutions of peroxide via homolytic cleavage with a pulsed UV laser (160, 161). Non-irradiated control samples have shown that incubation of the protein with such low concentrations of peroxide exhibit very little background oxidation. A seminal paper by Hambly and Gross described the pulse-labeling approach in combination with a continuous-flow reaction setup (160). This continuous-flow system is preferable to the static system of Aye *et al.* (161) because with a simple plug flow approximation, 'single-hit' conditions can be attained. 'Single-hit' refers to the regime in which each protein molecule only "sees" one labeling pulse. This can be achieved by adjusting the flow rate and laser frequency such that two adjacent irradiated volumes are separated by a non-irradiated volume (Figure 1.6). Oxidation has been shown to disrupt protein structure (162). A second labeling pulse may therefore modify side chains that were not initially solvent exposed, resulting in artifactual data. Plug flow constitutes a useful zero-order approximation. However, a more accurate picture of the dynamics within the reaction chamber takes into account laminar flow (163). Hambly and Gross also showed that the ·OH lifetime in solution can be tuned via the addition of radical scavengers, such as glutamine. This results in a labeling pulse as short as ~1 μs (160) which is shorter than oxidation-induced protein unfolding (164).
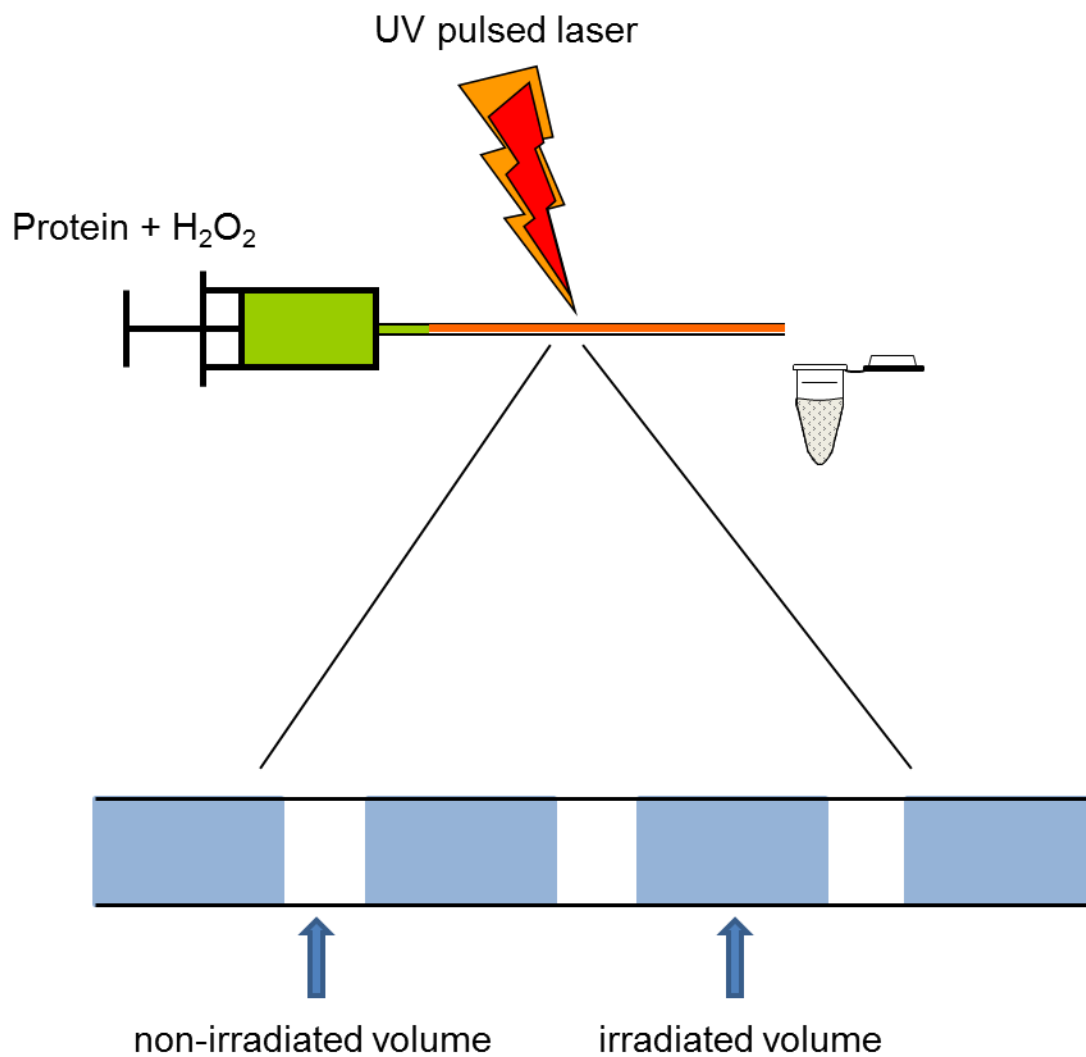
Figure 1.6. Photolysis experimental setup. Protein and peroxide are mixed and pumped through a UV transparent capillary onto which the laser is focused. Proper adjustment of flow rate and laser frequency leads to non-irradiated solution separating adjacent irradiated volumes.

·OH labeling by laser photolysis of peroxide has been used in many recent protein structural studies. Most seek to identify differences in solvent accessibility between multiple states of a protein. The interface of the carbohydrate-binding, homodimeric protein galectin-1 was characterized through comparison of labeling levels for the monomers and for the dimer. Results were shown to be in accord with theoretical solvent accessibilities calculated from MD simulations (140). Similarly, the conformational changes in calmodulin concomitant with peptide binding have been described (165). Pan *et al.* extended the methodology to look at exposed Met residues of the integral membrane protein bacteriorhodopsin (BR) (166).

The very short labeling pulse achieved in laser photolysis makes it ideally suited to investigate rapid, time-resolved changes in protein structure however there is a dearth of such studies in the literature. Following work that will be presented later in this thesis, Pan *et al.* used rapid mixing followed by oxidative labeling to explore the BR folding mechanism (141). A temperature-jump method has also recently been described that is capable of labeling proteins that fold on the sub-millisecond timescale (167).

1.5 Scope of Thesis

This work extends the photochemical oxidation method developed by Hambly and Gross (160) to study protein folding transitions through the combination of continuous-flow rapid mixing, pulsed labeling, and MS detection. The extremely short radical lifetime makes such a method ideally suited to study rapid structural changes. Prior to the work described here, ·OH labeling had only been utilized for equilibrium measurements.

We will describe the first application of pulsed oxidative labeling to characterize transient protein unfolding intermediates, along with a comprehensive data analysis strategy to deal with differing peptide reactivities (Chapter 2). The folding pathway of a small model protein is then characterized (Chapter 3). In both cases, the labeling data are correlated with optical measurements. The approach is extended to monitor the folding and assembly of a protein complex (Chapter 4). Finally, we apply oxidative labeling method to explore the folding of a large protein with direct implications in human disease (Chapter 5). These results are explicitly compared with hydrogen bonding data from HDX experiments.

The labeling approach introduced here provides an avenue for obtaining residue-resolved structural information regarding side-chain accessibility during rapid protein folding processes, with application to other biomolecular events as well. The combination of the resultant labeling data with those obtained from HDX allows for a comprehensive view of structural changes that occur during protein folding.

1.6 References

1.   Grow, E. J. and Wysocka, J. (2010) Flipping MLL1's Switch One Proline at a Time. *Cell* 141: 1108-1110.

2.   Fersht, A. R. (1999) *Structure and Mechanism in Protein Science*. (W. H. Freeman & Co., New York).

3.   Service, R. F. (2008) Problem Solved* (*sort of). *Science* 321: 784-786.

4.   Nelson, D. L. and Cox, M. M. (2000) *Lehninger Principles of Biochemistry*. (Worth Publishers, New York).

5.   Jensen, O. L. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8: 33-41.

6.   Mei, G., Di Venere, A., Rosato, N. and Finazzi-Agro, A. (2005) The importance of being dimeric. *FEBS J.* 272: 16-27.

7.   Rose, G. D., Gaselowitz, A. R., Lesser, G. J., Lee, R. H. and Zehfus, M. H. (1985) Hydrophobicity of Amino Acid Residues in Globular Proteins. *Science* 229: 834-838.

8.   Pace, C. N., et al. (2011) Contribution of Hydrophobic Interactions to Protein Stability. *J. Mol. Biol.* 408: 514-528.

9.   Rose, G. D. (1993) Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* 22: 381-415.

10.   Baldwin, R. L. (2007) Energetics of Protein Folding. *J. Mol. Biol.* 371: 283-301.

11.   Shi, Z., Krantz, B. A., Kallenbach, N. and Sosnick, T. R. (2002) Contribution of Hydrogen Bonding to Protein Stability Estimated from Isotope Effects. *Biochemistry* 41: 2120-2129.

12.   Strickler, S. S., et al. (2006) Protein Stability and Surface Electrostatics: A Charged Relationship. *Biochemistry* 45: 2761-2766.

13. Gallivan, J. P. and Dougherty, D. A. (1999) Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* 96: 9459-9464.

14. Crowley, P. B. and Golovin, A. (2005) Cation-π Interactions in Protein-Protein Interfaces. *Proteins: Struct. Funct. Bioinf.* 59: 231-239.

15. Pan, Y., Zhang, K., Qi, J., Springer, T. A. and Chen, J. (2010) Cation-π interaction regulates ligand-binding affinity and signaling of integrin $\alpha_4\beta_7$. *Proc. Natl. Acad. Sci. U.S.A.* 107: 21388-21393.

16. Wedemeyer, W. J., Welker, E., Narayan, M. and Scheraga, H. A. (2000) Disulfide Bonds and Protein Folding. *Biochemistry* 39: 4207-4216.

17. Kendrew, J. C., et al. (1958) A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* 181: 662-666.

18. Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science* 181: 223-230.

19. Levinthal, C. (1969) in *Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, eds J. T. P. DeBrunner and E. Munck (University of Illinois Press, Chicago), pp 22-24.

20. Levinthal, C. (1968) Are there pathways for protein folding? *J. Med. Phys.* 65: 44-45.

21. Kim, P. S. and Baldwin, R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Ann. Rev. Biochem.* 51: 459-89.

22. Fersht, A. R. (1997) Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7: 3-9.

23. Jackson, S. E. and Fersht, A. (1991) Folding of Chymotrypsin Inhibitor 2. 1. Evidence for a Two-State Transition. *Biochem.* 30: 10428-10435.

24. Dill, K. A. and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4: 10-19.

25. Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28: 18-25.

26. Gianni, S., et al. (2003) Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 100: 13286-13291.

27. Gianni, S., et al. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 104: 128-133.

28. Baldwin, R. L. (1994) Matching speed and stability. *Nature* 369: 183-184.

29. Lazaridis, T. and Karplus, M. (1997) "New View" of Protein Folding Reconciled with the Old Through Multiple Unfolding Simulations. *Science* 278: 1928-1931.

30. Ghosh, K., Ozkan, S. B. and Dill, K. A. (2007) The Ultimate Speed Limit to Protein Folding Is Conformational Searching. *J. Am. Chem. Soc.* 129: 11920-11927.

31. Onuchic, J. N., Luthey-Schulten, Z. and Wolynes, P. G. (1997) Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* 48: 545-600.

32. Shortle, D. and Ackerman, M. S. (2001) Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. *Science* 293: 487-489.

33. Plaxco, K. W. and Gross, M. (2001) Unfolded, yes, but random? Never ! *Nat. Struct. Biol.* 8: 659-660.

34. Lindorff-Larsen, K., Piana, S., Dror, R. O. and Shaw, D. E. (2011) How Fast-Folding Proteins Fold. *Science* 334: 517-520.

35. Gianni, S., Ivarsson, Y., Jemth, P., Brunori, M. and Travaglini-Allocatelli, C. (2007) Identification and characterization of protein folding intermediates. *Biophys. Chem.* 128: 105-113.

36.   Brockwell, D. J. and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Op. Struct. Biol.* 17: 30-37.

37.   Jackson, S. E. (1998) How do small single-domain proteins fold? *Fold. Des.* 3: R81-R91.

38.   Fersht, A. R. (2000) A kinetically significant intermediate in the folding of barnase. *Proc. Natl. Acad. Sci. U.S.A.* 97: 14121-14126.

39.   Nabuurs, S. M., Westphal, A. H. and van Mierlo, C. P. M. (2009) Extensive Formation of Off-Pathway Species during Folding of an α-β Parallel Protein Is Due to Docking of (Non)native Structure Elements in Unfolded Molecules. *J. Am. Chem. Soc.* 130: 16914-16920.

40.   Wildegger, G. and Kiefhaber, T. (1997) Three-state Model for Lysozyme Folding: Triangular Folding Mechanism With an Energetically Trapped Intermediate. *J. Mol. Biol.* 270: 294-304.

41.   Nishimura, C., Dyson, H. J. and Wright, P. E. (2005) Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc. Natl. Acad. Sci. U.S.A.* 102: 4765-4770.

42.   Daggett, V. and Fersht, A. (2003) The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4: 497-502.

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

44.   Aguzzi, A. and O'Connor, T. (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat. Rev. Drug Discov.* 9: 237-248.

45.   Gianni, S., et al. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat. Struct. Mol. Biol.* 17: 1431-1437.

46.   Connelly, S., Choi, S., Johnson, S. M., Kelly, J. W. and Wilson, I. A. (2010) Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses. *Curr. Opin. Struct. Biol.* 20: 54-62.

47. Woods, L. A., et al. (2011) Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nat. Chem. Biol.* 7: 730-739.

48. Perchiacca, J. M., Ladiwala, A. R. A., Bhattacharya, M. and Tessier, P. M. (2012) Structure-based design of conformation- and sequence-specific antibodies agains amyloid β. *Proc. Natl. Acad. Sci. U.S.A.* 109: 84-89.

49. Ivankov, D. N. and Finkelstein, A. V. (2004) Prediction of Protein Folding Rates from the Amino Acid Sequence-Predicted Secondary Structure. *Proc. Natl. Acad. Sci. U.S.A.* 101: 8942-8944.

50. Kryshtafovych, A., Fidelis, K. and Moult, J. (2011) CASP9 results compared to those of previous CASP experiments. *Proteins* 79 (Suppl 10): 196-207.

51. Koder, R. L., et al. (2009) Design and engineering of an O_2 transport protein. *Nature* 458: 305-309.

52. Qiu, L., Pabit, S. A., Roitberg, A. E. and Hagen, S. J. (2002) Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* 124: 12952-12953.

53. Roder, H., Maki, K. and Cheng, H. (2006) Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem. Rev.* 106: 1836-1861.

54. Gruebele, M. (1999) The Fast Protein Folding Problem. *Annu. Rev. Phys. Chem.* 50: 485-516.

55. Kelly, S. W., Jess, T. J. and Price, N. C. (2005) How to Study Protein by Circular Dichroism. *Biochim. Biophys. Acta* 1751: 119-139.

56. Royer, C. A. (2006) Probing Protein Folding and Conformational Transitions with Fluorescence. *Chem. Rev.* 106: 1769-1784.

57. Tew, D. J. and Bottomley, S. P. (2001) Probing the Equilibrium Denaturation of the serpin α_1-Antitrypsin with Single Tryptophan Mutants; Evidence for Structure in the Urea Unfolded State. *J. Mol. Biol.* 313: 1161-1169.

58.   Miyake-Stoner, S. J., et al. (2009) Probing Protein Folding Using Site-Specifically Encoded Unnatural Amino Acids as FRET Donors with Tryptophan. *Biochemistry* 48: 5953-5962.

59.   Schuler, B. and Eaton, W. A. (2008) Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18: 16-26.

60.   Mittermaier, A. K. and Kay, L. E. (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem. Sci.* 34: 601-611.

61.   Dobson, C. M. and Hore, P. J. (1998) Kinetic Studies of Protein Folding Using NMR Spectroscopy. *Nat. Struct. Biol.* 5: 504-507.

62.   Meinhold, D. W. and Wright, P. E. (2011) Measurement of protein unfolding/refolding kinetics and structural characterization of hidden intermediates by NMR relaxation dispersion. *Proc. Nat. Acad. Sci.* 108: 9078-9083.

63.   Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. and Kay, L. E. (2010) A Transient and Low-Populated Protein-Folding Intermediate at Atomic Resolution. *Science* 329: 1312-1316.

64.   Mok, K. H., et al. (2003) Rapid Sample-Mixing Technique for Transient NMR and Photo-CIDNP Spectroscopy: Applications to Real-Time Protein Folding. *J. Am. Chem. Soc.* 125: 12484-12492.

65.   Mok, K. H., et al. (2007) A pre-existing hydrophobic collapse in the unfolded state of an ultrafast folding protein. *Nature* 447: 106-109.

66.   Uzawa, T., et al. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13859-13864.

67.   McCammon, J. A., Gelin, B. R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature* 267: 585-590.

68.   Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* 19: 120-127.

69. Shaw, D. E., et al. (2010) Atomic Level Characterization of the Structural Dynamics of Proteins. *Science* 330: 341-346.

70. Kaltashov, I. A. and Eyles, S. J. (2005) *Mass Spectrometry in Biophysics*. (John Wiley and Sons, Inc., Hoboken, NJ).

71. Tanaka, K. (2003) The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angew. Chem. Int. Ed.* 42: 3861-3870.

72. Tanaka, K., Waki, H., Ido, Y., Akita, S. and Yoshida, Y. (1988) Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2: 151-153.

73. Karas, M. and Hillenkamp, F. (1988) Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons. *Anal. Chem.* 60: 2299-2301.

74. McDonnell, L. A. and Heeren, R. M. A. (2007) Imaging Mass Spectrometry. *Mass Spectrom. Rev.* 26: 606-643.

75. Dole, M., et al. (1968) Molecular beams of macroions. *J. Chem. Phys.* 49: 2240-2249.

76. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989) Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* 246: 64-71.

77. Rayleigh, L. (1882) On the Equilibrium of Liquid Conducting Masses charged with Electricity. *Phil. Mag.* 14: 184-186.

78. Kebarle, P. and Verkerk, U. H. (2009) Electrospray: From Ions in Solutions to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* 28: 898-917.

79. Iribarne, J. V. and Thomson, B. A. (1976) On the evaporation of small ions from charged droplets. *J. Chem. Phys.* 64: 2287-2294.

80.   Gamero-Castaño, M. and de la Mora, F. (2000) Kinetics of small ion evaporation from the charge and mass distribution of multiply charged clusters in electrosprays. *J. Mass Spectrom.* 35: 790-803.

81.   Ahadi, E. and Konermann, L. (2011) Ejection of solvated ions from electrosprayed methanol/water nanodroplets studied by molecular dynamics simulations. *J. Am. Chem. Soc.* 133: 9354-9363.

82.   Ahadi, E. and Konermann, L. (2012) Modeling the behavior of coarse-grained polymer chains in charged water droplets: Implications for the mechanism of electrospray ionization. *J. Phys. Chem. B* 116: 104-112.

83.   de la Mora, F. J. (2000) Electrospray Ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* 406: 93-104.

84.   Douglas, D. J. (2009) Linear Quadrupoles in Mass Spectrometry. *Mass Spectrom. Rev.* 28: 937-960.

85.   Cotter, R. J. (1997) *Time-of-Flight Mass Spectromety*. (American Chemical Society, Washington).

86.   Chowdhury, S. K., Katta, V. and Chait, B. T. (1990) Probing Conformational Changes in Proteins by Mass Spectrometry. *J. Am. Chem. Soc.* 112: 9012-9013.

87.   Konermann, L. (1998) Protein folding studied by electrospray ionization mass spectrometry. *Science Progress* 81: 123-140.

88.   Grandori, R. (2002) Detecting equilibrium cytochrome *c* folding intermediates by electrospray ionization mass spectrometry: Two partially folded forms populate the molten globule state. *Protein Sci.* 11: 453-458.

89.   Konermann, L., Collings, B. A. and Douglas, D. J. (1997) Cytochrome *c* Folding Kinetics Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *Biochemistry* 36: 5554-5559.

90.   Wilson, D. J. and Konermann, L. (2004) Mechanistic Studies on Enzymatic Reactions by Electrospray Ionization MS Using a Capillary Mixer with Adjustable Reaction Chamber Volume for Time-Resolved Measurements. *Anal. Chem.* 76: 2537-2543.

91.   Konermann, L., Pan, J., Simmons, D. A. and Wilson, D. J. (2007) in *The Encyclopedia of Mass Spectrometry*, eds M. L. Gross and R. M. Caprioli (Elsevier, Amsterdam), pp 802-810.

92.   Rob, T. and Wilson, D. J. (2009) A Versatile Microfluidic Chip for Millisecond Time-Scale Kinetic Studies by Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 20: 124-130.

93.   Zheng, X., Wintrode, P. L. and Chance, M. R. (2008) Complementary Structural Mass Spectrometry Techniques Reveal Local Dynamics in Functionally Important Regions of a Metastable Serpin. *Structure* 16: 38-51.

94.   Hvidt, A. and Linderstrom-Lang, K. (1954) Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. *Biochem. Biophys. Acta* 14: 574-575.

95.   Hvidt, A. and Nielsen, S. O. (1966) Hydrogen exchange in proteins. *Adv. Protein Chem.* 21: 287-386.

96.   Bai, Y., Milne, J. S., Mayne, L. and Englander, S. W. (1993) Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins: Struct. Funct. Genet.* 17: 75-86.

97.   Englander, S. W. and Kallenbach, N. R. (1984) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* 16: 521-655.

98.   Bai, Y. (2006) Protein Folding Pathways Studied by Pulsed- and Native-State Hydrogen Exchange. *Chem. Rev.* 106: 1757-1768.

99.   Roder, H., Elöve, G. A. and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335: 700-704.

100.   Katta, V. and Chait, B. T. (1991) Conformational Changes in Proteins Probed by Hydrogen-exchange Electrospray-ionisation Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 5: 214-217.

101.   Zhang, Z. and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass sectrometry: a new tool for protein structure elucidation. *Protein Sci.* 2: 522-531.

102.   Konermann, L., Pan, J., Liu, Y., (2011) Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* 40: 1224-1234.

103.   Chalmers, M. J., et al. (2006) Probing Protein Ligand Interactions by Automated Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* 78: 1005-1014.

104.   Dai, S. Y. and Fitzgerald, M. C. (2006) A Mass Spectrometry-Based Probe of Equilibrium Intermediates in Protein-Folding Reactions. *Biochemistry* 45: 12890-12897.

105.   Pan, Y., Brown, L. and Konermann, L. (2011) Hydrogen Exchange Mass Spectrometry of Bacteriorhodopsin Reveals Light-Induced Changes in the Structural Dynamics of a Biomolecular Machine. *J. Am. Chem. Soc.* 133: 20237-20244.

106.   Pan, J., Han, J., Borchers, C. H. and Konermann, L. (2010) Characterizing Short-Lived Protein Folding Intermediates by Top-Down Hydrogen Exchange Mass Spectrometry. *Anal. Chem.* 82: 8591-8597.

107.   Bieri, O. and Kiefhaber, T. (2001) Origin of Apparent Fast and Non-exponential Kinetics of Lysozyme Folding Measured in Pulsed Hydrogen Exchange Experiments. *J. Mol. Biol.* 310: 919-935.

108.   Dongré, A. R., Jones, J. L., Somogyi, Á. and Wysocki, V. H. (1996) Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* 118: 8365-8374.

109.   Demmers, J. A. A., Rijkers, D. T. S., Haverkamp, J., Killian, J. A. and Heck, A. J. R. (2002) Factors Affecting Gas-Phase Deuterium Scrambling in Peptide Ions and their Implications for Protein Structure Determination. *J. Am. Chem. Soc.* 124: 11191-11198.

110.   Ferguson, P. L., et al. (2007) Hydrogen/Deuterium Scrambling During Quadrupole-Time-of-Flight MS/MS Analysis of a Zinc-Binding Protein Domain. *Anal. Chem.* 79: 153-160.

111. Kruger, N. A., et al. (1999) Electron Capture versus energetic dissociation of protein ions. *Int. J. Mass Spectrom.* 182/183: 1-5.

112. Rand, K. D., Adams, C. M., Zubarev, R. A. and Jørgensen, T. J. D. (2008) Electron Capture Dissociation Proceeds with a Low Degree of Intramolecular Migration of Peptide Amide Hydrogens. *J. Am. Chem. Soc.* 130: 1341-1349.

113. Pan, J., Han, J., Borchers, C. H. and Konermann, L. (2008) Electron Capture Dissociation of Electrosprayed Protein Ions for Spatially-Resolved Hydrogen Exchange Measurements. *J. Am. Chem. Soc.* 130: 11574-11575.

114. Pan, J., Han, J., Borchers, C. H. and Konermann, L. (2009) Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 kDa Protein. *J. Am. Chem. Soc.* 131: 12801–12808.

115. Rand, K. D., Zehl, M., Jensen, O. N. and Jørgensen, T. J. D. (2009) Protein Hydrogen Exchange Measured at Single-Residue Resolution by Electron Transfer Dissociation Mass Spectrometry. *Anal. Chem.* 81: 5577-5584.

116. Mendoza, V. L. and Vachet, R. W. (2009) Probing Protein Structure by Amino Acid-specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 28: 785-815.

117. Suckau, D., Mak, M. and Przybylski, M. (1992) Protein surface topology-probing by selective chemical modification and mass spectrometric peptide mapping. *Proc. Natl. Acad. Sci. U.S.A.* 89: 5630-5634.

118. Kvaratskhelia, M., Miller, J. T., Budihas, S. R., Pannell, L. K. and Le Grice, S. F. J. (2002) Identification of specific HIV-1 reverse transcriptase contacts to the viral RNA:tRNA complex by mass spectrometry and a primary amine selective reagent. *Proc. Natl. Acad. Sci. U.S.A.* 99: 15988-15933.

119. Xu, Y., Falk, I. N., Hallen, M. A. and Fitzgerald, M. C. (2011) Mass Spectrometry- and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions. *Anal. Chem.* 83: 3555-3562.

120. Petrotchenko, E. and Borchers, C. H. (2010) Crosslinking Combined with Mass Spectrometry for Structural Proteomics. *Mass Spectrom. Rev.* 29: 862-876.

121. Kalisman, N., Adams, C. M. and Levitt, M. (2012) Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci. U.S.A.* 109: 2884-2889.

122. Wen, J., Zhang, H., Gross, M. L. and Blankenship, R. E. (2009) Membrane Orientation of the FMO Antenna Protein from *Chlorobaculum tepidum* as Determined by Mass Spectromety-based Footprinting. *Proc. Natl. Acad. Sci. U.S.A.* 106: 6134-6139.

123. Zhang, H., Wen, J., Huang, R. Y.-C., Blankenship, R. E. and Gross, M. L. (2012) Mass spectrometry-based carboxyl footprinting of proteins: Method evaluation. *Int. J. Mass Spectrom.* 213: 78-86.

124. Betts, M. J. and Russell, R. B. (2003) in *Bioinformatics for Geneticists*, eds M. R. Barnes and I. C. Gray (Wiley, West Sussex), pp 289-316.

125. Krishnan, B. and Gierasch, L. M. (2011) Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat. Struct. Mol. Biol.* 18: 222-226.

126. Ha, J.-H. and Loh, S. N. (1998) Changes in side chain packing during apomyoglobin folding characterized by pulsed thiol-disulfide exchange. *Nat. Struct. Biol.* 5: 730-737.

127. Jha, S. K. and Udgaonkar, J. B. (2007) Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* 282: 37479-37491.

128. Jha, S. K., Dasgupta, A., Malhotra, P. and Udgaonkar, J. B. (2011) Identification of Multiple Folding Pathways on Menellin Using Pulsed Thiol Labeling and Mass Spectrometry. *Biochemistry* 50: 3062-3074.

129. Gau, B., Garai, K., Frieden, C. and Gross, M. L. (2011) Mass spectrometry-based protein footprinting characterizes the structures of oligomeric apolipoprotein E2, E3, and E4. *Biochemistry* 50: 8117-8126.

130. Mendoza, V. L. and Vachet, R. W. (2008) Protein Surface Mapping Using Diethylpyrocarbonate with Mass Spectrometric Detection. *Anal. Chem.* 80: 2895-2904.

131. Jumper, C. C. and Schriemer, D. C. (2011) Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis. *Anal. Chem.* 83: 2913–2920.

132. Takamoto, K. and Chance, M. R. (2006) Radiolytic Protein Footprinting with Mass Spectrometry to Probe the Strucuture of Macromolecular Complexes. *Annu. Rev. Biophys. Biomol. Struct.* 35: 251-276.

133. Galas, D. J. and Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acid Research* 5: 3157-3170.

134. Zhong, M., Lin, L. and Kallenbach, N. R. (1995) A method for probing the topography and interactions of proteins: Footprinting of myoglobin. *Proc. Natl. Acad. Sci. U.S.A.* 92: 2111-2115.

135. Tullius, T. D. and Dombroski, B. A. (1986) Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to λ repressor and Cro protein. *Proc. Natl. Acad. Sci. U.S.A.* 83: 5469-5473.

136. Brenowitz, M., Chance, M. R., Dhavan, G. and Takamoto, K. (2002) Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical 'footprinting'. *Curr. Opin. Struct. Biol.* 12: 648-653.

137. Sclavi, B. (2008) Time-resolved footprinting for the study of the structural dynamics of DNA-protein interactions. *Biochem. Soc. Trans.* 36: 745-748.

138. Maleknia, S. D., Brenowitz, M. and Chance, M. R. (1999) Millisecond Radiolytic Modification of Peptides by Synchrotron X-rays Identified by Mass Spectrometry. *Anal. Chem.* 71: 3965-3973.

139. Jones, L. M., Sperry, J. B., Carroll, J. A. and Gross, M. L. (2011) Fast Photochemical Oxidation of Proteins for Epitope Mapping. *Anal. Chem.* 83: 7657-7661.

140. Charvatova, O., et al. (2008) Quantifying Protein Interface Footprinting by Hydroxyl Radical Oxidation and Molecular Dynamics Simulation: Application to Galectin-1. *J. Am. Soc. Mass Spectrom.* 19: 1692-1705.

141. Pan, Y., Brown, L. and Konermann, L. (2011) Kinetic Folding Mechanism of an Integral Membrane Protein Examined by Pulsed Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* 410: 146-158.

142. Garrison, W. M. (1987) Reaction mechanisms in the radiolysis of peptides,polypeptides, and proteins. *Chem. Rev.* 87: 381-398.

143. Xu, G. and Chance, M. R. (2007) Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* 107: 3514-3543.

144. Shcherbakova, I., Mitra, S., Beer, R. H. and Brenowitz, M. (2006) Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA,RNA and proteins. *Nucleic Acids Res.* 34: e48.

145. McClintock, C., Kertesz, V. and Hettich, R. L. (2008) Development of an Electrochemical Oxidation Method for Probing Higher Order Protein Structure with Mass Spectrometry. *Anal. Chem.* 80: 3304-3317.

146. Maleknia, S. D., Chance, M. R. and Downard, K. M. (1999) Electrospray-assisted Modification of Proteins: a Radical Probe of Protein Structure. *Rapid Commun. Mass Spectrom.* 13: 2352-2358.

147. Boys, B. L., Kuprowski, M. C., Noël, J. J. and Konermann, L. (2009) Protein Oxidative Modifications During Electrospray Ionization: Solution Phase Electrochemistry or Corona Discharge-Induced Radical Attack? *Anal. Chem.* 81: 4027-4034.

148. Chance, M. R. (2001) Unfolding of apomyoglobin examined by synchrotron footprinting. *Biochem. Biophys. Res. Comm.* 287: 614-621.

149. Xu, G. and Chance, M. R. (2005) Radiolytic Modification and Reactivity of Amino Acid Residues as Structural Probes for Protein Footprinting. *Anal. Chem.* 77: 4549-4555.

150. Xu, G., Takamoto, K. and Chance, M. R. (2003) Radiolytic Modification of Basic Amino Acid Residues in Peptides: Probes for Examining Protein-Protein Interactions. *Anal. Chem.* 75: 6995-7007.

151. Xu, G. and Chance, M. R. (2005) Radiolytic Modification of Sulfur-Containing Amino Acid Residues in Model Peptides: Fundamental Studies for Protein Footprinting. *Anal. Chem.* 77: 2437-2449.

152. Buxton, G. V., Greenstock, C. L., Helman, W. P. and Ross, A. B. (1988) Critical Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen Atoms and Hydroxyl Radicals (.OH/.O) in Aqueous Solution. *J. Phys. Chem. Ref. Data* 17: 513-886.

153. Bohon, J., Jennings, L. D., Phillips, C. M., Licht, S. and Chance, M. R. (2008) Synchrotron Protein Footprinting Supports Substrate Translocation by ClpA via ATP-Induced Movements of the D2 Loop. *Structure* 16: 1157-1165.

154. Angel, T. E., Gupta, S., Jastrzebska, B., Palczewski, K. and Chance, M. R. (2009) Structural Waters Define a Functional Channel Mediating Activation of the GPCR, Rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* 106: 14367-14372.

155. Sharp, J. S. and Tomer, K. B. (2007) Analysis of the Oxidative Damage-Induced Conformational Changes of Apo- and Holocalmodulin by Dose-Dependent Protein Oxidative Surface Mapping. *Biophys. J.* 92: 1682-1692.

156. Tong, X., Wren, J. C. and Konermann, L. (2008) γ-Ray-Mediated Oxidative Labeling for Detecting Protein Conformational Changes by Electrospray Mass Spectrometry. *Anal. Chem.* 80: 2222-2231.

157. Smedley, J. G., Sharp, J. S., Kuhn, J. F. and Tomer, K. B. (2008) Probing the pH-dependent prepore to pore transition of Bacillus anthracis protective antigen with differential oxidative protein footprinting. *Biochemistry* 47: 10694-10704.

158. Watson, C., et al. (2009) Pulsed Electron Beam Water Radiolysis for Submicrosecond Hydroxyl Radical Protein Footprinting. *Anal. Chem.* 81: 2496-2505.

159. Sharp, J. S., Becker, J. M. and Hettich, R. L. (2004) Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* 76: 672-683.

160. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

161. Aye, T. T., Low, T. Y. and Sze, S. K. (2005) Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* 77: 5814-5822.

162. Tong, X., Wren, J. C. and Konermann, L. (2007) Effects of Protein Concentration on the Extent of γ-Ray-Mediated Oxidative Labeling Studied by Electrospray Mass Spectrometry. *Anal. Chem.* 79: 6376-6382.

163. Konermann, L., Stocks, B. B. and Czarny, T. (2010) Laminar Flow Effects During Laser-Induced Oxidative Labeling For Protein Structural Studies by Mass Spectrometry. *Anal. Chem.* 82: 6667–6674.

164. Gau, B. C., Sharp, J. S., Rempel, D. L. and Gross, M. L. (2009) Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* 81: 6563–6571.

165. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I. and Gross, M. L. (2011) Fast Photochemical Oxidation of Proteins for Comparing Structures of Protein-Ligand Complexes: The Calmodulin-Peptide Model System. *Anal. Chem.* 83: 311-318.

166. Pan, Y. and Konermann, L. (2010) Membrane protein structural insights from chemical labeling and mass spectrometry. *Analyst* 135: 1191-1200.

167. Chen, J., Rempel, D. L. and Gross, M. L. (2010) Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* 132: 15502-15504.

Chapter 2 - Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry

2.1 Introduction

More than forty years after Anfinsen's seminal studies (1), the mechanisms of protein folding and misfolding remain a focal point of research (2-4). The structural characterization of partially structured intermediates remains one of the most important methods for gaining insights into folding processes. While it is possible to study some intermediates under equilibrium conditions (5-7), examination of these species in kinetic studies is a more direct approach (8). Unfortunately, experiments of this kind are complicated by the very short lifetimes (seconds to microseconds) of these conformers. In addition, many transient structures do not become strongly populated during folding (9). As a result, questions remain as to whether these elusive species are on- or off-pathway (10), whether they represent sequential or parallel pathways (11), whether they are an impediment to folding (12), or whether they speed up the conformational search of the polypeptide chain (13). Other questions are related to possible interactions of intermediates with protein cofactors (14). Transiently populated conformers have also been shown to be involved in the *un*folding of many proteins (15, 16). Optical spectroscopy in combination with rapid mixing (17) is a widely used approach for monitoring the kinetics of folding and unfolding. However, the information obtained from spectroscopic probes is typically limited to global structural features. A more

detailed characterization of intermediates has first become possible through the combination of heteronuclear NMR and pulsed hydrogen/deuterium exchange (HDX) (18, 19).

In recent years mass spectrometry-based techniques have emerged that complement many of the classical methods for studying protein structure and dynamics. For example, electrospray ionization (ESI) charge state distributions are now being routinely used for monitoring conformational changes (20-22). ESI-MS also allows folding transitions to be studied by measuring the extent of crown ether adduction to lysine side chains (23). In addition, ESI-MS provides direct information on protein-protein and protein-ligand interactions (24-27). Various laboratories have probed the structure of transient (un)folding intermediates by amide HDX pulse labeling with MS detection (28-30). Spatially-resolved information can be obtained in these experiments through the chromatographic separation of protein fragments after peptic digestion (31, 32).

Covalent labeling with MS detection represents another widely used approach (33). These experiments are based on the fact that the degree of solvent exposure modulates the reactivity of target sites on the protein. Sterically protected sites are labeled much less extensively than those that are freely accessible (34). For example, the folding mechanism of barstar has recently been studied by thiol pulse labeling of a series of Xaa → Cys constructs (35).

