AN ABSTRACT OF THE DISSERTATION OF

Yuan Heidi Zhang for the degree of Doctor of Philosophy in Chemistry presented on May 14, 2002.

Title: Mass Spectrometric Analysis of Proteins and Peptides: Elucidation of the Folding Pathways of Recombinant Human Macrophage Colony Stimulating Factor Beta.

Abstract approved: Redacted for Privacy

Max L. Deinzer

Recombinant human macrophage colony stimulating factor beta (rhm-CSFβ) is a glycoprotein that stimulates the proliferation, differentiation and survival of cells belonging to the monocyte-macrophage lineage. It contains nine inter-subunit and intra-subunit disulfide bonds and represents an excellent model system for studying disulfide bond formation during protein folding because the assembly of its monomeric subunits and the maturation of its biological activity depend on the progressive formation of the correct disulfide structure during *in vitro* folding. Knowledge obtained from these studies can be potentially useful in understanding the roles of disulfide bond formation during protein folding in general.

rhm-CSFβ was modified by partial reduction of disulfide bonds, yielding $CN^{157,159}$-modified rhm-CSFβ. The modification did not affect the biological activity, stability, or the overall conformation of the protein. However, the C-terminal regions near the modification sites were shown to exhibit faster deuterium exchange behavior as a result of the chemical modification, indicating that the C-terminal regions became more flexible. Folding kinetics of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ were shown to be essentially the same, suggesting that the modification did not affect the folding kinetics of the oxidized rhm-CSFβ.

The denatured and reduced rhm-CSFβ was refolded with the aid of a chemical oxidant. The data indicated that the *in vitro* folding rhm-CSFβ proceeded *via* multiple pathways involving monomeric and dimeric intermediates. Disulfide bond shuffling catalyzed by GSH/GSSG represented an important isomerization step in folding. A dimeric intermediate, D-SS8-cam2, was isolated and identified as a kinetic trap, perhaps requiring significant structural arrangement to convert to the native protein. The heterogeneous folding mixture detected by both disulfide bond quenching and H/D pulsed labeling indicate that rhm-CSFβ folding is a diffusion like process as described by the folding funnel model.

Mass Spectrometric Analysis of Proteins and Peptides: Elucidation of the Folding Pathways of Recombinant Human Macrophage Colony Stimulating Factor Beta

By

Yuan Heidi Zhang

A DISSERTATION

Submitted to

Oregon State University

in partial fulfillment of

the requirements for

the degree of

Doctor of Philosophy

Presented May 14 2002

Commencement June 2003

Doctor of Philosophy dissertation of Yuan Heidi Zhang presented on May 14, 2002

APPROVED:

Redacted for Privacy

Major professor, representing Chemistry

Redacted for Privacy

Head of the Department of Chemistry

Redacted for Privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes the release of my dissertation to any reader upon request.

Redacted for Privacy

Yuan Heidi Zhang, Author

ACKNOWLEDGMENTS

I thank Professor Max Deinzer for his guidance and support. I thank Professor Michael Schimerlik for the stimulating conversations, which have greatly deepened my understanding of protein science. I thank Dr. Claudia Maier and Dr. XuGuang Yan for the discussions regarding rhm-CSFβ and hydrogen deuterium exchange. I thank Dr. Mark Harder for calculating the theoretical number for disulfide isomers and assisting with the CD experiments. I thank Dr. C. Cogwill and Dr. J. Weaver (Chiron Corp. Emeryville, CA) for providing rhm-CSFβ and NFS-60 cell lines. I also thank Don Griffin, Brain Arbogast and Elisabeth Barofsky for the excellent technical assistance. This work was supported in part by NIEHS grant ES 00040.

I thank my family and friends for their constant love and support. I especially want to thank John Neff for his encouragement and patience. I also thank my friends and colleagues in the mass spectrometry lab for having provided a "home" away from home.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

LIST OF APPENDICES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

CD, circular dichroism

CDAP, 1-cyano-dimethylamino pyridinium salt

CID, collisional induced dissociation

ESI, electrospray ionization

FAB, fast atom bombardment

H/D, hydrogen deuterium

HPLC, high performance liquid chromatography

IAA, iodoacetic acid

IAM, iodoacetamide

MALDI, matrix assisted laser desorption ionization

MS, mass spectrometry

MS/MS, tandem mass spectrometry

MW, molecular weight

NMR, nuclear magnetic resonance

rhm-CSFβ, recombinant human macrophage colony stimulating factor beta

m-CSFR, macrophage colony stimulating factor receptor

TCEP, tris(2-carboxyethyl)phosphine

Mass Spectrometric Analysis of Proteins and Peptides: Elucidation of the Folding Pathways of Recombinant Human Macrophage Colony Stimulating Factor Beta

1 Introduction

1.1 Protein folding and disulfide bond formation

Proteins are made of amino acids whose sequence is dictated by the DNA sequence in the gene. To be biologically active, a protein must exhibit higher levels of structural organization. Therefore, it is important to understand how a protein folds from a random-coiled polypeptide to a well-defined structure, both from the standpoint of attaining scientific knowledge as well as designing effective therapeutic targets. How and why a protein adopts a specific three-dimensional conformation constitutes the protein folding problem (Anfinsen 1973). The ultimate goal of solving the protein folding problem is to be able to predict the folded structure of the protein from only a knowledge of its primary sequence.

The reversibility of protein unfolding and folding was demonstrated by Anfinsen with bovine pancreatic ribonuclease (RNase) (Anfinsen 1961). The native RNase was denatured in 8 M urea and β-mercaptoethanol; the unfolded and reduced RNase was readily converted to the native protein

with full biological activity. This experiment established that all the information necessary to define the native three-dimensional structure of a protein is contained in its primary sequence. However, for a polypeptide of 100 amino acid residues, the possible conformations are on the order of 10^{30} and would take many years to fold if the polypeptide were to sample all available conformations (Creighton 1992). How does the polypeptide chain search in a finite time through the extensive conformational space to find the native state simply based on its amino acid sequence? Since proteins fold much more quickly *in vitro*, some on the time scale of seconds, Levinthal concluded that proteins must fold by specific pathways characterized by folding intermediates (Levinthal 1968). Thus, the detection and structural characterization of folding intermediates are important to addressing the protein folding problem.

Often the intermediates along the folding pathways correspond to transient ensembles of micro-states. Consequently, they cannot be isolated or characterized in detail. However, in disulfide-containing proteins, the presence of a disulfide bond between two Cys residues requires that they be within a few angstroms of each other. Due to the linkage between the stabilities of the disulfide bonds and the protein conformation that favor them, the conformation of the protein can be effectively trapped by disulfide bond quenching, i.e. trapping of the folding intermediates with nascent

cysteinyl groups by alkylation. Whatever conformation helps promote the formation of a particular disulfide bond will itself be stabilized by the presence of that disulfide bond (Figure 1). Therefore, the trapped intermediates tend to adopt the conformations that favor the formation of these disulfide bonds, assuming that the trapping reaction does not alter the protein conformation. This method, termed oxidative folding, was developed by T.E.Creighton (Creighton 1977; Creighton 1978) and has found many applications in the elucidation of protein folding pathways (Eyles, Radford et al. 1994) (Ruoppolo 2000).

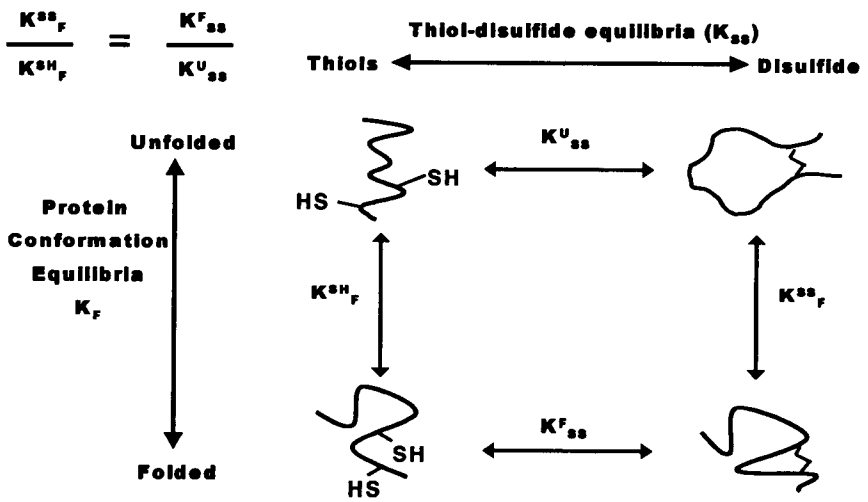


Figure 1. Disulfide bond formation and protein folding are interconnected.

The formation of a disulfide bond between two thiols is a two-electron oxidation reaction that requires an oxidant. While air oxygen may facilitate the formation of disulfide bonds, many oxidative folding studies utilize low molecular weight disulfides such as RSSR/RSH. During oxidation, the disulfide component (RSSR) of the redox buffer provides the oxidizing equivalents for protein disulfide formation. The thiol component (RSH) of the redox buffer serves to reduce the non-native disulfide bonds that may trap the protein in the incorrect disulfide bonded form. Thiol/disulfide exchange occurs *via* direct attack of the thiolate anion on one of the sulfurs involved in the disulfide bond. The rate constant for the reaction increases as the basicity of the thiolate increases (pKa increases) and as the basicity of the leaving group decreases. The rate constant also increases as the pH of the solution increases until the attacking thiol is predominantly in the thiolate form. The pKa of a typical cysteine sulfhydryl group is about 8.6 and may change depending on the local environment.

In order to isolate and structurally characterize the intermediates that are involved in oxidative folding, it is necessary to stop both the inter- and intramolecular thiol/disulfide exchange reactions and properly trap the intermediates. Criteria for a good trapping method requires that the trapping agent block quickly, completely and without modifying the protein

at sites other than the thiols (Rothwarf 1993). One method that has been used is to lower the solution pH to <2 by adding acid (Weissman 1991) (Figure 2). Quenching by acidification is rapid and occurs at the diffusion-controlled rate (10^9 $M^{-1}s^{-1}$). However, acidification does not completely stop the thiol/disulfide exchange. At pH 2, the thiol disulfide interchange still occurs at 10^{-6} of its rate at pH 8. Intramolecular disulfide rearrangement occurs at a rate of up to 10^5 s^{-1} with a half-life as short as 7 seconds at pH 2.