Considerable efforts are currently being undertaken to develop methods involving hydroxyl radicals (·OH) as a covalent probe (36). The attractiveness of ·OH for oxidative labeling stems from its small size, its non-selective reactivity, and the fact that it forms

stable covalent modifications that are retained during sample processing. Hydroxyl radicals react much faster with side chains than they do with the backbone (36), such that protein cleavage does not complicate the analysis. The extent to which individual side chains undergo oxidative labeling is determined by a combination of solvent accessibility and intrinsic reactivity (37). The sulfur-containing residues Cys and Met are most reactive, followed by the aromatic side chains Trp, Tyr, and Phe. Also His, Leu, Ile, Arg, Lys, Val, Pro, Gln, and Glu represent potential modification sites, whereas the remaining residues are less reactive (36). Although ·OH labeling proceeds through fairly complex mechanisms (38) the major pathways result in oxygen incorporation, leading to characteristic +16 Da adducts (39). Hydroxyl radicals can be generated in various ways, e.g., by using Fenton chemistry (40) or by approaches involving endogenous transition metal centers (41), γ- (42) and X-ray radiolysis of water (36), or photoinduced activation of H_2O_2 (43). Techniques involving an electrical discharge (44) or electrochemical flow cells (45) have also been described.

A particularly interesting strategy is the photolysis of H_2O_2 into ·OH radicals by a nanosecond-pulsed UV laser (46-48). Suitable radical scavengers can reduce the duration of the labeling pulse down to approximately 1μs (46). Although the technique seems ideally suited for the structural characterization of transient protein conformers, it has thus far only been applied for the structural characterization of proteins under equilibrium conditions (46-48).

Using myoglobin as a model system, this study marks the first application of laser-induced ·OH labeling for the characterization of short-lived protein conformers in kinetic experiments. Native holo-myoglobin (hMb, 17568 Da) adopts a compact globular

structure involving eight α-helices (A-H). The prosthetic heme group is bound in a hydrophobic pocket where it forms numerous noncovalent contacts with the protein. The central heme iron is coordinated on its proximal side by His93 of helix F, whereas the sixth (distal) coordination site is occupied by a water molecule that is hydrogen bonded to His64 of helix E (49). Acid-induced denaturation leads to extensive unfolding and disruption of the heme-protein contacts. Previous work has demonstrated that this conversion to apo-myoglobin (aMb, 16952 Da) is not a simple one-step process, but that considerable conformational changes precede the loss of heme (50-52). By using continuous-flow mixing in conjunction with laser-induced oxidative labeling and MS-based peptide mapping, this work provides detailed insights into the kinetic mechanism of acid-induced hMb denaturation.

2.2 Experimental

2.2.1 Materials

Equine skeletal muscle (met)myoglobin, bovine ubiquitin, glutamine and angiotensin I were obtained from Sigma (St. Louis, MO). LeuEnk and bradykinin were supplied by Bachem (King of Prussia, PA). All chemicals were used without further purification.

2.2.2 Optical Spectroscopy

UV-Vis stopped-flow experiments were carried out on a Biologic SFM 4S/Q (Molecular Kinetics, Indianapolis, IN) system operated in absorbance mode. Absorption

spectra were recorded on a Cary-100 Varian UV-Vis spectrophotometer (Palo Alto, CA). All solutions were identical to those used for oxidative labeling, except for the presence of H_2O_2.

2.2.3 Continuous-Flow Mixing and Radical Labeling

Kinetic protein unfolding experiments were performed with a custom-built three syringe continuous-flow setup (Figure 2.1), which represents a hybrid between the device described by Hambly (46, 47), and that employed in our earlier work (51). All connections were made of fused silica capillaries (TSP100170 Polymicro Technologies, Phoenix, AZ, inner diameter 100 μm), and mixing "tees" were manufactured as described previously.(51) Syringe 1 contained 40 μM native hMb, 600 mM NaCl, 60 mM glutamine and 10 mM sodium phosphate buffer at pH 6.5. This was combined with 0.2% (v/v) (*ca.* 80 mM) H_2O_2 from syringe 2 at mixer M1. Previous work has shown that the mere presence of H_2O_2 in this concentration range does not affect the structure and stability of proteins, as long as no oxidation reactions take place (53). Syringes 1 and 2 were advanced at 5 μL min^{-1} using a syringe pump (Harvard Apparatus, Boston, MA). After a 4 s pre-mixing step (9 cm distance), the contents of syringe 3 (5 mM HCl in H_2O) were added in a 1:1 ratio at mixer M2 to initiate unfolding of the protein within the reaction capillary. The pH downstream of the second mixer was 3.2 at a total flow rate of 20 μL min^{-1}. A KrF excimer laser (GAM EX 100/125, Orlando, FL) producing 18 ns pulses at 248 nm, 16 Hz, and 62.5 mJ per pulse was used to generate hydroxyl radicals through photolysis of H_2O_2 within the reaction capillary. The presence of 15 mM glutamine as a radical scavenger in the final mixture leads to quenching of the labeling
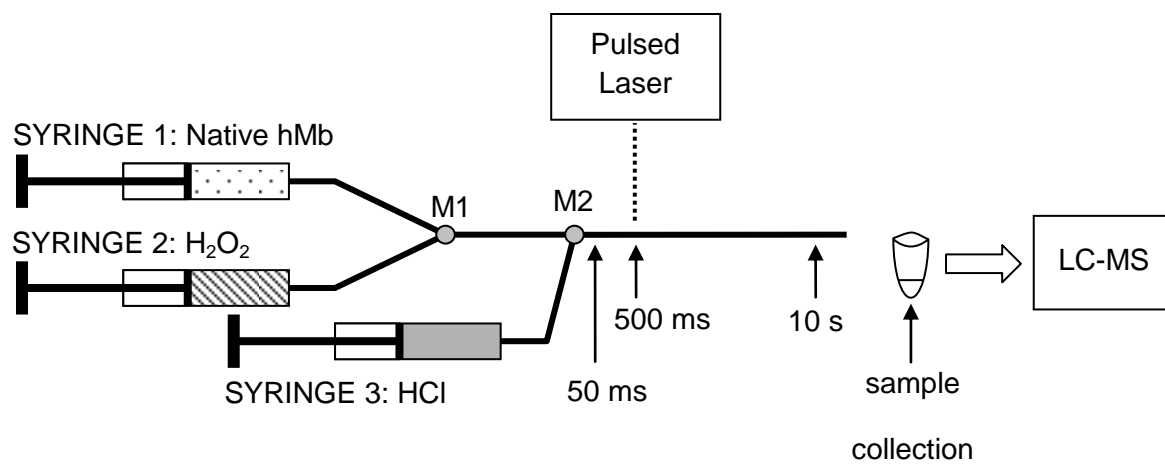
Figure 2.1. Schematic diagram of the 3-syringe continuous-flow rapid mixing setup used for all labeling experiments. Details are provided in the text.

reaction within *ca*. 1 μs (47). The laser beam was focused onto the capillary by means of a 500 mm convex lens, with a capillary-lens distance of 30 cm. UV transparent windows were created by removing the polyimide coating from the reaction capillary using a butane torch. The width of the laser spot at the capillary was determined to be *ca*. 2 mm, corresponding to a volume of about 16 nL being irradiated during each laser pulse. As previously discussed by Hambly and Gross (46, 47), artifacts caused by protein oxidative damage should not occur under these conditions due to (i) the extremely short duration of each labeling pulse, and (ii) the fact that no protein molecule is exposed to more than one pulse. Irradiation windows were located along the reaction capillary at distances downstream of mixer M2 corresponding to 50 ms (2.1 mm), 500 ms (2.1 cm), and 10 s (42 cm). The relationship between window position and reaction time is given by *distance* = (*flow rate*) x (*reaction time*) / (*cross-sectional area*). The reaction capillary was mounted on a customized sliding cart system that allowed for the reproducible switching between window positions. Samples for time zero were produced using the procedure outlined above, except that syringe 3 contained H_2O instead of HCl solution such that triggering of the unfolding reaction was avoided. The use of different solution conditions for time zero and for the kinetic samples is of potential concern, since the reactivity of hydroxyl radicals may be affected by the solvent environment (37). Control experiments were therefore carried out to compare the extent of covalent labeling for the model peptides bradykinin, LeuEnk, and angiotensin I. The oxidation levels of all three test compounds were found to be virtually indistinguishable with and without HCl (Figure 2.2), indicating that the ·OH intrinsic reactivity is very similar under both conditions. This conclusion is consistent with earlier γ-ray covalent labeling experiments
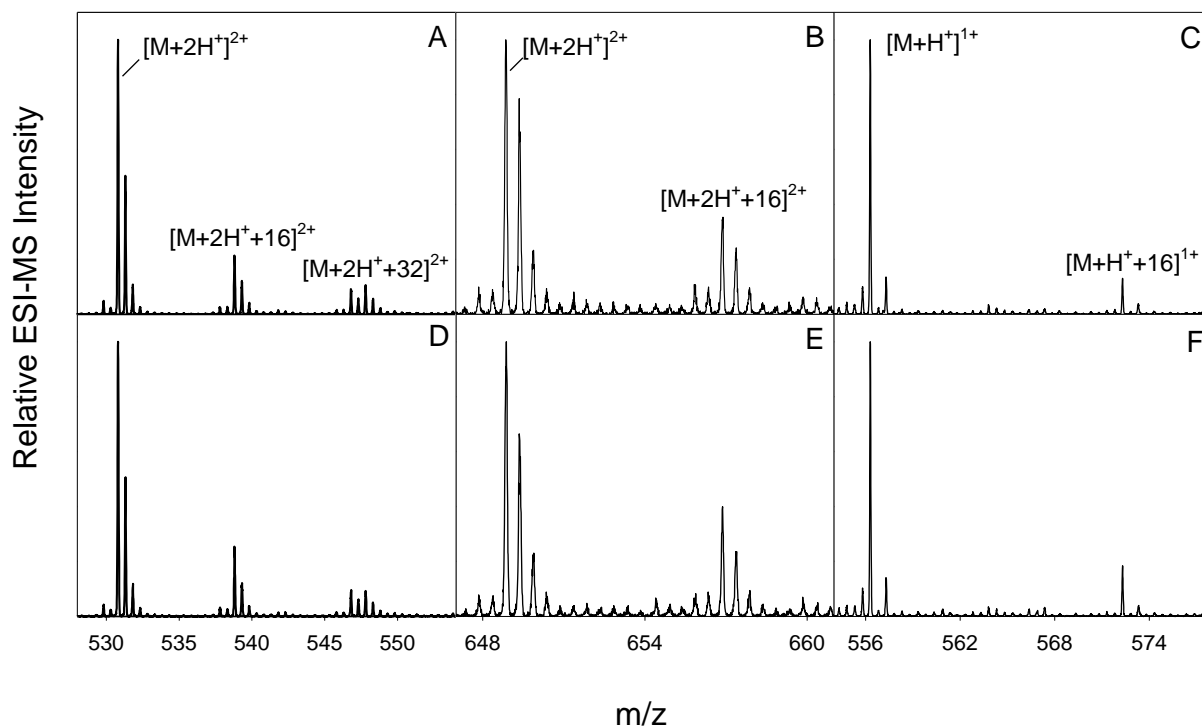
Figure 2.2. pH independence of peptide labeling. Bradykinin (A,D), angiotensin I (B,E), and leucine enkaphalin (C,F) were mixed with hydrogen peroxide in either phosphate buffer at pH 7 (A-C) or 10 mM hydrochloric acid (D-F) and exposed to laser irradiation.

(37). Based on these observations the inclusion of non-acidified time zero samples in our unfolding experiments is justified. hMb was manually mixed with HCl to characterize the endpoint of the reaction. The resulting solution was transferred into syringe 1 after 5 minutes. Syringe 3 contained glutamine and NaCl, such that the final composition of the mixture was the same as for the other kinetic time points.

The capillary outflow from a sufficient number of laser shots (typically about 250 μL) was collected in a microcentrifuge tube that contained 12 μL 170 mM phosphate buffer and 0.17 μM catalase at pH 6.8. Catalase was employed for inactivating residual H_2O_2, thereby avoiding secondary oxidation reactions (54). Part of each sample was retained for intact protein analysis, and the remainder (*ca*. 200 μL) was digested with trypsin for 24 h at 37 °C using a 1:20 (w/w) enzyme:myoglobin ratio. The digests were lyophilized and resuspended in 200 μL of an aqueous solution containing 120 μM bradykinin and 5 μM ubiquitin as internal intensity standards. These samples were flash-frozen in liquid nitrogen and stored at -80° C until further analysis.

2.2.4 LC/ESI-MS and Data Analysis

All experiments were performed on a Q-TOF Ultima API mass spectrometer (Waters, Milford, MA) equipped with a Z-spray electrospray ionization (ESI) source. Spectra were acquired in positive ion mode at a sprayer voltage of 3 kV and a desolvation temperature of 150 °C. TOF spectra were acquired in V-mode at a resolution of ca. 10,000 (fwhm). The mass spectrometer was coupled to a Waters 1525μ HPLC system employing either a C4 (Symmetry 300) 2.1 mm x 100 mm reversed-phase column for protein analysis, or a C18 (Symmetry 300) 2.1 mm x 100 mm column for peptide

analysis. Elution was carried out using a water/acetonitrile gradient in the presence of 0.05% trifluoroacetic acid at a flow rate of 100 μL min^{-1}. The identity of tryptic peptides was confirmed by tandem mass spectrometry experiments carried out in data-dependent acquisition mode. Spectra for the intact protein are presented as deconvoluted mass distributions, obtained by using the MaxEnt software supplied by the instrument manufacturer.

A number of control experiments had to be carried out in order to characterize the performance of the continuous-flow setup (Figure 2.3). Mb samples without both H_2O_2 and laser exposure showed minor satellite peaks representing trace impurity adducts, which are commonly observed in ESI-MS. The level of protein background oxidation caused by peroxide in the absence of irradiation was minimal, as can be seen by comparing the two "no laser" spectra in Figure 2.3. Laser exposure of native hMb in the presence of H_2O_2 (with water in syringe 3) leads to a series of + 16 Da adducts in the mass distributions. Included in Figure 2 is an overlay of three mass distributions obtained by irradiating the solution at the $t = 50$ ms, 500 ms, and 10 s time points downstream of mixer M2. Evidently, the three mass distributions are virtually indistinguishable, demonstrating that the kinetic time points are not affected by artifacts due to, *e .g.*, irreproducible positioning of the capillary within the laser beam, or differences in the light transmission of the three optical windows. In addition, these data confirm complete mixing downstream of M2, otherwise the oxidation levels of the three time points would be inconsistent due to differences in local protein concentration (55). Proper mixing was further confirmed by optical inspection of the tees in control measurements involving permanganate solutions and water (data not shown).
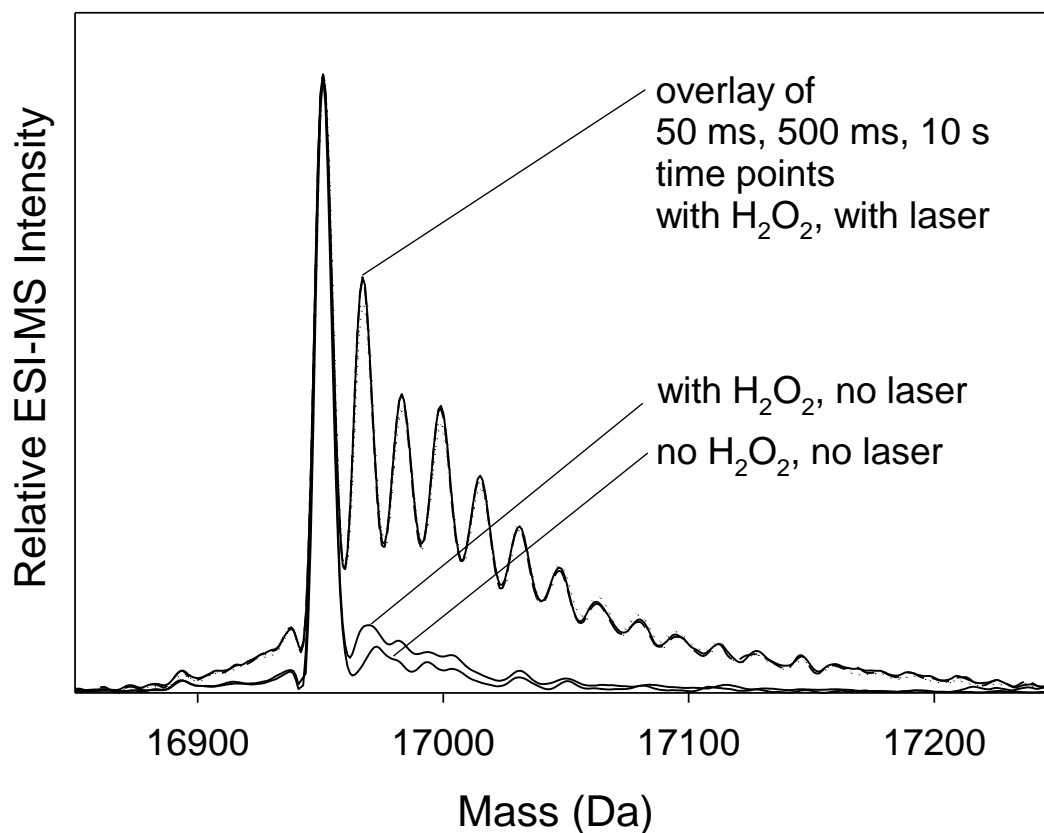
Figure 2.3. Deconvoluted ESI mass distributions of myoglobin, illustrating the results of various oxidative labeling control experiments. H_2O was used instead of HCl in syringe 3, i.e., unfolding was not triggered for any of the experiments shown here. Further details are provided in the text.

Similar to previous studies in other laboratories (36), the oxidation levels of specific protein regions were monitored by tracking the intensity of unmodified tryptic peptides as a function of time t. The reasons for choosing this strategy are outlined in the Results and Discussion section. Signal intensity values for each unmodified peptide were determined as ratio R of the peptide:bradykinin peak heights from the extracted ion chromatograms. This procedure of relating measured peptide intensities to the internal bradykinin standard compensates for possible sensitivity fluctuations of consecutive LC/MS analyses. The experimental error of the measured R values was found to be $\pm 5\%$, determined by comparing the bradykinin signal intensity to that of a second internal standard (ubiquitin) in a series of runs. A slight complication is the extremely high reactivity of the Met side chains in peptides T7 and T17, which resulted in very low signal intensities of the unoxidized species even at time zero. This behavior is consistent with the results of earlier work, which found extensive oxidation even for deeply buried Met residues (43). For the purpose of data analysis this problem can be overcome in the current study by simply considering the +16 Da versions of T7 and T17 as "unmodified" peptides.

The relative signal intensities measured for each peptide were used for calculating *normalized oxidation levels* according to

$$Normalized \;\; Oxidation \;\; Level = \frac{\ln R(t) - \ln R(0)}{\ln R(\infty) - \ln R(0)} \qquad (2\text{-}1)$$

$R(t)$ is the relative signal intensity of a particular unmodified peptide at unfolding time t, $R(0)$ is the corresponding value of the time zero control (native protein), and $R(\infty)$ is that of the final time point (5 min, after unfolding has gone to completion).

Except for the use of logarithms, Equation 2-1 mimics the framework commonly employed in HDX/MS measurements (56). Importantly, the incorporation of logarithms for the strategy used here ensures that the data analysis is independent of experimental parameters, e.g., the duration of ·OH exposure, the radical concentration, and the intrinsic reactivity of individual protein segments. Elimination of these factors removes possible ambiguities as to how they might affect the measured data. To see why these parameters cancel out we have to consider that the signal intensity R for each unmodified protein segment (equivalent to a tryptic peptide) depends on both the unfolding time t and the labeling time τ, so $R = R(t, \tau)$. During labeling the hydroxyl radical concentration decreases with a pseudo-first order rate constant a due to reactions with the glutamine scavenger and other solutes according to $[.OH](\tau) = [.OH]_0 \exp(-a\tau)$ (46). The oxidation kinetics are therefore governed by

$$\frac{dR(t,\tau)}{d\tau} = -k_{int}\alpha(t)[.OH]_0 \exp(-a\tau)R(t,\tau) \qquad (2\text{-}2)$$

where the second-order rate constant $k_{int} = \sum_{i=1}^{n} k_i$ represents the intrinsic reactivity of each segment as the result of n oxidative reaction pathways. A key element in Equation 2-2 is the conformational term $\alpha(t)$, which reflects the solvent exposure of the protein segment as a function of unfolding time, with $0 \le \alpha(t) \le 1$ (37). Complete protection from

radical attack corresponds to $\alpha(t) = 0$, whereas for full solvent exposure $\alpha(t) = 1$. Separation of variables yields

$$\int_{R_0}^{R(t,\infty)} \frac{dR(t,\tau)}{R(t,\tau)} = -k_{int}\alpha(t)[.OH]_0 \int_0^\infty \exp(-a\tau)d\tau \qquad (2\text{-}3)$$

where $R(t,\infty)$ represents the peptide intensity after oxidative labeling has gone to completion. Because $[\cdot OH]_0$ exceeds the protein concentration by at least two orders of magnitude (46), and because $t \gg a^{-1}$ Equation 2-3 can be treated under pseudo-first order conditions, resulting in

$$\ln R(t,\infty) - \ln R_0 = \frac{k_{int}}{a}\alpha(t)[.OH]_0\big(\exp(-a\,\infty) - \exp(-a\,0)\big) \qquad (2\text{-}4)$$

This is equivalent to

$$R(t) = R_0 \exp\big(-k_{int}\alpha(t)[.OH]_0 a^{-1}\big) \qquad (2\text{-}5)$$

where $R(t) = R(t,\infty)$. When substituting this expression into Equation 2-1 it is seen that

$$Normalized\ \ Oxidation\ \ Level = \frac{\alpha(t) - \alpha(0)}{\alpha(\infty) - \alpha(0)} \qquad (2\text{-}6)$$

In other words, the normalized oxidation level as defined in Equation 2-1 provides direct information on the solvent exposure of the individual protein regions. All other parameters that could potentially complicate the analysis (R_0, k_{int}, $[\cdot OH]_0$, a) have been eliminated.

2.3 Results and Discussion

2.3.1 Kinetics of Myoglobin Unfolding

Prior to studying the acid-induced denaturation of hMb by laser-induced oxidative labeling, the reaction kinetics were monitored by stopped-flow spectroscopy. Earlier studies of this process had been carried out in the absence of salts (51, 57), whereas the current work employs a relatively high ionic strength (150 mM NaCl). A pH jump from 6.5 to 3.2 under these conditions results in dramatic changes in the Soret region of the heme absorption spectrum (Figure 2.4A). The decay of the native heme-protein complex can be tracked at 409 nm, whereas accumulation of protein-free heme is observable as a slow absorption increase at 341 nm (Figure 2.4B). As reported previously (51, 57) the formation of an unfolding intermediate is clearly seen at 441 nm, where the kinetics exhibit a characteristic rise, followed by an absorption decrease. The shape of this trace reveals that the intermediate is strongly populated for reaction times around 500 ms. The data measured at 409 and 441 nm are well described by bi-exponential fits with apparent rate constants of about 10 s^{-1} and 1.4 s^{-1} (see figure caption for details). In contrast, the 341 nm trace exhibits a conspicuous "hook" at early reaction times, requiring the inclusion of a third exponential with 59 s^{-1}. This feature reveals the presence of an additional "early" intermediate, that is formed within tens of milliseconds after acid
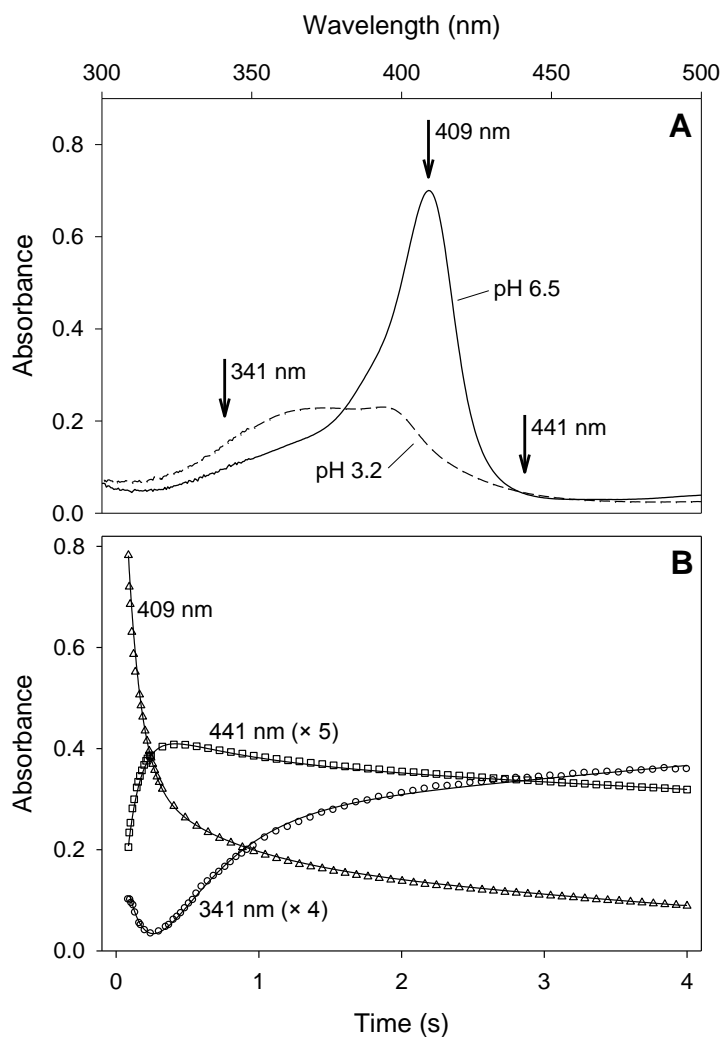
Figure 2.4. (A) Absorption spectra of native hMb at pH 6.5, and of the acid-denatured protein at pH 3.2. (B) Kinetics following a pH jump from 6.5 to 3.2 monitored by stopped-flow spectroscopy. Open symbols represent experimental data ("thinned" by a factor of 10 - 40). Data for 341 and 441 nm have been rescaled as indicated, data for 341 nm have additionally been shifted by -0.2 to reduce overlap. Solid lines are fits to the experimental data. Fitted parameters in s^{-1} are 341 nm: $k_1 = 59$, $k_2 = 8.3$, $k_3 = 1.9$; 409 nm: $k_2 = 11$; $k_3 = 1.5$; 441 nm: $k_2 = 10$; $k_3 = 1.4$.

exposure (58). In salt-free solution only the "late" (500 ms) intermediate is observable (51, 57). The presence of an additional kinetic species under the high salt conditions used here is consistent with the well-known fact that ionic strength effects can cause profound changes in the folding behavior of proteins (59).

2.3.2 Laser-Induced Oxidative Labeling

Unfolding time points were selected for oxidative labeling with the aim of characterizing the two unfolding intermediates that were detected in the stopped-flow experiments of Figure 2.4. Thus, the reaction mixture was exposed to a microsecond ·OH pulse at $t = 50$ ms or $t = 500$ ms. Data for $t = 10$ s were included for comparison, in addition to time zero and 5 min controls. Deconvoluted mass distributions of the intact protein are depicted in Figure 2.5. The relative contribution of covalent +16 Da adducts for $t = 50$ ms is much higher than for the native protein at $t = 0$, and the extent of labeling increases further with increasing reaction time. Also shown in Figure 2.5 is the fraction of unmodified protein, F_u, as a function of time. F_u is defined as A_u/A_{tot}, where A_u is the peak area corresponding to the unmodified protein, and A_{tot} is the total area of the unmodified protein plus that of all oxidation products (integrated between 16943 and 17100 Da) (36, 55). These data reveal that acid-induced unfolding goes to completion within ca. 10 s. The trends in Figure 2.5 reflect the fact that the overall solvent exposure of side chains increases as the protein unfolds, thereby enhancing the susceptibility of reactive sites to undergo oxidative labeling (37). However, very little structural information can be garnered from measurements on the intact protein. A more thorough characterization of the unfolding mechanism requires spatially resolved experiments.
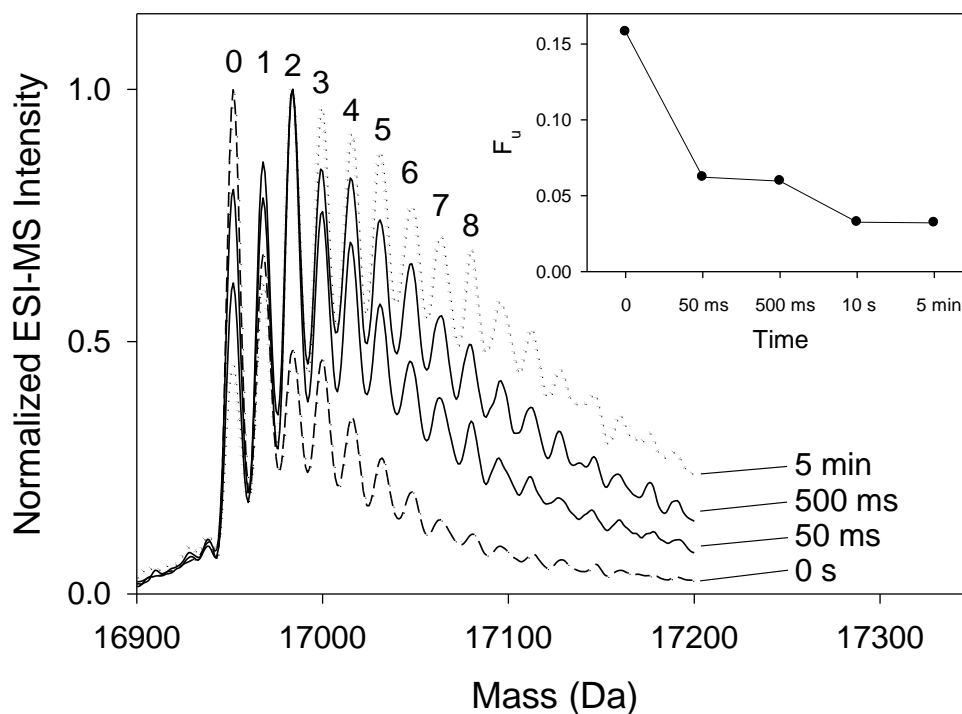
Figure 2.5. Deconvoluted ESI mass distributions obtained by exposing hMb to laser-induced oxidative labeling at different times after a pH jump from 6.5 to 3.2. Scaling was done by normalizing the most intense peak in each distribution to unity. Data for $t = 10$ s have been omitted to prevent cluttering. Notation: 0, 1, 2, ... indicate the number of incorporated oxygen atoms (+16 Da adducts). The inset shows the fraction of unmodified protein, F_u, calculated as $F_u = A_u / A_{tot}$, where A_u is the integrated area of the unmodified peak, and A_{tot} is the area of the entire mass distribution.

2.3.3 Peptide Mapping

Tryptic digestion of the protein following oxidative labeling results in a host of peptides which are labeled to various degrees. Ideally, it would be desirable to employ tandem mass spectrometry for monitoring the oxidation level at every reactive site as a function of unfolding time. Unfortunately, such a strategy is associated with considerable difficulties. "Singly" oxidized tryptic fragments represent a heterogeneous mixture of isobaric compounds, where collision-induced dissociation patterns and yields are strongly dependent on the location of oxidation sites (60). The problem is even more pronounced for peptides carrying multiple modifications. Quantification of differently labeled forms of the same peptide is difficult because the species do not co-elute due to oxidation-induced alterations in hydrophobicity, an effect that is coupled with changes in ionization efficiency (61). While most reaction paths result in the incorporation of oxygen atoms, there are also competing processes that lead to the formation of products different from the canonical +16 Da modifications (39). Tracking all of these species in a quantitative fashion is an exceedingly complex task. In addition, oxidation at Arg and Lys residues enhances the occurrence of missed cleavages, thereby modulating peptide signal intensities in ways that are difficult to predict.

In order to avoid all these complications we chose a more straightforward approach and monitored the depletion of unmodified peptides in the digests, instead of tracking the accumulation of oxidation products. Similar strategies were used in earlier ·OH footprinting work (36). For the experiments of this study peptide signal intensities *R* were measured relative to an internal standard, thereby compensating for possible sensitivity fluctuations of the LC/MS analysis procedure (see Experimental section).

Analogous to traditional HDX/MS experiments (31), the spatial resolution of the method used here is determined by the size of the hydrolysis products. Tryptic digestion of myoglobin resulted in 12 detectable peptides, covering 89% of the protein sequence (Figure 2.6). The few fragments not observed by LC/MS were small and hydrophilic, and thus likely did not adhere to the C18 column. Figure 2.7 exemplifies some of the data obtained using the described mapping strategy. Many peptides exhibit a behavior similar to T3 (Figure 2.7A), where the signal intensity decreases to less than 15% of its initial value within 500 ms. These peptides cover regions of the protein that become exposed early during the denaturation process. In contrast, T10 (Figure 2.7B) exhibits a much slower decay, which reflects a substantially higher protection from ·OH attack due to the retention of folded structure at $t = 500$ ms.

A comprehensive view of the structural changes following the pH jump is obtained by plotting normalized oxidation levels (Equation 2.1) of the 12 peptides as a function of unfolding time (Figure 2.8). Labeling differences are clearly apparent at 500 ms, where three peptides (T2, T10, and T13) exhibit oxidation levels that are considerably lower than those of the other protein regions. These three peptides also show below-average labeling at 50 ms; their behavior is highlighted by the blue boxes in Figure 2.8. It is interesting that T16 represents an additional strongly protected region at early times ($t = 50$ ms, red box in Figure 2.8). This peptide, however, is fully exposed at 500 ms. T7 exhibits unique kinetics in that it loses most of its protection immediately after onset of the reaction. The remaining peptides exhibit an intermediate behavior.

For the data depicted in Figure 2.8 all oxidation levels were calculated using the logarithmic expression of Equation 2.1. Instead, one might choose an alternative
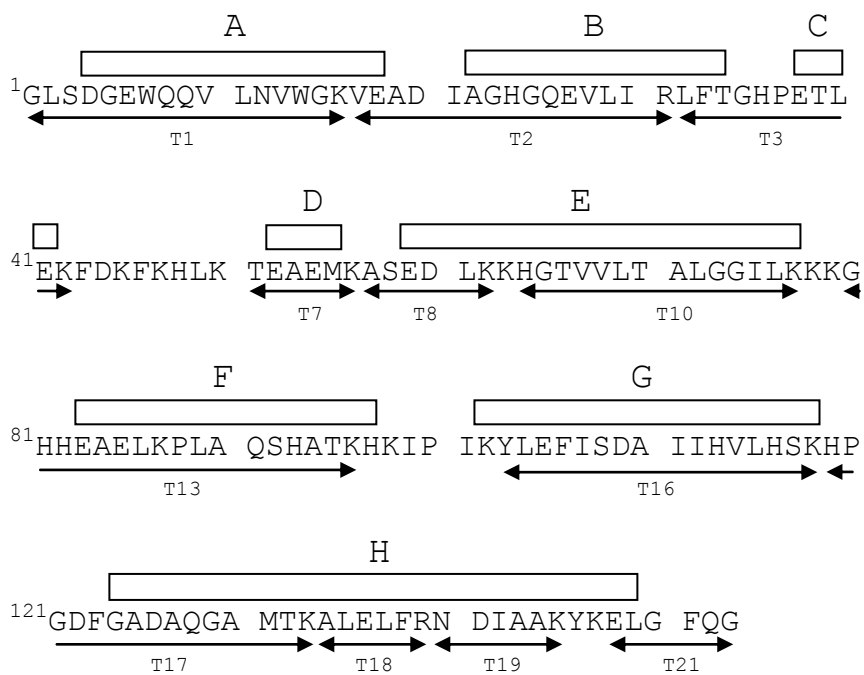
Figure 2.6. Amino acid sequence of horse myoglobin (pdb code 1WLA). α-helices are indicated as square boxes (A - H). Tryptic peptides (T1 - T21) used for data analysis are represented by double-headed arrows.
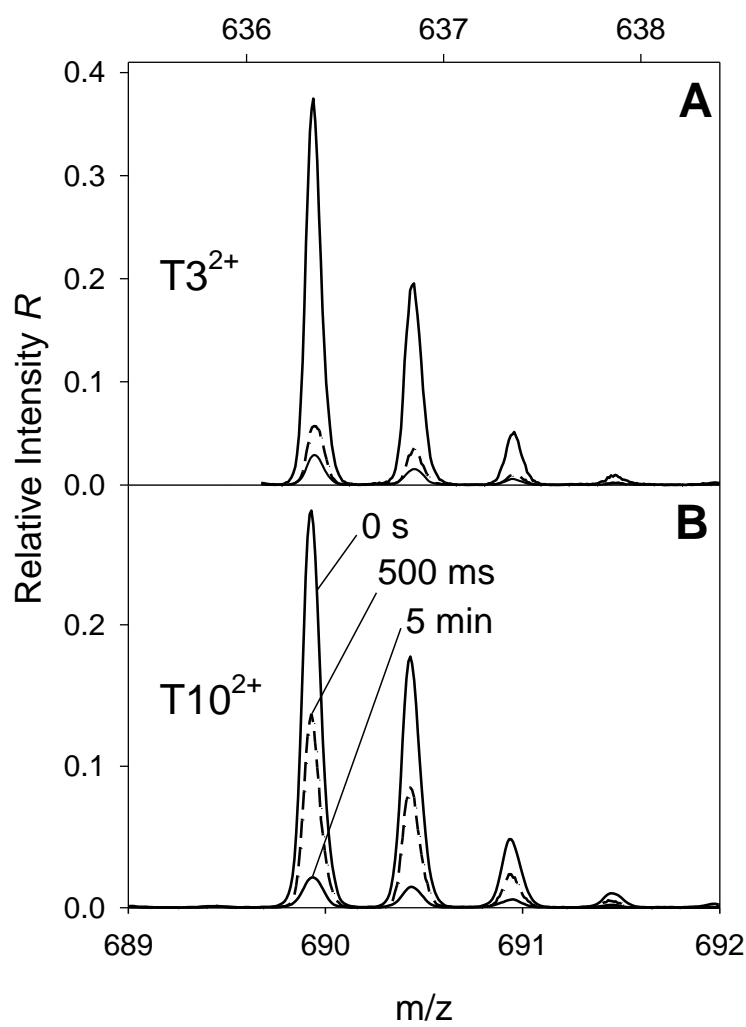
Figure 2.7. Relative intensities of the unoxidized tryptic peptides T3 (A) and T10 (B) for unfolding times of 0 s (solid line), 500 ms (dashed line), and 5 min (solid line).

approach where oxidation levels are defined more simply as $[R(t)-R(0)]$ / $[R(\infty)-R(0)]$. The temporal progressions obtained in this way are similar to those of Figure 2.8, and the conclusions reached would not change (data not shown). Nonetheless, we feel that the strategy based on Equation 2.1 is preferable because it facilitates the data interpretation in terms of solvent exposure, while other experimental factors are being eliminated (see *Experimental* section). The time profiles depicted in Figure 2.8 were found to be highly reproducible in repeat measurements.

2.3.4 Kinetic Unfolding Mechanism of hMb

The oxidation level progressions of Figure 2.8 reflect the temporal sequence of events during hMb unfolding. It is instructive to consider these data within the context of the native protein structure (Figure 2.9). For $t = 50$ ms many regions are already extensively unfolded (marked orange in Figure 2.9). However, partial protection is retained in helices B (T2, blue), E (T10, blue), F (T13, blue), and G (T16, red). Helices E and F are the main constituents of the heme binding pocket, and also G makes direct contact with the heme. Helix B is packed against E in the native protein. This BEFG core breaks down in a step-wise fashion. At $t = 500$ ms the side chains of helix G (red) are completely exposed to the solvent, whereas the BEF domain (comprising all the blue segments in Figure 2.9) still retains protection. These data are consistent with a kinetic unfolding pathway that involves an "early" and a "late" intermediate:
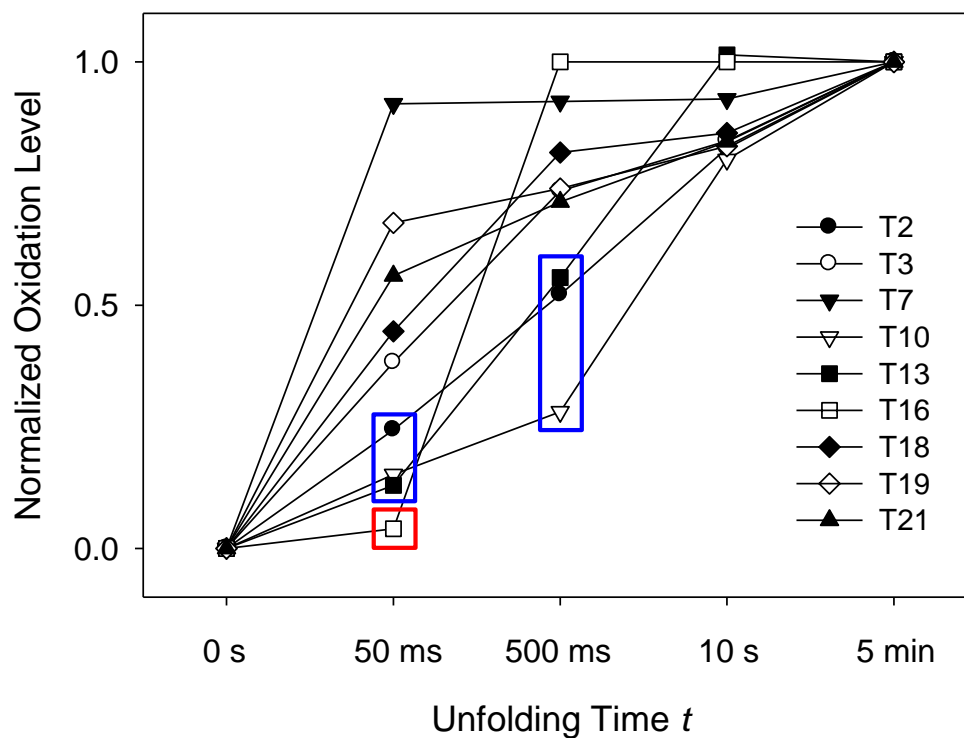
Figure 2.8. Normalized oxidation levels of tryptic peptides plotted as a function of unfolding time *t*, calculated based on Equation 1. Blue boxes highlight three peptides (T2, T10, and T13) that retain considerable protection from ·OH attack at 50 ms and 500 ms. The red box highlights T16, which is highly protected at 50 ms, but not at 500 ms. Three peptides are not shown to reduce cluttering of the Figure: T1 (with a time profile similar to T18), as well as T8 and T17 (both of which have time profiles similar to T3).
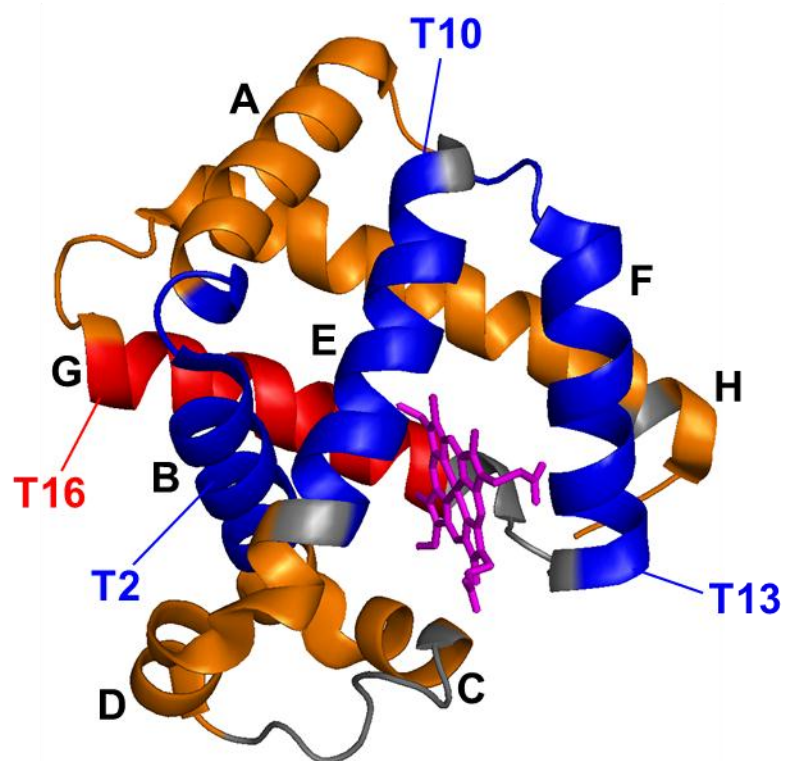
Figure 2.9. Crystal structure of native hMb (pdb code 1WLA) with its eight α-helices A-H. Regions that unfold first are shown in orange. Tryptic peptides T2, T10, and T13 represent regions that retain significant protection 500 ms after the onset of unfolding (marked in blue). T16 (red) corresponds to an additional protected region for $t = 50$ ms. The designation of red and blue segments matches that of Figure 7. Sections not covered by peptide mapping are depicted in gray. The heme group is shown in magenta.

[ABCDEFGH heme] → [aBcdEFGh heme] → [aBcdEFgh heme] → abcdefgh + heme

    native   early intermediate   late intermediate   unfolded

                  (~50 ms)   (~500 ms)
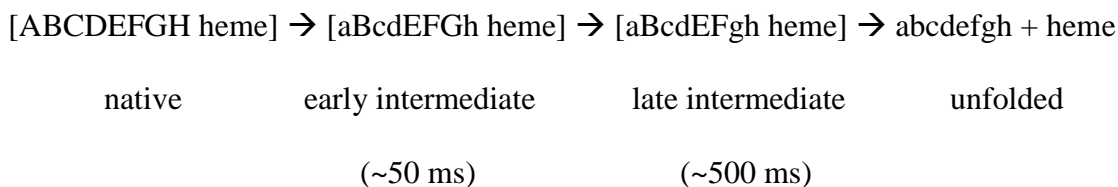
Upper case letters in this Scheme indicate partially structured α-helices, lower case letters represent unprotected regions, and square brackets stand for heme-bound states. The proposed scenario, where unfolding proceeds through two intermediates, is consistent with the stopped-flow results of Figure 2.3.

The retention of structural elements in the vicinity of the heme binding pocket strongly suggests that protein-cofactor interactions play a dominant role in determining the unfolding pathway of hMb at low pH. This view is supported by the fact that no intermediates are observed when unfolding is carried out at basic pH, where the heme is highly soluble and immediately dissociates from the protein (52). In contrast, the low solubility of heme under the acidic conditions used here (62) delays the release of the cofactor. As a result of residual porphyrin-protein interactions, regions close to the heme do not unfold as rapidly as those that are more remote. Further support for this mechanism comes from the observation that helix D (T7), representing an element that is far removed from the heme, is the first to lose all its protection (Figures 2.8, 2.9). The persistence of heme-protein interactions during the early stages of hMb unfolding had also been proposed earlier (50, 51). This work, however, provides the first direct proof of a heme-mediated core that is more resistant to acid-induced unfolding than the remainder of the protein.

2.4 Conclusions

This study demonstrates the first application of laser-induced oxidative labeling for monitoring the kinetics of a protein unfolding transition, and for characterizing the structures of short-lived protein conformers. Heme-protein interactions that are retained during the initial stages of unfolding appear to play a major role in determining the unfolding mechanism of hMb. The relatively stable BEF(G) core identified in this work for the unfolding of the holo-protein is totally different from the AGH motif that features so prominently in structural studies of aMb (7, 63). The observation that cofactors can have dramatic effects on protein energy landscapes and (un)folding mechanisms is consistent with previous reports on other systems (14).

From a more general perspective, it appears that laser-induced oxidative labeling represents a highly promising technique for mechanistic studies on protein conformational transitions. Considering the very short duration of the labeling pulse (on the order of 1 μs) (47), it should be possible to employ this method in combination with sub-millisecond mixing or photochemical triggering (64). Experiments of this kind could provide detailed structural insights into the very earliest stages of folding and unfolding. Work in this direction is currently ongoing in our laboratory and will be reported elsewhere.

2.5 References

1.   Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science* 181: 223-230.

2.   Daggett, V. and Fersht, A. (2003) The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.* 4: 497-502.

3.   Onuchic, J. N. and Wolynes, P. G. (2004) Theory of protein folding. *Curr. Opin. Struct. Biol.* 14: 70-75.

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

5.   Korzhnev, D. M., et al. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430: 586-590.

6.   Bai, Y., Sosnick, T. R., Mayne, L. and Englander, S. W. (1995) Protein Folding Intermediates: Native State Hydrogen Exchange. *Science* 269: 192-197.

7.   Hughson, F. M., Wright, P. E. and Baldwin, R. L. (1990) Structural Characterisation of a Partly Folded Apomyoglobin Intermediate. *Science* 249: 1544-1548.

8.   Gianni, S., Ivarsson, Y., Jemth, P., Brunori, M. and Travaglini-Allocatelli, C. (2007) Identification and characterization of protein folding intermediates. *Biophys. Chem.* 128: 105-113.

9.   Kato, H., Vu, N.-D., Feng, H., Zhou, Z. and Bai, Y. (2007) The Folding Pathway of T4 Lysozyme: An On-Pathway Hidden Folding Intermediate. *J. Mol. Biol.* 365: 881-891.

10.   Laurents, D. V., Bruix, M., Jamin, M. and Baldwin, R. L. (1998) A Pulse-Chase-Competition Experiment to Determine if a Folding Intermediate is On or Off-pathway: Application to Ribonuclease A. *J. Mol. Biol.* 283: 669-678.

11.   Bedard, S., Krishna, M. M. G., Mayne, L. and Englander, S. W. (2008) Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci. U. S. A.* 105: 7182-7187.

12.   Englander, S. W., et al. (1998) Fast and Slow Folding in Cytochrome c. *Acc. Chem. Res.* 31: 737-744.

13.   Wagner, C. and Kiefhaber, T. (1999) Intermediates can accelerate protein folding. *Proc. Natl. Acaed. Sci. USA* 96: 6716-6721.

14. Wittung-Stafshede, P. (2002) Role of Cofactors in Protein Folding. *Acc. Chem. Res.* 35: 201-208.

15. Sridevi, K. and Udgaonkar, J. B. (2003) Surface Expansion Is Independent of and Occurs Faster than Core Solvation during the Unfolding of Barstar. *Biochemistry* 42: 1551-1563.

16. Kiefhaber, T., Labhardt, A. M. and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375: 513-515.

17. Roder, H., Maki, K. and Cheng, H. (2006) Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem. Rev.* 106: 1836-1861.

18. Roder, H., Elöve, G. A. and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335: 700-704.

19. Udgaonkar, J. B. and Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335: 694-699.

20. Kaltashov, I. A. and Eyles, S. J. (2002) Studies of Biomolecular Conformations and Conformational Dynamics by Mass Spectrometry. *Mass Spectrom. Rev.* 21: 37-71.

21. Grandori, R., Matecko, I. and Muller, N. (2002) Uncoupled analysis of secondary and tertiary protein structure by circular dichroism and electrospray ionization mass spectrometry. *J. Mass Spectrom.* 37: 191-196.

22. Borysic, A. J. H., Radford, S. E. and Ashcroft, A. E. (2004) Co-populated Conformational Ensembles of b2-Microglobulin Uncovered Quantitatively by Electrospray Ionization Mass Spectrometry. *J. Biol. Chem.* 279: 27069-27077.

23. Liu, Z., Cheng, S., Gallie, D. R. and Julian, R. K. (2008) Exploring the Mechanism of Selective Noncovalent Adduct Protein Probing Mass Spectrometry Utilizing Site-Directed Mutagenesis To Examine Ubiquitin. *Anal. Chem.* 80: 3846-3852.

24. Heck, A. J. R. and Van den Heuvel, R. H. H. (2004) Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* 23: 368-389.

25. Smith, A. M., Jahn, T. R., Ashcroft, A. E. and Radford, S. E. (2006) Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Formation using Electrospray Ionization Mass Spectrometry. *J. Mol. Biol.* 364: 9-19.

26. Loo, J. A. (2005) in *The Encyclopedia of Mass Spectrometry*, eds M. L. Gross and R. M. Caprioli (Elsevier, Amsterdam), pp 289-299

27. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A. and Robinson, C. V. (2007) Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem. Rev.* 107: 3544-3567.

28. Heidary, D. K., Gross, L. A., Roy, M. and Jennings, P. A. (1997) Evidence for an obligatory intermediate in the folding of Interleukin-1β. *Nat. Struct. Biol.* 4: 725-731.

29. Tsui, V., et al. (1999) Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through an obligatory intermediate. *Protein Sci.* 8: 45-49.

30. Konermann, L. and Simmons, D. A. (2003) Protein Folding Kinetics and Mechanisms Studied by Pulse-Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 22: 1-26.

31. Wales, T. E. and Engen, J. R. (2006) Hydrogen Exchange Mass Spectrometry for the Analysis of Protein Dynamics. *Mass Spectrom. Rev.* 25: 158-170.

32. Chalmers, M. J., et al. (2006) Probing Protein Ligand Interactions by Automated Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* 78: 1005-1014.

33. Kaltashov, I. A. and Eyles, S. J. (2005) *Mass Spectrometry in Biophysics*. (John Wiley and Sons, Inc., Hoboken, NJ)

34. Chen, H., Schuster, M. C., Sfyroera, G., Geisbrecht, B. V. and Lambris, J. D. (2008) Solution Insights into the Structure of the Efb/C3 Complement Inhibitory Complex as Revealed by Lysine Acetylation and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 19: 55-65.

35. Jha, S. K. and Udgaonkar, J. B. (2007) Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* 282: 37479-37491.

36. Xu, G. and Chance, M. R. (2007) Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* 107: 3514-3543.

37. Tong, X., Wren, J. C. and Konermann, L. (2008) γ-Ray-Mediated Oxidative Labeling for Detecting Protein Conformational Changes by Electrospray Mass Spectrometry. *Anal. Chem.* 80: 2222-2231.

38. Garrison, W. M. (1987) Reaction mechanisms in the radiolysis of peptides,polypeptides, and proteins. *Chem. Rev.* 87: 381-398.

39. Takamoto, K. and Chance, M. R. (2006) Radiolytic Protein Footprinting with Mass Spectrometry to Probe the Strucuture of Macromolecular Complexes. *Annu. Rev. Biophys. Biomol. Struct.* 35: 251-276.

40. Sharp, J. S., Becker, J. M. and Hettich, R. L. (2003) Protein surface mapping by chemical oxidation: Structural analysis by mass spectrometry. *Anal. Biochem.* 313: 216-225.

41. Lim, J. and Vachet, R. W. (2003) Development of a Methodology Based on Metal-Catalyzed Oxidation Reactions and Mass Spectrometry To Determine the Metal Binding Sites in Copper Metalloproteins. *Anal. Chem.* 75: 1164-1172.

42. Sharp, J. S., Sullivan, D. M., Cavanagh, J. and Tomer, K. B. (2006) Measurement of Multisite Oxidation Kinetics Reveals an Active Site Conformational Change in Spo0F as a Result of Protein Oxidation. *Biochemistry* 45: 6260-6266.

43. Sharp, J. S., Becker, J. M. and Hettich, R. L. (2004) Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* 76: 672-683.

44. Wong, J. W. H., Maleknia, S. D. and Downard, K. M. (2003) Study of the Ribunuclease-S-Protein-Peptide Complex Using a Radical Probe and Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 75: 1557-1563.

45. McClintock, C., Kertesz, V. and Hettich, R. L. (2008) Development of an Electrochemical Oxidation Method for Probing Higher Order Protein Structure with Mass Spectrometry. *Anal. Chem.* 80: 3304-3317.

46. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

47. Hambly, D. M. and Gross, M. L. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myoglobin. *Int. J. Mass Spectrom.* 259: 124-129.

48. Aye, T. T., Low, T. Y. and Sze, S. K. (2005) Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* 77: 5814-5822.

49. Evans, S. V. and Brayer, G. D. (1990) High-resolution Study of the Three-dimensional Structure of Horse Heart Metmyoglobin. *J. Mol. Biol.* 213: 885-897.

50. Shen, L. L. and Hermans, J. (1972) Kinetics of Conformation Change of Sperm-Whale Myoglobin. I. Folding and Unfolding of Metmyoglobin following pH jump. *Biochemistry* 11: 1836-1841.

51. Konermann, L., Rosell, F. I., Mauk, A. G. and Douglas, D. J. (1997) Acid-Induced Denaturation of Myoglobin Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *Biochemistry* 36: 6448-6454.

52. Sogbein, O. O., Simmons, D. A. and Konermann, L. (2000) The Effects of pH on the Kinetic Reaction Mechanism of Myoglobin Unfolding Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 11: 312-319.

53. West, G. M., Tang, L. and Fitzgerald, M. C. (2008) Thermodynamic Analysis of Protein Stability and Ligand Binding Using a Chemical Modification- and Mass Spectrometry-Based Strategy. *Anal. Chem.* 80: 4175-4185.

54. Xu, G., Kiselar, J., He, Q. and Chance, M. R. (2005) Secondary Reactions and Strategies To Improve Quantitative Protein Footprinting. *Anal. Chem.* 77: 3029-3037.

55. Tong, X., Wren, J. C. and Konermann, L. (2007) Effects of Protein Concentration on the Extent of γ-Ray-Mediated Oxidative Labeling Studied by Electrospray Mass Spectrometry. *Anal. Chem.* 79: 6376-6382.

56. Smith, D. L., Deng, Y. and Zhang, Z. (1997) Probing the Noncovalent Structure of Proteins by Amide Hydrogen Exchange Mass Spectrometry. *J. Mass Spectrom.* 32: 135-146.

57. Shen, L. L. and Hermans, J. (1972) Kinetics of Conformation Change of Sperm-Whale Myoglobin. III. Folding and Unfolding of Apomyoglobin and the Suggested Overall Mechanism. *Biochemistry* 11: 1845-1849.

58. Berberan-Santos, M. N. and Martinho, J. M. G. (1990) The Integration of Kinetic Rate Equations by Matrix Methods. *J. Chem. Ed.* 67: 375-379.

59. Stellwagen, E., Olivieri, E. and Righetti, P. G. (2002) Salt-Promoted Protein Folding, Preferential Binding, or Electrostatic Screening? *Proteins: Struct. Funct. Gen.* 49: 147-153.

60. Srikanth, R., et al. (2007) Improved Sequencing of Oxidized Cysteine and Methionine Containing Peptides Using Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* 18: 1499-1506.

61. Cech, N. B. and Enke, C. G. (2000) Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.* 72: 2717-2723.

62.   Adams, P. A. (1977) The Kinetics of the Recombination Reaction between Apomyoglobin and Alkaline Haematin. *Biochem. J.* 163: 153-158.

63.   Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* 5: 148-155.

64.   Gruebele, M. (2005) in *Protein Folding Handbook*, eds J. Buchner and T. Kiefhaber (Wiley, New York), pp 454-490

Chapter 3 - Time-Dependent Changes in Side Chain Solvent Accessibility During Cytochrome *c* Folding Probed By Pulsed Oxidative Labeling and Mass Spectrometry

3.1 Introduction

Protein folding research is vital for the ongoing improvement of biomolecular structure prediction algorithms (1), and it provides valuable information regarding a number of diseases that are linked to misfolding and aggregation (2-4). Monitoring the temporal evolution of protein structure during folding is challenging because transiently populated conformers may not accumulate (5), a behavior that favors apparent two-state transitions (6). The rapid time scale of folding represents another experimental difficulty (7). Equilibrium studies sometimes allow the characterization of partially folded species that may resemble short-lived intermediates (8, 9), however, kinetic experiments remain the most direct approach for gaining insights into time-dependent structural changes. Rapid-mixing, as well as photochemical and temperature-jump events have been used as folding triggers in kinetic investigations (10). Short-lived folding intermediates have been detected and characterized using a range of spectroscopic techniques (11), and Φ-value analyses (12-14) yield information on transition states.

Cytochrome *c* (cyt *c*) serves as an important folding model system (15-18). The native protein adopts a compact globular fold and forms a hydrophobic binding pocket that accommodates the heme prosthetic group. Cyt *c* comprises three major helices (N-terminal, C-terminal, and 60's), two minor helical segments, and a short anti-parallel β-

sheet (residues 37-40 and 57-59). Other regions of the polypeptide chain are arranged in loops. The heme is covalently linked to C14 and C17 via thioether bonds, and the heme iron is axially coordinated by H18 and M80 on its proximal and distal sides, respectively (19, 20).

The presence of the heme allows aspects of the cyt c structure to be interrogated using simple spectroscopic probes (21). Soret absorption measurements provide information on the heme environment, and the iron-M80 bond gives rise to an absorption band at 695 nm (A_{695}) (22). Fluorescence quenching of W59 by the heme represents a measure for the overall protein compactness (23, 24). Heme ligands can be identified by resonance Raman spectroscopy (16, 24, 25). Other probes include circular dichroism spectroscopy for monitoring changes in helicity (26), and small-angle X-ray scattering which reports on the overall dimensions of folding intermediates (27, 28). Changes in protein compactness have also been monitored on the basis of electrospray ionization mass spectrometry (ESI-MS) charge state distributions (29, 30). Intramolecular distances can be probed by energy transfer measurements using protein constructs that carry additional chromophores (31). Pulsed hydrogen/deuterium exchange (HDX) in combination with NMR spectroscopy (9, 32-35) or ESI-MS (36, 37) represents an important tool for monitoring H-bonding patterns.

A wealth of information regarding the cyt c kinetic folding mechanism has been gathered over the years (15-18). The polypeptide chain undergoes a significant contraction during the first 100 µs (16, 28). This initial step may represent a barrier-limited (38) hydrophobic collapse with little secondary structure formation (26), or a nonspecific response to the change in solvent conditions (39). Formation of the three

major helices occurs on a time scale of several milliseconds (32, 38). Concomitantly, a "N/C intermediate" is formed that exhibits native-like hydrophobic contacts between the N- and C-terminal helices (40, 41). Although this intermediate is thought to represent an on-pathway species (42), its structure can be modulated by the solvent conditions. At neutral pH the distal (M80) binding site on the heme iron becomes coordinated by a non-native ligand, requiring the occurrence of a rate-limiting dissociation step in the $2 - 10$ s^{-1} range (43) before folding can proceed. H33 has been shown to be the main culprit, whereas H26 plays a minor role (43). Some misligation by the N-terminus may occur as well (42, 44). These misligation barriers on the folding pathway cause the N/C intermediate to accumulate, thereby facilitating its structural characterization (32). After barrier crossing, the protein folds to its native state within ~1 s (32, 34), although some molecules may still undergo transient aggregation or proline mis-isomerization (42). Misligation can be suppressed by conducting the experiments at pH ≤ 5 (45, 46) where H33 and H26 are predominantly protonated. Folding under these conditions is dramatically accelerated (46, 47). Misligation may also be suppressed through addition of free imidazole (31, 46).

Despite these previous insights, several aspects of the cyt *c* folding mechanism remain incompletely understood (48, 49). These include the significance of sub-millisecond conformational changes (39), and the question of parallel folding vs. a sequential foldon-mediated pathway (50, 51). Also, uncertainties remain regarding the number, origin, and structures of folding intermediates, as well as their degree of structural heterogeneity (31, 32). No experimental method is capable of tracking the exact conformational changes for every single residue during folding. On the contrary, most

spectroscopic techniques only provide relatively global information. Even apparently specific probes such as A_{695} can be ambiguous because of background interferences (22). HDX, while being exquisitely sensitive for monitoring the formation of H-bonds, usually yields only indirect information on tertiary interactions and solvent accessibilities (52). For these reasons, multiple probes that report on complementary structural aspects are generally required for a thorough characterization of folding mechanisms (48).

The aim of the current work is to examine the kinetic mechanism of cyt *c* folding from a hitherto unexplored perspective, i.e., by directly monitoring the solvent accessibility of individual residues in a time-dependent manner. A brief hydroxyl radical (·OH) pulse is applied at selected time points during folding, thereby inducing side chain oxidative modifications that can be probed by ESI-MS. Analogous to other covalent labeling strategies (53), these experiments are based on the premise that the oxidation susceptibility of each residue is modulated by its degree of solvent exposure, subject to a hierarchy of intrinsic reactivities (54). Numerous labeling reagents have been described (55, 56), but with few exceptions (57) the reaction rates of those species are far too slow for kinetic studies. Oxidative modifications, in contrast, are formed on a microsecond time scale when using a pulsed UV laser for ·OH production by photodissociation of H_2O_2 (58-60). ·OH can react with more than half of the twenty amino acids, such that detailed conformational data are obtainable (61-65). A considerable number of earlier studies have employed ·OH labeling for characterizing protein structures at equilibrium (58-65), but the potential of this approach for kinetic experiments remains largely untapped (Chapter 2). The present work demonstrates the use of laser-induced ·OH labeling for monitoring conformational changes of cyt *c* during pH-jump-induced

folding. The solvent accessibility data obtained in this way are complementary to information previously obtained using other experimental approaches.

3.2 Materials and Methods

3.2.1 Materials

Oxidized (ferri) horse heart cyt *c* (104 residues, 12360 Da), horse muscle myoglobin, catalase, glutamine, and bromocresol purple were obtained from Sigma (St. Louis, MO). Bradykinin was supplied by Bachem (King of Prussia, PA). All chemicals were used without further purification.

3.2.2 Optical Spectroscopy

UV-Vis stopped-flow experiments were carried out on a Biologic SFM 4S/Q (Molecular Kinetics, Indianapolis, IN) system operated in fluorescence mode, employing 2-syringe mixing in a 1:1 volume ratio. Excitation was at 280 nm, and the integrated fluorescence emission was collected for $\lambda_{em} > 320$ nm using a cut-on filter. Syringe 1 contained 20 μM cyt *c*, 30 mM glutamine, 300 mM NaCl and 10 mM HCl at pH 2. Syringe 2 contained 10 mM NaOH and 0.2 M phosphate buffer, adjusted to result in either pH 5 or pH 7 after mixing. Both syringes were advanced at 1.4 mL s^{-1}.

3.2.3 Continuous-Flow Mixing and Oxidative Labeling

Folding experiments with oxidative labeling were performed using a custom built two-syringe continuous-flow device described previously (Chapter 2) with minor

modifications. Syringe 1 contained 20 μM cyt c, 30 mM glutamine, 300 mM NaCl, and HCl at pH 2. Folding of the protein was triggered by combining this solution with NaOH and 0.2 M phosphate buffer from syringe 2 at a capillary mixer in a 1:1 volume ratio, resulting in a final pH of 7, and a protein concentration of 10 μM. Transient aggregation phenomena have previously been shown to be negligible in this concentration range.(42) Syringe 2 also contained 0.1% (v/v) (*ca*. 40 mM) H_2O_2. Both syringes were advanced simultaneously at 70 μL min^{-1} using a syringe pump (Harvard Apparatus, Boston, MA). A KrF excimer laser (GAM EX 100/125, Orlando, FL) producing 18 ns pulses at 248 nm, 92 Hz and 63 mJ pulse^{-1} was used to generate ·OH by photolysis of H_2O_2 within the reaction capillary. Glutamine acts as a radical scavenger that quenches the labeling reaction on the time scale of 1 μs (59). Oxidative labeling was performed at different time points after initiation of folding by irradiating the reaction mixture at suitable positions downstream of the mixer. On the basis of a plug flow approximation (66, 67) it can be determined that reaction times of 10 ms, 0.1 s, and 1 s correspond to distances between mixer and radiation spot of 3 mm, 3 cm, and 30 cm, respectively. Effective mixing and reproducible labeling was confirmed at all three time points using myoglobin as a test system, following a procedure outlined previously (Chapter 2). In addition, the mixer performance was verified optically in pH jump experiments on bromocresol purple (35). To characterize the endpoint of the reaction, the contents of syringe 1 were manually mixed with NaOH and phosphate buffer in a microcentrifuge tube. The resulting solution was transferred into syringe 1 after 5 minutes. Syringe 2 contained H_2O_2 in water, such that the final composition of the mixture was the same as for the earlier time points.

250 µL portions of capillary outflow were collected 3 s after mixing in a microcentrifuge tube that contained 12 µL 170 mM phosphate buffer and 0.68 µM catalase at pH 7. Catalase was employed for inactivating residual H_2O_2, thereby avoiding secondary oxidation reactions (68). Part of each sample was retained for intact protein analysis, and the remainder (*ca*. 225 µL) was digested with trypsin for 24 h at 37 °C using a 1:20 (w/w) enzyme:protein ratio. The digests were lyophilized and resuspended in 200 µL of an aqueous solution containing 20 µM bradykinin and 5 µM ubiquitin as internal intensity standards (69, Chapter 2). These samples were flash-frozen in liquid nitrogen and stored at -80 °C until further analysis.

3.2.4 LC/ESI-MS

All experiments were performed on a Q-TOF Ultima API mass spectrometer (Waters, Milford, MA) equipped with a Z-spray ESI source. Spectra were acquired in positive ion mode at a sprayer voltage of 3 kV and a desolvation temperature of 200 °C. TOF spectra were acquired in V-mode at a resolution of ca. 10,000 (fwhm). The mass spectrometer was coupled to a Waters Acquity UPLC system employing either a C4 (BEH300) 2.1 mm x 50 mm reversed-phase column for protein analysis, or a C18 (BEH300) 2.1 mm x 100 mm column for peptide analysis. Elution was carried out using a water/acetonitrile gradient in the presence of 0.1% formic acid at a flow rate of 100 µL min[-1]. The identity of tryptic peptides was confirmed by tandem mass spectrometry experiments carried out in data-dependent acquisition mode. Spectra for the intact protein are presented as deconvoluted mass distributions, obtained by using the MaxEnt software supplied by the instrument manufacturer.

Oxidation sites for each tryptic peptide were determined in off-line MS/MS experiments by directing the UPLC eluent into a TriVersa NanoMate (Advion, Ithaca, NY) chip-based ionization source. The incoming flow was split such that ~0.3% was sent directly to the mass spectrometer while the remainder was collected in 1 minute fractions in a 96-well plate. After fraction collection the plate was flash-frozen, and its contents were lyophilized. The samples in each well were then resuspended in 50 µL 2% aqueous acetic acid such that all samples could be sprayed under identical source conditions from the chip-based nanosprayer (1.6 kV, 0.3 psi).

3.2.5 Fraction Unmodified and Background Correction

Relative signal intensities of unlabeled tryptic peptides $R(t)$ were generated by dividing the UPLC/MS peak heights of each peptide by the peak height of the bradykinin internal standard for each folding time t (69, Chapter 2). Two types of $R(t)$ values have to be distinguished: (i) $R^{app}(t)$, the apparent value measured after oxidative labeling; (ii) $R^{bgr}(t)$ the corresponding value for samples that had undergone exactly the same treatment, except that the H_2O_2-containing solutions were not exposed to laser irradiation. Background corrected "fraction unmodified" values $F_u^{corr}(t)$ for each peptide can then be determined as (70)

$$F_u^{corr}(t) = \frac{R^{app}(t)}{R^{bgr}(t)} \tag{3-1}$$

3.2.6 Data Analysis

Previous work from our laboratory employed a strategy where solvent

accessibility values were reported on a relative scale, using data obtained for the unfolded and the refolded protein as reference points (Chapter 2). As discussed above, labeling of cyt c at pH 2 under the conditions used here resulted in excessive oxidation and backbone cleavage. This behavior precludes the acquisition of useful data for the unfolded protein and necessitates the use of a modified data analysis strategy. The value of $F_u(t)$ for a specific protein segment (represented by a tryptic peptide) can be modeled as (Chapter 2)

$$F_u(t) = \exp\left(-k_{int}\alpha(t)[.OH]_0\, a^{-1}\right)$$
(3-2)

where the second-order rate constant $k_{int} = \sum_{j=1}^{n} k_j$ represents the intrinsic reactivity of the segment as the result of n oxidative reaction pathways, $[\cdot OH]_0$ is the initial hydroxyl radical concentration generated by the laser pulse, and $a^{-1} \approx 1$ µs (59) is the effective labeling time. The most salient term for the interpretation of folding experiments is $\alpha(t)$ which represents the solvent accessibility (more accurately, the $\cdot OH$ accessibility) of the protein segment at folding time t. A complete lack of solvent accessibility due to burial of the segment inside the protein corresponds to $\alpha = 0$. On the other hand, segments that are completely unprotected due to extensive unfolding are characterized by $\alpha = 1$. For the current work we express the relative solvent accessibility $\alpha_{rel}(t)$ of each protein segment by comparing it to that of the refolded protein at $t = 5$ min according to

$$\alpha_{rel}(t) = \frac{\alpha(t)}{\alpha(5 \text{ min})}$$
(3-3)

On the basis of equation 3.2 it is seen that $\alpha_{rel}(t)$ can be obtained directly from the measured $F_u(t)$ values as

$$\alpha_{rel}(t) = \frac{\ln F_u^{corr}(t)}{\ln F_u^{corr}(5 \text{ min})} \tag{3-4}$$

$\alpha_{rel}(t) = 1$ applies in the case of a segment that has the same solvent accessibility as in the refolded protein, whereas $\alpha_{rel}(t) > 1$ represents an accessibility that is greater than for $t = 5$ min.

3.3 Results and Discussion

3.3.1 Background Oxidation

·OH-induced oxidative labeling gives rise to the formation of covalent +16 Da adducts that can be easily identified in the ESI mass distribution of a protein. Less abundant reaction products corresponding to other mass shifts may be formed as well (61). Ideally, protein oxidation would occur exclusively during the ~1 μs labeling pulse (59, 60), but a certain level of background oxidation is usually unavoidable. This spurious labeling has to be minimized as much as possible. Background oxidation can potentially occur (i) during isolation and storage of the protein (71), and (ii) upon contact with H_2O_2 in the reaction mixture before and after application of the laser pulse (68). Mass analysis of the protein after refolding in H_2O_2-free solution reveals that the first of these contributions is negligible for the experiments of this work (Fig. 3.1A, solid line).
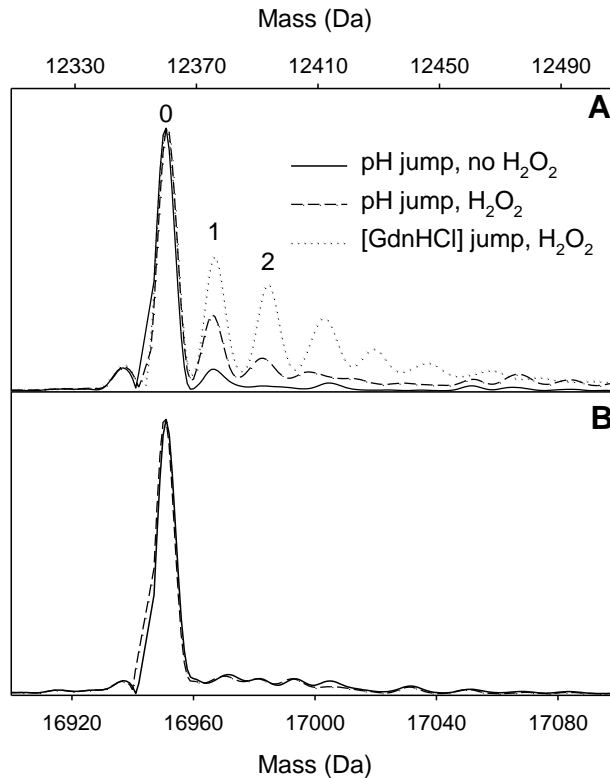
Figure 3.1. (A) Deconvoluted ESI mass distributions of cyt *c*, demonstrating the extent of background oxidation in the absence of UV laser irradiation. Solid line: pH-jump-induced folding (pH 2 → pH 7) in the absence of H_2O_2. Dashed line: pH-jump-induced folding (pH 2 → pH 7) in the presence of 0.05% (v/v) H_2O_2. Dotted line: folding triggered by a 4.2 M → 0.7 M GdnHCl jump in the presence of 0.005% (v/v) H_2O_2. (B) pH jump-induced folding as in panel A, but using myoglobin as test protein. Notation: 0, 1, 2, ... indicate the number of incorporated oxygen atoms (+16 Da adducts).