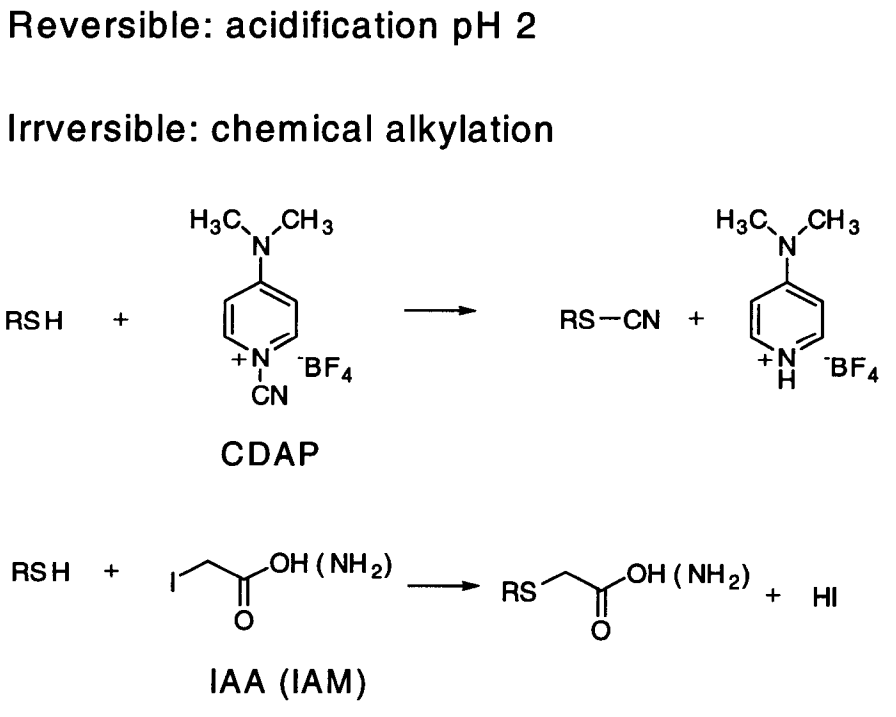


Figure 2. Some commonly used trapping methods.

Alternatively, the thiol groups can be trapped irreversibly by alkylation. Once all free thiol groups are blocked, the trapped species are stable and may be analyzed in detail. Traditionally, irreversible alkylation has been accomplished by adding iodoacetamide (IAM or CAM) or iodoacetic acid (IAA or CAA). However, rearrangement of disulfides during trapping has been observed (Rothwarf 1991; Weissman 1991). Modification of other functional groups by the high concentration of iodoacetamide has also been reported (Torella 1994). Cyano-dimethylamino-pyridinium salt (CDAP) at low pH offers a number of advantages over IAM and IAA (Wakselman 1976; Koyama 1994; Wu, Yang et al. 1998). Cyanylation occurs under acidic conditions where the rate of sulfhydryl/disulfide exchange is lowered thereby limiting disulfide bond shuffling. Additionally, cyanylation is selective and modification at sites other than cysteine residues is rare. However, cyanylated proteins and peptides may undergo β-elimination under basic conditions.

The structural elucidation of successive folding intermediates is made possible by the disulfide bond quenching method. However, the oxidative folding of a protein containing only a few disulfide bonds can be quite complex. For a protein with six cysteine residues that form three disulfide bonds, there are 15 ways the first disulfide bonds can form, 45

possible two disulfide bonded species, and 15 different pairings of the cysteine residues to generate molecules with three disulfide bonds in the native structure. The complexity rapidly increases when more disulfides are involved. In studies carried out by Baldwin and Weissman, various folding intermediates involving non-native disulfides have been identified in the oxidative folding of bovine pancreatic trypsin inhibitor (Kim 1990; Weissman 1992). The pathways constructed using these and other intermediates were interesting in that they didn't occur *via* a simple sequential pathway. Instead, parallel folding pathways were identified.

1.2 Recombinant human macrophage colony stimulating factor β

Macrophage colony stimulating factor (M-CSF) is a glycoprotein that stimulates the proliferation, differentiation and survival of cells belonging to the monocyte-macrophage lineage both *in vivo* and *in vitro* (Stanley 1994). For instance, M-CSF can induce M-CSF dependent proliferation of fibroblasts (Roussel 1989) or hematopoietic cells (Kato 1990). M-CSF can also stimulate the differentiation towards monocyte and macrophage lineage of myeloid progenitor cell lines such as FDC-P1 (Rohschneider 1989) or NFS-60 (Pawlak 1999). The key role of M-CSF in monocytic

development has been demonstrated in mutant mice that lack functional M-CSF and are deficient in macrophages but can be cured by the injection of M-CSF (Wiktor-Jedrejczak 1991).

The effects of M-CSF are mediated by its binding to its receptor (M-CSFR), which is an integral transmembrane glycoprotein and functions as a ligand-activated protein tyrosine kinase (Bourette 2000) (Figure 3).

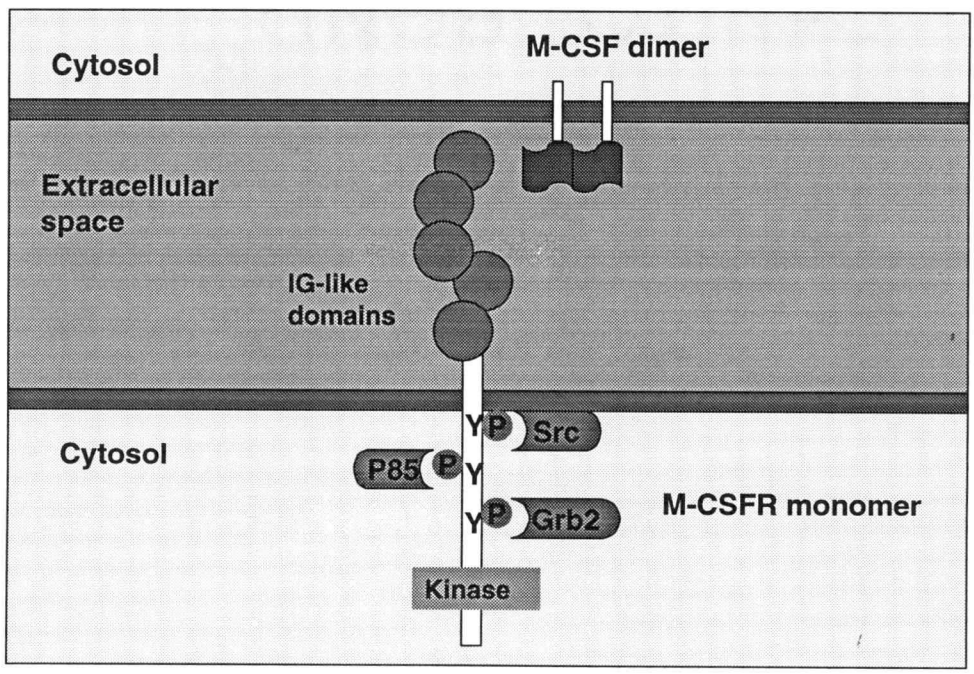


Figure 3. rhm-CSFβ triggers the development of cells belonging to monocyte-macrophage lineage.

M-CSF is the only known ligand for M-CSFR. Ligand binding to the extracellular region induces dimerization of M-CSFR. Transphosphorylation of several cytoplasmic tyrosine residues is followed by subsequent interaction of the phosphorylated cytoplasmic domain of M-CSFR with several 'primary' adaptor proteins, each signaling along specific pathways (Csar 2001). The best characterized signal pathway is the one initiated by the Grb2-SOS activation of Ras, leading to MAP kinase activation (Treisman 1996). Another major pathway is initiated with binding of Src to the activated phosphorylated M-CSFR, relieving constraints on the Src kinase activity (Moarefi 1997).

The original interest in M-CSF stemmed from the hope that M-CSF may find clinical application in diseases characterized by suboptimal production of specific cell types such as in the treatment of infectious diseases, various forms of cancer and the management of bone marrow transplants. However, M-CSF was shown to exhibit much less proliferation on human bone marrow cells than on comparable populations of murine bone marrow cells (Cerretti 1988). Contrary to its previously reported role in bone marrow development, M-CSF does not seem to have any mitogenic effect on bone marrow cells of human origin (Tushinski 1982).

Human M-CSF is encoded by a single gene that maps to chromosome 5q33.1; the primary messenger RNA (mRNA) transcript of this gene undergoes alternative splicing to generate multiple mature mRNAs encoding different forms of membrane-bound M-CSF precursors. The three M-CSF polypeptides differ in length: α contains 256 aa (Kawasaki 1985), β consists of 554 aa (Wong 1987), and γ is made up of 438 aa (Cerretti 1988). All three proteins share a common domain structure consisting of an amino-terminal signal peptide followed by a segment of hydrophobic amino acids that serves as the membrane anchor for the precursor and a short carboxy-terminal tail. It has been proposed that M-CSF is synthesized as a membrane bound precursor which is then proteolytically cleaved, releasing different forms of the soluble growth factor (Rettenmier 1987; Stein 1990).

M-CSFβ undergoes both asparagine N- and O-lined glycosylation and is rapidly cleaved from the transmembrane domain. It is a homodimer maintained by three inter-subunit and six intra-subunit disulfide bonds (Figure 4).

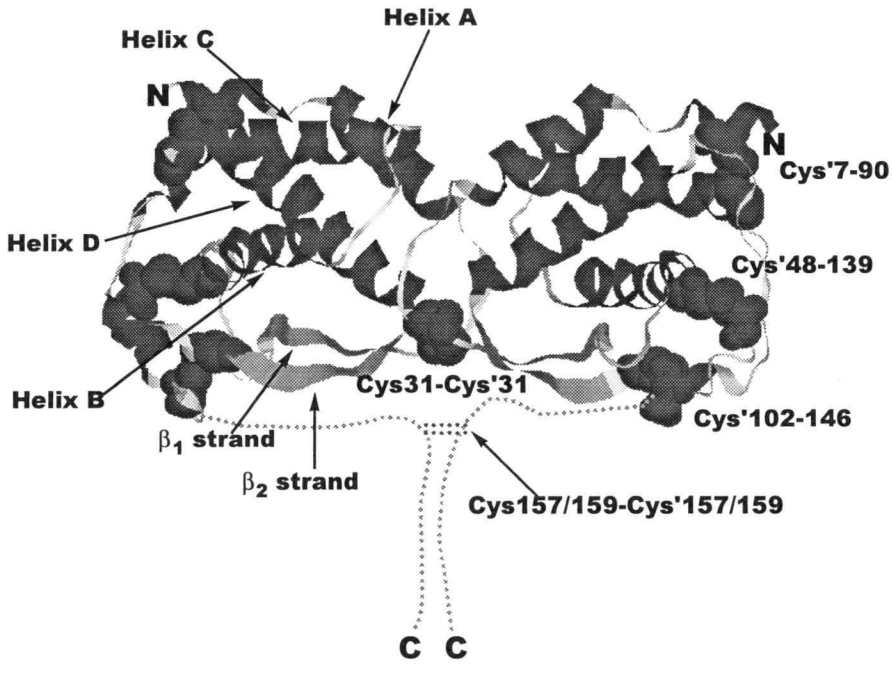


Figure 4. Schematic representation of rhm-CSFβ based on x-ray crystal structure of m-CSFα (Pandit, 1992). The unstructured C-terminal regions (aa157-221) and the disulfide bridges linking Cys157/159-Cys'157/159 are shown as dotted lines.

The disulfide bridges were assigned using enzymatic digestion and fast atom bombardment mass spectrometry (FAB-MS) (Glocker, Arbogast et al. 1993). The inter-subunit disulfide bridges hold the dimer together and form symmetrical bonds in which Cys 31 and Cys157/159 from one monomer unit were linked to the corresponding cysteines of the second monomer. The intra-subunit disulfide bonds were found between Cys4-Cys90, Cys48-Csy139 and Cys102-Cys146. Mutation experiments have shown that the

Cys 4, 31, 48, 90, 102 and 139 in each chain were essential for biological activity (Kawasaki 1990). A crystal structure of M-CSFα suggested that these cysteine residues were important for structure formation or stability rather than being directly involved in receptor recognition (Pandit 1992). Two intermolecular disulfide bonds in the C-terminal region were found between Cys157/159 and Cys'157/159. It is unclear whether Cys157 forms a disulfide bond with Cys'157 or Cys'159. Nonetheless, it is evident that these cysteine residues form symmetrical disulfide bonds that are in close proximity in three-dimensional space.