Most previous cyt *c* studies employed guanidinium hydrochloride (GdnHCl) as denaturant. Initial test experiments for this work therefore used a GdnHCl concentration jump as folding trigger, without laser irradiation. Unfortunately, background oxidation under these conditions is very significant (Fig. 3.1A, dotted line). This behavior is attributed to the well-known peroxidase activity of cyt *c* that causes ·OH-mediated self-oxidation in the presence of H_2O_2, without requiring UV photons (72, 73). Previous studies have demonstrated that peroxidase activity is enhanced in the presence of GdnHCl (74, 75). In an effort to reduce background oxidation, we next tested the use of a pH jump from 2 to 7 as folding trigger (24). Undesired oxidation under these GdnHCl-free conditions is greatly suppressed, despite the use of a tenfold higher H_2O_2 concentration (Fig. 3.1A, dashed line). A low level of peroxidase-mediated cyt *c* background labeling persists even in the pH-jump experiments, evident from a comparison with a myoglobin control for which virtually no oxidation is observed (Fig. 3.1B). These myoglobin data also confirm the absence of artifactual protein oxidation during storage of the flash-frozen samples (76). Myoglobin does not possess peroxidase activity under the conditions used here due to heme precipitation at pH 2 (77). The findings of Fig. 1A prompted us to base the cyt *c* refolding experiments of this work on a pH jump protocol, rather than using a change in [GdnHCl] as folding trigger. The remaining low-level inadvertent oxidation was dealt with by using a background correction protocol (Methods Section, eq. 3-1) (70).

3.3.2 Stopped-Flow Kinetics

Prior to discussing the results of oxidative labeling, it is instructive to examine the pH jump behavior of cyt c by fluorescence spectroscopy. As reported previously (78, 79), equilibrium unfolding occurs with a transition midpoint around pH 2.5 (Fig. 3.2A). At pH 2, the protein has been shown to be extensively disordered (80). The greatly increased heme-W59 distance in acid-unfolded cyt c disrupts energy transfer quenching, thereby forming the basis for the fluorescence intensity changes seen in Fig. 3.2 (23, 38). Both H18 and M80 become displaced from their ligation sites at pH 2 (80), and the heme iron adopts a 5-coordinate form with water as the fifth ligand (16).

A pH jump from 2 to 5 causes a ~85% burst in W59 fluorescence, followed by a decay that is well described by apparent rate constants on the order of 100 s^{-1}, 10 s^{-1}, and 1 s^{-1} (see Fig. 3.2B caption for details). Considerably slower folding kinetics with a reduced burst amplitude of ~75% are observed for a pH 2 $\rightarrow$ pH 7 jump (Fig. 3.2B). This difference manifests itself in a *ca.* three-fold amplitude increase for the 10 s^{-1} and 1 s^{-1} phases. The slower folding kinetics for the pH 2 $\rightarrow$ pH 7 jump can be attributed to heme misligation that causes the N/C intermediate to accumulate, as discussed in the Introduction (32, 34, 40-47). The cyt c behavior displayed here is very similar to earlier kinetic data (24, 34), and the purpose of highlighting these stopped-flow data here is to emphasize the consistency of our pH-jump conditions with earlier GdnHCl results. In the following, we will focus on folding of the initially acid-unfolded protein at pH 7.
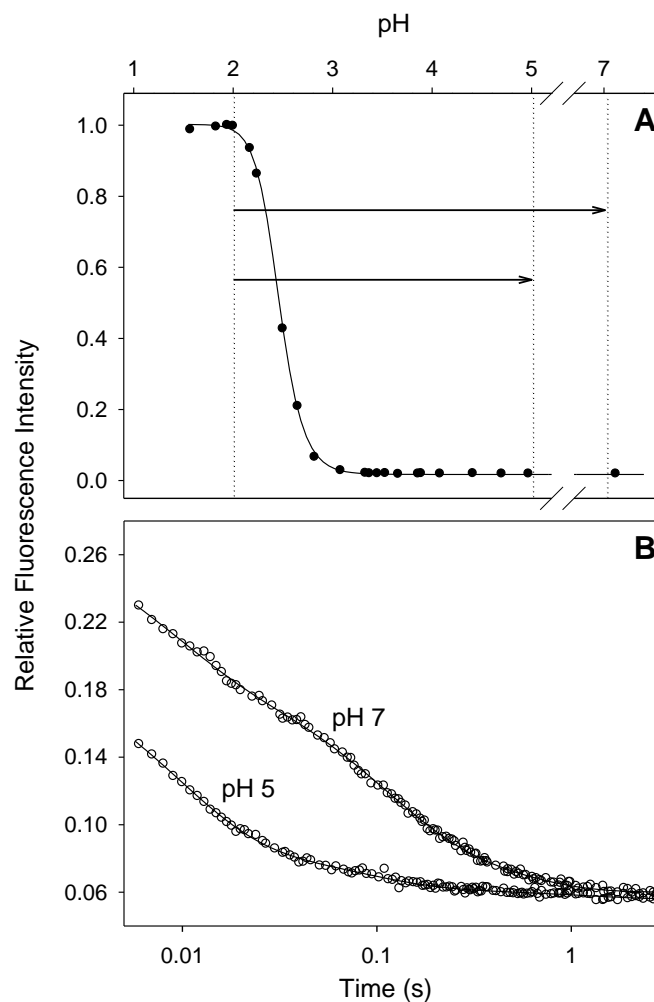
Figure 3.2. (A) Acid-induced equilibrium unfolding of cyt *c* monitored by tryptophan fluorescence spectroscopy. Vertical dotted lines indicate initial and final conditions of kinetic pH jump measurements. (B) Folding kinetics following a pH jump from 2 to 7, and from 2 to 5, monitored by stopped-flow spectroscopy. The data were normalized to the fluorescence intensity of unfolded cyt *c* at pH 2. Solid lines are tri-exponential fits to the experimental data. Apparent rate constants (and amplitudes) at pH 7 are: $k_1 = 130$ s^{-1} (0.086), $k_2 = 10$ s^{-1} (0.098), $k_3 = 2.3$ s^{-1} (0.038). At pH 5: $k_1 = 111$ s^{-1} (0.12), $k_2 = 13$ s^{-1} (0.028), $k_3 = 0.91$ s^{-1} (0.0071).

3.3.3 Laser-Induced Oxidative Labeling

Cyt *c* was exposed to laser-induced oxidative labeling at different time points (10 ms, 0.1 s, and 1 s) following a pH 2 → pH 7 jump, using a custom-built continuous-flow mixing device. Deconvoluted mass distributions of the intact protein are depicted in Fig. 3.3. Data for a 5 min native control are included as well. The extent of covalent labeling is highest for 10 ms, and decreases with increasing reaction time. These trends reflect the overall decrease in solvent accessibility as the protein folds, thereby reducing the susceptibility of side chains to undergo labeling (81).

Unfortunately, it was not possible to obtain satisfactory reference data for the unfolded initial state at pH 2. Labeling under these conditions causes severe oxidation, resulting in mass distributions with poor signal-to-noise ratio (not shown). In comparison, the 10 ms time point in Fig. 3.3 represents a significantly lower oxidation level. This behavior attests to the substantial collapse of the polypeptide chain early during folding, which is also evident from the burst in fluorescence quenching (Fig. 3.2B) (16, 28). In addition, some of the protection at 10 ms may be a remnant of residual structure in the unfolded state (82). Based on the extensive oxidation of the protein at pH 2, however, it appears that acid-denatured cyt *c* is quite disordered (80), implying that the relative contribution of compact conformers in the pH 2 ensemble is low.

3.3.4 Peptide Mapping and Tandem Mass Spectrometry

The data of Fig. 3.3 reveal the occurrence of dramatic time-dependent changes in global protein structure. A more detailed characterization of the cyt *c* folding mechanism requires tracking the behavior of individual residues and protein segments. Prerequisite
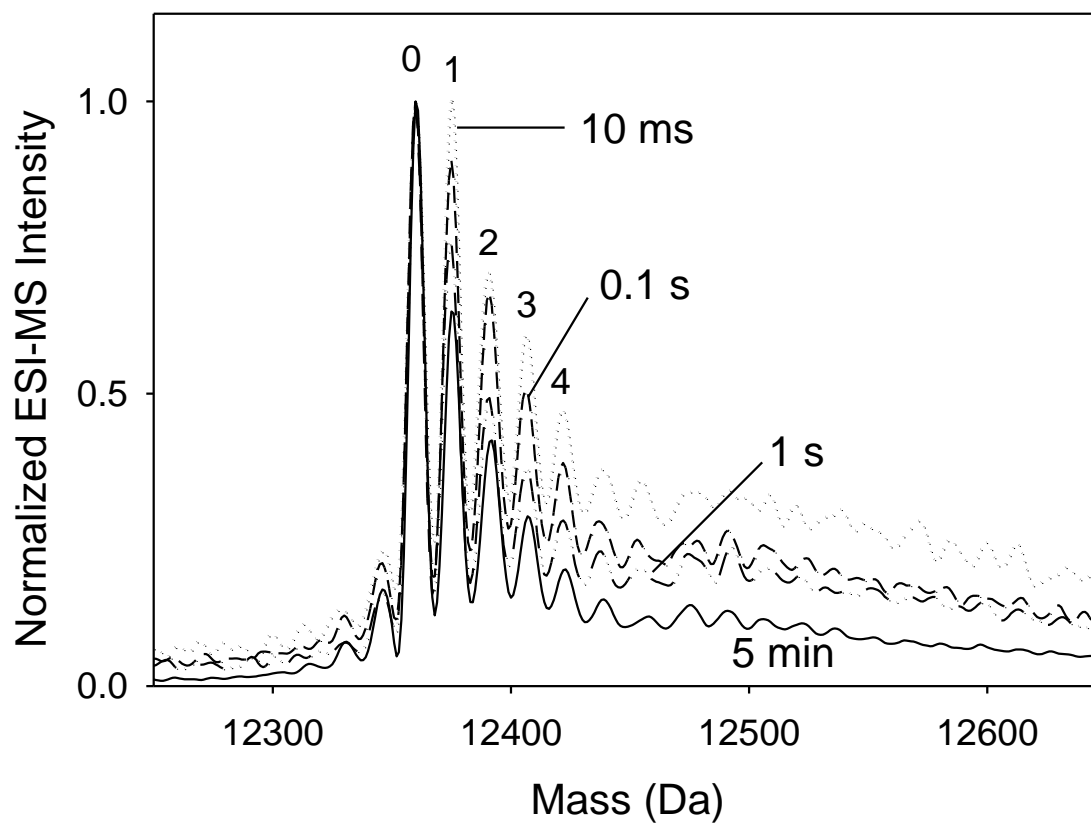
Figure 3.3. Deconvoluted ESI mass distributions of intact cyt *c* obtained by laser-induced oxidative labeling at different time points following a jump from pH 2 to pH 7. Scaling was done by normalizing the unoxidized peak in each distribution to unity.

for this strategy is the identification of modification sites by peptide mapping and tandem MS (MS/MS). Tryptic digestion of cyt c resulted in 11 detectable peptides that cover 84% of the protein sequence (Fig. 3.4). The few unobservable fragments are small and hydrophilic, and thus likely do not adhere to the C18 column.

Typical MS/MS data of tryptic peptides after labeling at 10 ms are exemplified in Fig. 3.5. Collision-induced dissociation (83) of (T20 + 16) results in unmodified y_1 and y_2 ion signals. Fragments y_3 to y_6 are shifted by 16 mass units, thereby pinpointing Y97 as the oxidation site in this peptide (Fig. 3.5A). The situation is somewhat more complex for the heme containing species T5 (Fig. 3.5B). Observation of a (b4 + heme + 16) fragment, along with unmodified free heme reveals incorporation of an oxygen atom in the range of residues 14 to 17. Based on the intrinsically high reactivity of thioether groups, these data strongly suggest the occurrence of oxidative modifications at C14 and/or C17, rather than at the fairly nonreactive A15 and Q16 (54). T8 appears with a +5 Da modification which is typical for ·OH-induced opening of an imidazole ring (61). MS/MS confirms H33 to be the site of modification in this peptide (Fig. 3.5C). Other oxidation sites were identified in a similar fashion, namely F46 and Y48 (T10), W59 (T12), M65 and Y67 (T13), Y74 (T15), M80, as well as I81 (T16). These results reflect the high intrinsic reactivities of aromatic and sulfur-containing side chains (61). Many of the oxidation sites found here have previously been identified in γ-radiolysis experiments on native cyt c (84). Peptides T1 (GDVEK) and T22 (ATNE) did not provide unambiguous results, preventing the determination of oxidation sites in these two segments. Low level oxidation at residues other than those listed above cannot be excluded.
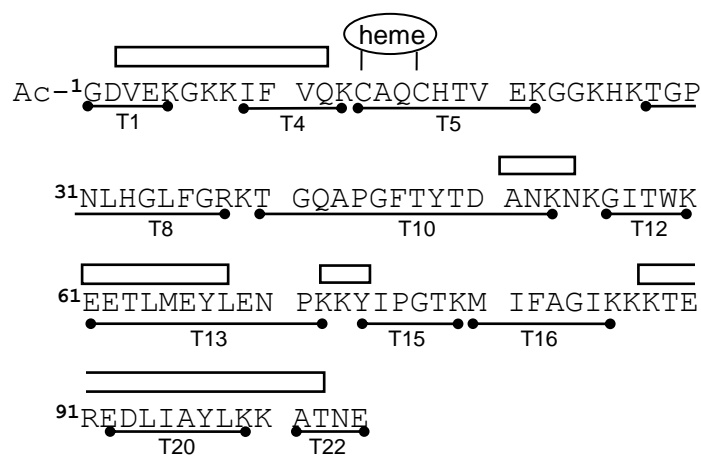
Figure 3.4. Amino acid sequence of horse cyt *c* (19). Rectangles indicate helical regions. Also shown are tryptic peptides (T1, T4, ...) used for the analyses of this work. The N-terminal acetyl group is denoted as Ac.
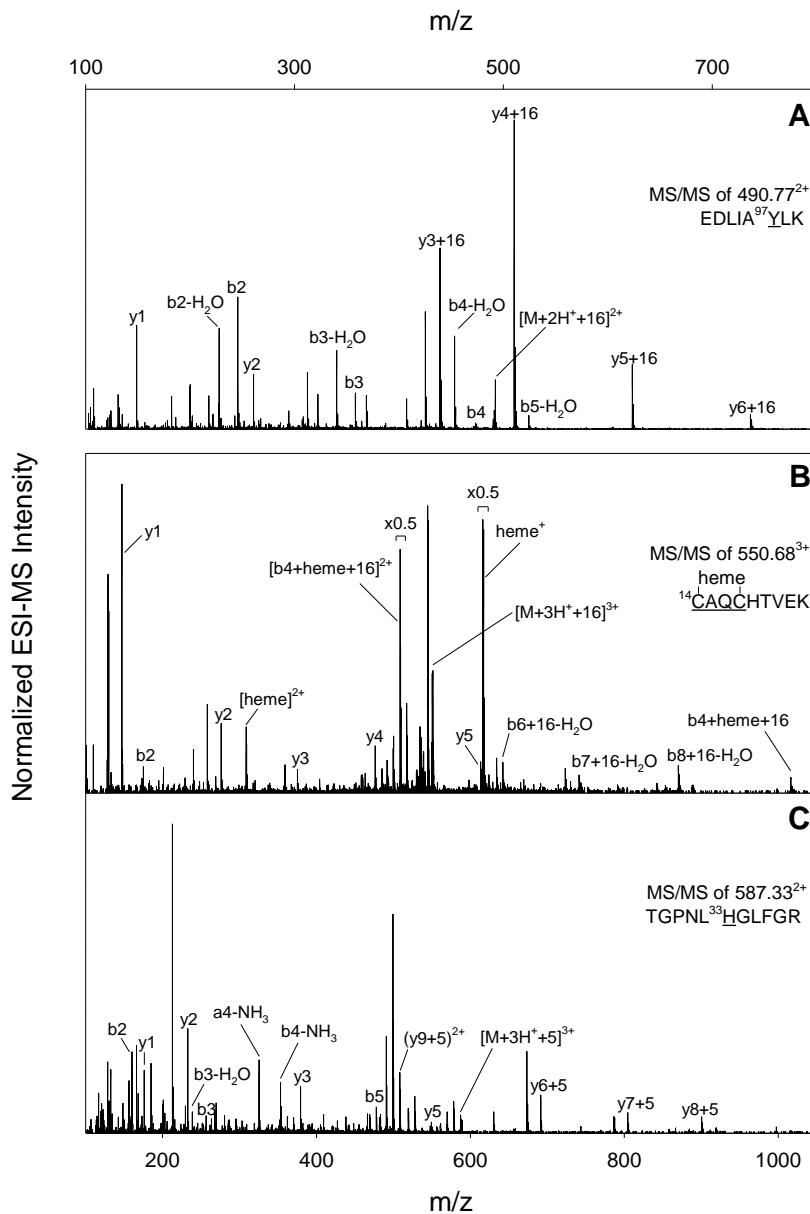
Figure 3.5. Tandem mass spectra of selected oxidatively modified peptides. (A) MS/MS of [T20 + 16]; (B) MS/MS of [T5 + 16]; (C) MS/MS of [T8 + 5]. Peaks are labeled using standard *b* and *y* ion notation. Also indicated in each panel is the peptide sequence, oxidation sites are underlined.

3.3.5 Relative Solvent Accessibilities

It is customary to report the degree of oxidation in ·OH labeling experiments as *fraction unmodified* (F_u) (61, 81). An F_u value of unity represents a complete lack of oxidation, whereas in the case of extensive labeling $F_u \ll 1$. The current work employs the notation F_u^{corr} to emphasize that the reported data have been corrected for background oxidation. Plots of F_u^{corr} *vs.* folding time t for all eleven tryptic peptides reveal large differences (Fig. 3.6A). It is tempting to immediately contemplate the implications of these data in a structural context, but prior to taking this step a few additional considerations are required.

The oxidation behavior of each segment (represented by a tryptic peptide) is governed by a factor $\alpha(t)$ that reflects its solvent accessibility at folding time t, and a term k_{int} describing its intrinsic reactivity. Accordingly, the rate constant k_{ox} of a segment for reaction with ·OH may be expressed as $k_{ox} = \alpha(t) \times k_{int}$ (61, 81, 85). The value of $\alpha(t)$ can range from unity for full solvent exposure, down to $\alpha(t) = 0$ for regions that do not undergo oxidation as a result of complete steric shielding. The rate constant k_{int} of each segment is determined by its amino acid composition. The value of this rate constant reflects the reactivity in the hypothetical case of complete exposure, i.e., for $\alpha(t) = 1$. For an *n*-residue segment k_{int} may be approximated as $k_{\text{int}} = \sum_{j=1}^{n} k_j$, where k_j are the rate constants of individual residues for reaction with ·OH.

Values of k_{int}, estimated on the basis of published free residue data (54, 61), are plotted for each tryptic peptide in Fig. 3.6B. Not surprisingly, the peptide with the largest k_{int} (T5) exhibits the lowest F_u^{corr} values in Fig. 3.6A. This behavior highlights the fact that the different intrinsic reactivities of the various peptides can be an impediment for
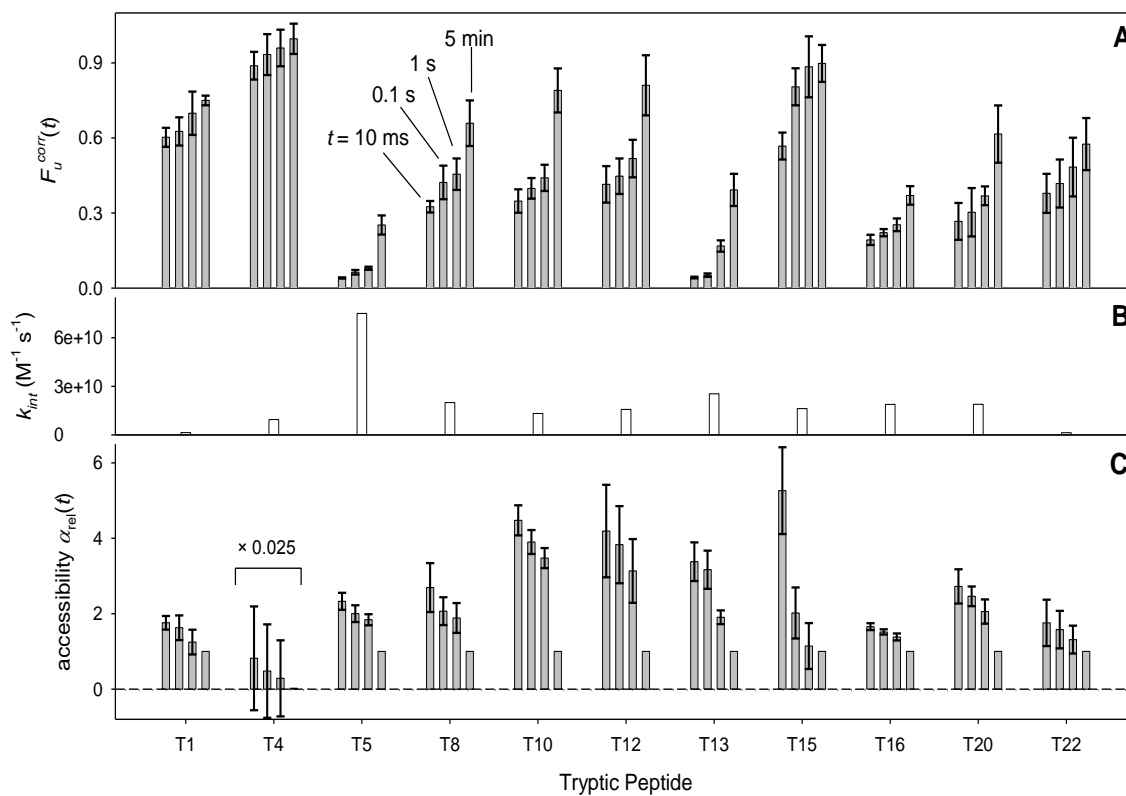
Figure 3.6. (A) Oxidation behavior of tryptic peptides during folding at pH 7, expressed as fraction unmodified (F_u). Groups of bars represent data obtained for folding times of t = 10 ms, 0.1 s, 1 s, and 5 min. (B) Intrinsic reactivity k_{int} for each peptide, estimated as indicated in the text. (C) Solvent accessibility $\alpha_{rel}(t)$ of individual peptides, relative to the refolded protein at t = 5 min. Data points were obtained as an average of three independent experiments, each of them with its own background correction. Error bars indicate standard deviations.

the data interpretation, unless they are being dealt with properly. We emphasize that the k_{int} data of Fig. 3.6B represent an approximation because they are based on k_j values that will be somewhat different in a polypeptide context from those tabulated for the free residues (54, 61). To circumvent possible ambiguities the current work does *not* rely on published k_j values for extracting structural information. Instead, the $F_u^{corr}(t)$ progressions of Fig. 3.6A are used for calculating solvent accessibilities $\alpha_{rel}(t)$ relative to the 5 min native control (Methods Section, eqs. 3-3, 3-4). This strategy ensures that the k_{int} values cancel out, such that the labeling behavior of the different peptides can be compared on a uniform scale. After this transformation, $\alpha_{rel}(t) = 1$ represents the case where a segment exhibits the same solvent accessibility as in the native protein, and $\alpha_{rel}(t) > 1$ for segments that are more accessible than in the native state.

Relative accessibilities $\alpha_{rel}(t)$ for all tryptic peptides are depicted in Fig. 3.6C. In every single case α_{rel} is highest for the 10 ms time point, and then decreases to unity for $t = 5$ min. Yet, considerable differences are apparent for the temporal profiles of the individual segments. Early during folding the relative accessibilities of regions located towards the center of the protein sequence (T10 - T15) are severalfold higher than in the native state. In contrast, N- and C-terminal regions exhibit fairly low α_{rel} values for all time points studied. The calculated error of $\alpha_{rel}(t)$ becomes unacceptably large for peptides that have $F_u^{corr}(5$ min$)$ very close to unity. This behavior is caused by a mathematical singularity during the logarithmic normalization of eq. 3-4 at $\ln(1) = 0$. Fortunately, for the data set considered here this problem is encountered only for a single peptide (T4).

3.3.6 Conformational Changes during Folding

For interpreting the measured solvent accessibility data it is convenient to map the $\alpha_{rel}(t)$ values of Fig. 3.6C to the crystal structure of native cyt c (19). The resulting pictorial representations (Fig. 3.7) take into account the fact that the $\alpha_{rel}(t)$ progressions are dominated by the behavior of relatively few oxidizable residues, most of which were identified by MS/MS (Fig. 3.5). The side chains of these structural probes are highlighted in Fig. 3.7 using a color scheme that ranges from highly accessible (red, $\alpha_{rel}(t) > 3.5$) to native-like (blue, $\alpha_{rel}(t) < 2$). Only T1 and T22 have been colored in their entirety because the oxidation sites for these two segments could not be uncovered.

Previous work has demonstrated the accumulation of an N/C intermediate during the first few milliseconds of cyt c folding at neutral pH (34, 36, 40, 41), which then remains the dominant kinetic species up to at least 100 ms (32). Consistent with those earlier findings, oxidative labeling at 10 ms (Fig. 3.7A) and 100 ms (Fig. 3.7B) reveals very low solvent accessibilities for T1 and T22 as marked by the blue color, suggesting a stabilization of the N- and C-helices by mutual docking. Regions that are remote from the chain termini retain extensive solvent exposure during this time, as seen by the very high α_{rel} values of F46, Y48, and Y74 (red in Fig. 3.7A). Interestingly, W59 also remains highly accessible up to at least 100 ms (red in Figs. 3.7A, B) despite the *ca.* 85% drop in fluorescence intensity (Fig. 3.2B) which implies a native-like heme-W59 orientation and distance. The persistence of a high solvent accessibility for W59 under these conditions suggests the presence of major packing defects in the collapsed core (26) of the N/C intermediate.
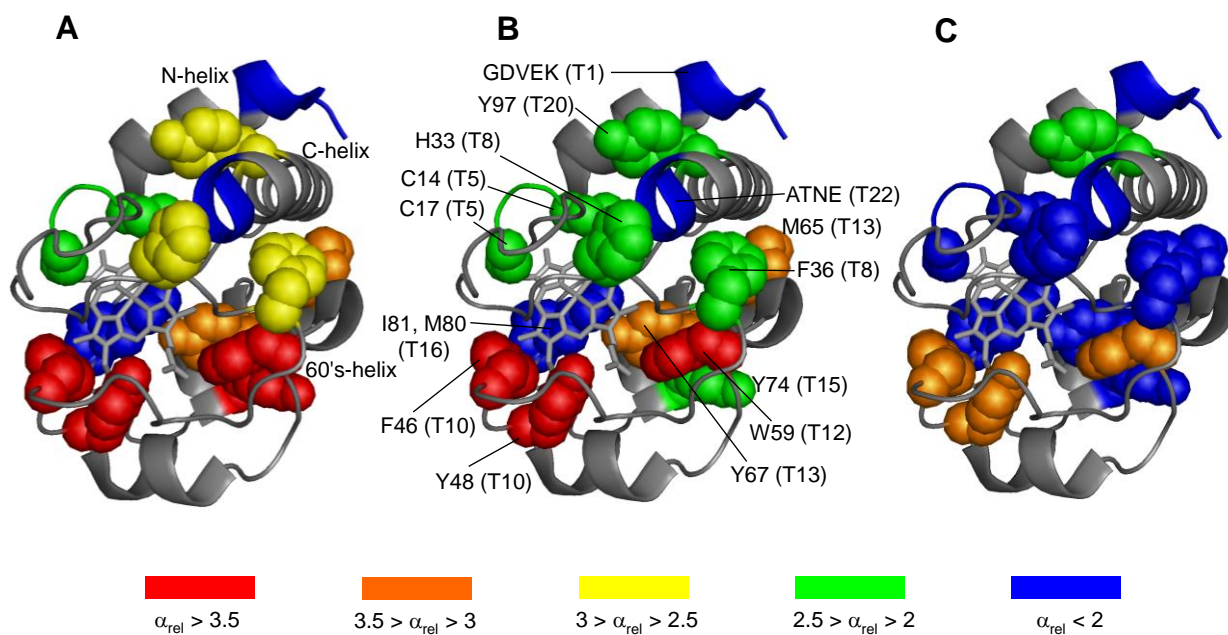
Figure 3.7. Structural changes of cyt *c* during refolding at 10 ms (A), 0.1 s (B), and 1 s (C) after a pH jump from 2 to 7. Major oxidation sites are color-coded according to the relative solvent accessibility $\alpha_{rel}(t)$ of the corresponding peptides (based on the data in Fig. 6C), using the color scheme outlined along the bottom of the figure.

The interactions between the N- and C-helices in native cyt c are mediated by an intricate network of hydrophobic contacts (19, 40). Earlier studies have not unequivocally determined whether all aspects of this network become fully developed already at the N/C intermediate stage. Y97 represents a key residue close to the center of the helix interface. Fig. 3.7A shows this residue in yellow, highlighting the fact that its solvent accessibility at 10 ms is significantly higher than in the native state. Inspection of Figs. 3.7B, C reveals that Y97 retains an elevated (green) α_{rel} value up to at least 1 s. These observations point to the occurrence of structural reorganization events at the N/C helix interface throughout much of the folding process.

Previous work (32, 34, 42-44) has shown that during the first ~ 100 ms of folding H33 is the dominant iron ligand on the distal side of the heme, where it occupies the position held by M80 in the native protein. The occurrence of this misligation implies that H33 has to reach around to the distal face of the porphyrin plane. Interestingly, this conformation is associated with very low accessibilities for both the displaced M80 ligand and the adjacent I81 (blue in Fig. 3.7A, B). We attribute the considerable protection of these residues to side chain crowding and concomitant exclusion of solvent on the distal side of the heme, caused by encroachment of the non-native ligand. H33 remains fairly accessible at 10 ms (yellow in Fig. 3.7A). This pattern is consistent with a structural arrangement where one face of the imidazole ring pushes against M80 and I81, while the other face remains solvent accessible. As the transient H33-iron bond gets disrupted, the imidazole ring can move to its native position where it interacts with the free terminus of the C-helix. It appears that that this move facilitates the gradual consolidation of the N/C helix interface. This interpretation is supported by similarities in

the accessibility progressions of H33 and Y97, both of which switch from yellow to green on the same time scale (Fig. 3.7A, B).

Earlier studies proposed that cyt *c* folds along a foldon-mediated pathway (50, 51), where the initial formation of N/C helical interactions (foldon 1) is followed by sequential open-to-closed transitions of four additional foldons, until ultimately the native state has been reached. For the protein orientation used in Fig. 3.7 this stepwise structural consolidation is thought to occur in a "top to bottom" fashion, where regions close to the top (N and C-helices) fold first and regions close to the bottom fold last (50, 51). Overall, the solvent accessibility patterns uncovered in this work are in agreement with those earlier proposals. Specifically, each of the three time points displayed in Fig. 3.7 exhibits a $\alpha_{rel}(t)$ polarity, where residues with low accessibilities tend to be located towards the top of the figure, and highly exposed side chains closer to the bottom. The one exception to this trend is the low solvent accessibility of M80 and I81, but as discussed this phenomenon is attributed to distal crowding caused by H33 heme misligation.

Another contributing factor to the polarity in solvent accessibility could be the positioning of the protein's four proline residues (P30, P44, P71, and P76). Proline isomerization can be rate-determining during folding (42, 86, 87), and it is interesting to note that all four Pro residues are located in the "bottom" part of the native cyt *c* structure, in close spatial proximity to residues that retain the highest solvent accessibility at 1 s (orange in Fig. 3.7C).

3.4 Conclusions

This work marks the first application of laser-induced oxidative labeling in combination with rapid mixing for exploring the kinetic mechanism of protein folding. An earlier pulsed ·OH labeling study from our group focused on protein denaturation, did not involve MS/MS mapping of oxidation sites, and employed a continuous-flow device with a lower time resolution (Chapter 2).

Oxidative labeling provides information on the solvent accessibility of individual residues, reflecting the degree of side chain packing and burial. The insights obtained in this way are complementary to those from stopped-flow spectroscopy and pulsed HDX, the latter being the method of choice for probing H-bonding patterns during folding (9, 32-35). Some information on the solvent accessibility of individual foldons can be gained from *m*-values measured by native state HDX (51, 88). In contrast, the approach used here directly probes the degree of exposure for specific side chains under non-equilibrium conditions. Pulsed ·OH labeling can be conducted at neutral pH, whereas pulsed HDX generally requires non-physiological (highly basic) conditions which may lead to experimental ambiguities as pointed out by others (32, 89). Structural artifacts caused by covalent modifications of the protein are negligible due to the rapid time scale of the oxidation reactions (59, 66). Instead of using a pulsed laser for the production of ·OH by photolysis of dilute H_2O_2 it might also be possible to conduct kinetic experiments on purely aqueous solutions using a pulsed electron beam (64).

The solvent accessibility patterns uncovered here during cyt *c* folding are consistent with formation and docking of the N- and C-terminal helices as early as 10 ms after a pH 2 → 7 jump. However, structural consolidation of the helix/helix interface

extends up to at least 1 s. Initial heme misligation leads to distal encroachment of H33. This induces a very low solvent accessibility for M80 and I81 early during folding, even though the M80-iron coordination bond has not formed yet. The hydrophobically collapsed core of the protein exhibits packing defects that persist up to at least 100 ms, as seen by the labeling behavior of W59.

In addition to the pH 2 $\rightarrow$ 7 investigations discussed above, we also attempted to characterize the cyt c folding behavior by oxidative labeling in pH 2 $\rightarrow$ 5 jump experiments where misligation is suppressed (data not shown). Unfortunately, the limited time resolution of the mixing setup used here precluded meaningful measurements under those much faster conditions. However, it should be possible to combine the labeling technique used here with other types of continuous-flow mixers that provide access to the sub-millisecond range (90). Work in this direction is currently ongoing in our laboratory and will be reported elsewhere.

3.5 References

1. Ben-David, M., et al. (2009) Assessment of CASP8 structure predictions for template free targets. *Proteins* 77: 50-65 Suppl. 9.

2. Luheshi, L. M., Crowther, D. C. and Dobson, C. M. (2008) Protein Misfolding and Disease: From the Test Tube to the Organism. *Curr. Op. Chem. Biol.* 12: 25-31.

3. Bernstein, S. L., et al. (2009) Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* 1: 326-331.

4. Rumfeldt, J. A. O., Lepock, J. R. and Meiering, E. M. (2009) Unfolding and Folding Kinetics of Amyotrophic Lateral Sclerosis-Associated Mutant Cu,Zn Superoxide Dismutases. *J. Mol. Biol.* 385: 278-298.

5. Kato, H., Feng, H. and Bai, Y. (2007) The Folding Pathway of T4 Lysozyme: The High Resolution Structure and Folding of a Hidden Intermediate. *J. Mol. Biol.* 365: 870-880.

6. Bachmann, A. and Kiefhaber, T. (2001) Apparent Two-state Tendamistat Folding is a Sequential Process Along a Defined Route. *J. Mol. Biol.* 306: 375-386.

7. Yang, W. Y. and Gruebele, M. (2003) Folding at the speed limit. *Nature* 423: 193-197.

8. Eliezer, D., Yao, J., Dyson, H. J. and Wright, P. E. (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* 5: 148-155.

9. Raschke, T. M. and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules under native conditions. *Nat. Struct. Biol.* 4: 298-304.

10. Gruebele, M. (2005) in *Protein Folding Handbook*, eds J. Buchner and T. Kiefhaber (Wiley, New York), pp 454-490.

11. Gianni, S., Ivarsson, Y., Jemth, P., Brunori, M. and Travaglini-Allocatelli, C. (2007) Identification and characterization of protein folding intermediates. *Biophys. Chem.* 128: 105-113.

12. De Los Rios, M. A., et al. (2006) On the precision of experimentally determined protein folding rates and Φ-values. *Protein Sci.* 15: 553-563.

13. Sanchez, I. E. and Kiefhaber, T. (2003) Origin of Unusual Φ-values in Protein Folding: Evidence Against Specific Nucleation Sites. *J. Mol. Biol.* 334: 1077-1085.

14. Fersht, A. R. (2004) Relationship of Leffler (Bronsted) α values and protein folding Φ values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. U.S.A.* 101: 14338-14342.

15. Shastry, M. C. R., Sauder, J. M. and Roder, H. (1998) Kinetic and Structural Analysis of Submillisecond Folding Events in Cytochrome c. *Acc. Chem. Res.* 31: 717-725.

16. Yeh, S.-R., Han, S. and Rousseau, D. L. (1998) Cytochrome c Folding and Unfolding: A Biphasic Mechanism. *Acc. Chem. Res.* 31: 727-736.

17. Englander, S. W., et al. (1998) Fast and Slow Folding in Cytochrome c. *Acc. Chem. Res.* 31: 737-744.

18. Telford, J. R., Wittung-Stafshede, P., Gray, H. B. and Winkler, J. R. (1998) Protein Folding Triggered by Electron Transfer. *Acc. Chem. Res.* 31: 755-763.

19. Bushnell, G. W., Louie, G. V. and Brayer, G. D. (1990) High-resolution Three-dimensional Structure of Horse Heart Cytochrome c. *J. Mol. Biol.* 214: 585-595.