1.3 Hydrogen deuterium exchange (H/D)

Hydrogen deuterium exchange is a powerful technique that has been used to study protein structure (Zhang 2001), dynamics (Gregory 1986; Resing 1998), folding (Miranker, Robinson et al. 1993) and protein-ligand interactions (Anderegg 1995; Kragelund 1995; Wang 1997; Mandell 2001). Hydrogen deuterium exchange involves the replacement of labile protein-bound hydrogen with solvent hydrogen of a different isotopic composition, often deuterium. Proteins contain three types of exchangeable hydrogens:

those located on the amino acid side chains, those at the N- and the C-termini, and those at the peptide amide linkages (NHs) (Figure 5).

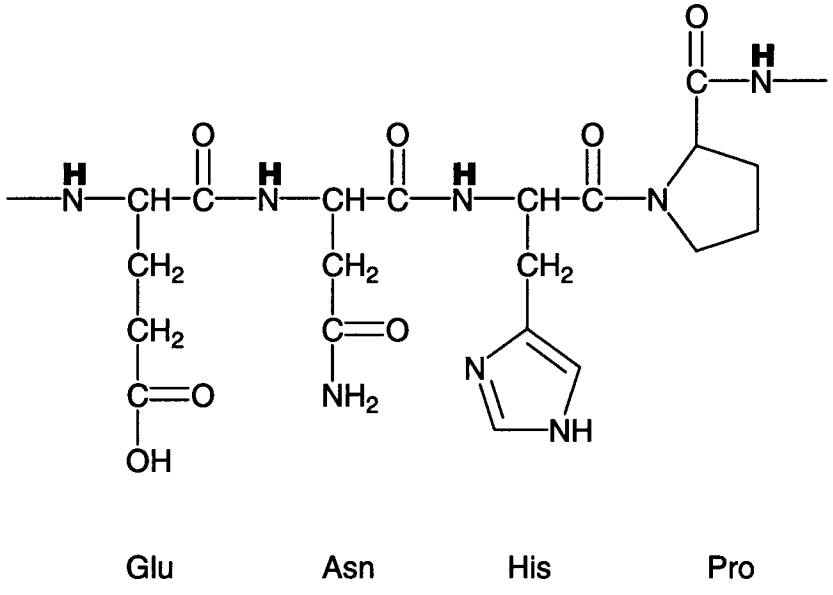


Figure 5. Amide hydrogens that undergo isotopic exchange.

Among these hydrogens, only those located at the peptide amide linkages exchange at a rate that can be readily measured (Englander 1984). Every amino acid residue, except Pro, has one hydrogen at the peptide amide linkage. Consequently, NHs can form a continuous string of sensors

extending the entire length of a polypeptide chain, providing important structural information.

Amide hydrogens can undergo exchange *via* three mechanisms (Perrin 1989): (A) base catalysis which requires the abstraction of the amide hydrogen by hydroxide ion and subsequent deuteration of the amide nitrogen by proton/deuterium exchange from the solvent; (B) acid catalysis which involves the deuteration of the amide nitrogen followed by deprotonation by solvent; and (C) imidic acid mechanism which involves the deuteration of the carbonyl oxygen and subsequent removal of the amide proton by solvent yielding the imidic acid and then the amide (Figure 6).

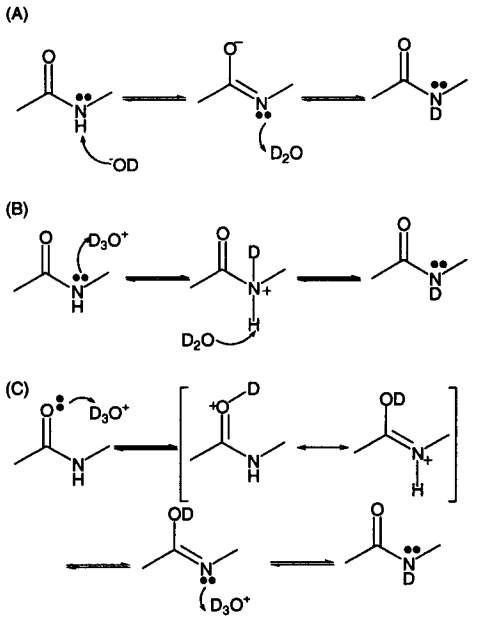


Figure 6. Details of the chemical exchange reaction of amide hydrogen. (A) Base catalysis. (B) Acid catalysis. (C) Imidic acid mechanism.

The base catalysis mechanism dominates at pHs greater than 3. At pHs less than 3, the imidic acid mechanism has been shown to dominate over the N-protonation acid catalysis mechanism due to better charge delocalization in the transition state (Perrin 1982). The exchange rate constant for a freely exposed NH can be expressed as

$$K_{ex}=k_{OH}[OH^-] + k_H[H^+] + k_o \qquad \text{Equation 1}$$

where k_{OH}, k_H, and k_o are the rate constants for the base-catalyzed, acid catalyzed exchange and direct exchange with water, respectively. Detailed studies of amide H/D exchange in the model polypeptide polyalanine indicate that k_{OH} and k_H have values of 1.12X10^10 $M^{-1}min^{-1}$ and 41.7 $M^{-1}min^{-1}$, respectively, at 20°C and low concentrations of salt (Bai 1993). The isotopic exchange rate constants for exposed amide hydrogens of polyalanine, polar side-chain hydrogens, as well as hydrogens at the N- and C-termini are given as a function of pH (Figure 7).

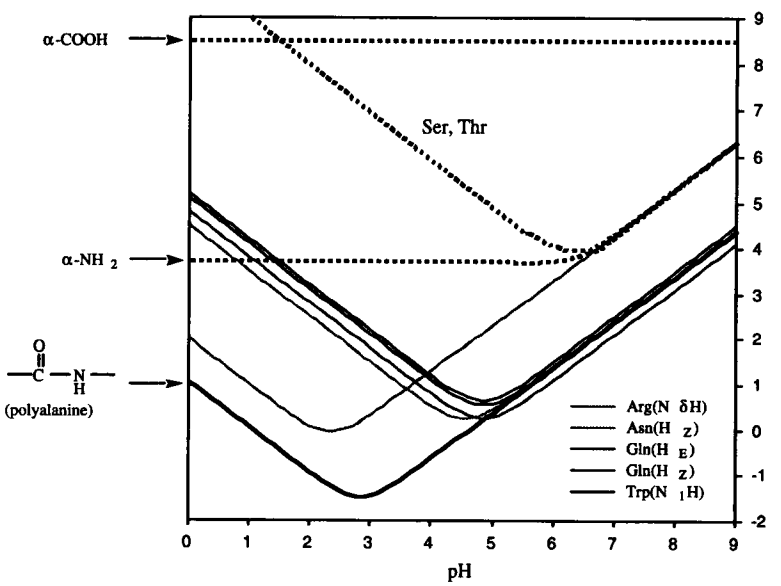


Figure 7. Amide H/D exchange rate constant as a function of pH.

Direct exchange with H_2O is usually ignored because k_o has a value of 0.03 $M^{-1}min^{-1}$. The exchange rate of NH in an unfolded polypeptide is the slowest at pH 2.5~3 and increases 10-fold for each pH unit to give a linear correlation (Figure 8).

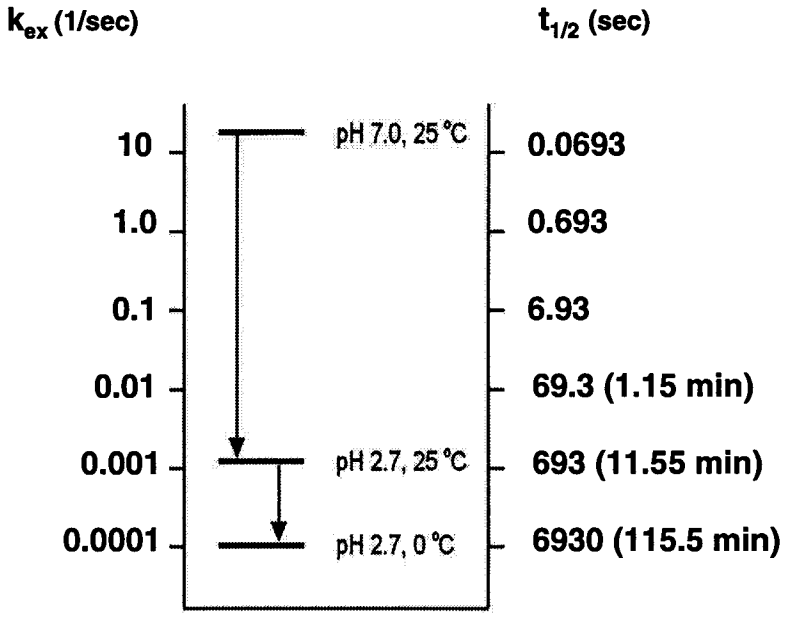


Figure 8. Amide hydrogen exchange rate is related to pH and temperature.

Isotopic exchange of NH is also sensitive to temperature. The rate constants follow the Arrhenius equation:

$$\ln k = \ln A - E_a/RT \qquad \text{Equation 2}$$

where k is the rate coefficient, A is a constant, E_a is the activation energy, R is the universal gas constant, and T is the temperature (in degrees Kelvin). Thus, to predict exchange rate constants at different temperatures the following equation can be used:

$$k_{ex,T2} = k_{ex,T1}\exp(-E_a(1/T_2-1/T_1)/R) \qquad \text{Equation 3}$$

where $k_{ex,T2}$ and $k_{ex,T1}$ are the exchange rate constants at temperatures T_1 and T_2 (K), R is the gas constant and the activation energies E_a for the acid, base, and water catalyzed exchange are 14 kcal/mol, 17 kcal/mol and 19 kcal/mol (Bai 1993), respectively. Thus, with every 10°C increase in temperature, the isotopic exchange rate can increase by 3-fold.

The dependence of isotopic exchange rate on pH and temperature provides the basis for quenching isotopic exchange, thereby facilitating detailed analysis of labeled proteins (Zhang and Smith 1993). At pH 7 and 25°C, the half-lives of the exposed NHs are in the range of 0.05~0.01 s, whereas at 0°C and pH 2.5, the half-lives of the exposed NHs range from 1 to 2 hr. In addition to the effects of temperature and pH, solvent composition also affects the isotopic exchange rates of amide hydrogens.

Because the equilibrium constant K_w for water decreases in the presence of organic solvents, the rates for isotopic exchange decrease. Consequently, the pH_{min} shifts to higher values with increasing fraction of organic solvents.