20. Cianetti, S., Negrerie, S., Vos, M. H., Martin, J.-L. and Kruglik, S. G. (2004) Photodissociation of Heme Distal Methionine in Ferrous Cytochrome c Revealed by Subpicosecond Time-Resolved Resonance Raman Spectroscopy. *J. Am. Chem. Soc.* 126: 13932-13933.

21. Winkler, J. R. (2004) Cytochrome c Folding Dynamics. *Current Opinion in Chemical Biology* 8: 169-174.

22. Kaminsky, L. S., Miller, V. J. and Davison, A. J. (1973) Thermodynamic Studies of the Opening of the Heme Crevice of Ferricytochrome c. *Biochemistry* 12: 2215-2221.

23. Tsong, T. Y. (1976) Ferricytochrome c Chain Folding Measured by the Energy Transfer of Tryptophan 59 to the Heme Group. *Biochemistry* 15: 5467-5473.

24. Zhong, S., Rousseau, D. L. and Yeh, S.-R. (2004) Modulation of the Folding Energy Landscape of Cytochrome c with Salt. *J. Am. Chem. Soc.* 126: 13934-13935.

25. Döpner, S., Hildebrandt, P., Rosell, F. I. and Mauk, A. G. (1998) Alkaline Conformational Transitions of Ferricytochrome c Studied by Resonance Raman Spectroscopy. *J. Am. Chem. Soc.* 120: 11246-11255.

26.   Akiyama, S., Takashi, S., Ishimori, K. and Morishima, I. (2000) Stepwise formation of α-helices during cytochrome c folding. *Nat. Struct. Biol.* 6: 514-520.

27.   Bilsel, O. and Matthews, C. R. (2006) Molecular Dimensions and Their Distributions in Early Folding Intermediates. *Curr. Opin. Struct. Biol.* 16: 86-93.

28.   Akiyama, S., et al. (2002) Conformational Landscape of Cytochrome c Folding Studied by Microsecond-Resolved Small-Angle X-Ray Scattering. *Proc. Natl. Acad. Sci.* 99: 1329-1334.

29.   Konermann, L., Collings, B. A. and Douglas, D. J. (1997) Cytochrome *c* Folding Kinetics Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *Biochemistry* 36: 5554-5559.

30.   Grandori, R. (2002) Detecting equilibrium cytochrome *c* folding intermediates by electrospray ionization mass spectrometry: Two partially folded forms populate the molten globule state. *Protein Sci.* 11: 453-458.

31.   Pletneva, E. V., Gray, H. B. and Winkler, J. R. (2005) Snapshots of cytochrome *c* folding. *Proc. Natl. Acad. Sci. U.S.A.* 102: 18397-18402.

32.   Krishna, M. M. G., Lin, Y., Mayne, L. and Englander, S. W. (2003) Intimate View of a Kinetic Protein Folding Intermediate: Residue-resolved Structure, Interactions, Stability, Folding and Unfolding Rates, Homogeneity. *J. Mol. Biol.* 334: 501-513.

33.   Udgaonkar, J. B. and Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335: 694-699.

34.   Roder, H., Elöve, G. A. and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335: 700-704.

35.   Uzawa, T., et al. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13859-13864.

36.   Yang, H. and Smith, D. L. (1997) Kinetics of Cytochrome c Folding Examined by Hydrogen Exchange and Mass Spectrometry. *Biochemistry* 36: 14992-14999.

37.   Konermann, L. and Simmons, D. A. (2003) Protein Folding Kinetics and Mechanisms Studied by Pulse-Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 22: 1-26.

38. Shastry, M. C. R. and Roder, H. (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat. Struct. Biol.* 5: 385-392.

39. Sosnick, T. R., Shtilerman, M. D., Mayne, L. and Englander, S. W. (1997) Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci. USA* 94: 8545-8550.

40. Colon, W., Elöve, G. A., Wakem, L. P., Sherman, F. and Roder, H. (1996) Side Chain Packing of the N- and C-Terminal Helices Plays a Critical Role in the Kinetics of Cytochrome c Folding. *Biochemistry* 35: 5538-5549.

41. Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J. and Roder, H. (1993) A Noncovalent Peptide Complex as a Model for an Early Folding Intermediate of Cytochrome c. *Biochemistry* 32: 10271-10276.

42. Krishna, M. M. G., Lin, Y. and Englander, S. W. (2004) Protein Misfolding: Optional Barriers, Misfolded Intermediates, and Pathway Heterogeneity. *J. Mol. Biol.* 343: 1095-1109.

43. Colon, W., Wakem, L. P., Sherman, F. and Roder, H. (1997) Identification of the Predominant Non-Native Histidine Ligand in Unfolded Cytochrome c. *Biochemistry* 36: 12535-12541.

44. Hammack, B., Shubhada, G. and Bowler, B. E. (1998) Cytochrome c Folding Traps Are Not Due Solely to Histidine-heme Ligation: Direct Demonstration of a Role for N-terminal Amino Group-Heme Ligation. *J. Mol. Biol.* 275: 719-724.

45. Brems, D. N. and Stellwagen, E. (1983) Manipulation of the Observed Kinetic Phases in the Refolding of Denatured Ferricytochrome c. *J. Biol. Chem.* 258: 3655-3660.

46. Elöve, G. A., Bhuyan, A. K. and Roder, H. (1994) Kinetic Mechanism of Cytochrome c Folding: Involvement of the Heme and Its Ligands. *Biochemistry* 33: 6925-6935.

47. Sosnick, T. R., Mayne, L., Hiller, R. and Englander, S. W. (1994) The barriers in protein folding. *Nature Struct. Biol.* 1: 149-156.

48. Yeh, S.-R. and Rousseau, D. L. (2000) Hierarchical folding of cytochrome c. *Nat. Struct. Biol.* 7: 443-445.

49. Jennings, P. A. (1998) Speeding along the protein folding highway, are we reading the signs correctly? *Nat. Struct. Biol.* 5: 846-848.

50. Prabhu, N. P., Kumar, R. and Bhuyan, A. K. (2004) Folding barrier in horse cytochrome c: Support for a classical folding pathway. *J. Mol. Biol.* 337: 195-208.

51. Maity, H., Maity, M., Krishna, M. M. G., Mayne, L. and Englander, S. W. (2005) Protein folding: The stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. U.S.A.* 102: 4741-4746.

52. Chetty, P. S., et al. (2009) Helical structure and stability in human apolipoprotein A-I by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* 106: 19005-19010.

53. Chen, H., Schuster, M. C., Sfyroera, G., Geisbrecht, B. V. and Lambris, J. D. (2008) Solution Insights into the Structure of the Efb/C3 Complement Inhibitory Complex as Revealed by Lysine Acetylation and Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* 19: 55-65.

54. Buxton, G. V., Greenstock, C. L., Helman, W. P. and Ross, A. B. (1988) Critical Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen Atoms and Hydroxyl Radicals (.OH/.O) in Aqueous Solution. *J. Phys. Chem. Ref. Data* 17: 513-886.

55. Mendoza, V. L. and Vachet, R. W. (2009) Probing Protein Structure by Amino Acid-specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 28: 785-815.

56. Fitzgerald, M. C. and West, G. M. (2009) Painting Proteins with Covalent Labels: What's In the Picture? *J. Am. Soc. Mass Spectrom.* 20: 1193-1206.

57. Jha, S. K. and Udgaonkar, J. B. (2007) Exploring the Cooperativity of the Fast Folding Reaction of a Small Protein Using Pulsed Thiol Labeling and Mass Spectrometry. *J. Biol. Chem.* 282: 37479-37491.

58. Aye, T. T., Low, T. Y. and Sze, S. K. (2005) Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry. *Anal. Chem.* 77: 5814-5822.

59. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

60. Hambly, D. M. and Gross, M. L. (2009) in *Protein Mass Spectrometry*, eds J. P. Whitelegge (Elsevier, Amsterdam), pp 151-177.

61. Xu, G. and Chance, M. R. (2007) Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* 107: 3514-3543.

62.　McClintock, C., Kertesz, V. and Hettich, R. L. (2008) Development of an Electrochemical Oxidation Method for Probing Higher Order Protein Structure with Mass Spectrometry. *Anal. Chem.* 80: 3304-3317.

63.　Smedley, J. G., Sharp, J. S., Kuhn, J. F. and Tomer, K. B. (2008) Probing the pH-dependent prepore to pore transition of Bacillus anthracis protective antigen with differential oxidative protein footprinting. *Biochemistry* 47: 10694-10704.

64.　Watson, C., et al. (2009) Pulsed Electron Beam Water Radiolysis for Submicrosecond Hydroxyl Radical Protein Footprinting. *Anal. Chem.* 81: 2496-2505.

65.　Maleknia, S. D. and Downard, K. (2001) Radical Approaches to Probe Protein Structure, Folding, and Interactions by Mass Spectrometry. *Mass Spectrom. Rev.* 20: 388-401.

66.　Gau, B. C., Sharp, J. S., Rempel, D. L. and Gross, M. L. (2009) Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* 81: 6563–6571.

67.　Konermann, L. (1999) Monitoring Reaction Kinetics by Continuous-Flow Methods: The Effects of Convection and Molecular Diffusion Under Laminar Flow Conditions. *J. Phys. Chem. A* 103: 7210-7216.

68.　Xu, G., Kiselar, J., He, Q. and Chance, M. R. (2005) Secondary Reactions and Strategies To Improve Quantitative Protein Footprinting. *Anal. Chem.* 77: 3029-3037.

69.　Smith, A. M., Jahn, T. R., Ashcroft, A. E. and Radford, S. E. (2006) Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Formation using Electrospray Ionization Mass Spectrometry. *J. Mol. Biol.* 364: 9-19.

70.　Pan, Y., Brown, L. and Konermann, L. (2009) Mapping the Structure of an Integral Membrane Protein under Semi-Denaturing Conditions by Laser-Induced Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* 394: 968-981.

71.　Boys, B. L., Kuprowski, M. C. and Konermann, L. (2007) Symmetric Behavior of Hemoglobin α- and β- Subunits during Acid-Induced Denaturation Observed by Electrospray Mass Spectrometry. *Biochemistry* 46: 10675-10684.

72.　Kim, N. H. and Kang, J. H. (2006) Oxidative Damage of DNA Induced by the Cytochrome *c* and Hydrogen Peroxide System. *J. Biochem. Mol. Biol.* 39: 452-456.

73.   Florence, T. M. (1985) The Degradation of Cytochrome *c* by Hydrogen Peroxide. *J. Inorg. Biochem.* 23: 131-141.

74.   Diederix, R. E. M., Ubbink, M. and Canters, G. W. (2002) Peroxidase Activity as a Tool for Studying the Folding of *c*-Type Cytochromes. *Biochemistry* 41: 13067-13077.

75.   Baldwin, D. A., Marques, H. M. and Pratt, J. M. (1987) Hemes and Hemoproteins. 5: Kinetics of the Peroxidatic Activity of Microperoxidase-8: Model for the Peroxidase Enzymes. *J. Inorg. Biochem.* 30: 203-217.

76.   Hambly, D. M. and Gross, M. L. (2009) Cold Chemical Oxidation of Proteins. *Anal. Chem.* 81: 7235-7242.

77.   Adams, P. A. (1976) The Kinetics and Mechanism of the Recombination Reaction between Apomyoglobin and Haemin. *Biochem. J.* 159: 371-376.

78.   Babul, J. and Stellwagen, W. (1972) Participation of the Protein Ligands in the Folding of Cytochrome c. *Biochemistry* 7: 1195-1200.

79.   Knapp, J. A. and Pace, C. N. (1971) Guanidine Hydrochloride and Acid Denaturation of Horse, Cow, and Candida krusei Cytochromes c. *Biochemistry* 13: 1289-1294.

80.   Stellwagen, E. and Babul, J. (1975) Stabilisation of the Globular Structure of Ferricytochrome c by Chloride in Acidic Solvents. *Biochemistry* 23: 5135-5140.

81.   Tong, X., Wren, J. C. and Konermann, L. (2008) γ-Ray-Mediated Oxidative Labeling for Detecting Protein Conformational Changes by Electrospray Mass Spectrometry. *Anal. Chem.* 80: 2222-2231.

82.   Pletneva, E. V., Gray, H. B. and Winkler, J. R. (2005) Many faces of the Unfolded State: Conformational Heterogeneity in Denatured Yeast Cytochrome c. *J. Mol. Biol.* 345: 855-867.

83.   Roepstorff, P. (1984) Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomedical Mass Spectrometry* 11: 601.

84.   Nukuna, B. N., Sun, G. and Anderson, V. E. (2004) Hydroxyl Radical Oxidation of Cytochrome c by Aerobic Radiolysis. *Free Radical Biol. Med.* 37: 1203-1213.

85.   Konermann, L., Tong, X. and Pan, Y. (2008) Protein Structure and Dynamics Studied by Mass Spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom.* 43: 1021-1036.

86. Brandts, J. F., Brennan, M. and Lin, L. (1977) Unfolding and refolding occur much faster for a proline-free protein than for most proline-containing proteins. *Proc. Natl. Acad. Sci. USA* 74: 4178-4181.

87. Houry, W. A. and Scheraga, H. A. (1996) Nature of the Unfolded State of Ribonuclease A: Effect of Cis-Trans X-Pro Peptide Bond Isomerisation. *Biochemistry* 35: 11719-11733.

88. Myers, J. K., Pace, C. N. and Schotz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* 4: 2138-2148.

89. Bieri, O. and Kiefhaber, T. (2001) Origin of Apparent Fast and Non-exponential Kinetics of Lysozyme Folding Measured in Pulsed Hydrogen Exchange Experiments. *J. Mol. Biol.* 310: 919-935.

90. Roder, H., Maki, K. and Cheng, H. (2006) Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem. Rev.* 106: 1836-1861.

Chapter 4 - Temporal Development of Protein Structure During S100A11 Folding and Dimerization Probed by Oxidative Labeling and Mass Spectrometry

4.1 Introduction

Protein folding is a spontaneous process that is driven by the free energy difference between the native and the unfolded states (1). The mechanisms of folding have been a focal point of research for many years (2-6). Common features begin to emerge (7, 8), aided by a "new view" that envisions conformational trajectories as biased conformational diffusion on a funneled energy landscape (9-11). Interest in protein folding is fuelled by the realization that not all polypeptide chains fold correctly under all conditions. Misfolding and aggregation have been implicated in a considerable number of diseases (12).

The bulk of the existing protein folding literature focuses on the behavior of monomeric polypeptide chains. Yet, the majority of cellular proteins are oligomers, consisting of two or more noncovalently bound subunits (13). Relatively little is known about the folding and assembly of these multi-subunit systems. The transition from disordered monomers to a native complex involves the formation of both intra- and intermolecular contacts, thereby adding another layer of complexity to the protein folding problem (14, 15). Homodimers are of particular interest because they constitute the most common type of quaternary structure (16). Studies conducted at equilibrium can provide

glimpses on certain aspects of coupled folding/binding processes (17). The temporal sequence of events, however, can only be explored in time-resolved experiments (18).

Akin to investigations on monomeric proteins (19-21), the detection and structural characterization of kinetic intermediates provides key insights into the folding/assembly mechanisms of protein complexes. In many instances these transient species facilitate the formation of native multi-subunit systems (22, 23). In other cases intermediates may represent kinetic traps (24), or branching points that can lead to aggregates (25). Regardless of their mechanistic role, a thorough understanding of intermediates is a prerequisite for deciphering the self-assembly of multi-component systems.

The canonical view for folding/assembly reactions of dimeric proteins is that individual subunits first fold and then associate to yield the native quaternary structure (26). This type of process may be described by a three-state mechanism $2\ M_U \rightarrow 2\ M_F \rightarrow D_N$, where M_U and M_F represent unfolded and folded monomers, respectively, and D_N is the native dimer. For example, superoxide dismutase (27) and HIV-1 protease (28) display this kinetic behavior. However, other scenarios are possible as well. Some proteins first establish intermolecular contacts, before the subunits attain their native fold (e.g., $2\ M_U \rightarrow D_I \rightarrow D_N$, where D_I is a dimeric intermediate). Even apparent two-state behavior ($2\ M_U \rightarrow D_N$), as well as more complicated mechanisms have been observed (reviewed in ref. (13)).

A general problem with kinetic protein folding investigations is the rapid time scale of these processes. Lifetimes of folding/assembly intermediates are often on the order of milliseconds to seconds. Stopped-flow circular dichroism (CD) and fluorescence spectroscopy are standard approaches for studies in this area (27, 28). Unfortunately,

these global optical probes cannot provide detailed structural information. Pulsed hydrogen/deuterium exchange (HDX) has been widely used for exploring the kinetics of hydrogen bond formation in monomeric proteins (20, 21), but the application of this technique to oligomeric systems is scarce (29). Moreover, the non-physiological pH (or pD) during pulsed HDX can lead to non-trivial phenomena (20, 30).

The current work demonstrates the application of an alternative technique, pulsed oxidative labeling, for monitoring structural changes during folding and assembly of a dimeric protein. This strategy involves protein exposure to a brief pulse of hydroxyl radical ($\cdot$OH) at selected time points after initiation of the reaction. Oxidative labeling causes covalent modifications at solvent accessible side chains, usually resulting in +16 Da mass shifts that are detectable by mass spectrometry (MS) (31). Buried regions are protected from $\cdot$OH attack, such that the oxidation pattern is determined by the protein structure and interactions at the time of the labeling pulse. Of the many methods for $\cdot$OH production (31), the photolysis of dilute H_2O_2 by a pulsed UV laser is particularly suitable for time-resolved studies, because it provides a labeling period as short as 1 µs (32). In contrast to many other covalent probes (33), the reactivity of $\cdot$OH is quite nonspecific such that multiple types of residues can be monitored (31). Solvent accessibility patterns measured in this way are complementary to HDX data (34) and, advantageously, oxidative modifications can be introduced at neutral pH. The potential of $\cdot$OH labeling for time-resolved investigations is only beginning to emerge (Chapter 3) (35-38), and never before has this approach been applied to rapid self-assembly processes of multi-subunit proteins.

The S100 family comprises a number of highly conserved EF-hand calcium binding proteins (39) that are involved in various processes such as tumor suppression (40), cytoskeleton rearrangement (41), and DNA repair (42). S100A11, like many other S100 proteins, forms a compact homodimer (Fig. 4.1A) (43). Each subunit consists of 101 residues that fold into four helices, termed I-IV (Fig. 4.1B). The calcium binding sites are formed by helix-loop-helix motifs of helices I and II, as well as III and IV. The hydrophobic dimerization interface comprises elements of helices I and IV from both subunits. Folding and dimerization occur spontaneously for both the calcium-bound and the apo-forms (43). The dissociation constant of calcium-free S100A11 is in the sub-μM range at physiological ionic strength (44).

The acid-denatured state of S100A11 has been characterized by electrospray (ESI) MS, CD, and NMR spectroscopy (45). These techniques revealed that acid exposure induces complete dimer dissociation. The free subunits exhibit a greatly diminished NMR chemical shift dispersion at pH 2, indicating major perturbations in tertiary structure. However, the preservation of some tertiary interactions could not be excluded. In addition, the monomeric state at pH 2 retains native-like helicity.

Refolding and assembly of the native dimer can be triggered by a pH jump. Time-resolved ESI-MS has previously been used to examine the mechanism of this process at the intact protein level (45). It was found that the reaction proceeds through a burst-phase monomeric intermediate. Pulsed HDX suggested the formation of additional transient species, although the isotope exchange data were convoluted and did not provide clear insights into the reaction mechanism (45). These difficulties may be attributable to the retention of native-like secondary structure in acid-denatured S100A11, which makes it
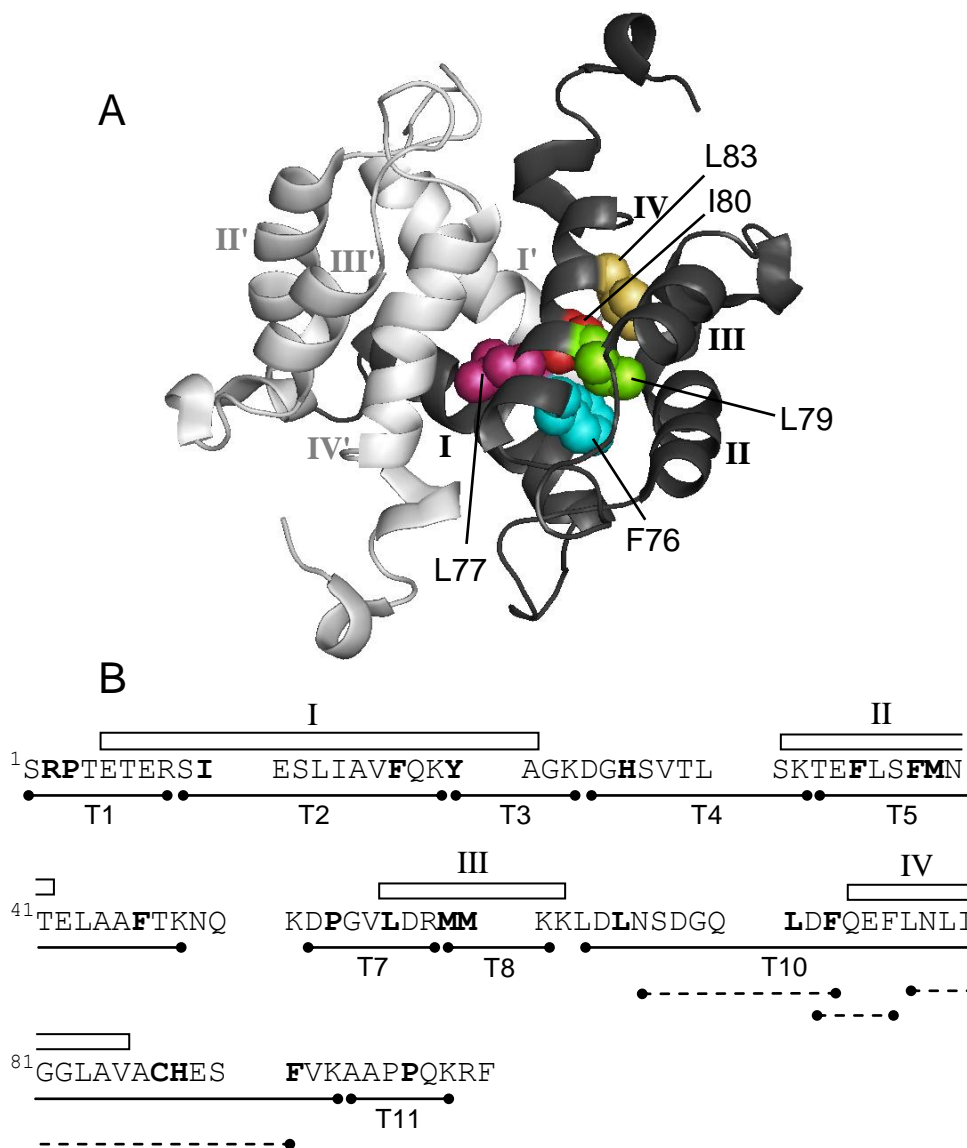
Figure 4.1. (A) Structure of the native S100A11 homodimer in its calcium-free form (44) with α-helices I-IV (I'-IV' for the second subunit). Highlighted residues are involved in a hydrophobic cluster that persists in the unfolded state (see text). (B) Sequence and α-helices. Tryptic peptides (T1-T11) are indicated by solid lines. Selected chymotryptic peptides are indicated as dashed lines. Detected oxidation sites are marked in bold.

challenging to differentiate kinetic species on the basis of their hydrogen bonding characteristics.

In this work we demonstrate that folding and dimerization of apo-S100A11 are accompanied by major changes in side chain solvent accessibility. Monitoring these changes by pulsed oxidative labeling with MS detection provides detailed insights into the kinetic reaction mechanism. In addition to the burst-phase monomeric intermediate, the process involves a semi-native dimeric species. This dimer intermediate exhibits disordered peripheral regions while the protein-protein interface appears to be native-like. The methodology used here should also be suitable for monitoring other biochemical self-assembly processes.

4.2 Materials and Methods

4.2.1 Materials

Recombinant rabbit S100A11 was expressed and purified as described (43). The pseudo-wild type C9S variant was used to avoid possible obstacles presented by intermolecular disulfide bridging (46). The expected molecular mass for the monomer is 11 281 Da, which matches the measured value within experimental error of $\pm$ 1 Da. This mass reflects the natural isotope abundance, whereas previous MS studies from our laboratory used uniformly N15-labeled samples (45). Protein cartoons refer to the first frame of pdb file 1NSH (43), depicting the most representative structure of the NMR

ensemble. Folding/assembly experiments were carried out on the apo-protein, i.e., without adding Ca^{2+} salts. Bovine liver catalase, bovine pancreas α-chymotrypsin, glutamine, and bromocresol purple were obtained from Sigma (St. Louis, MO). Bradykinin was supplied by Bachem (King of Prussia, PA) and sequencing-grade trypsin was purchased from Promega (Madison, WI). All chemicals were used without further purification.

4.2.2 Continuous-Flow Mixing and Oxidative Labeling

Kinetic protein folding experiments with oxidative labeling were performed using a two-syringe continuous-flow mixing device similar to that employed for previous work (Chapter 3). Both syringes were advanced simultaneously at 70 μL min^{-1} using a syringe pump (Harvard Apparatus, Boston, MA). Syringe 1 contained acid-denatured S100A11 (50 μM, expressed on a monomer basis), 30 mM glutamine, 300 mM NaCl, and HCl at pH 2. Protein folding and dimerization was triggered by combining this solution with NaOH and 0.2 M phosphate buffer from syringe 2 (pH 7.5) at a homemade capillary mixer in a 1:1 volume ratio, resulting in a final pH of 7, and a protein concentration of 25 μM. Under these conditions virtually all of the S100A11 will dimerize after equilibration (44). The outlet of the mixer was connected to a reaction capillary with i.d. 100 μm. Syringe 2 also contained 0.4% (v/v) (*ca*. 160 mM) H_2O_2. A KrF excimer laser (GAM EX 100/125, Orlando, FL) producing 18 ns pulses at 248 nm, 92 Hz, 63 mJ, and an irradiation spot width of ca. 2 mm was used to generate ·OH by photolysis of H_2O_2 within the reaction capillary. Glutamine acts as a radical scavenger that quenches the labeling reaction on a time scale of 1 μs (32). Following initiation of folding, oxidative labeling

was performed at different time points by irradiating the reaction mixture at suitable positions downstream of the mixer. Average reaction times of 10 ms, 0.2 s, and 0.8 s correspond to distances between mixer and irradiation spot of 3 mm, 6 cm, and 24 cm, respectively (47). Effective mixing and reproducible labeling was confirmed as described previously (Chapters 2, 3). In addition, the mixer performance was verified optically in pH jump experiments on bromocresol purple (21).

250 µL portions of capillary outflow for each irradiated sample were collected ~1 s after mixing in microcentrifuge tubes that contained 13 µL ~150 mM phosphate buffer and 1.4 µM catalase at pH 7. Catalase was employed for deactivating residual H_2O_2, thereby suppressing undesired secondary oxidation (48, 49). The extent of secondary oxidation was further reduced by flash-freezing all samples in liquid N_2 immediately after collection, and by storage at -80 °C until further analysis. Inspection of the dashed spectrum in Fig. 4.2D confirms that the combination of these measures reduces background oxidation to very low levels.

Part of each sample was retained for intact protein measurements, and the remainder (*ca*. 225 µL) was digested with trypsin at pH 7 and 37 °C for 24 h using a 1:20 (w/w) enzyme:protein ratio. The digests were lyophilized and resuspended in 200 µL of water containing 10 µM bradykinin as an internal intensity standard (Chapters 2, 3)(50). Non-irradiated control samples were collected for every measurement by using the same solution conditions and mixing sequence, except that the laser was switched off.

Unfolded ($t = 0$) samples were prepared by mixing the contents of syringe 1 with H_2O_2 in HCl, and the solution was irradiated at pH 2. The labeled samples were then collected in a tube containing NaOH and phosphate buffer, resulting in a pH 7 mixture.

This allowed for catalase deactivation of residual peroxide, followed by tryptic digestion analogous to the time-resolved experiments described above. Alternatively, $t = 0$ samples were digested overnight at pH 7 with chymotrypsin using a 1:60 (w/w) enzyme:protein ratio at 30 °C to aid in oxidation site determination within helix IV. Earlier work revealed that oxidation rates are not strongly affected by the acidity of the solution, thereby allowing a direct comparison of samples labeled at pH 2 and pH 7 (Chapter 2). For studying the $t = 5$ min endpoint of the reaction, the contents of syringe 1 were manually mixed with NaOH and phosphate buffer. The resulting solution was transferred into syringe 1 after 5 minutes. Syringe 2 contained H_2O_2 in water, such that the final composition of the mixture during labeling was the same as for all other time points.

4.2.3 LC/ESI-MS

All experiments were performed on a Q-TOF Ultima API mass spectrometer (Waters, Milford, MA) equipped with a Z-spray ESI source that was operated at 3 kV. The instrument was coupled to a Waters Acquity UPLC system employing either a C4 (BEH300) 2.1 mm x 50 mm reversed-phase column for intact protein analyses, or a C18 (BEH300) 2.1 mm x 100 mm column for peptide measurements. Elution was carried out using a water/acetonitrile gradient in the presence of 0.1% formic acid at a flow rate of 100 μL min^{-1} and an injection volume of 10 μL. The identity of tryptic peptides was confirmed by MS/MS. Spectra for the intact protein (Fig. 4.2) are presented as deconvoluted mass distributions, obtained by using the MaxEnt software supplied by the instrument manufacturer (51).

Oxidation sites for each tryptic peptide were determined in off-line MS/MS

experiments by directing the UPLC eluent into a TriVersa NanoMate (Advion, Ithaca, NY) chip-based nanoESI source. The incoming flow was split such that ~0.3% was sent directly to the mass spectrometer while the remainder was collected in 1 minute fractions in a 96-well plate. After fraction collection the plate was flash-frozen, and its contents were lyophilized. The samples in each well were then resuspended in 50 µL 2% aqueous acetic acid such that all of them could be sprayed under identical source conditions (1.6 kV ESI voltage, 0.3 psi nitrogen backing pressure).

4.3.4 Data Analysis

A key parameter for quantifying the extent of oxidation within any given protein segment (represented by a tryptic peptide) is the fraction unmodified F_u, defined as (31)

$$F_u = \frac{I_u}{I_u + I_{ox}}$$
(4-1)

where I_u and I_{ox} are the LC-ESI/MS peak intensities (in units of counts s^{-1}) of the unmodified peptide and its oxidation product(s), respectively. For the current work the determination of I_u was based on the most intense isotope peak for each peptide, which happens to be the monoisotopic signal for most protein segments. Integrated isotope cluster areas may be used instead of intensities, but the latter approach has the advantage that it avoids complications in differentiating between low-abundance isotopologues and baseline noise.

Care was taken to suppress the extent of secondary oxidation as much as possible (see above). Nonetheless, low levels of background oxidation cannot be ruled out. A background correction was therefore applied to all data. Let us first consider the

hypothetical case where normalized peak intensities R_u and R_{ox} are used instead of I_u and I_{ox}, where $R_u + R_{ox} = 1$ for each spectrum. In this case it can stated that (52)

$$F_u^{corr} = \frac{R_u^{app}}{R_u^{app} + R_{ox}^{app} - R_{ox}^{bgr}} \qquad (4\text{-}2)$$

where F_u^{corr} reflects the actual extent of laser-induced labeling, corrected for background oxidation. The superscript *app* refers to spectra measured after laser exposure, whereas *bgr* signifies the corresponding non-irradated background (control) data. With $(1 - R_{ox}^{bgr})$ $= R_u^{bgr}$ it is found that

$$F_u^{corr} = \frac{R_u^{app}}{R_u^{bgr}} \qquad (4\text{-}3)$$

Because the actual measured peak intensities I_u are proportional to the corresponding normalized values R_u one can also express the last equation as

$$F_u^{corr} = \frac{I_u^{app}}{I_u^{bgr}} \qquad (4\text{-}4)$$

Unfortunately, F_u^{corr} values determined on the basis of eq. 4-4 are vulnerable to possible run-to-run sensitivity drifts of the mass spectrometer. This problem is circumvented in the current work by using a slightly modified strategy. A constant amount of the internal standard bradykinin was added to all samples, which appears as a signal I_{BK} in all of the data sets. Relative peak intensities r_u for each unmodified tryptic peptide can then be determined as $r_u = I_u/I_{BK}$. These r_u values are directly comparable across different runs, because any sensitivity drifts cancel out as they affect both the tryptic peptides and the internal standard. Background-corrected fraction unmodified values were therefore

determined in the present work as

$$F_u^{\ corr} = \frac{r_u^{\ app}}{r_u^{\ bgr}} \qquad (4\text{-}5)$$

An additional advantage of using eq. 4-5 is that it remains valid even in cases where certain oxidation products of a peptide go undetected. Also, eq. 4-5 does not rely on any assumptions regarding the ionization efficiencies of peptides and their oxidation products.

Normalized oxidation levels (*NOL*) for each peptide at refolding time *t* were calculated according to (Chapter 2)

$$NOL(t) = \frac{\ln F_u^{\ corr}(t) - \ln F_u^{\ corr}(5\,\text{min})}{\ln F_u^{\ corr}(0) - \ln F_u^{\ corr}(5\,\text{min})} \qquad (4\text{-}6)$$

NOL = 1 reflects the case where a protein segment has the same solvent accessibility as in the acid-unfolded (*t* = 0) state. Conversely, segments that are protected to the same extent as in the refolded (*t* = 5 min) protein are characterized by *NOL* = 0. Error bars in the *NOL* data of Fig. 4.4 represent standard deviations of triplicate measurements, each with its own background correction.

4.3 Results and Discussion

4.3.1 Pulsed Oxidative Labeling

Structural changes taking place during S100A11 folding and assembly were monitored by exposing the protein to a microsecond ·OH pulse at various time points during the reaction. To facilitate comparisons with previous ESI-MS measurements (45) the protein was labeled at either $t = 10$ ms, 200 ms, or 800 ms following a pH jump from 2 to 7. Native control samples, corresponding to $t = 5$ min, were included to represent the reaction endpoint. Mass distributions recorded at the intact protein level reveal that the transition from denatured monomers to the native dimer is associated with dramatic changes in solvent accessibility. The extent of covalent +16 Da adduction is highest for $t = 10$ ms (Fig. 4.2A). As indicated above, the protein is known to form a burst-phase monomeric intermediate at this time point, although the structure of this species remains unexplored (45). With increasing reaction time the labeling level gradually decreases (Figure 4.2B, C), reflecting the progressive protection of side chains from the solvent. At 800 ms (Fig. 4.2C) dimerization has gone to completion (45). The transition from this time point to the refolded protein ($t = 5$ min, Fig. 4.2D) is nonetheless accompanied by a further decrease in the extent of labeling.

The mass distributions of Fig. 4.2 were acquired using a denaturing LC/MS eluent. For this reason all spectra correspond to monomeric S100A11, regardless of the oligomerization state at the time of labeling. The data obtained in this way reflect changes in solvent accessibility averaged over all monomer and dimer conformers that are present at a given reaction time. Also included in Fig. 4.2D are spectra for two control samples. The dotted line represents native S100A11 that had not undergone a
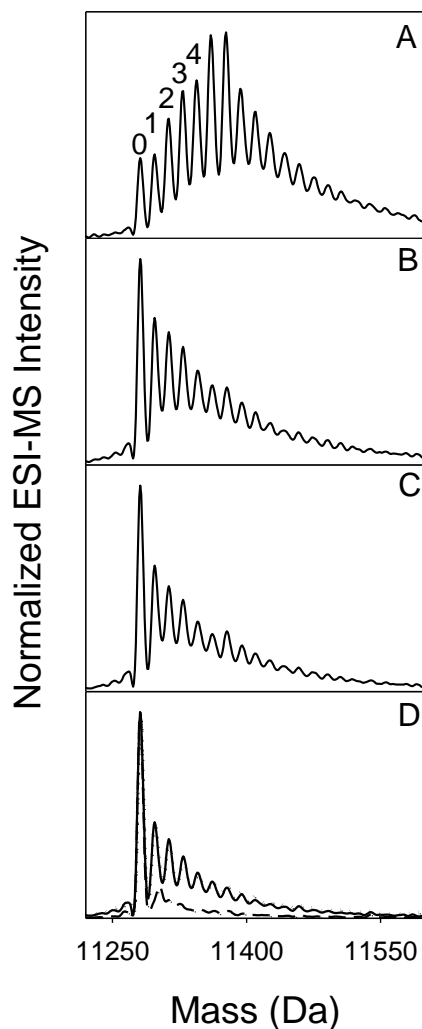
Figure 4.2. Deconvoluted mass distributions obtained by pulsed oxidative labeling at different time points during folding. (A) 10 ms, (B) 200 ms, (C) 800 ms, (D) 5 min. In (A) the unoxidized protein is marked as "0". Oxidative +16 Da modifications are marked as 1, 2, etc. Included in (D) is the spectrum of a native control that was labeled without prior unfolding (dotted line). Also shown (D, dashed line) is the spectrum of an unlabeled control.

denaturation/refolding cycle prior to labeling. The resulting mass distribution is virtually indistinguishable from that of the refolded protein at $t = 5$ min, confirming that formation of the native complex indeed goes to completion in our kinetic experiments. Fig. 4.2D also shows the mass distribution of a sample that did not get exposed to laser irradiation. Evidently, only a low level of background oxidation occurs for this non-labeled control. The time-resolved data discussed below were nonetheless subjected to a background correction, as described in *Materials and Methods*.