In polypeptides with random structures where few intramolecular hydrogen bonds exist and the amide sites are exposed to the solvent, NHs exchange rapidly. In folded proteins, the rates at which NHs undergo isotopic exchange appear to depend on whether the hydrogens are participating in intramolecular hydrogen bonding (i.e. in α-helices or β-sheets), and on the extent to which the hydrogens are shielded from the solvent (Englander 1984), as well as on the depth of the amide residues (Chakravarty 1999; Yan 2002). Isotopic exchange rate constants for NHs in folded proteins differ by many orders of magnitude. NHs in folded proteins or peptides exchange slowly because folded polypeptide chains involve a complex network of intramolecular hydrogen bonds, and because the folded structures limit the solvent exposure of NHs. The half-lives for isotopic exchange, which depend on the structure and the environment of the NH, typically span the range from milliseconds to months at pH 7.

Hydrogen exchange from folded proteins can be described by two proposed mechanisms (Li 1999): exchanging from the folded state and exchanging from the unfolded state (Figure 9).

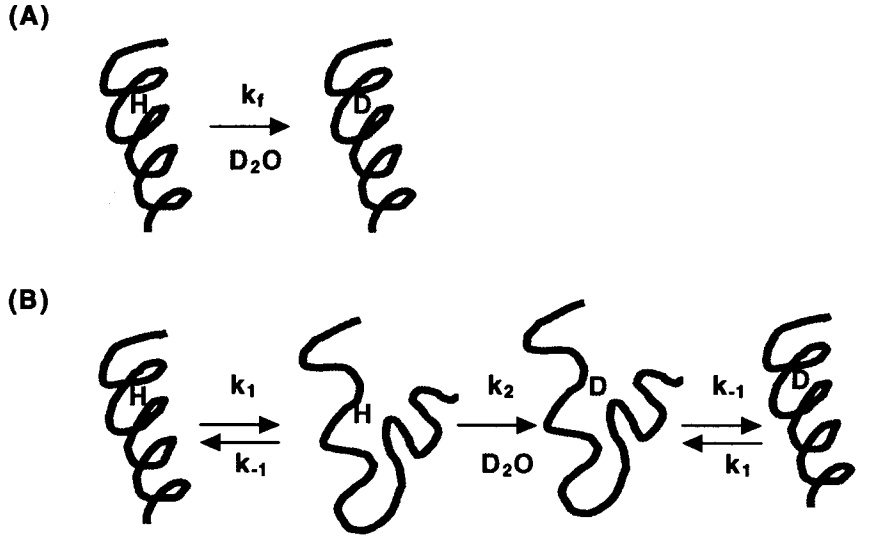


Figure 9. Pictorial presentation of the underlying principle of peptide amide hydrogen exchange. (A) Hydrogen deuterium exchange from the folded state. (B) Hydrogen deuterium exchange from the unfolded state.

The total exchange rate, K_{ex}, is the sum of exchange rate from the folded state, k_f, and the exchange rate from the unfolded state, k_u:

$$K_{ex} = k_f + k_u \qquad \text{Equation 4}$$

Amide hydrogens that exchange directly from the folded state are expected to dominate for NHs located on peptide bonds near the surface or open

channels within the folded protein. If the protein is folded at the instant of hydrogen exchange, the rate constant for isotopic exchange is designated as $k_f=\beta k$, where β is the probability of D_2O being present at the site, k is the exchange rate in an unstructured polypeptides and its value can be calculated. Amide hydrogens exchanging from the unfolded state are often involved in intramolecular hydrogen bonding. They may undergo local unfolding, which involves a rapid unfolding and refolding of small regions, as well as global unfolding that involves the entire protein. In this mechanism, the isotopic exchange can only take place after intramolecular hydrogen bonds are broken and the NHs become exposed to the deuterated solvent. When a protein is unfolded at the instant of hydrogen exchange, the rate constant for isotope exchange, k_{ex}, is related to k_1, k_{-1}, and k_2, as indicated in the following equation (Deng and Smith 1998)

$$K_{ex} = (k_1k_2)/(k_{-1}+k_2) \qquad \text{Equation 5}$$

When the folding rate is much greater than the isotope exchange rate ($k_{-1}>>k_2$), a region unfolds and refolds many times before isotope exchange is complete at all peptide linkages in the local region. This kinetic limit (EX2 kinetics) leads to a random distribution of deuterium in the unfolding region among all molecules comprising the sample. The mass spectra of peptides

derived from regions in which deuteriums are distributed randomly among all molecules exhibit one envelope of isotope peaks. Most proteins undergo EX2-type kinetics under physiological conditions. When the refolding rate is much smaller than the isotope exchange rate ($k_{-1}<<k_2$), a local region becomes completely deuterated the first time it unfolds (EX1 kinetics). In this case, the mass spectra exhibit two envelopes of isotope peaks, one corresponding to the unexchanged species, the folded state, and the other representing the exchanged species, the unfolded state.

There are two types of hydrogen exchange experiments, namely continuous labeling and pulsed labeling (Deng, Zhang et al. 1999). In the continuous labeling experiments, the protein is exposed to D_2O for a certain time period, while individual molecules fluctuate between folded and unfolded states. Molecules that are or become unfolded during the labeling time become deuterated, and molecules that have not yet unfolded during this time remain protonated. Thus, the deuterium level in a protein labeled continuously represents the sum of all molecules that unfold during the deuteration time, which may range from msec to days. In pulsed labeling experiments, the exposure of the protein to D_2O is short relative to the scale of the folding/unfolding event. Since little folding/unfolding occurs during the labeling step, the deuterium levels resulting from the pulsed labeling experiments indicate the instantaneous populations of folded and unfolded

molecules. The experimental condition for the labeling pulse must be carefully selected to ensure that all labile protons can exchange with solvent deuterium and all inaccessible protons remain shielded from D_2O. The optimal duration for exposure of the protein to D_2O depends primarily on pH, temperature, the amino acid sequence as well as the kinetics of the process of interest. Pulsed labeling hydrogen exchange experiments have been used to study rapid kinetics of various reactions taking place on the time scale of msec (Pinheiro 1997; Coyle 1999; Tsui, Garcia et al. 1999). These experiments require the use of an automated mixing apparatus such as the BioLogic quenched flow system (Appendix 1).

Most hydrogen exchange experiments involve the exchange of protein-bound protons for solvent deuteriums, known as the deuterium exchange-in approach or hydrogen/deuterium (H/D) exchange. In these experiments, labeled deuteriums represent amide sites that are accessible to the deuterated solvent at the time of exchange. This approach is useful for proteins that are difficult to prepare in the fully deuterated form. However, it suffers from back exchange during HPLC analysis. Therefore, this method is most appropriate for comparing structures or conformations between different proteins. Some hydrogen exchange experiments employ fully deuterated proteins and involve the exchange of protein-bound deuteriums for solvent protons, termed the deuterium exchange-out

experiments or deuterium/hydrogen (D/H) exchange. Deuterium labels in these experiments represent the amide sites that remain inaccessible at the time of exchange. In D/H experiments, there is a smaller deuterium loss during HPLC analysis. However, some proteins can be difficult to prepare in the fully deuterated form in which case they can only be studied by the deuterium exchange-in method.

1.4 Electrospray ionization mass spectrometry (ESI-MS)

Proteins that have undergone H/D exchange can be analyzed by nuclear magnetic resonance (NMR), which is considered the reference technique for measuring H/D exchange (Woodward 1999). This is because NMR is a residue-specific technique, revealing proton or deuterium occupancy at each amide site averaged over all molecules. However, hydrogen exchange studies monitored using NMR are restricted to large quantities (protein concentration of 1-2 mM) of highly soluble proteins with medium molecular weights (< 30,000 Da).

Mass spectrometry (MS) has become an important method for studying hydrogen exchange in proteins (Zhang and Smith 1993). This approach joins solution phase hydrogen exchange and mass spectrometric

measurement of the isotope composition of proteins and peptides. It offers some important advantages, including the ability to detect peptides and proteins of low quantity (subnanomolar), and to analyze proteins, protein-protein complexes and protein-nucleic acid complexes with molecular weights greater than 300,000 Da (Gale 1995). In addition, MS can distinguish populations of protein molecules that differ in molecular weight (Miranker, Robinson et al. 1993; Arrington 1999). Recently, it has been demonstrated that MS can provide amide site-specific information about H/D exchange in cytochrome c (Deng 1999) and thioredoxin (Kim 2001).

In ESI-MS, electrospray serves as an ionization source as well as an atmospheric pressure liquid inlet system for a mass spectrometer. The sample solution is sprayed through a capillary that is charged with high electrical potential (3~5 kV), yielding a mist of droplets mixed with vapor at atmospheric pressure. Through the vaporization of solvent, the density of electrical charge on the droplets increases until multiply charged ions form. When multiply charged ions are formed, the apparent mass of an ion is much smaller than its true mass ($m/z < m$ when $z > 1$). The ESI technique has found wide application in the analysis of proteins (Przybyski 1996), nucleic acids (Przybyski 1995), lipids (Murphy 2001) and noncovalently associated biomolecules (Gale 1995).

When applied to proteins, ESI-MS produces a series of charged ions. Generally, ESI mass spectra show a steady increase of these "charge states" for proteins with increasing molecular weights. The number of charged groups in solution has been found to be a determining factor for the maximum number of charges on macromolecular ions. Many ESI-MS experiments are carried out in acidic solutions, the distribution of charge states may be attributed to the accessibility of basic amino acid residues for protonation. Significant shifts of the ion distribution have been observed from a comparison of mass spectra of the native proteins and the disulfide bond reduced proteins (Loo 1990; Winger 1992). These studies have shown that the mass spectra obtained for the disulfide-bonded proteins favored substantially lower charge states than unfolded proteins in which the disulfide bonds have been reduced. Chowdhury, Katta and Chait have investigated electrosprayed cytochrome c as a function of solution pH (Chowdhury 1990). They found that as the solution pH decreased and the protein became unfolded, the distribution shifted from lower to higher charges. Other factors such as temperature, solvent composition and ionic strength, which lead to protein unfolding, can also influence protein conformation in solution. Loo examined the changes in ESI charge states for lyzozyme that was observed upon varying the fraction of several organic solvents (Loo 1991). When the fraction of organic solvent increased, a

shift to higher charge states was found indicating that the protein became denatured in solution. In these cases, the three dimensional conformations in solution are reflected in the structures of the gas-phase molecular ions. There is also evidence suggesting that gas phase protein ions retain the distribution of conformations that are present in solution (Loo 1994). The evidence indicates that both the properties of the solution and the nature of the gas phase ions can influence the ion structures in the final stages of ESI.

Most H/D exchange studies used to probe protein conformations have been carried out in solution (Smith 1994). Recently, H/D exchange studies were used to probe gas phase ion conformations (Suckau 1993; McLafferty 1998). McLafferty and co-workers used fourier transform ion cyclotron resonance (FT-ICR) MS to investigate the conformations of cytochrome c. In their experiment, the trapped ions were exposed to dilute concentrations of D_2O vapor and exchange was monitored for extended time periods (Suckau 1993). Under these conditions, cyt c was found to exist as multiple conformational states that were stable for hours, while in solution H/D exchange of cyt c involved rapid equilibration among the conformers. Comparison of the H/D exchange data between the solution and gas phases should provide information on the role of water in protein

structures, particularly where conformational changes take place as in folding pathways.

Because electrospray ionization is a soft ionization method, it does not produce fragment ions as readily as some other ionization techniques do. Consequently, various tandem mass spectrometry techniques have been developed to provide fragment ions, including collision induced dissociation (CID) (Loo 1988), infrared multiphoton dissociation (IRMPD) (Little 1994), blackbody infrared dissociation (BIRD) (Price 1996), surface induced dissociation (SID) (Chorush 1995), inelastic electron collision (electron impact excitation of ions from organics, EIEIO) (Wang 1990), and electron capture dissociation (ECD) (Zubarev 2000).

During CID experiments, precursor ions are selected in MS1, dissociated by collision activation in a region between MS1 and MS2, and scanned in MS2 (Figure 10).

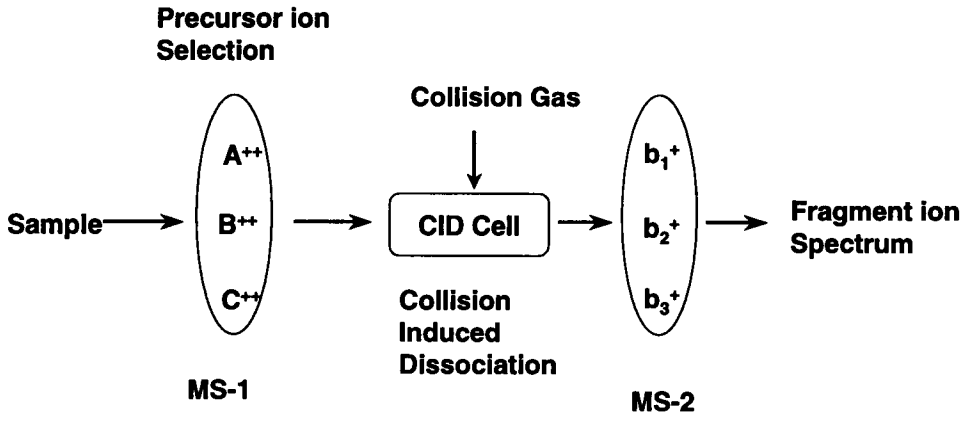


Figure 10. Collision induced dissociation used to identify fragment ions.

Peptides are fragmented along the backbone in specific places, yielding diagnostic ion series such as A_n, C_n, X_n, and Z_n (Figure 11).

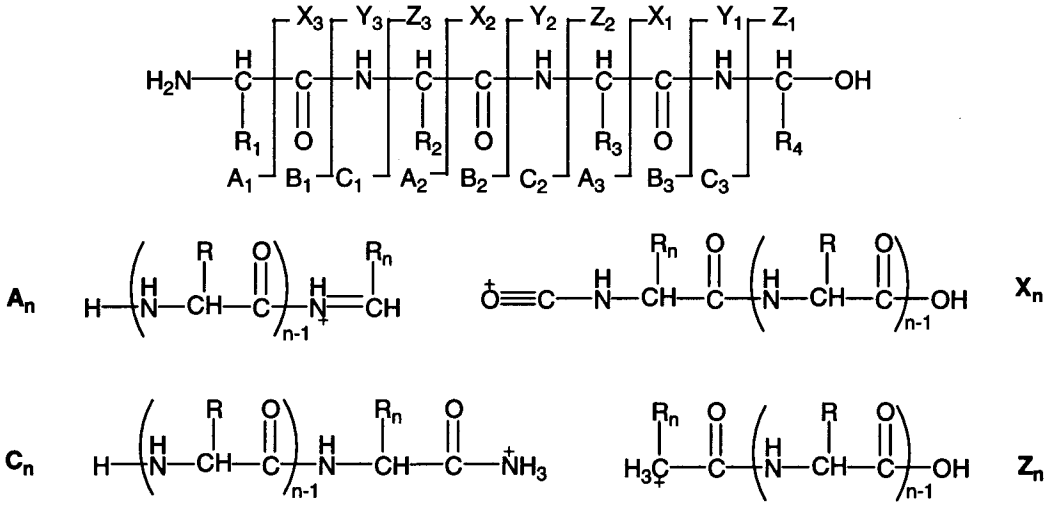


Figure 11. Fragment ions resulted from cleavage of peptide backbones.

The number of amides undergoing deuterium exchange in a protein or peptide can be calculated by subtracting the m/z value of the ion peak in the protonated form from the m/z value of the corresponding ion peak in the deuterated form and multiplied by the appropriate charge state. At individual amino acid residues, the deuterium content is obtained by comparing the b and the y ions in the deuterated and the nondeuterated forms.

Although similar analysis of the y ions often yield deuterium levels consistent with that found using the b ions, there are discrepancies suggesting that intramolecular H/D scrambling may occur during CID experiments for the y ions (Deng 1999). In the formation of y ions, the amide bond is cleaved and two hydrogen atoms are transferred to the C-terminal fragment (Figure 12).

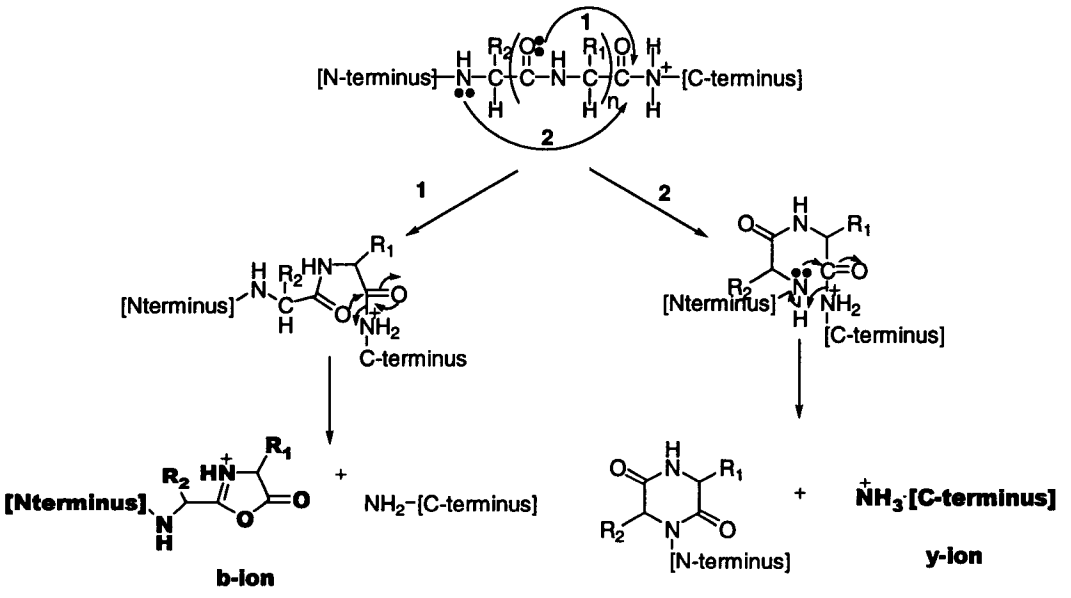


Figure 12. Proposed mechanisms for the formation of b and y ions.

One of these hydrogens presumably is from the protonating solvent and would therefore be expected to be a deuterium atom. The other hydrogen comes from an amide hydrogen attached to the N-terminal side of the cleavage site (Kenny 1992). This may lead to local scrambling of deuterium label. In contrast, the formation of b ions does not involve the transfer of any ambiguous hydrogen (Anderegg 1994). Thus, only b ions are used for studying amide site-specific deuterium exchange. There is evidence that electron capture dissociation (ECD) can fragment peptides into a series of c and z ions, which may also be appropriate for studying site-specific H/D exchange. However, fragmentation by ECD alone yields sequence information only for the terminal region peptides.

There are various methods for introducing samples into the ESI-MS: continuous flow infusion and on line separation systems such as liquid chromatography (LC) and capillary electrophoresis (CE). In H/D exchange experiments, the deuterium contents of proteins or peptides are often determined by LC-ESI-MS. During chromatography, deuterium can undergo back-exchange with protons from the HPLC solvent. Several experimental procedures can be adopted to minimize this back exchange. First, the pH and the temperature of the protein solution can be adjusted so that NH exchange rate is at a minimum (pH 2.5 and 0°C). Second, the HPLC injector and the column can be submerged in ice/water slurry, and

the transfer syringe pre-cooled on ice before use. Third, chromatography can be optimized for speed rather than resolution. Under these conditions, model studies have shown that the half-life for isotopic exchange at peptide amide linkages is 30~120 min (Bai 1993; Smith 1998).

To adjust for any unavoidable deuterium loss during analysis, a completely deuterated protein sample (100%) and an unlabeled protein sample (0%) are used as reference points (Zhang and Smith 1993). The deuterium content of partially deuterated peptides can be obtained from the following expression:

$$D = [(m-m_{0\%})/(m_{100\%}- m_{0\%})] \times N \qquad \text{Equation 6}$$

where D is the deuterium content in a particular peptide and $m_{0\%}$, m, and $m_{100\%}$ are the average molecular weights of the same peptide in the zero-time control partially deuterated sample and fully deuterated control respectively. N is the total number of peptide amide hydrogens in the peptide.

It should be noted that Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry has been successfully used for analyzing H/D exchange in protein complexes (Mandell 1998; Mandell 1998).

1.5 Research goals

rhm-CSFβ represents an excellent model system for studying disulfide bond formation during protein folding because the assembly of its monomeric subunits and the maturation of its biological activity depend on the progressive formation of the correct disulfide structure during *in vitro* folding (Glocker, Arbogast et al. 1994). The folding of rhm-CSFβ can be studied by isolating various disulfide-bonded species during reductive unfolding and oxidative folding. Furthermore, the intermediates can be structurally and energetically characterized and their locations on the unfolding or folding pathways defined. Knowledge obtained from these studies can be potentially useful in understanding the roles of disulfide bond formation during protein folding in general.

The specific goals of this project were to (1) selectively remove disulfide bonds by chemical reduction, characterize the structures of the protein derivatives, and access the bioactivity of the altered protein structures; (2) investigate the roles of specific disulfide bonds on protein folding; and (3) identify the intermediates in the oxidative folding pathways and their rates of formation and decomposition. In this project, H/D exchange was used to identify changes in relative solvent accessibility, which was related to structural changes as a result of disulfide bond

modifications. H/D and D/H exchange pulsed labeling experiments were used to detect the presence of transient intermediates during protein folding. ESI-MS was used to elucidate protein structures, identify protein modification sites, and analyze hydrogen deuterium exchange data.

2 Structural Comparison of Recombinant Human Macrophage Colony Stimulating Factor β (rhm-CSFβ) and Its Derivative Using Hydrogen Deuterium Exchange and Electrospray Ionization Mass Spectrometry

2.1 Abstract

Hydrogen deuterium exchange, monitored by electrospray ionization mass spectrometry, has been employed to characterize structural features of a derivative of recombinant human macrophage colony stimulating factor beta (rhm-CSFβ) in which two of the nine disulfide bridges (Cys157/Cys159-Cys'157/Cys'159) were selectively reduced and alkylated. Removal of these two disulfide bridges did not affect the biological activity of the protein. Similarities between CD and fluorescence spectra for rhm-CSFβ and its derivative indicate that removing the disulfide bonds did not strongly alter the overall three-dimensional structure of rhm-CSFβ. However, differences between deuterium exchange data of the intact proteins indicate that more NHs underwent fast deuterium exchange in the derivative than in rhm-CSFβ. Regions located near the disulfide bond removal site were shown to exhibit faster deuterium exchange behavior in the derivative than in rhm-CSFβ.