4.3.2 Peptide Mapping and Oxidation Site Determination

Ascertaining structural details about the folding and assembly process requires spatially-resolved labeling experiments. To identify the side chains that govern the protein oxidation behavior we conducted MS/MS-based peptide mapping on acid-denatured S100A11 that had been pulse-labeled. Tryptic digestion resulted in nine peptides (denoted as T1, T2, etc. in Fig. 4.1B) that cover 95% of the sequence. The remaining 5% correspond to two short segments (T6 and T9) that were not detected in our mapping procedure. Fig. 4.3 exemplifies typical MS/MS data, obtained by subjecting T2+16 to collision-induced dissociation. Analysis of the tandem mass spectrum reveals that the chosen precursor mass encompasses two isobaric peptides, both of which carry a single oxidative modification. The presence of +16 Da peaks for y_3 to y_9, along with unmodified y_1 and y_2 ions identifies F17 as one of the oxidation sites. In addition, the spectrum shows unmodified y_1 through y_9 ions, in combination with a b_2+16 signal. These signals, along with the high intrinsic reactivity of Ile relative to Ser (31), pinpoint I10 as the second site of oxidation. In a similar fashion, tryptic peptide mapping and
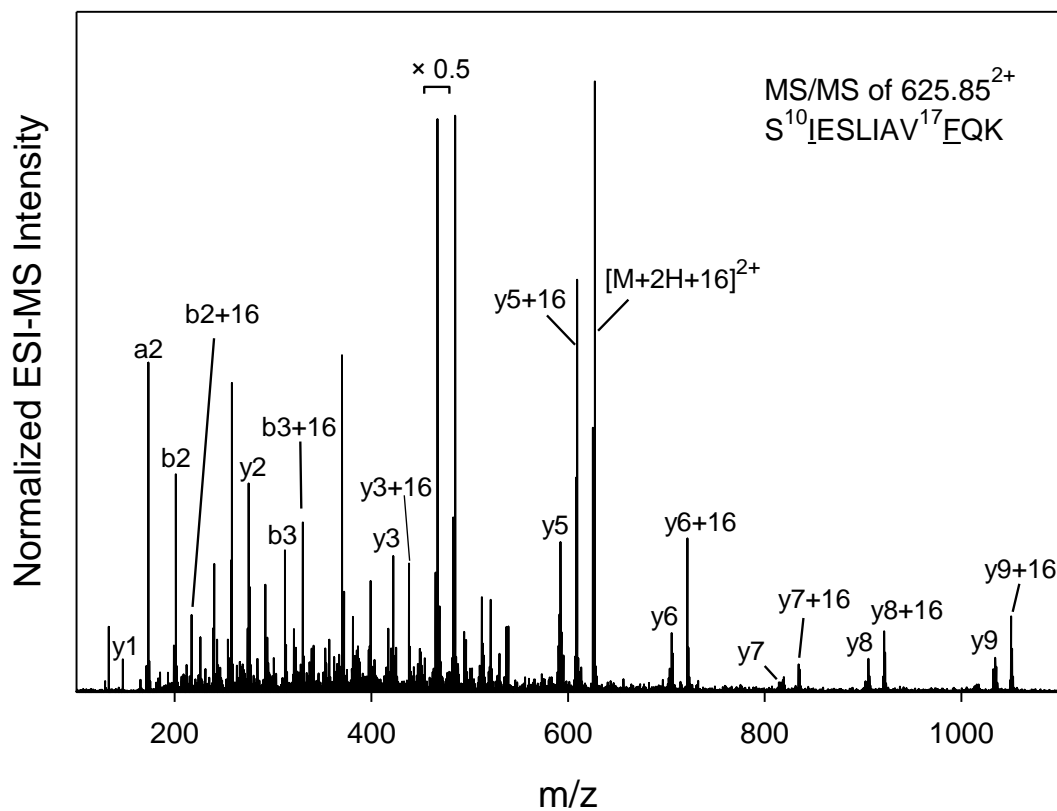
Figure 4.3. Representative MS/MS spectrum of [T2+16]. Peaks are marked using standard *b* and *y* ion notation. Detected oxidation sites (I10 and F17) are underlined.

MS/MS of the remaining peptides revealed 17 additional oxidation sites. The considerable size of one particular peptide (T10, 31 residues) made it challenging to determine possible oxidative modifications in the vicinity of helix IV. Chymotrypsin digestion was therefore carried out, which covers the T10 region with three smaller fragments (Fig. 4.1B). Two more oxidation sites, L71 and F73, were detected in this way. Taken together, the 21 identified modification sites (highlighted in bold, Fig. 4.1B) form a comprehensive set of structural probes that report on changes in solvent accessibility during S100A11 folding and dimerization. This group of residues encompasses F, M, L, P, H, R, Y, I, and C, all of which have been shown to be highly reactive in earlier ·OH labeling experiments on different proteins (31). Oxidation at these sites was found to be associated with +16 Da modifications under the conditions of our work. The formation of additional low abundance products cannot be excluded.

4.3.3 Denatured Monomeric State

As noted earlier, monomeric S100A11 at pH 2 does not adopt a random coil conformation but retains residual structure (45), not unlike many other denatured proteins (53, 54). Clues to the location of structured elements in S100A11 come from the positioning of oxidized and non-oxidized residues in the free subunits (Fig. 4.1B). Phenylalanine exhibits one of the highest intrinsic reactivities with ·OH (31). Thus, it is not surprising that numerous oxidized Phe residues were detected. Of the seven phenylalanines covered by peptide mapping, only one (F76) is protected from oxidation. F76 is in close vicinity to several other intrinsically reactive side chains that also remain unlabeled (L77, L79, I80, L83). These five residues are part of helix IV, and in native

S100A11 they form a hydrophobic core that is shielded from the solvent (highlighted in Fig. 4.1A). The fact that all five residues remain unlabeled at pH 2 indicates that they continue to be protected from the aqueous environment in the acid-denatured state (55). Because L77 and I80 are positioned at the dimerization interface, their lack of oxidation at pH 2 suggests that the acid-denatured monomers possess a hydrophobic core which is structurally somewhat different from that of the native dimer. A similar situation exists in helix I where L13, I14 and V16, residues also located at the dimer interface, are protected from oxidation at pH 2. The most likely scenario is re-packing of the nonpolar side chains of helix I (L13, I14, V16) against those of helix IV (L77, I80). In other words, our data suggest that the monomeric protein at pH 2 compensates the loss of *inter*molecular hydrophobic contacts by forming *intra*molecular contacts between helices I and IV. The consequences of such partially non-native structural features for the assembly mechanism will be considered below.

4.3.4 Time-Dependent Peptide Solvent Accessibilities

A comparison of the time-dependent changes in solvent accessibility for the various protein segments becomes possible by displaying their normalized oxidation levels (*NOL*) (Fig. 4.4, eq. 4-6 of *Materials and Methods*). *NOL* = 1 represents the case where a segment has the same accessibility as in acid-denatured S100A11, whereas *NOL* = 0 signifies that protection has reached the native state level. Fractional *NOL* values correspond to solvent accessibilities in-between these two cases. In Fig. 4.4 peptides are grouped into three categories according to the appearance of their *NOL* profiles. Also shown are the locations of individual peptides in the context of the native S100A11 dimer
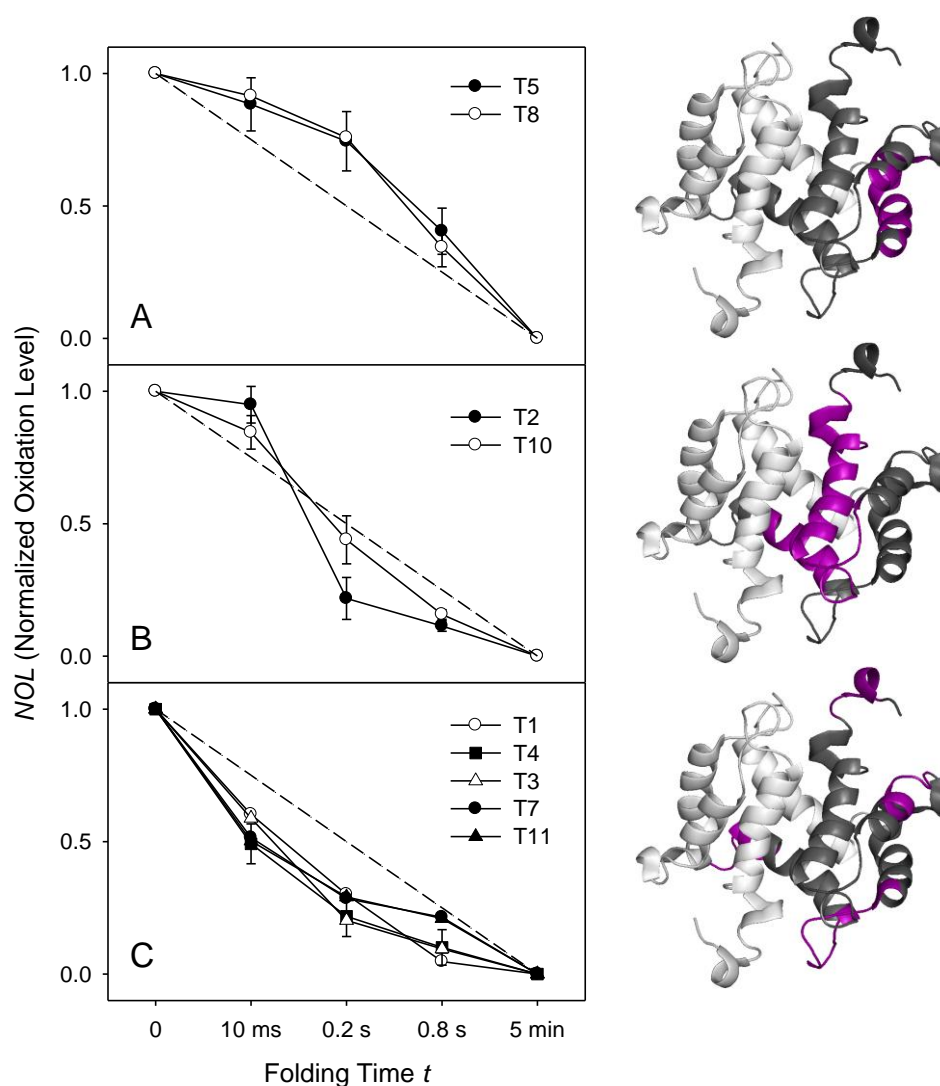
Figure 4.4. Normalized oxidation levels (*NOL*) of protein segments (tryptic peptides T1, T2, ...) as a function of folding time *t*. Dashed lines are included to facilitate data comparison, they do not represent a theoretically expected time course. Cartoons highlight the locations of peptides (purple) within one subunit of the native dimer. All data points represent averages of three independent measurement. Error bars indicate standard deviations. In (C), error bars are shown for T4 only, while others have been omitted for clarity. Note that the time axis is not linear.

structure. Panel A represents two segments that retain relatively high oxidation levels at t = 10 ms, 200 ms, and 800 ms. In comparison, the peptides in Fig. 4.4C have significantly lower *NOL* values at all three time points. The remaining two segments, T2 and T10 (Fig. 4.4B) show an interesting behavior where *NOL* at t = 10 ms is quite high, and comparable to the panel A data set. At longer reaction times these two peptides switch towards the low *NOL* values exhibited by the segments in panel C. This transition from high to low oxidation is particularly pronounced for T2, which is among the most protected segments at the 200 and 800 ms time points. Inspection of the structural cartoons of Fig. 4.4 reveals that T2 and T10 encompass the dimerization interface, i.e., helices I and IV.

4.3.5 Folding and Assembly of S100A11

The structural changes taking place during folding and dimerization were visualized by mapping the *NOL* data of Fig. 4.4 onto the native protein structure, using a five-step color scheme to signify the extent of oxidation at reactive side chains (Fig. 4.5). For the acid-denatured monomers all 21 oxidation sites are shown in red (*NOL* = 1), whereas the five non-oxidizable residues involved in the partially non-native hydrophobic core are displayed in blue (Fig. 4.5A).

After 10 ms S100A11 is known to adopt a monomeric intermediate (45). A major *NOL* drop to values around 0.5 at this time point has occurred for Y20, H26, P53, L56, and P97, all of which are colored yellow in Fig. 4.5B. When inspecting the locations of these sites it is apparent that they are distant from the dimerization interface. The labeling behavior throughout the rest of the protein at 10 ms remains close to that of the acid-
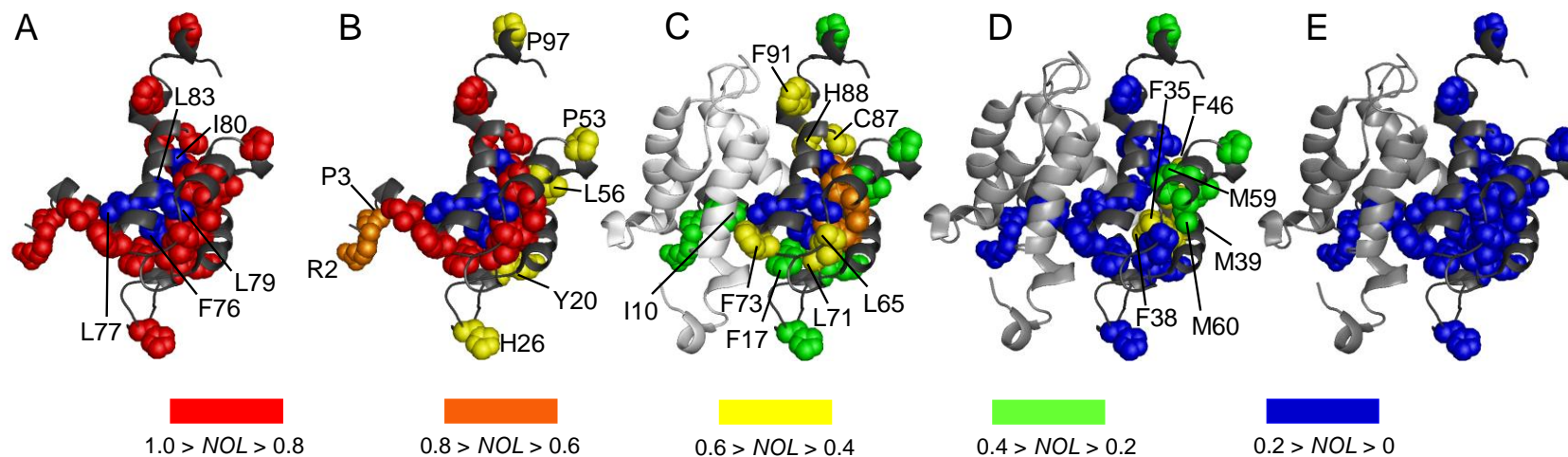
Figure 4.5. Structural changes during folding and dimerization measured by pulsed oxidative labeling. (A) acid-denatured protein ($t = 0$), (B) 10 ms, (C) 200 ms, (D) 800 ms, (E) native protein at $t = 5$ min. Normalized oxidation levels (*NOL*) are visualized using a five-level color code. The onset of dimerization is indicated in (C) by showing the second subunit in light gray. Dimerization is complete for (D), (E). Oxidation sites are highlighted for one subunit only.

denatured state, reflecting the lack of protein-protein interactions that keeps many residues at the dimerization interface solvent exposed.

Further protection at Y20, H26, P53, L56, and P97 is evident at 200 ms. In Fig. 4.5C these residues are depicted in green. At the same time, notable protection starts to appear for residues at the dimerization interface, including I10, F17, F73, H88, and F91. The labeling behavior at these sites is consistent with partial dimerization, as previously seen by time-resolved ESI-MS at this time point (45). An overall compaction starts to affect the entire protein at 200 ms, evident from the lack of "red" side chains in Fig. 4.5C. The onset of dimerization is symbolized by displaying the second subunit in light gray.

Subunit association is known to be complete after 800 ms (45). Oxidative labeling at this time point results in *NOL* values close to zero for many side chains (blue in Fig. 4.5D), indicating that their solvent accessibilities are native-like. These residues include all those in the vicinity of the dimerization interface (helices I and IV), consistent with steric shielding from ·OH attack due to mutual protection of the two subunits. Interestingly, however, side chains in helices II and III retain significant exposure at 800 ms (F35, F38, M39, F46, M59, M60, green and yellow, Fig. 4.5D). This behavior implies the presence of a transient intermediate at 800 ms that is dimeric, yet retains non-native elements. This dimer represents a previously unidentified species on the kinetic folding/assembly pathway of S100A11. It exhibits strong protection throughout helices I and IV and in the dimerization interface. In contrast, the helix II/III regions retain elevated solvent accessibilities. Structural consolidation of these remaining non-native regions ultimately leads to the fully folded S100A11 complex (Figure 4.5E).

It is interesting to note that a dimeric intermediate has also been observed for the closely related S100B, albeit under equilibrium conditions. Tryptophan substitutions suggested involvement of A6 and T81 of S100B in subunit interactions (56). These positions correspond to S9 and A86, respectively, in the C9S variant of S100A11 used here (57). Neither S9 nor A86 were found to be labeled in our experiments due to their low intrinsic reactivities (31). However, the neighboring I10, C87 and H88 were found to be oxidizable (Fig. 4.1B). At a folding time of 10 ms, these three residues exhibit very high labeling levels (red in Fig. 4.5B). In contrast, at 800 ms they are strongly protected (blue in Fig. 4.5D). These observations provide further support that S100A11 exists as a monomer intermediate at 10 ms, while forming a dimeric intermediate at 800 ms.

4.4 Conclusions

The current work provides insights into the structural changes taking place during a biomolecular self-assembly process. Using the notation outlined in the *Introduction*, the transition from acid-denatured monomeric S100A11 to the native dimer may be portrayed by a simplified pathway scheme

$$2\,M_U \rightarrow 2\,M_I \rightarrow D_I \rightarrow D_N$$
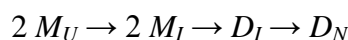
Describing kinetic folding/assembly processes in terms of pathways has a long history (13, 27, 28). However, past descriptions of this type were based largely on spectroscopic data that provided limited structural information. Here we demonstrate that pulsed ·OH

labeling yields distinct conformational features of the participating species by reporting on side chain solvent accessibilities as a function of reaction time.

Earlier kinetic studies on S100A11 had already presented evidence for the formation of a monomeric burst-phase intermediate M_I (45). The oxidative labeling data of this work reveal that solvent exclusion in this transient conformer mainly affects elements that are distant from the dimerization interface (Fig. 4.5B). At a later stage S100A11 forms a dimeric species D_I with native-like protein-protein interactions (helices I/IV), whereas regions in the periphery (helices II/III) remain partially disordered. A similar species has been shown to be populated in equilibrium experiments on S100B (56). The observation that S100A11 forms both a monomeric and a dimeric intermediate in our kinetic experiments implies that folding and protein-protein binding are intertwined processes, where one cannot go to completion without the other. This finding contrasts earlier proposals, where the presence of fully folded monomers is considered to be a general prerequisite for the assembly of native protein complexes (26).

It is interesting to consider why dimerization of S100A11 takes place on a time scale of hundreds of milliseconds, whereas some other systems can establish protein-protein contacts much more rapidly. For example, formation of a dimeric intermediate occurs close to the diffusion limit for the Trp repressor from *E. coli* (58). Our results suggest that the relatively slow association kinetics of S100A11 are caused by the partially non-native hydrophobic core of the monomeric subunits. Yeast two-hybrid assays on S100A4 have revealed that residues of helix IV corresponding to F76, L77, L79, I80, and L83 in S100A11 play a key role for the dimerization process (41). Partial shielding of these residues by elements of helix I as indicated by oxidative labeling

therefore represents an initial impediment to subunit association. Whether the subsequent dimerization step follows an induced-fit scenario, or whether it involves equilibrium shifts of conformational populations (14) cannot be decided from our data.

The solvent accessibility maps of Fig. 4.5 are akin to frames of a "molecular movie" that depicts structural changes during folding and assembly of a protein complex. The earliest possible reaction time point that could be monitored with the system employed here is around 10 ms. Based on the short (~1 μs) duration of the ·OH labeling pulse, it should be possible to extend the time frame of these experiments to much faster processes by using improved mixing schemes or temperature-jump devices (36). Work in this direction is currently ongoing in our laboratory, as well as elsewhere (35). It is hoped that these efforts will help pave the way towards a better structural and mechanistic understanding of rapid biomolecular processes.

4.5 References

1. Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science* 181: 223-230.

2. Shaw, D. E., et al. (2010) Atomic Level Characterization of the Structural Dynamics of Proteins. *Science* 330: 341-346.

3. Bartlett, A. I. and Radford, S. E. (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* 16: 582-588.

4. Fersht, A. R. (2008) From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* 9: 650-654.

5. Englander, S. W., Mayne, L. and Krishna, M. M. G. (2008) Protein folding and misfolding: mechanism and principles. *Quart. Rev. Biophys.* 40: 287-326.

6. Wittung-Stafshede, P., Lee, J. C., Winkler, J. R. and Gray, H. B. (1999) Cytochrome b562 folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein. *Proc. Natl. Acad. Sci. U.S.A.* 96: 6587-6590.

7. Baker, D. (2000) A surprising simplicity to protein folding. *Nature* 405: 39-42.

8. Gianni, S., et al. (2003) Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* 100: 13286-13291.

9. Lazaridis, T. and Karplus, M. (1997) "New View" of Protein Folding Reconciled with the Old Through Multiple Unfolding Simulations. *Science* 278: 1928-1931.

10. Dill, K. A. and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4: 10-19.

11. Weinkam, P., Zong, C. and Wolynes, P. G. (2005) A funneled energy landscape for cytochrome c directly predicts the sequential folding route inferred from hydrogen exchange experiments. *Proc. Natl. Acad. Sci. U.S.A.* 102: 12401-12406.

12. Aguzzi, A. and O'Connor, T. (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat. Rev. Drug Discov.* 9: 237-248.

13. Rumfeldt, J. A. O., Galvagnion, C., Vassall, K. A. and Meiering, E. M. (2008) Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* 98: 61-84.

14.   Boehr, D. D., Nussinov, R. and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5: 789-796.

15.   Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G. and Kiefhaber, T. (2011) Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. U.S.A.* 108: 3952-3957.

16.   Mei, G., Di Venere, A., Rosato, N. and Finazzi-Agro, A. (2005) The importance of being dimeric. *FEBS J.* 272: 16-27.

17.   Habazettl, J., Reiner, A. and Kiefaber, T. (2009) NMR Structure of a Monomeric Intermediate on the Evolutionarily Optimized Assembly Pathway of a Small Tremerization Domain. *J. Mol. Biol.* 389: 103-114.

18.   Mulder, A. M., et al. (2010) Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit. *Science* 330: 673-677.

19.   Brockwell, D. J. and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Op. Struct. Biol.* 17: 30-37.

20.   Krishna, M. M. G., Lin, Y., Mayne, L. and Englander, S. W. (2003) Intimate View of a Kinetic Protein Folding Intermediate: Residue-resolved Structure, Interactions, Stability, Folding and Unfolding Rates, Homogeneity. *J. Mol. Biol.* 334: 501-513.

21.   Uzawa, T., et al. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13859-13864.

22.   Stump, M. R. and Gloss, L. M. (2008) Mutational Analysis of the Stability of the H2A and H2B Histone Monomers. *J. Mol. Biol.* 384: 1369-1383.

23.   Topping, T. B., Hoch, D. A. and Gloss, L. M. (2004) Folding Mechanism of FIS, the Interwined, Dimeric Factor for Inversion Stimulation. *J. Mol. Biol.* 335: 1065-1081.

24.   Bedard, S., Krishna, M. M. G., Mayne, L. and Englander, S. W. (2008) Protein folding: Independent unrelated pathways or predetermined pathway with optional errors. *Proc. Natl. Acad. Sci. U.S.A.* 105: 7182-7187.

25.   Galani, D., Fersht, A. R. and Perrett, S. (2002) Folding of the Yeast Prion Protein Ure2: Kinetic Evidence for Folding and Unfolding Intermediates. *J. Mol. Biol.* 315: 213-227.

26. Seckler, R. (2000) in *Mechanisms of Protein Folding*, eds R. H. Pain (Oxford University Press, Oxford), pp 279-308.

27. Svensson, A.-K. E., Bilsel, O., Kondrashkina, E., Zitzewitz, J. A. and Matthews, C. R. (2006) Mapping the Folding Free Energy Surface for Metal-free Human Cu,Zn Superoxide Dismutase. *J. Mol. Biol.* 364: 1084-1102.

28. Noel, A. F., et al. (2009) The Folding Free-Energy Surface of HIV-1 Protease: Insights into the Thermodynamic Basis for Resistance to Inhibitors. *J. Mol. Biol.* 387: 1002-1016.

29. Pan, H. and Smith, D. L. (2003) Quaternary Structure of Aldolase Leads to Differences in Its Folding and Unfolding Intermediates. *Biochemistry* 42: 5713-5721.

30. Bieri, O. and Kiefhaber, T. (2001) Origin of Apparent Fast and Non-exponential Kinetics of Lysozyme Folding Measured in Pulsed Hydrogen Exchange Experiments. *J. Mol. Biol.* 310: 919-935.

31. Xu, G. and Chance, M. R. (2007) Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* 107: 3514-3543.

32. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

33. Mendoza, V. L. and Vachet, R. W. (2009) Probing Protein Structure by Amino Acid-specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 28: 785-815.

34. Zheng, X., Wintrode, P. L. and Chance, M. R. (2008) Complementary Structural Mass Spectrometry Techniques Reveal Local Dynamics in Functionally Important Regions of a Metastable Serpin. *Structure* 16: 38-51.

35. Chen, J., Rempel, D. L. and Gross, M. L. (2010) Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* 132: 15502-15504.

36. Gruebele, M. (2010) Weighing Up Protein Folding. *Nature* 468: 640-641.

37. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M. and Woodson, S. A. (1998) RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting. *Science* 279: 1940-1943.

38. Kiselar, J. G. and Chance, M. R. (2010) Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J. Mass Spectrom.* 45: 1373-1382.

39. Fritz, G., Botelh, H. M., Morozova-Roche, L. A. and Gomes, C. M. (2010) Natural and amyloid self-assembly of S100 proteins: structural basis of functional diversity. *FEBS J.* 277: 4578-4590.

40. Ohuchida, K., et al. (2006) S100A11, A Putative Tumor Suppressor Gene, Is Overexpressed in Pancreatic Carcinogenesis. *Clin. Cancer Res.* 12: 5417-5422.

41. Tarabykina, S., et al. (2001) The Dimerization Interface of the Metastasis-associated Protein S100A4 (Mts1). *J. Biol. Chem.* 276: 24212-24222.

42. Murzik, U., et al. (2008) Rad54 Targeting to DNA Double-Strand Break Repair Sites Requires Complex Formation with S100A11. *Mol. Biol. Cell* 19: 2926-2935.

43. Dempsey, A. C., Walsh, M. P. and Shaw, G. S. (2003) Unmasking the annexin I interaction from the structure of apo-S100A11. *Structure* 11: 887-897.

44. Marlatt, N. M., Boys, B. L., Konermann, L. and Shaw, G. S. (2009) Formation of Monomeric S100B and S100A11 Proteins at Low Ionic Strength. *Biochemistry* 48: 1954-1963.

45. Pan, J. X., Rintala-Dempsey, A., Li, Y., Shaw, G. S. and Konermann, L. (2006) Folding Kinetics of the S100A11 Protein Dimer Studied by Time-Resolved Electrospray Mass Spectrometry and Pulsed Hydrogen-Deuterium Exchange. *Biochemistry* 45: 3005-3013.

46. Rety, S., et al. (2000) Structural basis of the Ca2+-dependent association between S100C (S100A11) and its target, the N-terminal part of annexin I. *Structure* 8: 175-184.

47. Konermann, L., Stocks, B. B. and Czarny, T. (2010) Laminar Flow Effects During Laser-Induced Oxidative Labeling For Protein Structural Studies by Mass Spectrometry. *Anal. Chem.* 82: 6667–6674.

48. Xu, G., Kiselar, J., He, Q. and Chance, M. R. (2005) Secondary Reactions and Strategies To Improve Quantitative Protein Footprinting. *Anal. Chem.* 77: 3029-3037.

49. Hambly, D. M. and Gross, M. L. (2009) Cold Chemical Oxidation of Proteins. *Anal. Chem.* 81: 7235-7242.

50.   Smith, A. M., Jahn, T. R., Ashcroft, A. E. and Radford, S. E. (2006) Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Formation using Electrospray Ionization Mass Spectrometry. *J. Mol. Biol.* 364: 9-19.

51.   Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A. and Skilling, J. (1992) Disentangling Electrospray Spectra with Maximum Entropy. *Rapid Commun. Mass Spectrom.* 6: 707-711.

52.   Pan, Y., Brown, L. and Konermann, L. (2009) Mapping the Structure of an Integral Membrane Protein under Semi-Denaturing Conditions by Laser-Induced Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* 394: 968-981.

53.   Plaxco, K. W. and Gross, M. (2001) Unfolded, yes, but random? Never ! *Nat. Struct. Biol.* 8: 659-660.

54.   Ratcliffe, K. and Marqusee, S. (2010) Identification of Residual Structure in the Unfolded State of Ribonuclease H1 from the Moderately Thermophilic *Chlorobium tepidum*: Comparison with Thermophilic and Mesophilic Homologues. *Biochemistry* 49: 5167-5175.

55.   Sharp, J. S., Becker, J. M. and Hettich, R. L. (2004) Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Anal. Chem.* 76: 672-683.

56.   Shaw, G. S., Marlatt, N. M., Ferguson, P. L., Barber, K. R. and Bottomley, S. P. (2008) Identification of a Dimeric Intermediate in the Unfolding Pathway for the Calcium-Binding Protein S100B. *J. Mol. Biol.* 382: 1075-1088.

57.   Marenholz, I., Heizmann, C. W. and Fritz, G. (2004) S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomeclature). *Biochem. Biophys. Res. Comm.* 322: 1111-1122.

58.   Gloss, L. M. and Matthews, C. R. (1998) Mechanism of Folding of the Dimeric Core Domain of *Escherichia coli* Trp Repressor: A Nearly Diffusion-Limited Reaction Leads to the Formation of an On-Pathway Dimeric Intermediate. *Biochemistry* 37: 15990-15999.

Chapter 5 - Folding Mechanism of α_1-Antitrypsin to Its Metastable Active State: Insights from Pulsed Oxidative Labeling and Mass Spectrometry

5.1 Introduction

Determining the mechanisms by which proteins attain their biologically active conformation from an initial disordered strcucture remains an ardently pursued goal. Interest in this area stems in large part from the numerous diseases that have protein *mis*folding implicated as a causative factor (1). Studying protein folding in a cellular context is complicated by factors such as molecular crowding, chaperone activity, and co-translational events (2). Computational methods have improved dramatically in recent years, but remain limited to very fast folding proteins (3) and require experimental validation. *In vitro* studies using simplified experimental systems have provided a wealth of insight (4). One of the most direct routes to understanding folding mechanisms involves the structural characterization of transient intermediates formed along the pathway to the native state (5). Unfortunately, such experiments are complicated by the fleeting nature of intermediates which precludes the application of classical high resolution techniques. Optical readouts are widely used for kinetic folding experiments, but the extent of structural information obtainable in this way is usually limited (6). Pulsed hydrogen/deuterium exchange (HDX) is a powerful tool for monitoring secondary structure by reporting on hydrogen bond formation in a spatially-resolved fashion (7, 8).

Mass spectrometry (MS)-based techniques are becoming widely used for studying protein structural changes (9). Mapping protein solvent accessibility at amino acid resolution is possible via covalent labeling methods (10). The combination of MS with hydroxyl radical (·OH) labeling is a particularly powerful approach (11). These experiments rely on the premise that solvent-exposed side chains readily react with ·OH, whereas burial dramatically reduces the extent of labeling. The lifetime of the ·OH in solution can be reduced to ~1 us through the addition of suitable scavengers (12), implying the absence of structural artifacts casused by oxidation-induced conformational changes under "single hit" conditions (13). Most ·OH labeling applications to date have investigated structural changes under equilibrium conditions (14). Due to the short duration of the labeling pulse, ·OH-based methods are well suited for time-resolved experiments. However, this particular application is only beginning to be realized (15, 16).

Members of the serine protease inhibitor (serpin) superfamily, of which α_1-antitrypsin (α_1AT) is the archetype, share a common structural core comprising three β-sheets and eight or nine α-helices (Figure 5.1). Serpins are involved in the regulation of proteolytic cascades, such as those taking place during inflammation and thrombosis (17). Point mutations can give rise to serpin polymerization via a domain swap mechanism, leading to various disease phenotypes (18). The serpin inhibitory function is based on a suicide substrate mechanism. Cleavage of the reactive center loop (RCL) by a target protease results in a long-lived complex that has Ser195 of the protease covalently linked to Met358 of α_1AT via an ester bond. This is followed by insertion of the cleaved RCL into β-sheet A as a new strand (A4), in between strands A3 and A5. Loop insertion
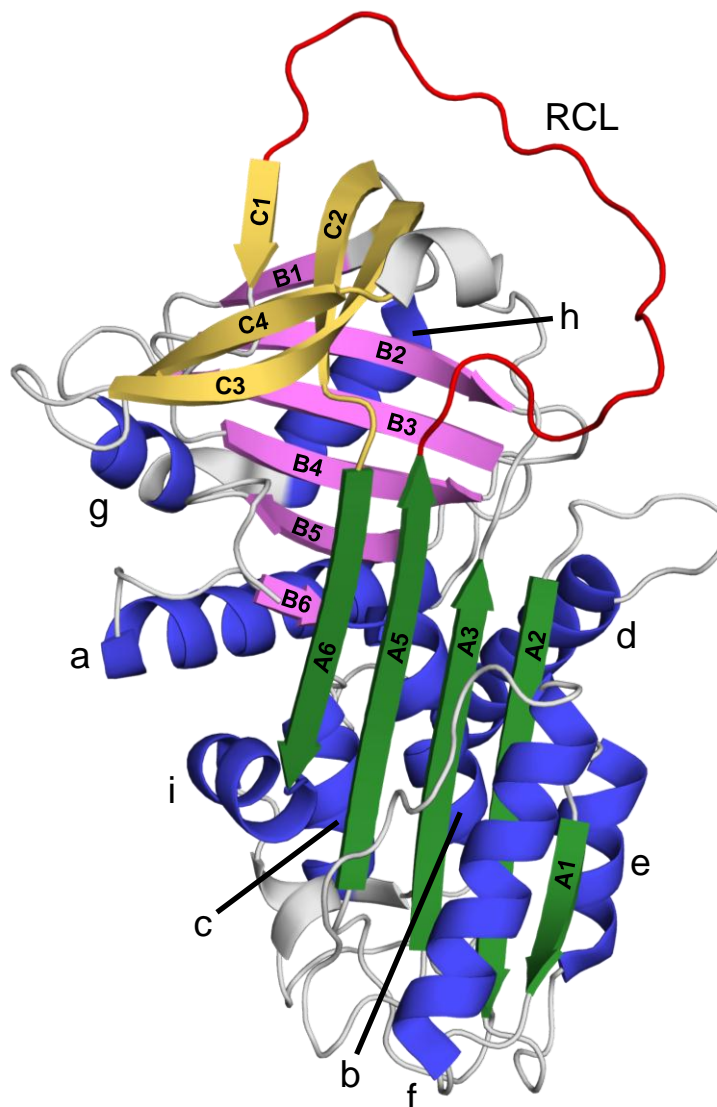
Figure 5.1. X-ray crystal structure of human α₁-antitrypsin in the metastable (active) conformation (PDB ID: 1QLP (19)) α-helices (a-i) are shown in blue and β-sheets A, B, and C are depicted in green, purple, and yellow respectively. Component strands in each β-sheet are labeled X1, X2, etc. where X = A, B, C. The reactive center loop (RCL) is shown in red.

is concomitant with translocation of the protease to the opposite pole of α_1AT (top to bottom of Figure 5.1)(20). The loop-inserted conformation shows an increased thermodynamic stability compared to active α_1AT, revealing that the active conformation represents a *meta*stable state (21). Importantly, the RCL can also insert without prior cleavage to give an enzymatically inactive ("latent") state (22). It is an interesting question how α_1AT folds to its metastable, active state while avoiding its thermodynamically preferred, loop-inserted latent conformation (17).

While there exist many equilibrium studies on α_1AT folding in the literature (23-25), accounts of the time-dependent folding mechanism are sparse. Most small proteins fold to their native state on the order of several seconds or less, however folding of α_1AT is remarkably slow. Kim and Yu have employed ANS fluorescence to show that the initial fast phase of α_1AT folding, collapse of a hydrophobic core, occurs *ca.* 5 s post denaturant dilution (26). Intrinsic fluorescence from Trp194 and Trp238 was also used to define an intermediate phase (τ = 200-500 s) and a slow phase (τ = 1000-3000 s). A recent HDX study by Tsutsui *et al.* proposed a folding mechanism in which early formation of strands C1 and B4 anchor the RCL, rendering it incapable of inserting into β-sheet A (27).

In this work, we complement the results of Tsutsui *et al.* (27) monitoring temporal changes in solvent accessibility of side chains during α_1AT folding using pulsed ·OH labeling. The earliest stages of folding involve a near-global collapse and the formation of a central folding nucleus which sequesters regions of β-sheet A. Such quarantine disfavors loop insertion and allows strand C1 to form and restrict RCL movement, as proposed by Tsutsui *et. al.* (27). While formation of secondary structure lags behind the

formation of tertiary contacts early on, good agreement is seen between the HDX and ·OH labeling data at late folding times. A late kinetic folding intermediate is populated consisting of a solvent-inaccessible core very similar to a previously characterized equilibrium polymerogenic intermediate (28).

5.2 Materials and Methods

5.2.1 Materials

Recombinant human α_1-antitrypsin was expressed and purified as described (29). R101H and E376D substitutions were utilized to facilitate purification. The expected molecular mass for the monomer is 44 291 Da, which matches the measured value within experimental error of $\pm$ 1 Da. Bovine liver catalase, bovine pancreas α-chymotrypsin, glutamine, leucine enkephalin, and bromocresol purple were obtained from Sigma (St. Louis, MO). Bradykinin was supplied by Bachem (King of Prussia, PA) and sequencing-grade trypsin was purchased from Promega (Madison, WI). All chemicals were used without further purification.