2.2 Introduction

Correctly folded protein is fundamental in the mechanism and control of a wide range of cellular processes. A thorough understanding of the complex process of protein folding requires the identification of the cooperative interactions among various folding events. This is intrinsically difficult because these events take place quickly and the transient intermediates cannot be readily obtained. However, in disulfide-coupled folding processes, folding intermediates can be isolated by selective reduction of disulfide bonds followed by alkylation of free cysteine residues. In this way, intermediates can be structurally characterized using a variety of physical techniques such as X-ray crystallography, NMR, CD and fluorescence spectroscopy. Recent developments combining hydrogen deuterium exchange and electrospray ionization mass spectrometry (H/D ESI-MS) has provided an additional tool for the characterization of protein-folding intermediates (Zhang and Smith 1993). It is non-perturbing, sensitive and offers sufficient structural and temporal resolution for identifying conformational changes resulting from ligand binding (Kragelund 1995; Wang 1997) and chemical modification (Resing 1998).

Macrophage colony-stimulating factor is the principle regulator of monocyte and macrophage development (Stanley 1997). The interaction of

m-CSF and its receptor, m-CSFR, induces proliferation of fibroblasts, or of hematopoietic cells, which results in differentiation towards the monocyte/macrophage lineage of myeloid progenitor cell lines (Bourette 2000). The recombinant human m-CSFβ, was expressed from *Escherichia coli* as a truncated form (aa 4-218, 49KDa) of one of the three human cDNA clones (Pandit 1992). It is biologically active only in the homo-dimeric form that is maintained by three inter- and six intra-molecular disulfide bridges (Figure 4). The systematic reduction of these disulfide bonds in rhm-CSFβ can reveal the specific role that each disulfide bond plays in maintaining the overall protein structural integrity as well as the biological activity. The disulfide bridges were assigned using enzymatic digestion and fast atom bombardment mass spectrometry (Glocker, Arbogast et al. 1993). Two intermolecular disulfide bonds in the C-terminal region were found between Cys157/159 and Cys'157/159. It is unclear whether Cys157 forms a disulfide bond with Cys'157 or Cys'159. Nonetheless, it is evident that these cysteine residues form symmetrical disulfide bonds that are in close proximity in the three-dimensional space. The present study focuses on the effect of removing Cys157/159-Cys'157/159 on the three-dimensional structure of rhm-CSFβ using the H/D-ESI-MS approach.

2.3 Results

2.3.1 Chemical modification.

Mild reduction at pH 3.5 was applied to rhM-CSFβ to achieve selective reduction of Cys157/159-Cys'157/159. The low pH was used in order to avoid disulfide bond scrambling. The reduced cysteine thiols were then alkylated with CDAP at pH 3.5. In contrast to what was reported in the literature (Wu 1997), TCEP was found to interfere with cyanylation and must be removed prior to the addition of CDAP to the partially reduced protein. Chromatographic separation by RP-HPLC provided a homogenous species whose molecular weight was determined to be 49,130 Da (Figure 13). The partially reduced cyanylated protein has a MW 100.6 Da greater in comparison to rhM-CSFβ (49,029.4 Da), a mass increase that corresponds to an incorporation of 4 cyano groups. The charge state distributions in ESI-MS for the native and the $CN^{157,159}$-modified rhm-CSFβ appear to be different (Figure 13).

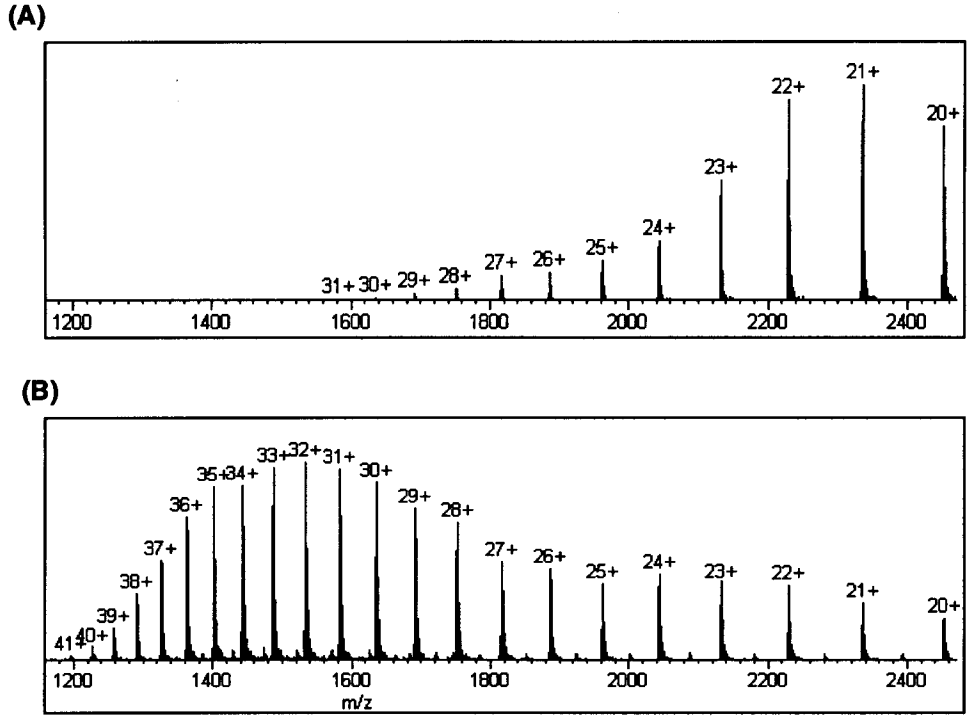


Figure 13. Mass spectra of rhm-CSFβ (A) and $CN^{157,159}$-modified rhm-CSFβ (B).

The native protein shows a distribution of charge states ranging from 28+ to 20+, with 22+ being the most intense peak. In contrast, the modified protein gives a distribution with higher charge states and a maximum around 32+.

2.3.2 Identification of modification sites.

To identify the sites of cyanylation, the modified protein was subjected to base hydrolysis (Wu 1997) which cleaved the N-terminal peptide bonds of the cyanylated cysteinyl residues and formed a cyclic ring structure. After cleavage at the cyanylated cysteinyl residues, the truncated peptide chains still linked by the remaining disulfide bonds were reduced by excess TCEP. The mass spectra indicated the presence of polypeptides 6,517.5 Da and 17840.8 Da (Figure 14).

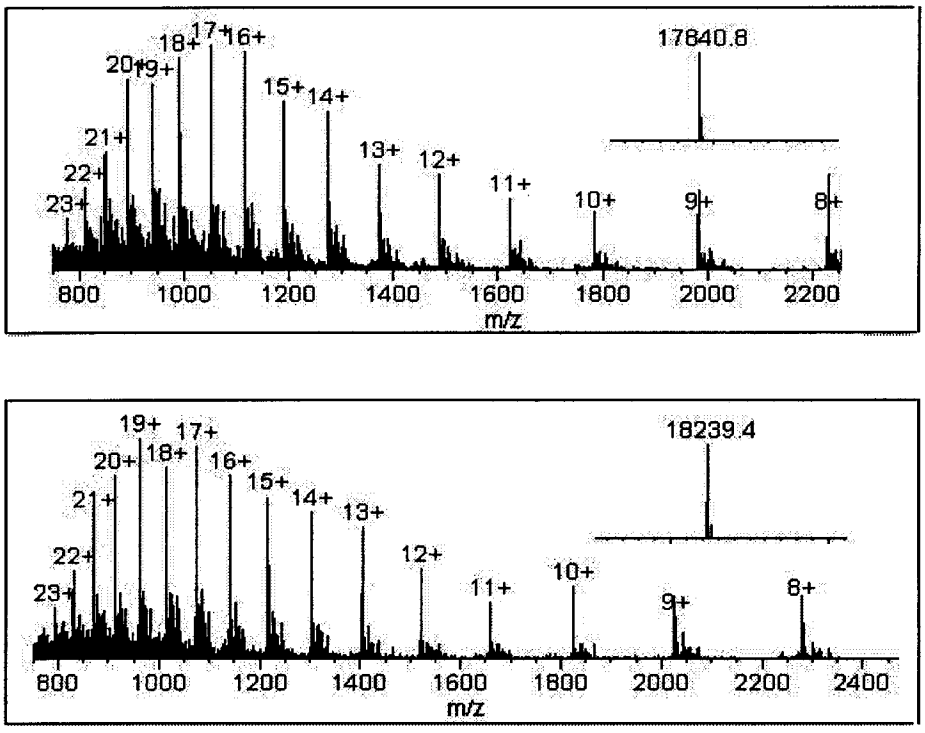


Figure 14. Mass spectra for protein intermediates in the reductive unfolding pathway. (A) Mass spectrum of polypeptide aa 4-156 (MW 17840. 8 Da) after base hydrolysis and reduction. (B) Mass spectrum of polypeptide aa 4-156 after complete reduction and carboxyamidomethylation (MW 18239.4 Da, $CAM^{7,31,48,90,102,139,146}$).

After complete reduction with TCEP, the nascent thio groups were alkylated with iodoacetamide. The mass spectra indicated there were species with molecular weights 6,518 Da and 18,239.4 Da (Figure 14). By comparing the predicted and experimental masses, the polypeptide with mass of 6,517 Da was identified as aa 159-221 with Cys^{159} cyanylated. The polypeptide with mass 17,840.8 Da was identified as aa 4-156 with no

alkylation, and the polypeptide with mass 18,239.4 was determined to be aa 4-156 with 7 cysteinyl residues carboxyamidomethylated. Peptide fragment aa 157-158 was not found; however, cyanylation must have taken place at Cys 157 based on the MW of the intact, modified protein. These results indicated that Cys 157 and 159 had been cyanylated and that the disulfide bonds linking these cysteinyl residues in the dimer had been reduced. The tryptic digestion of the completely reduced and carboxyamidomethylated protein further verified this conclusion. CID MS/MS experiments confirmed that Cys 157 and Cys 159 were the only cyanylated cysteinyl residues and all other cysteines were carboxyamidomethylated.

2.3.3 Biological activity.

M-CSF-dependent NFS-60 assay was used to compare the specific biological activity of the modified protein to that of rhm-CSFβ. The doubling time of cell growth that is stimulated by $CN^{157,159}$-modified rhm-CSFβ was identical to the doubling time of cell culture stimulated by rhm-CSFβ (Figure 15). Thus, the removal of Cys157/159-Cys'157/159 did not affect the biological activity of rhm-CSFβ.