5.2.2 Continuous-flow Mixing and Oxidative Labeling

Kinetic protein folding experiments with oxidative labeling were performed using a two-syringe continuous-flow mixing device similar to that employed for previous work (Chapter 4). For early (0.5 and 7 s) reaction times, the mixer was used for both folding initiation and labeling. Later reaction times (2 min, 10 min, 30 min, and 24 hr) employed

off-line, manual mixing to trigger folding followed by labeling using the mixing device.

For 0.5 and 7 s time points, syringes 1 and 2 were advanced at 2.5 and 47.5 μL min^{-1}, respectively, using a syringe pump (Harvard Apparatus, Boston, MA). Syringe 1 contained 200 μM α_1AT denatured in 6 M GdnHCl, 10 mM phosphate buffer (pH 7.8), and 50 mM NaCl. Protein folding was triggered by combining this solution with phosphate buffer (pH 7.8) from syringe 2 at a homemade capillary mixer in a 1:19 volume ratio, resulting in a final denaturant concentration of 0.3 M and a protein concentration of 10 μM. The outlet of the mixer was connected to a reaction capillary with i.d. 100 μm. Syringe 2 also contained 15.8 mM glutamine and 0.105% (v/v) (*ca.* 30 mM) H$_2$O$_2$. A KrF excimer laser (GAM EX 750, Orlando, FL) producing 18 ns pulses at 248 nm, 72 Hz, 37 mJ, and an irradiation spot width of ca. 1 mm was used to generate $\cdot$OH by photolysis of H$_2$O$_2$ within the reaction capillary. Glutamine acts as a radical scavenger that quenches the labeling reaction on a time scale of 1 μs (12). Following initiation of folding, oxidative labeling was performed at different time points by irradiating the reaction mixture at suitable positions downstream of the mixer. Average reaction times of 0.5 s and 7 s correspond to distances between mixer and irradiation spot of 5.3 cm and 74.2 cm, respectively. Effective mixing and reproducible labeling was confirmed as described previously (Chapters 2 and 3). In addition, the mixer performance was verified optically in pH jump experiments on bromocresol purple in the presence of GdnHCl (7).

For 2, 10, and 30 min time points, 200 μM α_1AT denatured in 6 M GdnHCl was diluted into 10 mM phosphate buffer (pH 7.8) to trigger folding. Resultant was a denaturant concentration of 0.4 M and a final protein concentration of 13.34 μM. For

each labeling time, a fraction of this mixture was loaded into syringe 1. The second syringe contained 10 mM phosphate buffer (pH 7.8), 60 mM glutamine and 0.4% (v/v) H_2O_2. Syringes 1 and 2 were advanced at 75 and 25 μL min^{-1}, respectively, using a syringe pump (Harvard Apparatus, Boston, MA) resulting in mixing at the homemade capillary mixing device. The solution was irradiated for 2 min with the KrF excimer laser at a window 74.2 cm downstream of the mixing point. Because folding was initiated via offline mixing, measured folding times were actually between t and $t + 2$ min.

Efficient absorption of the 248 nm photons by GdnHCl precluded the labeling of an unfolded sample in 6 M denaturant. An acid-denatured sample was not feasible due to the polymerization propensity of $\alpha_1 AT$ at low pH (30). Due to the long lag phase during refolding, the 0.5 s sample served as the 'unfolded' control. A refolded sample, however, was included to mark the endpoint of the folding process. This sample was prepared in the same manner as the 2, 10, and 30 min samples and then was loaded into syringe 1 and subjected to laser irradiation after 24 hr of folding. Refolded $\alpha_1 AT$ under the conditions used here has been shown to maintain 70% inhibitory activity, indicating the majority remains in the monomeric, active conformation (27).

200 μL portions of capillary outflow for each irradiated sample were collected in microcentrifuge tubes that contained 15 μL 10 mM phosphate buffer and 1.4 μM catalase at pH 7.8. Catalase was employed for deactivating residual H_2O_2, thereby suppressing undesired secondary oxidation (31, 32). The extent of secondary oxidation was further reduced by flash-freezing all samples in liquid N_2 immediately after collection, and by storage at -80 °C until further analysis.

Each sample was dialyzed overnight against 10 mM phosphate buffer (pH 7.8) for

further GdnHCl dilution to facilitate protease digestion. The resulting solutions were split, with half being digested with trypsin at pH 7.8 and 37 °C for 24 h using a 1:20 (w/w) enzyme:protein ratio and half being digested with chymotrypsin at pH 7.8 and 30 °C for 24 hr using a 1:60 (w/w) enzyme:protein ratio. The digests were lyophilized and resuspended in 50 μL of water containing 0.5 μM bradykinin as an internal intensity standard (33)(Chapter 4). Non-irradiated control samples were collected for every measurement by using the same solution conditions and mixing sequence, except that the laser was switched off.

5.2.3 LC/ESI-MS

Peptide mapping experiments were performed on a Q-TOF Ultima API mass spectrometer (Waters, Milford, MA) equipped with a Z-spray ESI source that was operated at 3 kV. The instrument was coupled to a Waters Acquity UPLC system employing a C18 (BEH300) 2.1 mm x 100 mm column for peptide measurements. Elution was carried out using a water/acetonitrile gradient in the presence of 0.1% formic acid at a flow rate of 100 μL min^{-1} and an injection volume of 10 μL. The identity of tryptic and chymotryptic peptides was confirmed by MS/MS.

Oxidation sites for each tryptic and chymotryptic peptide were determined using a Synapt mass spectrometer (Waters, Milford, MA), operating in either data-dependent acquisition (DDA) mode or in MS^{E} mode, coupled to a Waters Acquity UPLC system. For DDA experiments, fragmentation of the two most abundant eluent precursor ions was performed in the trap using optimized collision energies. Five rounds of iterative exclusion were used to ensure maximal sequence coverage (34) and data were analyzed

using Peaks (BSI Solutions, Waterloo, ON) (35). For MS^E experiments, the trap was maintained at 6 V during the low-energy scans and was ramped from 14 to 34 V in the high energy scans. Leucine enkephalin (m/z 556.2771^{1+}) was used as a lock mass. Data were analyzed using the ProteinLynxGlobalServer software package provided by the manufacturer. All oxidation site assignments were verified manually.

5.2.4 Data Analysis

In accordance with our previous studies, oxidation levels of specific protein segments were observed by following the intensity of unmodified tryptic peptides as a function of folding time, t (Chapters 3 and 4). Relative signal intensities of unlabeled proteolytic peptides, $R_u^{app}(t)$, were generated by dividing the UPLC/MS peak heights of each peptide by the peak height of the bradykinin internal standard for each folding time t (33)(Chapter 2). Peak heights were utilized rather than peak areas to negate complications associated with separating low abundance isotopologues from baseline noise. The resulting signal intensities were corrected for trace amounts of background oxidation by

$$R_u^{corr} = \frac{R_u^{app}}{R_u^{bgr}}$$
(5-1)

where R_u^{bgr} is the bradykinin corrected peak height of the corresponding protein segment found in the unlabeled control sample. Normalized oxidation levels (NOLs) for each peptide were calculated using the corrected relative signal intensities according to

$$NOL = \frac{\ln R_u^{corr}(t) - \ln R_u^{corr}(24hr)}{\ln R_u^{corr}(0.5s) - \ln R_u^{corr}(24hr)} \qquad (5\text{-}2)$$

$R_u^{corr}(t)$ is the background corrected, relative signal intensity of a particular unmodified peptide at folding time t, $R_u^{corr}(0.5\ s)$ is the corresponding value of the "denatured" control ($t = 0.5$ s), and $R_u^{corr}(24\ hr)$ is that of the refolded control sample ($t = 24$ hr). Equation 5-2 provides direct information on the solvent exposure of individual tryptic peptides at a given folding time (Chapter 2). Error bars shown in Fig. 5.3 were calculated using the standard deviation of triplicate measurements.

5.3 Results and Discussion

5.3.1 Peptide Mapping and Oxidation Site Determination

α_1AT was initially unfolded in 6 M guanidinium hydrochloride (GdnHCl) at pH 7.8 for 2 hr. Folding was initiated via a denaturant dilution from 6 M to 0.3 M GdnHCl. This ensured that the protein was far below the folding transition midpoint (23). We subjected α_1AT to oxidative labeling at 7 s, 2 min, 10 min, and 30 min after GdnHCl dilution. To mark the endpoint of the reaction, a sample corresponding to $t = 24$ hr was also prepared. The strong absorption of the 248 nm photons by concentrated GdnHCl precluded the labeling of fully denatured α_1AT. However, the fact that early collapse of the protein is so slow (26) justifies the use of $t = 0.5$ s samples as a "pseudo-unfolded" reference state.

To garner detailed insights into protein structural changes accompanying folding, spatially resolved labeling data are required. Tryptic digestion of α_1AT after $\cdot$OH exposure resulted in 28 observable peptides. Two such segments (spanning residues 40-135) were exceedingly long, necessitating the usage of chymotrypsin to obtain 7 shorter peptides. With this combination of trypsin and chymotrypsin, 33 peptides covering 93% of the sequence were detected (Figure 5.2). The remainder of the sequence comprises small, hydrophilic segments that likely did not adhere to the C18 column. Residue-level mapping of oxidation sites was performed via MS/MS experiments on a pulse labeled 0.5 s refolding sample. Typical spectra utilized for oxidation site determination are shown in Figure 5.3. The presence of b_5-H_2O+16 and y_4+16 peaks in Fig 5.3A, coupled with the b_4-H_2O peak and absence of a y_4 peak, lead to the assignment of Y160 as the modified residue. The interpretation of the spectrum shown in Fig. 5.3B is more complex. This results from the simultaneous fragmentation of two isobaric peptides. The y_2+48 ion indicates that C232 has been fully oxidized to the sulfonic acid. The b_3+16 peak indicates a modification between L224 and M226. The y_9+64 peak leads to M226 as an oxidation site but the presence of a y_9+48 ion necessitates a second site carrying a +16 Da modification. Since oxidation of glycine leads to backbone cleavage, the remaining oxidation site is localized to L224. A total of 69 side chains were found to be oxidized, which is considerably more than a previous study of native α_1AT (36). The locations of the oxidation sites within the α_1AT structure are mapped in Figure 5.4.

```
                                              a                        B6
 1EDPQGDAAQK  TDTSHHDQDH  PTFNKITPNL  AEFAFSLYRQ  LAHQSNSTNI
    P1             P2            P3            P4

            b                      c                      d
51FFSPVSIATA  FAMLSLGTKA  DTHDEILEGL  NFNLTEIPEA  QIHEGFQELL
       P5            P6                    P7

                  A2                  e              A1
101HTLNQPDSQL  QLTTGNGLFL  SEGLKLVDKF  LEDVKKLYHS  EAFTVNFGDT
        P8           P9         P10          P11          P12

         f                              A3
151EEAKKQINDY  VEKGTQGKIV  DLVKELDRDT  VFALVNYIFF  KGKWERPFEV
      P13       P14   P15                 P16              P17

       C4              C3            B1             B2          B3
201KDTEEEDFHV  DQVTTVKVPM  MKRLGMFNIQ  HCKKLSSWVL  LMKYLGNATA
        P18           P19          P20           P21          P22

            g          h               C2               A6
251IFFLPDEGKL  QHLENELTHD  IITKFLENED  RRSASLHLPK  LSITGTYDLK
         P23              P24           P25              P26

      i                                        A5         Reactive
301SVLGQLGITK  VFSNGADLSG  VTEEAPLKLS  KAVHKAVLTI  DEKGTEAAGA
       P27              P28                       P29

  Center Loop    C1              B4              B5
351MFLEAIPMSI  PPEVKFNKPF  VFLMIDQNTK  SPLFMGKVVN  PTQK
       P30              P31          P32      P33
```

Figure 5.2. Amino acid sequence of the α$_1$-AT construct used in this work. Secondary structure elements are depicted in colors corresponding to those in Fig. 5.1. Proteolytic peptides used for kinetic analysis are shown as double-headed line segments and labeled sequentially (P1, P2, etc.). Solid lines indicate peptides from tryptic digests and chymotryptic peptides are shown as dotted.
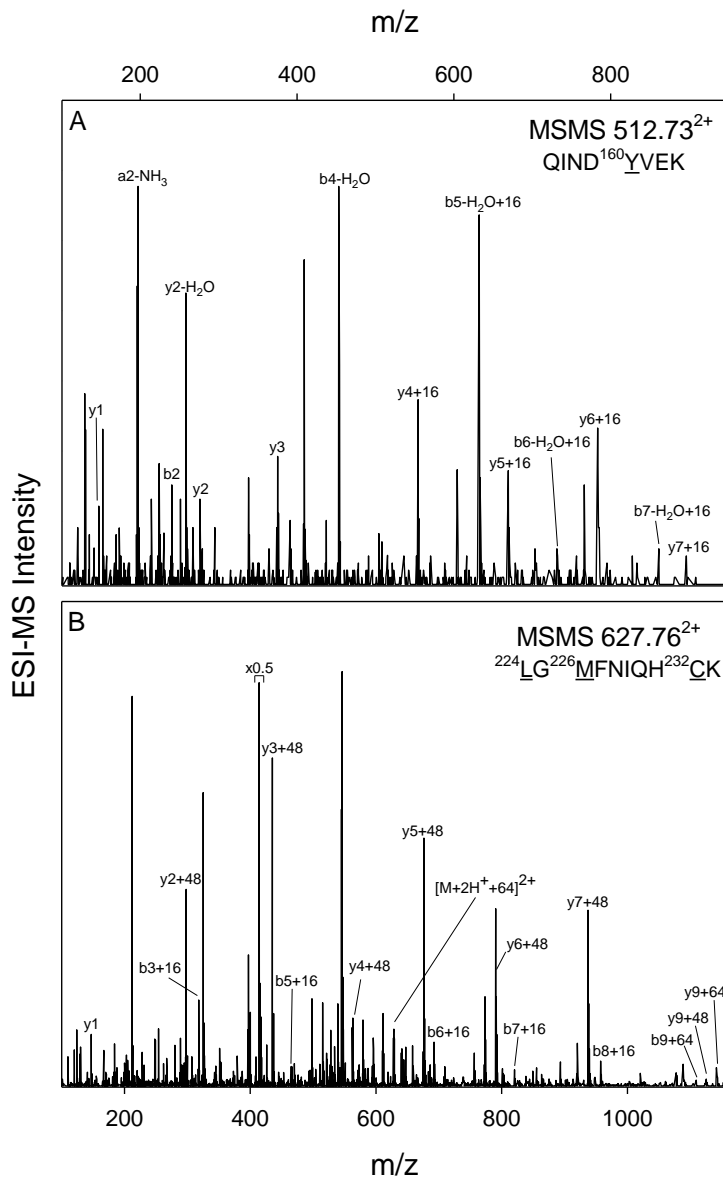
Figure 5.3. Tandem-MS spectra of oxidatively modified proteolytic α_1AT peptides. Most major peaks are labeled using standard *b-* and *y-ion* notation. (A) MSMS of [P13+16]. (B) MSMS of [P20+64].
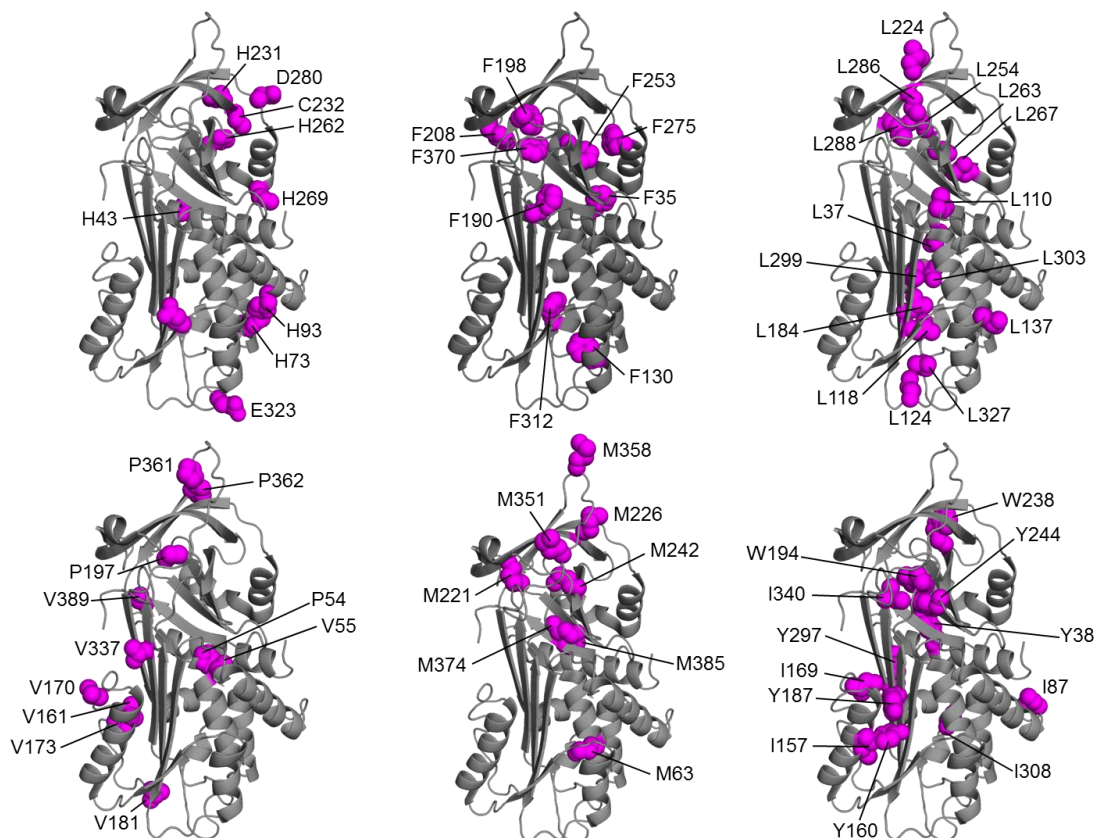
Figure 5.4. Amino acid side chains determined to be oxidized in denatured α1AT mapped onto the native crystal structure (PDB ID: 1QLP (19)). Sites are shown as magenta spheres and broadly grouped according to amino acid type.

5.2.2 Time-Dependent Peptide Solvent Accessibilities

Determination of temporal changes in peptide solvent accessibilities is facilitated by examining their normalized oxidation levels (NOLs). NOL = 1 signifies solvent accessibility identical to that found in the 0.5 s reference sample, while NOL = 0 indicates an accessibility matching that of the refolded α_1AT. Solvent accessibility levels intermediate to these two endpoints are characterized by $0 < $ NOL $ < 1$. The resulting temporal NOL profiles can be broadly clustered by oxidation behavior. Segregation of curves into multiple panels (A-H) was done to reduce clutter and does not imply 8 distinct kinetic regimes. Figure 5.5A represents protein segments that become significantly protected after 7 s of folding and progress rapidly to native-like NOL values. Intermediate NOL progressions are depicted in panels B-G. Of note are P2 and P22 (Fig. 3C) which attain over half of their native state solvent accessibility after 2 min of folding, but show only modest decreases at 30 min. The seemingly stepwise decrease in accessibility of the protein segments in Fig. 5.5E is also notable. After a rapid collapse between 0.5 and 7 s, very little change in accessibility is observed in the 2 min sample. Similar plateaus in oxidation level can be seen between 10 and 30 min of folding. The peptides shown in panels F and G, while exhibiting temporally distinct protection patterns, are those with the highest solvent accessibility after even 30 min, whereas nearly all other segments have reached native-like levels. Interestingly, P31 (Fig. 5.5H) exhibits the highest solvent accessibility as late as 10 min after onset of folding, but is among the most native-like segments at 30 min.

Interpretation of the conformational changes during α_1AT folding is facilitated by mapping the NOL values from Figure 5.5 onto the crystal structure of the protein (Figure
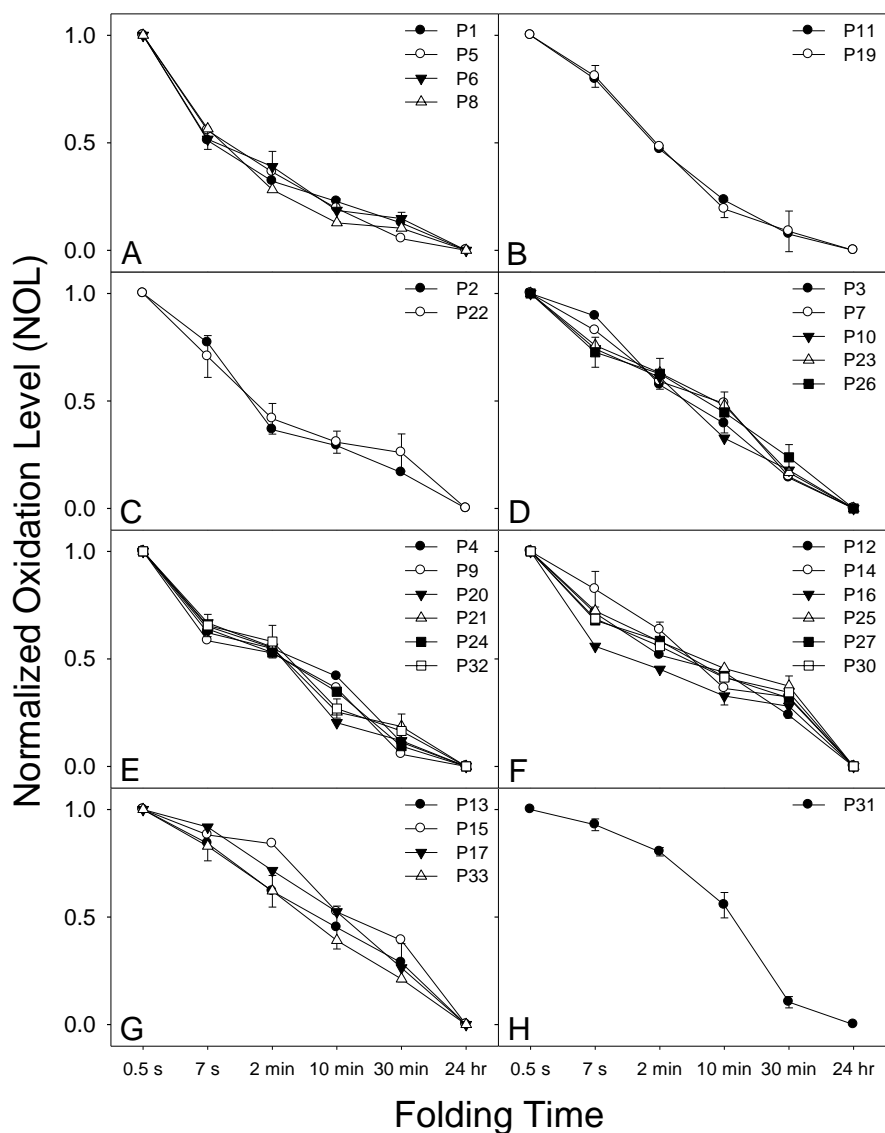
Figure 5.5. Normalized oxidation levels (NOLs) of α_1-AT proteolytic peptides as a function of folding time. Peptides are broadly grouped according to kinetic behavior. All data points are averages of triplicate measurements. Error bars are shown for only one peptide per panel for clarity. In panel F, two traces have been omitted to reduce clutter: P18 and P29, which exhibit virtually identical behavior to P30 and P16, respectively.

5.6, A-D). While peptide oxidation behavior was attributed to the modification of a subset of reactive residues (Fig. 5.2, bold), entire proteolytic peptides are colored in Fig. 5.6 to reduce clutter. All modified residues that were detected are shown with respect to their locations within the crystal structure in Fig. 5.4. For comparative purposes, recent kinetic HDX data of Wintrode and coworkers (27) are also mapped onto the α_1AT crystal structure (Fig. 5.6, E-H).

5.2.3 The Denatured State

To fully understand the folding pathway of a protein, it is instructive to characterize the initial state. A previous tyrptophan fluorescence study reported residual structure around W238 in the urea denatured state of α_1AT (24). In the native state, the side chain of W238 sits in a hydrophobic pocket consisting of I229 and H231 from strand B1 and V364 from strand C1. I229 and V364 have been shown to be critical for proper α_1AT folding (37). H231 and W238 were found to be oxidized after only 0.5 s of folding (Fig. 5.4) and continue to exhibit extensive solvent accessibility after 7 s (Fig. 5.6A). This finding implies residual water access in the hydrophobic core as a result of incomplete side chain packing.

5.2.4 α_1AT Folding Mechanism

For the following discussion it is helpful to conduct a side-by-side comparison of our ·OH labeling data with the HDX results of Tsutsui *et al.*(27)(Fig. 5.6). As seen in Fig. 5.6A, a near-global decrease in solvent accessibility has occurred at $t = 7$ s, indicating a rapid collapse of the protein chain. This observation is in agreement with previous
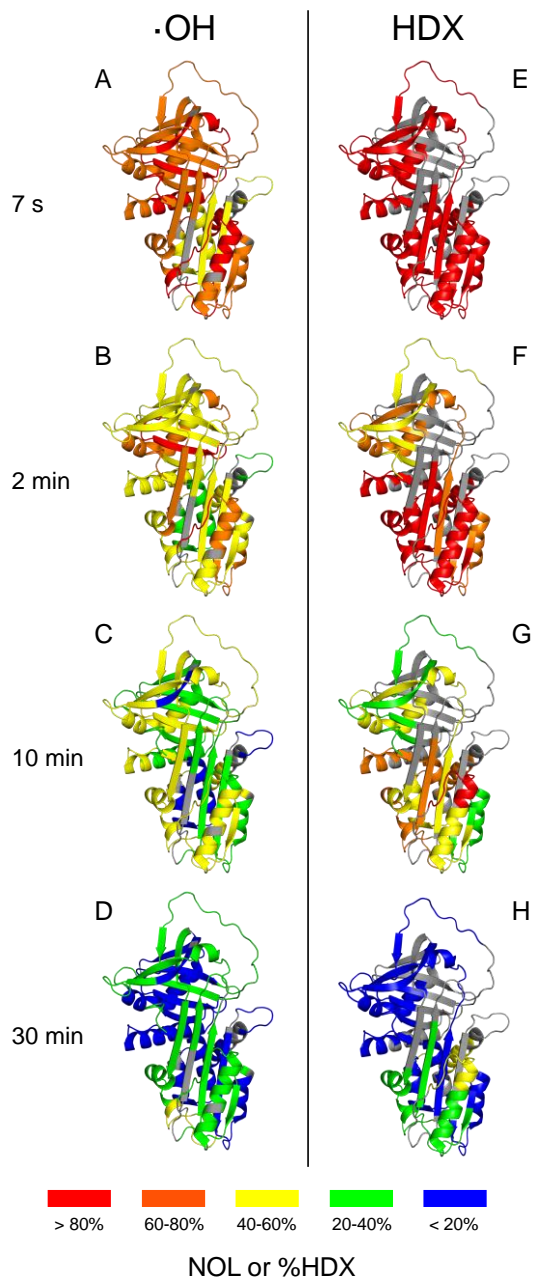
Figure 5.6. Structural changes during α_1AT folding probed with pulsed oxidative ($\cdot$OH) labeling after (A) 7 s, (B) 2 min, (C) 10 min, and (D) 30 min of folding. Hydrogen/deuterium exchange (HDX) levels at corresponding folding times from ref. (27) are shown on the right (E-H) for comparison. Both NOLs and HDX levels are depicted using a five color code.

reports (26). It has been shown that collapse precedes secondary structure formation (38). Interestingly, HDX reveals no stable hydrogen bonding at this time point (Fig. 5.6E)(27). While the specificity of protein collapse remains up for debate (39), the formation of a structural folding core (Fig. 5.6A, yellow) is concomitant with the incipient collapse seen for α_1AT. The core consists of protein segments comprising helices b and c, and strands A2 and A3. These elements are non-contiguous (Fig. 5.2) but it is apparent from the structure in Fig. 5.6A that they form multiple tertiary contacts. Since no stable secondary structure exists on this time scale (Fig. 5.6E), we postulate that the early formation of tertiary contacts between helices b/c and A2 and A3 allow the strands to be sequestered, which disfavors RCL association with β-sheet A. This is crucial since the thermodynamically preferred conformation of α_1AT has the RCL inserted into sheet A, although such a state is enzymatically inactive.

A further decrease in accessibility is observed for much of the protein sequence after 2 min of folding (Fig. 5.6B). Notable exceptions include strand B4 along with helices g and h, segments which were previously proposed to be significantly involved in the residual structure found in the denatured state (24). This provides further evidence that the residual structure contains non-native side chain packing. Helices b and c have attained further protection from modification while A2 and A3 exhibit levels similar to those found in the 7 s sample, indicating that the helices form the foundation of the folding core seen in Fig. 5.6A. While B4 is among the most solvent exposed regions after 2 min it, along with C1, are the first to acquire stable secondary structure (Fig. 5.6F, yellow). Tsutsui *et. al.* propose that the early formation of these elements locks the RCL in a conformation that is unable to insert into the yet unformed A-sheet, signifying the

crucial step in maintaining α_1AT in its metastable form (27). Previous work has also shown strand C1 to be critical to maintaining the active, native structure of α_1AT (37, 40).

The interpretation that helices b and c form a folding nucleus gains traction when the NOL values for the 10 min folding time point are structurally mapped (Fig. 5.6C). Helices b and c have reached native-like NOL levels (blue in Fig. 5.6C), while those of the majority of all other segments remain elevated. From the orientation shown in Figure 5.6C, a back-to-front folding sequence becomes apparent. While helices b and c are native-like, strands A2 and A3 exhibit slightly elevated accessibilities (coloured green). These two strands form tertiary contacts early during folding with helices b and c. Ultimately helix f (coloured yellow) packs on the front side of β-sheet A, but after 10 min displays significantly higher accessibility than A2 and A3 indicating that such packing is far from native. From the orientation shown in Fig. 5.6C, the temporal folding of the bottom half of α_1AT can be said to progress via a back-to-front pathway where the packing of helices b and c is native-like, A2 and A3 are slightly less native, and helix f is even more disordered.

Fig. 5.6D further exemplifies the proposed back-to-front folding progression as the majority of the back side of the protein exhibits native-like accessibility (colored blue), while those elements packed onto the front show slightly elevated accessibility levels (green). It also becomes apparent that β-sheet B is virtually native after 30 min of folding, while sheets A and C continue to exhibit non-native accessibilities. A possible explanation for these data is the antiparallel nature of sheet B. Antiparallel sheets have a tendency to fold faster than parallel sheets (41). Sheet B is the sole continuously

antiparallel sheet in active α_1AT; sheet A only becomes antiparallel once the RCL inserts as the sixth strand. Helix f also displays elevated solvent accessibility at 30 min. Conformational flexibility in this region has been shown to be indispensible to the serpin inhibitory mechanism as it sterically blocks protease translocation in the native state (42). I157 and V161, both of which were found to be oxidized, have been shown to be crucial to α_1AT stability by forming hydrophobic contacts with strands A2 and A3 (43). The non-native accessibility in this region indicates that the packing interface has not yet fully formed. While hydrogen bond formation lagged behind tertiary packing for much of the folding progression, both the backbone and side chains display native-like structure in many regions after 30 min (Fig. 5.6D, H). Further structural consolidation of all these elements lead ultimately to native-like packing in the refolded sample.

5.2.5 Implications for α_1AT Aggregation

Serpin polymerization has been linked to diseases such as emphysema, cirrhosis and dementia (44). A crystal structure of a self-terminating dimer of antithrombin has shed light into the polymerization mechanism and a potential polymerogenic folding intermediate has been suggested based on limited proteolysis (45). In the intermediate state, it is proposed that the contiguous segment from strand A6 to strand C1, consisting of A5, helix i and the RCL (Fig. 5.2), is unfolded. This region can then insert as strands A4 and A5 in a second monomer and so on. The same contiguous elements postulated to be unstructured in the polymerogenic intermediate, helix i-A5-RCL-C1, exhibit elevated solvent accessibilities and non-native hydrogen bonding in the kinetic intermediate formed at 30 min (Fig. 5.6D, H). Krishnan and Gierasch have utilized PEGylation to

probe the solvent accessiblity of engineered Cys residues within α_1AT as a function of denaturant concentration (28). In the intermediate state populated near 1.5 M GdnHCl, mutated Cys sites exhibited a range of accessibilities. Figure 5.7A shows regions that were determined to be solvent inaccessible (dark gray) and those that were found to be easily modified (white). The structural core of the 30 min kinetic folding intermediate characterized in our study (Fig. 5.7B, blue) bears remarkable similarity to the equilibrium species. Resemblance between equilibrium and kinetic intermediates has previously been observed for small proteins such as ribonuclease H (46) and apomyoglobin (47). Krishnan and Gierasch also investigated the solvent accessibility of the polymerogenic Z-form of α_1AT (containing an E342K mutation) and reported similar solvent accessibilities to those of the native protein (28). Although wild-type α_1AT is not known to polymerize under physiological conditions, when destabilized it may populate an intermediate with similar topology to that of the Z-form. Our findings allow for the proposal that a comparable folding intermediate may also be populated during kinetic refolding.

While it has long been known that mutated α_1AT forms polymers *in vivo* (48), it has recently been proposed that the strand A4/A5 domain swap may not be the true mechanism, but rather a consequence of *in vitro* chemical denaturation (18). This assessment builds on previous work which showed that antibodies specific for α_1AT polymers formed *in vivo* react with polymers fromed via heating, but not by incubation with GdnHCl (48). Yamasaki *et. al.* crystallized an α_1AT trimer that reacted strongly with antibody and revealed a C-termainal domain (C1/B4/B5) swap (18). The serpin folding pathway is thought to pass through a branch point at which an intermediate capable of folding to the native state or forming polymers is populated. It has been proposed that the
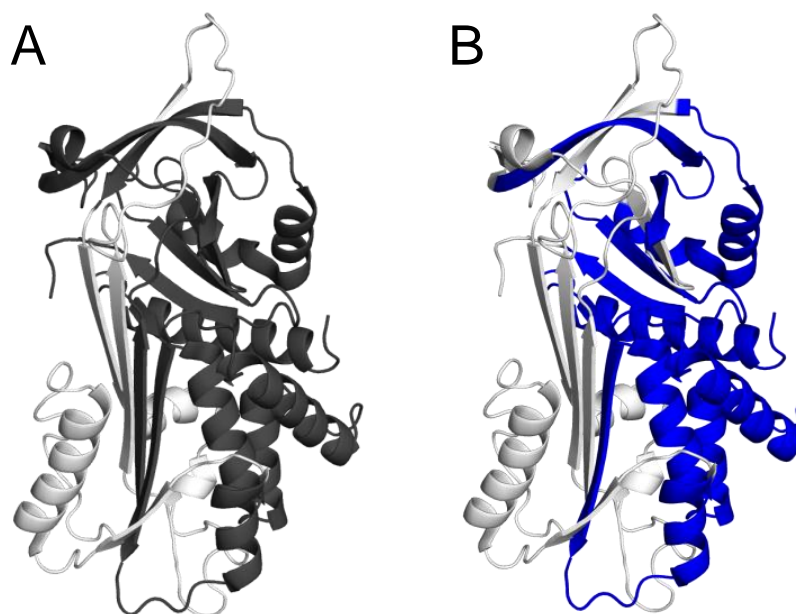
Figure 5.7. Comparison of structural cores of α_1AT folding intermediates exhibiting native-like solvent accessibility. (A) Structural core, depicted in dark gray, of the equilibrium intermediate of α_1AT populated at 1.5 M GdnHCl probed by PEGylation of engineered cysteine residues by Krishnan and Gierasch (28). (B) Structural core, shown in blue, of the kinetic intermediate of α_1AT populated after 30 min of folding, probed by pulsed oxidative labeling. These regions correspond to those colored blue in Figure 5.6D. In both (A) and (B) solvent accessible regions are shown in white.

outcome of the intermediate is determined by a competition between burying the C-terminus into the core and completion of β-sheet A (18). Our ·OH labeling data suggest that helices b/c sequester strands A2 and A3 via tertiary contacts early during folding (Fig. 5.6A). This association prevents further formation of the A sheet and thus renders the RCL incapable of insertion. This interpretation is complemented by HDX data which show the C-terminus is among the first regions to form stable hydrogen bonds (Fig. 5.6F)(27). After 30 min of folding, the C-terminus shows both native-like solvent accessibility and hydrogen bonding while much of β-sheet A exhibits non-native levels (Fig. 5.6D, H). We propose that the early segregation of strands A2 and A3 allow time for the C-terminus to fold and begin insertion into the protein core. Since the formation of β-sheet A is compromised, RCL insertion is blocked. This virtually negates polymerization via C-terminal domain swapping and promotes the folding of α_1AT to its metastable, active state.

5.4 Conclusions

Folding to a metastable state, rather than the thermodyamic product, is essential for serpin function. Previous kinetic studies have indicated that α_1AT folding is quite protracted. Tryptophan fluorescence and HDX data reveal that α_1AT only attains its active structure after ~1 h (26, 27). This work utilized hydroxyl radical labeling to gain insights into changes in side chain solvent accessibility as a function of folding time. Comparison of our ·OH labeling results with previously published HDX data (27) allows the proposal of a mechanism by which α_1AT becomes kinetically trapped in its metastable state.

An initial structural collapse occurs within the first ~10 s (26) although this transpires in the absence of any stable hydrogen bonding (27). We find that the early collapse is accompanied by significant solvent protection within helices b and c. Through tertiary contacts, these early-forming helices sequester strands A2 and A3. This collapse of β-sheet A renders it unavailable for RCL insertion. The solvent accessibility of many parts of β-sheet A remains elevated even after 30 min of folding. A structural core, characterized by native-like solvent accessibility, becomes apparent on a similar timescale. This core shows high similarity to that of a previously characterized polymerogenic equilibrium unfolding intermediate of α_1AT (28).