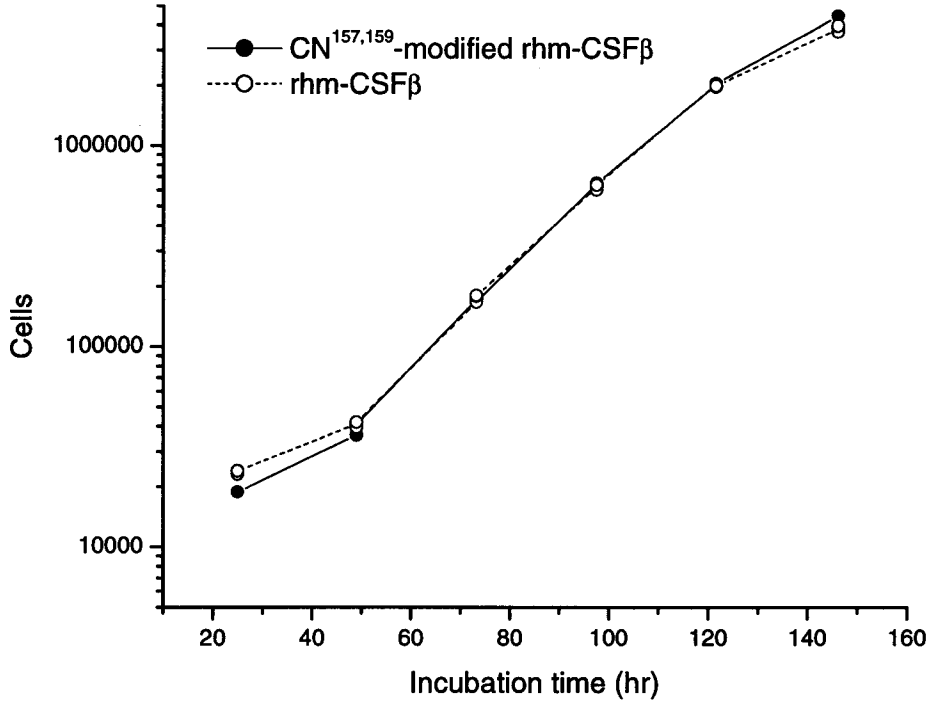


Figure 15. Biological activity of $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ determined by M-NFS-60 cell assay. Doubling time was determined to be 14.9 hr rhm-CSFβ and 15.08 hr for $CN^{157,159}$-modified rhm-CSFβ.

2.3.4 Circular and fluorescence spectroscopy.

The CD spectrum for $CN^{157,159}$-modified rhm-CSFβ closely resembles that of rhm-CSFβ which exhibits a positive maximum near 190nm and a negative minimum near 208nm (Figure 16). These CD spectra reflect that the structural integrity of the α-helical bundle is

maintained in the $CN^{157,159}$-modified rhm-CSFβ. Fluorescence emission spectra obtained by excitation at 280nm and 290nm for $CN^{157,159}$-rhm-CSFβ essentially overlap with those for rhm-CSFβ (Figure 17) demonstrating little structural change resulted from the chemical modification.

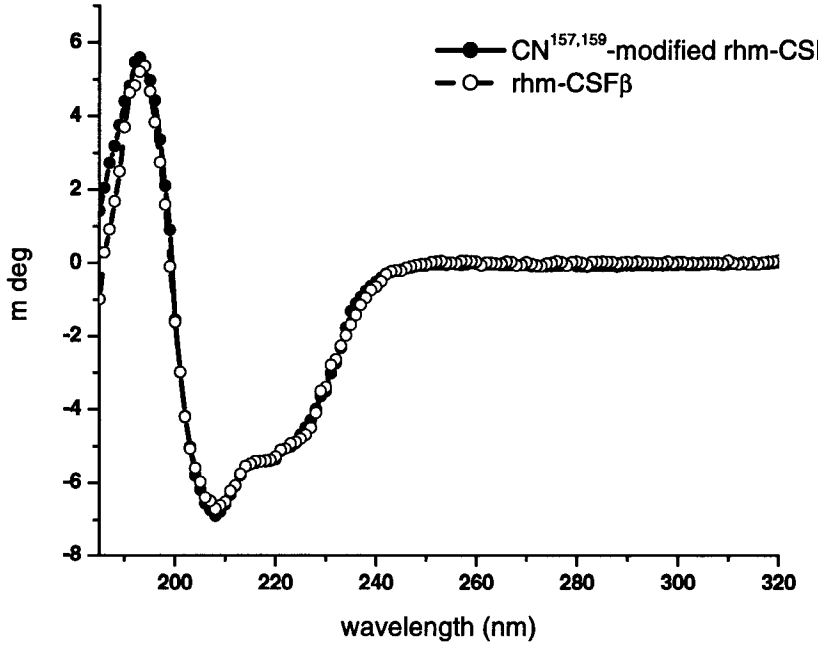


Figure 16. CD spectra of rhm-CSFβ and CN157,159-modified rhm-CSFβ in 10mM ammonium phosphate buffer.

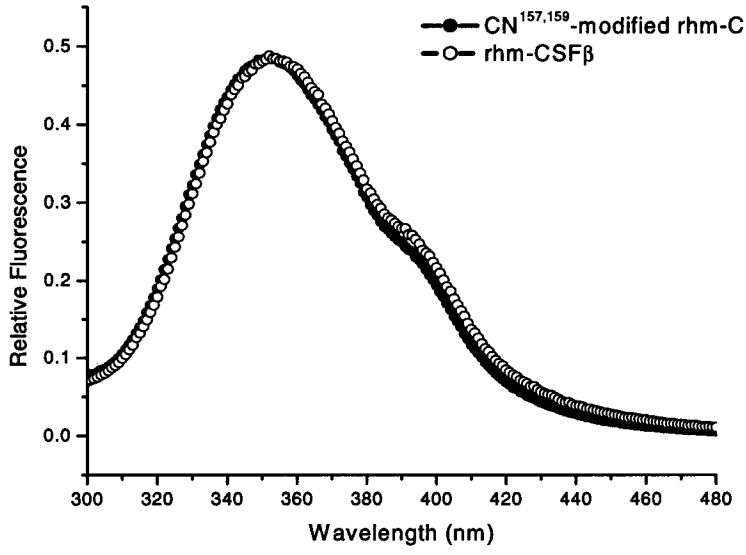


Figure 17. Fluorescence emission spectra of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ in 10mM ammonium phosphate buffer at pH 6.8.

2.3.5 Hydrogen deuterium exchange.

To probe detailed structural changes caused by removing Cys157/159-Cys'157/159, both $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ were incubated in D_2O for 4 s to 70 hr. The deuterium level in both proteins increased with incubation time until approximately 320 of the NHs had been replaced with deuterium (Figure 18).

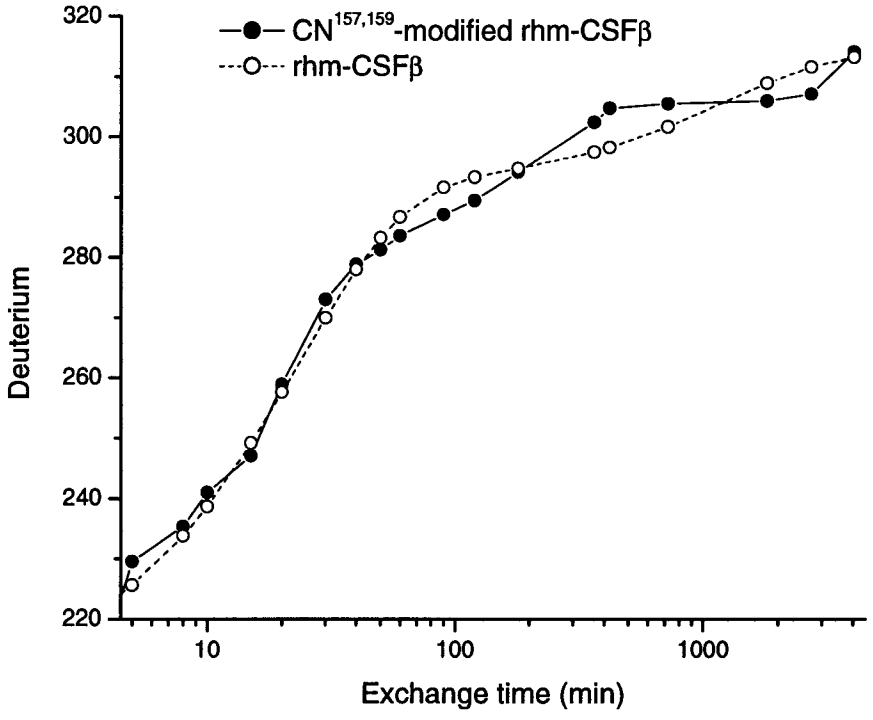


Figure 18. Dependence of deuterium in $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ on the exchange time (4s to 70h).

The change in deuterium level with incubation time was fitted to equation 2 to estimate rate constants for isotope exchange at peptide amide linkages (Table 1). These data reveal that the removal of Cys157/159-Cys'157/159 resulted in an increase of 19 fast exchanging NHs ($k>10min^{-1}$), a decrease of 9 fast exchanging NHs ($k\sim1min^{-1}$) and a decrease of 5 slow exchanging NHs ($k\sim0.01min^{-1}$). The shift in population in these rate categories indicates a general increase in the amount of

deuterium exchange as the disulfide bonds are removed resulting in a looser three-dimensional structure of the protein.

Table 1. Distribution of rate constants for isotope exchange at peptide amide linkages in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ .

	Number of Amide Hydrogens			
	$k > 10\ min^{-1}$	$k \sim 1\ min^{-1}$	$k \sim 0.01\ min^{-1}$	$k < 0.001\ min^{-1}$
rhm-CSFβ	153.34±5.40	57.25±5.90	50.72±4.6	142.71±15.9
$CN^{157,159}$- rhm-CSFβ	172.6±4.0	47.4±3.9	45.9±2.7	138.1±10.6

To locate the specific regions that became more flexible after the removal of Cys157/159-Cys'157/159, deuterium-labeled $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ were digested by pepsin prior to mass spectrometric analysis. Peptic digestion produced 45 and 38 overlapping peptides which cover >90% of the backbone in $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ respectively. The deuterium levels in each peptide were determined by analyzing the doubly charged fragments using CID MS/MS. The change in deuterium levels with incubation time was fitted to equation 2 to estimate rate constants for individual peptides (Table 2).

Table 2. Distribution of rate constants for isotope exchange at peptide amide bonds in rhm-CSFβ (N) and $CN^{157,159}$-modified rhm-CSFβ (M) (k, min^{-1}).

Segment	Fastest (>10)	Fast	Slow	Slowest (<0.001)
4-19 N	10.27	0.350 (0.15)	0.71 (0.024)	3.67
M	10.35	0.63 (0.04)	0.70 (0.008)	3.32
37-55	5.81	2.38 (0.39)	3.25 (0.035)	5.56
	5.84	3.31 (0.29)	2.62 (0.018)	5.23
56-62	0.65	1.33 (4.76)	1.15 (0.113)	2.87
	0.38	1.50 (0.54)	1.22 (0.004)	2.90
63-76	5.18	1.24 (0.22)	2.98 (0.031)	2.60
	4.80	1.37 (0.74)	3.54 (0.058)	2.29
77-82	1.31	0	1.22 (0.015)	2.47
	0.86	0.63 (1.47)	1.16 (0.011)	2.35
83-105	5.33	5.03 (0.44)	4.95 (0.016)	5.69
	5.19	5.75 (0.45)	5.30 (0.009)	4.76
106-113	1.73	0.80 (0.37)	1.18 (0.030)	2.29
	1.54	0.86 (0.68)	0.76 (0.098)	2.84
114-143	11.7	6.02 (0.46)	2.92 (0.014)	8.36
	11.68	5.8 (0.5)	4.09 (0.013)	7.43
151-189	25.11	0.27 (0.50)	0.22 (0.006)	4.4
161-189	22.15	0.37 (1.00)	0.23 (0.002)	0
	22.30	0.85 (1.88)	0.30 (0.018)	0
151-160	6.30	0.37 (1.35)	0.51 (0.015)	0.82
190-221	24.56	0.24 (0.36)	0.48 (0.001)	2.72
	25.94	1.07 (1.35)	0.11 (0.10)	0.88

Local structural differences between the two proteins can be identified by comparing deuterium levels and kinetic rates for the corresponding peptides. As seen with whole proteins, most of the peptides exhibit fairly similar deuterium exchange behavior. This is particularly true in the N-terminal region (aa4-149) where exchange kinetics of peptides

$CN^{157,159}$-modified rhm-CSFβ are identical to those in rhm-CSFβ. The deuterium level found in the segment covering residues 4-19 (Figure 19) as a function of the time the intact proteins were incubated in D_2O was fitted to a series of exponential functions.