Tsutsui *et al.* monitored secondary structure formation during α_1AT folding with HDX (27). The formation of strand C1, on a timescale of minutes, was proposed to anchor the RCL in an open conformation which is unable to insert into sheet A. Our ·OH labeling data suggest that early contacts between helices b/c and strands A2/A3 disfavor RCL insertion by delaying structural consolidation of sheet A. These interactions provide the time necessary for strand C1 to form. Hydrogen bond formation lags behind side chain protection for much of the folding pathway. However, unconsolidated hydrogen bonding after 30 min occurs in regions which also exhibit elevated solvent accessibility.

Overall, our study demonstrates the ability of ·OH labeling to characterize partially structured species along the folding pathway of α_1AT. The combination of complementary labeling techniques (·OH labeling and HDX) can provide a comprehensive view of structural changes occuring during folding.

5.5 References

1. Chiti, F. and Dobson, C. M. (2006) Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* 75: 333-366.

2. Gershenson, A. and Gierasch, L. M. (2011) Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.* 21: 32-41.

3. Lindorff-Larsen, K., Piana, S., Dror, R. O. and Shaw, D. E. (2011) How Fast-Folding Proteins Fold. *Science* 334: 517-520.

4. Bartlett, A. I. and Radford, S. E. (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* 16: 582-588.

5. Brockwell, D. J. and Radford, S. E. (2007) Intermediates: ubiquitous species on folding energy landscapes? *Curr. Op. Struct. Biol.* 17: 30-37.

6. Culik, R. M., Serrano, A. L., Bunagan, M. R. and Gai, F. (2011) Achieving Secondary Structural Resolution in Kinetic Measurements of Protein Folding: A Case Study of the Folding Mechanism of Trp-cage. *Angew. Chem. Int. Ed.* 50: 10884-10887.

7. Uzawa, T., et al. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U.S.A.* 105: 13859-13864.

8. Pan, J., Han, J., Borchers, C. H. and Konermann, L. (2010) Characterizing Short-Lived Protein Folding Intermediates by Top-Down Hydrogen Exchange Mass Spectrometry. *Anal. Chem.* 82: 8591-8597.

9. Ben-Nissan, G. and Sharon, M. (2011) Capturing protein structural kinetics by mass spectrometry. *Chem. Soc. Rev.* 40: 3627-3637.

10. Mendoza, V. L. and Vachet, R. W. (2009) Probing Protein Structure by Amino Acid-specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* 28: 785-815.

11. Xu, G. and Chance, M. R. (2007) Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* 107: 3514-3543.

12. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

13. Gau, B. C., Sharp, J. S., Rempel, D. L. and Gross, M. L. (2009) Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* 81: 6563–6571.

14. Wang, L. and Chance, M. R. (2011) Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Anal. Chem.* 83: 7234-7241.

15. Chen, J., Rempel, D. L. and Gross, M. L. (2010) Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* 132: 15502-15504.

16. Konermann, L., Pan, Y. and Stocks, B. B. (2011) Protein folding mechanisms studied by pulsed oxidative labeling and mass spectrometry. *Curr. Opin. Struct. Biol.* 21: 634-640.

17. Gettins, P. G. W. (2002) Serpin Structure, Mechanism, and Function. *Chem. Rev.* 102: 4751-4803.

18. Yamasaki, M., Sendall, T. J., Pearce, M. C., Whisstock, J. C. and Huntington, J. A. (2011) Molecular basis of α_1-antitrypsin deficiency revealed by the structure of a domain-swapped trimer. *EMBO Rep.* 12: 1011-1018.

19. Elliot, P. R., Pei, X. Y., Dafforn, T. R., and Lomas, D. A. (2000) Topography of a 2.0 Å Structure of α_1-Antitrypsin Reveals Targets for Rational Drug Design to Prevent Conformational Disease. *Protein Sci.* 9: 1274-1281.

20. Huntington, J. A., Read, R. J. and Carrell, R. W. (2000) Structure of a serpin-protease complex shows inhibition by deformation. *Nature* 407: 923-926.

21. Whisstock, J. C. and Bottomley, S. P. (2006) Molecular gymnastics: serpin structure, folding and misfolding. *Curr. Opin. Struct. Biol.* 16: 761-768.

22. Lomas, D. A., Elliott, P. R., Chang, W.-S. W., Wardell, M. R. and Carrell, R. W. (1995) Preparation and Characterization of Latent α_1-Antitrypsin. *J. Biol. Chem.* 270: 5282-5288.

23. James, E. L., Whisstock, J. C., Gore, M. G. and Bottomley, S. P. (1999) Probing the Unfolding Pathway of α_1-antitrypsin. *J. Biol. Chem.* 274: 9482-9488.

24. Tew, D. J. and Bottomley, S. P. (2001) Probing the Equilibrium Denaturation of the serpin α_1-Antitrypsin with Single Tryptophan Mutants; Evidence for Structure in the Urea Unfolded State. *J. Mol. Biol.* 313: 1161-1169.

25. Tsutsui, Y. and Wintrode, P. L. (2007) Cooperative Unfolding of a Metastable Sperpin to a Molten Globule Suggests a Link Between Functional and Folding Energy Landscapes. *J. Mol. Biol.* 371: 245-255.

26.   Kim, D. and Yu, M.-H. (1996) Folding Pathway of Human α_1-antitrypsin: Characterization of an Intermediate That Is Active but Prone to Aggregation. *Biochem. Biophys. Res. Comm.* 226: 378-384.

27.   Tsutsui, Y., Dela Cruz, R. G. and Wintrode, P. L. (2012) The folding mechanism of the metastable serpin α_1-antitrypsin. *Proc. Natl. Acad. Sci. U.S.A.* in press: doi/10.1073/pnas.1109125109.

28.   Krishnan, B. and Gierasch, L. M. (2011) Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat. Struct. Mol. Biol.* 18: 222-226.

29.   Tsutsui, Y., Liu, L., Gershenson, A. and Wintrode, P. L. (2006) The Conformational Dynamics of a Metastable Serpin studied by Hydrogen Exchange and Mass spectrometry. *Biochemistry* 45: 6561-6569.

30.   Devlin, G. L., Chow, M. K. M., Howlett, G. J. and Bottomley, S. P. (2002) Acid Denaturation of α_1-Antitrypsin: Characterization of a Novel Mechanism of Serpin Polymerization. *J. Mol. Biol.* 324: 859-870.

31.   Xu, G., Kiselar, J., He, Q. and Chance, M. R. (2005) Secondary Reactions and Strategies To Improve Quantitative Protein Footprinting. *Anal. Chem.* 77: 3029-3037.

32.   Hambly, D. M. and Gross, M. L. (2009) Cold Chemical Oxidation of Proteins. *Anal. Chem.* 81: 7235-7242.

33.   Smith, A. M., Jahn, T. R., Ashcroft, A. E. and Radford, S. E. (2006) Direct Observation of Oligomeric Species formed in the Early Stages of Amyloid Formation using Electrospray Ionization Mass Spectrometry. *J. Mol. Biol.* 364: 9-19.

34.   Bendall, S. C., et al. (2009) An Enhanced Mass Spectrometry Approach Reveals Human Embryonic Stem Cell Growth Factors in Culture. *Mol. Cell. Proteomics* 8: 421-432.

35.   Lacerda, C. M. R., Xin, L., Rogers, I. and Reardon, K. F. (2008) Analysis of iTRAQ data using Mascot and Peaks quantification algorithms. *Brief. Funct. Genomic. Proteomic.* 7: 119-126.

36.   Zheng, X., Wintrode, P. L. and Chance, M. R. (2008) Complementary Structural Mass Spectrometry Techniques Reveal Local Dynamics in Functionally Important Regions of a Metastable Serpin. *Structure* 16: 38-51.

37.   Im, H., Woo, M.-S., Hwang, K. Y. and Yu, M.-H. (2002) Interactions Causing the Kinetic Trap in Serpin Protein Folding. *J. Biol. Chem.* 277: 46347-46354.

38.   Sadqi, M., Lapidus, L. J. and Munoz, V. (2003) How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. U.S.A.* 100: 12117-12122.

39.   Dasgupta, A. and Udgaonkar, J. B. (2010) Evidence for Initial Non-specific Polypeptide Chain Collapse During the Refolding of the SH3 Domain of PI3 Kinase. *J. Mol. Biol.* 403: 430-445.

40.   Bottomley, S. P., et al. (2001) The role of strand 1 of the C β-sheet in the structure and function of α_1-antitrypsin. *Protein Sci.* 10: 2518-2524.

41.   Dinner, A. R. and Karplus, M. (1998) A metastable state in folding simulations of a protein model. *Nat. Struct. Biol.* 5: 236-241.

42.   Gettins, P. G. W. (2002) The F-helix of serpins plays an essential, active role in the proteinase inhibition mechanism. *FEBS Lett.* 523: 2-6.

43.   Cabrita, L. D., Dai, W. and Bottomley, S. P. (2004) Different Conformational Changes within the F-Helix Occur during Serpin Folding, Polymerization, and Proteinase Inhibition. *Biochemistry* 43: 9834-9839.

44.   Lomas, D. A. and Carrell, R. W. (2002) Serpinopathies and the conformational dementias. *Nat. Rev. Genet.* 3: 759-768.

45.   Yamasaki, M., Li, W., Johnson, D. J. D. and Huntington, J. A. (2008) Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* 455: 1255-1258.

46.   Raschke, T. M. and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules under native conditions. *Nat. Struct. Biol.* 4: 298-304.

47.   Nishimura, C., Dyson, H. J. and Wright, P. E. (2005) Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments. *Proc. Natl. Acad. Sci. U.S.A.* 102: 4765-4770.

48.   Lomas, D. A., Evans, D. L., Finch, J. T. and Carrell, R. W. (1992) The mechanism of Z α_1-antitrypsin accumulation in the liver. *Nature* 357: 605-607.

49.   Ekeowa, U. I., et al. (2010) Defining the mechanism of polymerization in the serpinopathies. *Proc. Natl. Acad. Sci. U. S. A.* 107: 17146-17151.

Chapter 6 – Conclusions

6.1 Summary

The detailed characterization of transient folding intermediates provides mechanistic insights into the structural transitions of proteins however such experiments are complicated by the short lifetimes of these species in solution. A novel method for studying protein folding reactions was introduced in this work. Continuous-flow rapid mixing was utilized in conjunction with pulsed oxidative labeling and mass spectrometry. This allowed for changes in side-chain solvent accessibility to be measured in a temporal fashion. The experiments yield data that are complementary to those acquired by HDX studies and other experimental approaches. The applicability of pulsed oxidative labeling for studying protein unfolding (Chapter 2), protein folding (Chapters 3 and 5), and coupled folding/assembly processes (Chapter 4) was demonstrated.

The aim of the work in Chapter 2 was to assess the applicability of pulsed oxidative labeling combined with rapid mixing for studying a protein unfolding reaction. A rigorous data analysis method was devised such that oxidation patterns of peptides with differing reactivities with ·OH could be compared on the basis of solvent accessibility. Time-resolved labeling data in conjunction with stopped-flow absorbance measurements provided insights into the unfolding mechanism of myoglobin. Structural characterization of two folding intermediates revealed that the heme binding pocket remains relatively intact even after 500 ms, which was in agreement with a previous proposal (1). This was the first instance in which pulsed oxidative labeling had been utilized to study a protein unfolding transition in a time-resolved fashion.

The folding kinetics of acid-denatured cytochrome *c* have been shown to be dependent on the final buffer pH, with a misfolded intermediate structure populated when refolding takes place at neutral pH (2). The intention of Chapter 3 was to structurally characterize folding intermediates along the folding pathway of cyt *c* and compare the results to previous studies. The time resolution of the rapid mixing setup was improved five-fold compared to the work in Chapter 2. This improvement allowed for the observation of heme misligation from a distant histidine. The misligating residue blocks solvent accessibility of the native methionine residue early during folding. It was also seen that the N- and C-terminal helices became protected from oxidation within 10 ms of folding, which correlated with an intermediate structure characterized by HDX measurements (3). Although fluorescence measurements indicated the folding was largely complete, the oxidation kinetics revealed packing defects in the hydrophobic core persisting up to 1 s. This study marked the first utilization of pulsed oxidative labeling in conjunction with rapid mixing to examine a protein folding pathway.

Folding and binding are closely intertwined processes for protein complexes (4). The main objective of the work in Chapter 4 was to gain mechanistic insights into the folding and assembly pathway for a 22 kDa homodimer. S100A11 folding had been studied previously with time-resolved ESI-MS (5), however very few in depth structural details were garnered. The application of pulsed oxidative labeling allowed for the temporal changes in solvent accessibility of 21 side chains to be monitored. Oxidation patterns revealed the presence of a non-native hydrophobic core within the monomeric subunits in the unfolded state. A monomeric intermediate was identified after 10 ms of folding, followed by the onset of dimerization after 200 ms. Binding was complete,

indicated by oxidation levels within the interface comparable to the native state, after 800 ms although elevated accessibility was observed in regions distant from the binding surface. These data indicated the population of a dimeric folding intermediate, challenging the conventional paradigm in which dimerization ensues only after complete subunit folding. Such a finding is supported by studies on various proteins with alternative folding mechanisms (6). This study was the first to show sufficient structural detail could be obtained by pulsed oxidative labeling to resolve folding and binding events of a protein complex.

The native state of a protein has long been thought to represent the thermodynamic product of the folding reaction, populating the global free energy minimum (7). Recently, the notion that native, functional proteins are only metastable has been gaining momentum (8). This has been spurred by the finding that even globular proteins can form aggregates under physiological conditions (9). For the serine protease inhibitor α_1AT, metastability is required for function. Previous work utilized HDX to monitor backbone hydrogen bonding as a function of folding time (10) and the work in Chapter 5 aimed to elucidate complementary information regarding side-chain solvent accessibility. An early collapse of the protein structure after 7 s of folding was accompanied by significant solvent protection in helices b and c. These fast forming elements sequestered strands from β-sheet A preventing formation of the reactive loop (RCL) insertion site. These interactions allow for β-sheets B and C to form and lock the RCL into position away from the A-sheet, which exhibits non-native solvent accessibility even after 30 min of folding. This series of tertiary contacts allows α_1AT to fold into its active state rather than its thermodynamic product. A solvent inaccessible core was

formed after 30 min and displayed a striking resemblance to a previously characterized polymerogenic equilibrium intermediate (11).

6.2 Future Work

6.2.1 Sub-millisecond Folding Studies

While mechanistic details garnered from protein folding studies have been plentiful, most experiments have been limited in time resolution to the millisecond regime. However, many folding events occur on a sub-millisecond or faster timescale (12, 13). Studies probing such fast folding events typically employ a rapid folding trigger, such as a temperature (T) jump (14), ultra-rapid mixing (15), or photochemical methods (16). Unfortunately, the most commonly used detectors report only on global structural features.

Computer simulations provide the gold standard for atomic-level resolution of folding events on a rapid timescale (17). However, structural predictions from molecular dynamics only gain relevance when evaluated in conjunction with suitable experimental data (17, 18). With a protein folding "speed limit" proposed to be on the order of a few microseconds (19, 20), new detection methods with high structural and temporal resolution are required. Pulsed oxidative labeling conducted under suitable conditions utilizes a labeling pulse of ~1 μs (21). ·OH labeling has been used in combination with a T-jump system to probe the folding of barstar on the submillisecond timescale (22, 23). Such an experiment unfortunately relies on the protein of interest being unfolded with chemical denaturants at temperatures that can reach far below 0 °C. Pulsed ·OH labeling

combined with suitable ultra-rapid mixing devices (15, 24), is potentially a more universal method to provide detailed structural insights into even the earliest stages of protein folding.

6.2.2 Comparison of Native and "Diseased" Protein Folding

A mutation in a coding sequence of DNA often leads to the incorporation of an incorrect, or deletion of a correct, amino acid into a growing protein chain. Such a mutation has the potential to change the structure and/or function of the translated protein. Many diseases are resultant of a protein *mis*folding and either losing functionality or gaining aberrant activity (25). An examples of this type of 'single mutation' disease is cystic fibrosis (ΔF508 in the cystic fibrosis transmembrane conductance regulator). A second type of misfolding disorder results from protein aggregation. Diseases of this type, such as Alzheimer's, Parkinson's, and prion diseases, are characterized by the amyloid plaques of misfolded protein that become deposited over time (26). Aside from the pathogenicity of the amyloids themselves, aggregating proteins can upset cellular proteostasis by overwhelming repair machinery, rendering the cell unable to mitigate the effects of additional misfolded proteins (27-30).

Determination of how the folding of mutant and aggregating proteins differs from that of their wild-type counterparts represents a crucial step on the path to therapeutic intervention. Insights into the differential folding mechanisms could be garnered from studies utilizing the radical labeling method described herein. Such results could uncover misfolding intermediates that may be used as drug targets (31) to reverse, or avoid, diseased phenotypes.

6.2.3 Chaperone-mediated Folding

Protein folding has been shown to be a spontaneous process *in vitro* (7). However, *in vivo*, many unfolded or misfolded proteins require assistance in attaining their native, functional fold. This role is often filled by proteins and protein complexes called chaperones (32) and they perform not by correctly folding a target protein, but by preventing incorrect interactions within and between polypeptides. Their action generally increases the yield of a folding reaction but can also affect the rate. There are two main classes of chaperones, each with distinct function and structural features. The heat-shock proteins (HSPs) are monomeric (MW ~ 40-100 kDa) and are involved in *de novo* folding, refolding and assembly of protein complexes. Chaperonins are large, double-ring complexes (MW ~ 0.8-1 MDa) that globally encompass target proteins for folding (32).

A recent study showed that the introduction of the chaperone SecB induced changes in the folding pathway of the maltose-binding protein (33). Alternatively, previous work by Radford and coworkers utilized tryptophan fluorescence and pulsed HDX-MS to show that lysozyme refolding was accelerated by the chaperonin GroEL but the folding mechanism remained unchanged (34). Structural characterization of chaperone-mediated folding pathways should also be possible with pulsed oxidative labeling, and using a small HSP rather than a chaperonin would simplify data analysis as all proteins in solution will be labeled. Slight experimental alterations need to be made to account for the presence of the chaperone as the extent of labeling has been shown to be dependent on total protein concentration (35). A significant hurdle for such an experiment continues to be the maintenance of an unfolded target protein and a folded chaperone protein under the same solution conditions. Examples in the literature are

sparse and generally involve the use of mildly denaturing conditions (36) or destabilized mutant proteins (37). Addressing this issue for a protein of interest could pave the way for structural studies by oxidative labeling on its chaperone-assisted folding.

6.2.4   *Co-translational Protein Folding*

All cellular proteins are translated on the ribosome. Much effort has recently been put forth to determine whether folding can occur before the entire protein has been released from the ribosome (38). As an example, covalent modification of engineered cysteine residues has indicated that side-chain interaction with the ribosomal exit tunnel may have implications for folding (39). Such co-translational folding studies indicate that protein folding reactions which occur on the ribosome can be distinct from those observed in the more commonplace *in vitro* study of bulk solution (40, 41).

Studying co-translational folding kinetics of proteins *in vitro* would be a difficult task due to the requirement for all the cellular translational components. A modified approach was recently described that explores the folding status of translation-arrested nascent polypeptide chains of different lengths attached to ribosomes (42). Eichmann *et al.* used NMR spectroscopy to probe the degree to which the SH3 domain of α-spectrin folds after various intervals of translation. Utilizing a similar method for production and purification of ribosome-stalled nascent chain complexes would provide an avenue to probe the solvent accessibility of partially folded ribosome-bound protein chains by ·OH labeling.
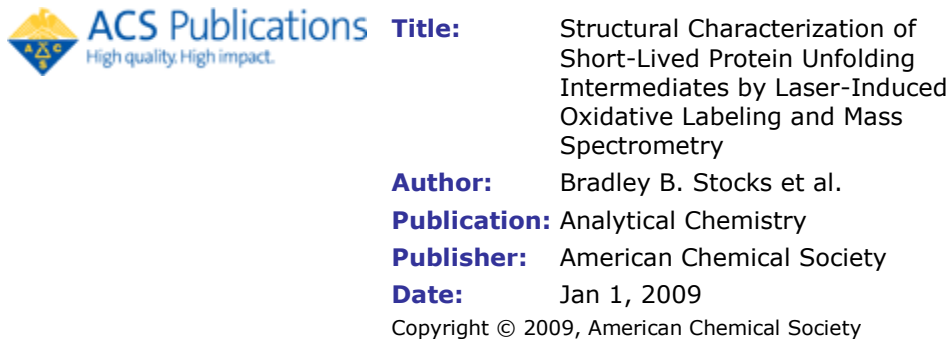
6.3 References

1. Konermann, L., Rosell, F. I., Mauk, A. G. and Douglas, D. J. (1997) Acid-Induced Denaturation of Myoglobin Studied by Time-Resolved Electrospray Ionization Mass Spectrometry. *Biochemistry* 36: 6448-6454.

2. Elöve, G. A., Bhuyan, A. K. and Roder, H. (1994) Kinetic Mechanism of Cytochrome c Folding: Involvement of the Heme and Its Ligands. *Biochemistry* 33: 6925-6935.

3. Hoang, L., Bedard, S., Krishna, M. M. G., Lin, Y. and Englander, S. W. (2002) Cytochrome c folding pathway: Kinetic native-state hydrogen exchange. *Proc. Natl. Acad. Sci. U.S.A.* 99: 12173-12178.

4. Daggett, V. and Fersht, A. R. (2009) Protein folding and binding: moving into unchartered territory. *Curr. Op. Struct. Biol.* 19: 1-2.

5. Pan, J. X., Rintala-Dempsey, A., Li, Y., Shaw, G. S. and Konermann, L. (2006) Folding Kinetics of the S100A11 Protein Dimer Studied by Time-Resolved Electrospray Mass Spectrometry and Pulsed Hydrogen-Deuterium Exchange. *Biochemistry* 45: 3005-3013.

6. Rumfeldt, J. A. O., Galvagnion, C., Vassall, K. A. and Meiering, E. M. (2008) Conformational stability and folding mechanisms of dimeric proteins. *Prog. Biophys. Mol. Biol.* 98: 61-84.

7. Anfinsen, C. B. (1973) Principles that Govern the Folding of Protein Chains. *Science* 181: 223-230.

8. Thirumalai, D. and Reddy, G. (2011) Are native proteins metastable? *Nat. Chem.* 3: 910-911.

9. Baldwin, A. J., et al. (2011) Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* 133: 14160-14163.

10. Tsutsui, Y., Dela Cruz, R. G. and Wintrode, P. L. (2012) The folding mechanism of the metastable serpin α_1-antitrypsin. *Proc. Natl. Acad. Sci. U.S.A.* in press: doi/10.1073/pnas.1109125109.

11. Krishnan, B. and Gierasch, L. M. (2011) Dynamic local unfolding in the serpin α-1 antitrypsin provides a mechanism for loop insertion and polymerization. *Nat. Struct. Mol. Biol.* 18: 222-226.

12. Gruebele, M. (1999) The Fast Protein Folding Problem. *Annu. Rev. Phys. Chem.* 50: 485-516.

13. Mayor, U., et al. (2003) The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* 421: 863-867.

14. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A. and Hofrichter, J. (2006) Sub-microsecond Protein Folding. *J. Mol. Biol.* 359: 546-553.

15. Roder, H., Maki, K. and Cheng, H. (2006) Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem. Rev.* 106: 1836-1861.

16. Jones, C. M., et al. (1993) Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. U.S.A.* 90: 11860-11864.

17. Shaw, D. E., et al. (2010) Atomic Level Characterization of the Structural Dynamics of Proteins. *Science* 330: 341-346.

18. Cellmer, T., Buscaglia, M., Henry, E. R., Hofrichter, J. and Eaton, W. A. (2011) Making connections between ultrafast protein folding kinetics and molecular dynamics simulations. *Proc. Nat. Acad. Sci. U.S.A.* 108: 6103-6108.

19. Kubelka, J., Hofrichter, J. and Eaton, W. A. (2004) The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* 14: 76-88.

20. Abel, C. J., Goldbeck, R. A., Latypov, R. F., Roder, H. and Kliger, D. S. (2007) Conformational Equilibrium Time of Unfolded Protein Chains and the Folding Speed Limit. *Biochemistry* 46: 4090-4099.

21. Hambly, D. M. and Gross, M. L. (2005) Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale. *J. Am. Soc. Mass Spectrom.* 16: 2057-2063.

22. Chen, J., Rempel, D. L. and Gross, M. L. (2010) Temperature Jump and Fast Photochemical Oxidation Probe Submillisecond Protein Folding. *J. Am. Chem. Soc.* 132: 15502-15504.

23. Gruebele, M. (2010) Weighing Up Protein Folding. *Nature* 468: 640-641.

24. Chan, C.-K., et al. (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. U.S.A.* 94: 1779-1784.

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

26. Aguzzi, A. and O'Connor, T. (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat. Rev. Drug Discov.* 9: 237-248.

27. Bates, G. P. (2006) One Misfolded Protein Allows Others to Sneak By. *Science* 311: 1385-1386.

28.	Balch, W. E., Morimoto, R. I., Dillin, A. and Kelly, J. W. (2008) Adapting Proteostasis for Disease Intervention. *Science* 319: 916-919.

29.	Olzscha, H., et al. (2011) Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. *Cell* 144: 67-78.

30.	Alavez, S., Vantipalli, M. C., Zucker, D. J. S., Klang, I. M. and Lithgow, G. J. (2011) Amyloid-binding compounds maintain protein homeostasis during ageing and extend lifespan. *Nature* 472: 226-230.

31.	Woods, L. A., et al. (2011) Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nat. Chem. Biol.* 7: 730-739.

32.	Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475: 324-332.

33.	Bechtluft, P., et al. (2007) Direct Observation of Chaperone-Induced Changes in a Protein Folding Pathway. *Science* 318: 1458-1461.

34.	Coyle, J. E., et al. (1999) GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat. Struct. Biol.* 6: 683-690.

35.	Tong, X., Wren, J. C. and Konermann, L. (2007) Effects of Protein Concentration on the Extent of γ-Ray-Mediated Oxidative Labeling Studied by Electrospray Mass Spectrometry. *Anal. Chem.* 79: 6376-6382.

36.	Mok, Y.-K., Kay, C. M., Kay, L. E. and Forman-Kay, J. (1999) NOE Data Demonstrating a Compact Unfolded State for an SH3 Domain Under Non-denaturing Conditions. *J. Mol. Biol.* 289: 619-638.

37.	Mayor, U., Gunter Grossmann, J., Foster, N. W., Freund, S. M. V. and Fersht, A. R. (2003) The Denatured State of Engrailed Homeodomain under Denaturing and Native Conditions. *J. Mol. Biol.* 333: 977-991.

38.	Kramer, G., Boehringer, D., Ban, N. and Bukau, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* 16: 589-597.

39.	Lu, J., Hua, Z., Kobertz, W. R. and Deutsch, C. (2011) Nascent Peptide Side Chains Induce Rearrangements in Distinct Locations of the Ribosomal Tunnel. *J. Mol. Biol.* 411: 499-510.

40.	O'Brien, E. P., Christodoulou, J., Vendruscolo, M. and Dobson, C. M. (2011) New Scenarios of Protein Folding Can Occur on the Ribosome. *J. Am. Chem. Soc.* 133: 513–526.

41.   Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. and Bustamante, C. (2011) The Ribosome Modulates Nascent Protein Folding. *Science* 334: 1723-1727.

42.   Eichmann, C., Preissler, S., Riek, R. and Deuerling, E. (2010) Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 107: 9111-9116.

Appendix I – Permissions

Copyright Clearance Center

RightsLink®

ACS Publications
High quality. High impact.

Title:	Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry
Author:	Bradley B. Stocks et al.
Publication:	Analytical Chemistry
Publisher:	American Chemical Society
Date:	Jan 1, 2009

Copyright © 2009, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

ELSEVIER LICENSE

TERMS AND CONDITIONS

This is a License Agreement between Bradley B Stocks ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier   Elsevier Limited

                  The Boulevard,Langford Lane

                  Kidlington,Oxford,OX5 1GB,UK

Registered Company Number 1982084

Customer name   Bradley B Stocks

Customer address   University of Western Ontario

                       London, ON N6A5B7

License number   2866170504786

License date   Mar 11, 2012

Licensed content publisher   Elsevier

Licensed content publication   Journal of Molecular Biology

Licensed content title   Time-Dependent Changes in Side-Chain Solvent Accessibility during Cytochrome c Folding Probed by Pulsed Oxidative Labeling and Mass Spectrometry

Licensed content author   Bradley B. Stocks,Lars Konermann

Licensed content date   30 April 2010

Licensed content volume number   398

Type of Use   reuse in a thesis/dissertation

ELSEVIER LICENSE

TERMS AND CONDITIONS

This is a License Agreement between Bradley B Stocks ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier   Elsevier Limited

                       The Boulevard,Langford Lane

                       Kidlington,Oxford,OX5 1GB,UK

Registered Company Number 1982084

Customer name   Bradley B Stocks

Customer address   University of Western Ontario

                         London, ON N6A5B7

License number   2866170829460

License date   Mar 11, 2012

Licensed content publisher   Elsevier

Licensed content publication   Journal of Molecular Biology

Licensed content title   Temporal Development of Protein Structure during S100A11 Folding and Dimerization Probed by Oxidative Labeling and Mass Spectrometry

Licensed content author   Bradley B. Stocks,Atoosa Rezvanpour,Gary S. Shaw,Lars Konermann

Licensed content date   17 June 2011

Licensed content volume number   409

Type of Use   reuse in a thesis/dissertation

Curriculum Vitae

Bradley B. Stocks

Ph.D. Candidate

Department of Biochemistry
The University of Western Ontario
London, Ontario, N6A 5C1, Canada

Research Group Home Page: http://publish.uwo.ca/~konerman

Education

Degree	University	Department	Year
Ph.D.	The University of Western Ontario	Biochemistry	2007-present
B.Sc. with Distinction	University of Windsor	Chemistry and Biochemistry	2001-2005

Employment History

Date	Position	Department	Institution
2010-present	Teaching Assistant	Biochemistry	UWO
2007-present	Research Assistant (Ph.D.)	Biochemistry	UWO
2006	GC-MS Operator/Analyst	Organics	ALS Environmental (Vancouver, BC)
2002	Teaching Assistant	Mathematics	University of Windsor

Honors, Awards and Scholarships

Date	Award Name	Type	Location of Tenure	Amount
2011	Nellie Farthing Fellowship	Institutional (Research)	The University of Western Ontario	$3000
2011	Graduate Thesis Travel Award	Institutional (Research)	The University of Western Ontario	$656

2010	CSMS Student Travel Award	National (Research)	The University of Western Ontario	$2000
2010	Nellie Farthing Fellowship	Institutional (Research)	The University of Western Ontario	$3000
2010	NSERC Canada Graduate Scholarship	National (Research)	The University of Western Ontario	$35000/yr (2 years)
2010	Graduate Thesis Travel Award	Institutional (Research)	The University of Western Ontario	$770
2009	Ontario Graduate Scholarship	Provincial (Research)	The University of Western Ontario	$15000
2009	Graduate Thesis Travel Award	Institutional (Research)	The University of Western Ontario	$950
2007	Schulich Special Scholarship	Institutional (Academic)	The University of Western Ontario	$5000/yr (2 years)
2001-2005	President's Honour Roll	Institutional (Academic)	University of Windsor	N/A
2001	Continuing Entrance Scholarship	Institutional (Academic)	University of Windsor	$8000 (over 4 yrs)

Publications in Peer Reviewed Journals

Protein Folding Mechanisms Studied by Pulsed Oxidative Labeling and Mass Spectrometry. Lars Konermann, Yan Pan, and **Bradley B. Stocks** (2011) *Curr. Opin. Struct. Biol.* 21, 634-640.

Temporal Development of Protein Structure during S100A11 Folding and Dimerization Probed by Oxidative Labeling and Mass Spectrometry. **Bradley B. Stocks**, Atoosa Rezvanpour, Gary S. Shaw, and Lars Konermann (2011) *J. Mol. Biol. 409*, 669-679.

Laminar Flow Effects during Laser-Induced Oxidative Labeling for Protein Structural Studies by Mass Spectrometry. Lars Konermann, **Bradley B. Stocks**, and Tomasz Czarny (2010) *Anal. Chem. 82*, 6667-6674.

Mass Spectrometry Combined with Oxidative Labeling for Exploring Protein Structure and Folding. Lars Konermann, **Bradley B. Stocks**, Yan Pan, and Xin Tong (2010) *Mass Spectrom. Rev. 29*, 651-667.

Time-Dependent Changes in Side Chain Solvent Accessibility during Cytochrome *c* Folding Probed by Pulsed Oxidative Labeling and Mass Spectrometry. **Bradley B. Stocks** and Lars Konermann (2010) *J. Mol. Biol. 398*, 362-373.

Structural Characterization of an Integral Membrane Protein in its Natural Lipid Environment by Oxidative Methionine Labeling and Mass Spectrometry. Yan Pan, **Bradley B. Stocks**, Leonid Brown, and Lars Konermann (2009) *Anal. Chem. 81,* 28-35.

Structural Characterization of Short-Lived Protein Unfolding Intermediates by Laser-Induced Oxidative Labeling and Mass Spectrometry. **Bradley B. Stocks** and Lars Konermann (2009) *Anal. Chem. 81*, 20-27.

Conference Presentations

Poster – Investigating the Folding of the Metastable Serpin α_1-Antitrypsin by Pulsed Oxidative Labeling and ESI-MS. **Bradley B. Stocks**, Patrick Wintrode, and Lars Konermann. *Gordon Research Conference: Biological Molecules in the Gas Phase and in Solution.* Andover, NH. (2011, Poster Award: First Prize)

Poster - Structural Characterization of Short-Lived Intermediates during Protein Folding and Assembly by Pulsed Oxidative Labeling and ESI-MS. **Bradley B. Stocks**, Patrick Wintrode, and Lars Konermann. *59th ASMS Conference on Mass Spectrometry.* Denver, CO. (2011)

Oral – Investigating the Folding and Assembly of a Protein Complex by Pulsed Oxidative Labeling and ESI-MS/MS. **Bradley Stocks**, Gary Shaw, and Lars Konermann. *23rd International Workshop on Tandem Mass Spectrometry*, Lake Louise, AB. (2010)

Oral – Exploring the Mechanisms of Protein Folding and Subunit Assembly by Pulsed Oxidative Labeling and ESI-MS. **Bradley B. Stocks** and Lars Konermann. *58th ASMS Conference on Mass Spectrometry.* Salt Lake City, UT. (2010)

Poster - Laser-Induced Oxidative Labeling of Proteins for Probing Folding Kinetics and Mechanisms. **Bradley B. Stocks** and Lars Konermann. *57th ASMS Conference on Mass Spectrometry*. Philadelphia, PA. (2009)

Poster - Probing the Structure of Short-Lived Protein Folding Intermediates by Hydroxyl Radical-Mediated Oxidative Labeling and ESI-MS. **Bradley B. Stocks** and Lars Konermann. *56th ASMS Conference on Mass Spectrometry.* Denver, CO. (2008)

Conference Proceedings and Abstracts

Conservation and Variation in the Folding and Dynamics of Serpins: Comparative Studies of α_1-Antitrypsin and Neuroserpin. Patrick L. Wintrode, Anindya Sarkar, **Bradley Stocks**, and Lars Konermann. *6th International Symposium on the Chemistry and Biology of Serpins.* Chapel Hill, NC. (2011)

Deciphering Protein Folding Mechanisms by Mass Spectrometry Using Top-Down ECD and Radical Labeling. Lars Konermann, Jingxi Pan, Jun Han, Christoph Borchers, **Bradley Stocks** and Yan Pan. *8th Uppsala Conference on Electron Capture and Transfer Dissociation.* Villars-sur-Ollon, Switzerland. (2011)

Progress in Top-down and Bottom-up Mass Spectrometry for Structural Studies of Proteins in Solution. Lars Konermann, Jingxi Pan, Yan Pan, **Bradley B. Stocks**, Christoph H. Borchers and Jun Han. *22nd International Workshop on Tandem Mass Spectrometry.* Lake Louise, AB. (2009)

Elucidating Folding Mechanisms of Globular and Membrane Proteins by Laser-Induced Oxidative Labeling and ESI-MS. Yan Pan, **Bradley B. Stocks**, Leonid Brown, and Lars Konermann. *18th International Mass Spectrometry Conference.* Bremen, Germany. (2009)

Structural Characterization of an Integral Membrane Protein by Oxidative Methionine Labeling and Mass Spectrometry. Yan Pan, **Bradley B. Stocks,** Leonid Brown and Lars Konermann. *57th ASMS Conference on Mass Spectrometry.* Philadelphia, PA. (2009)

Protein Folding Studied by Hydrogen Exchange, Covalent Labeling, and Tandem Mass Spectrometry. Lars Konermann, Yan Pan, Jingxi Pan, and **Bradley B. Stocks**. *21st International Workshop on Tandem Mass Spectrometry*, Lake Louise, AB. (2008)

Protein Dynamics Studied by Hydrogen Exchange and Radical Probe Mass Spectrometry. L. Konermann*, Y.-H. Liu, **B. B. Stocks** and X. Tong. *91st Canadian Chemistry Conference.* Edmonton, AB. (2008)

Hydroxyl Radical Labeling and ESI-MS/MS for Characterizing Protein Structures in Solution. Lars Konermann, Xin Tong, **Bradley Stocks**, and Yuhong Liu. *20th International Workshop on Tandem Mass Spectrometry*, Lake Louise, AB. (2007)