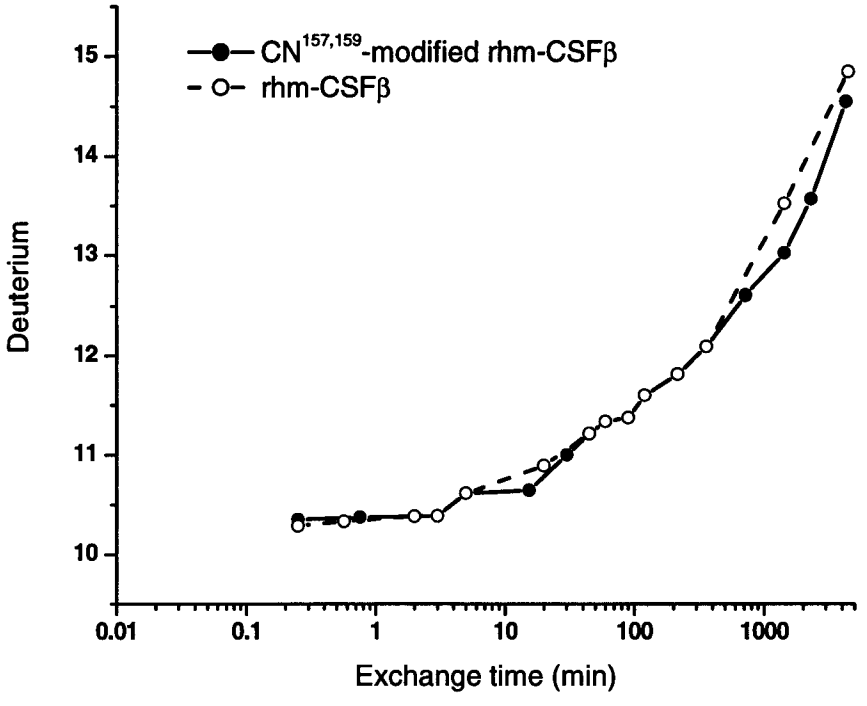


Figure 19. The deuterium levels found in segments covering aa 4-19 in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ are shown as a function of the exchange time (4s to 40h).

The results indicate that in rhm-CSFβ, 10.27 amide hydrogens in the segment 4-19 have isotope exchange rate constants greater than $10min^{-1}$, 0.35 amide hydrogens of 0.15 min^{-1}, 0.71 amide hydrogens of $0.024min^{-1}$, and 3.67 amide hydrogens of exchange rate constants smaller than $0.001min^{-1}$ (Table 2). In comparison, the same segment derived from $CN^{157,159}$-modified rhm-CSFβ had 10.35 amide hydrogens which exchanged with rates greater than $10min^{-1}$, 0.63 at $0.04min^{-1}$, 0.7 at $0.008min^{-1}$, and 3.32 with rate constants that were less than $0.001min^{-1}$. The deuterium exchange rates are independent of whether disulfide bonds linking Cys157/159-Cys'157/159 are present. Similar analyses were done for other segments covering residues 4 to 153 derived from both proteins. Since no difference in amide hydrogen exchange rates was detected in this region, these results suggest that removing these disulfide bonds has little effect on the structure in the N-terminal region.

The C-terminal region of $CN^{157,159}$-modified rhm-CSFβ (aa151-221) showed a higher population of very fast and fast exchanging NHs than in rhm-CSFβ. This is not surprising because Cys157, 159 are located in this area. The deuterium level found in the segment derived from the N-terminal side of Cys157,159 (aa 151-160) in $CN^{157,159}$-rhm-CSFβ was fitted to equation 2. In this segment, 6.3 NHs underwent deuterium exchange with rate constants greater than $10min^{-1}$, 0.37 NHs with k of $1.35min^{-1}$ and

0.51 NHs with k of 0.015min^{-1} (Table 2). The remaining 0.82 amide hydrogens had deuterium exchange rate constants less than 0.001min^{-1}. This exact peptide was not found in the peptic digest of rhm-CSFβ. However, information on deuterium levels in this segment can be deduced by analyzing two segments covering residues 151-189 and 161-189. The difference between deuterium levels for these two segments yields information regarding the segment aa151-160 in rhm-CSFβ. Results from such a comparison (Table 3) show that in the segment aa151-160 of rhm-CSFβ, 2.23 NHs underwent very fast or fast exchange with exchange rate constants greater than 0.5min^{-1} and 4.4 amide hydrogen exchanged at rate constants less than 0.001min^{-1}.

Table 3. Distribution of rate constants for isotope exchange at peptide amide bonds in rhm-CSFβ and CN^{157,159}-modified rhm-CSFβ (k, min^{-1}).

	Very fast	Fast	Slow	Very slow
rhm-CSFβ 151-189	25.11 (>10)	0.27 (0.50)	0.22 (0.006)	4.4 (<0.001)
rhmCSFβ 161-189	22.15 (>10)	0.37 (1.00)	0.23 (0.002)	0 (<0.001)
rhm-CSFβ 151-160	2.96 (>10)	0 (~1)	0 (~0.002)	4.4 (<0.001)
modified 151-160	6.30 (>10)	0.37 (1.35)	0.51 (0.015)	0.82 (<0.001)
Difference	3.34 (>10)	0.37 (~1)	0.51 (~0.01)	-3.58 (<0.001)

It is significant that the deuterium exchange pattern in this segment of rhm-CSFβ is different from the same segment derived from $CN^{157,159}$-rhm-CSFβ. Most importantly, 4 amide hydrogens which exchanged with small rate constants in rhm-CSFβ became large in the modified protein, thereby showing that local structural changes resulted from removing Cys157/159-Cys'157/159. The deuterium level data found in the segment residues 191-221 (Figure 20) was fitted to equation 2.

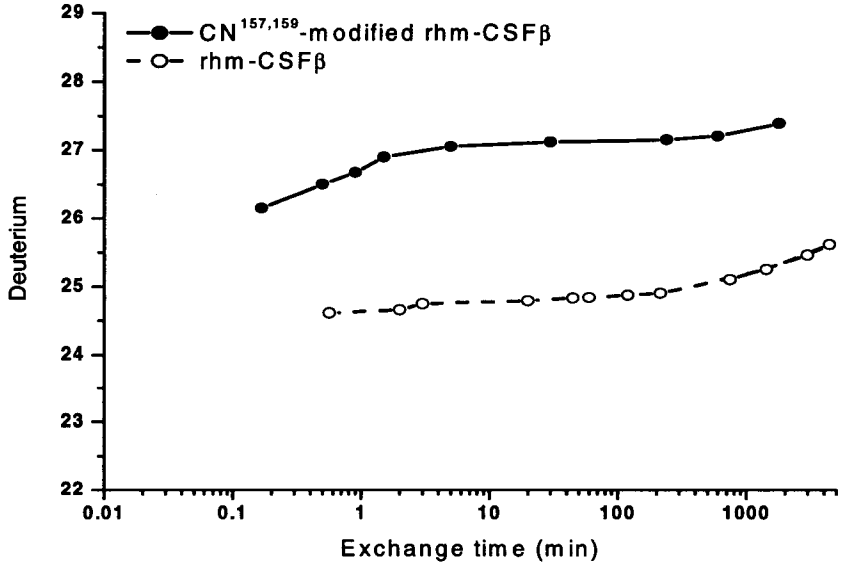


Figure 20. Deuterium levels found in segments covering aa190-221 in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ are shown as a function of the exchange time (4s to 40h).

Data analysis shows that in the segment derived from the modified protein, 25.94 amide hydrogens have isotope exchange rate constants greater than $10min^{-1}$, 1.07 amide hydrogens with k of 1.35 min^{-1}, 0.11 amide hydrogens with k of $0.01min^{-1}$, and 0.88 amide hydrogens with exchange rate constants smaller than $0.001min^{-1}$. In contrast, in the same segment derived from rhm-CSFβ, 24.56 amide hydrogens exchanged with rate constants greater than $10min^{-1}$, 0.24 at $0.36min^{-1}$, 0.48 at $0.001min^{-1}$, and 2.72 exchanged with rate constants smaller than $0.001min^{-1}$ (Table 2). Clearly, the deuterium levels in the segment of the modified protein are greater than in rhm-CSFβ, suggesting that at least some part of the structure of the former is more open. However, the difference is quite small which indicates that the overall conformation in this area is not significantly different for the two proteins. Such analysis for other peptides covering aa 151-221 shows similar results: nearly all NHs in this region underwent fast isotope exchange in both proteins, suggesting very little protection against deuterium exchange. In addition, peptides derived from the modified protein consist of more fast exchanging NHs than those from rhm-CSFβ, indicating that the C-terminal region in the modified protein is more flexible than in rhm-CSFβ.

To demonstrate a possible correlation between amide hydrogen exchange rates and intramolecular bonding NHs in α helices and β sheets,

the number of peptide amide hydrogens participating in hydrogen bonding was estimated for each segment of rhm-CSFβ using available x-ray data. These results (Table 4) expressed as the percent of hydrogen bonding Hs in each segment are the percent of NHs with exchange rate constants less than 0.01 min^{-1}. It is evident that the slowest exchanging amide hydrogens are located in segments in which the amide hydrogens participate in intramolecular hydrogen bonding.

Table 4. Percent of slow exchanging ($k<0.05min^{-1}$) amide hydrogens in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ compared with percent of hydrogen bonding amide hydrogens identified by X-ray crystal data of m-CSFα.

Segment	rhm-CSFβ	$CN^{157,159}$-rhm-CSFβ	α or β (%)
4-19	29.2	31.2	46.7
37-55	51.8	46.2	58.8
56-62	67.0	52.0	100
63-76	46.5	56.9	25.0
77-82	73.8	70.2	100
83-105	48.4	45.7	36.4
106-113	57.8	60.0	100
114-128	71.4	75.2	100
114-143	59.7	59.7	69.0

2.4 Discussion

The removal of disulfide bonds linked by Cys157/159-Cys'157/159 was shown not to affect the biological activity. The overall three-

dimensional structure of the protein was not significantly affected as demonstrated by CD and fluorescence spectroscopy. However, some structural variations between the native and the modified rhm-CSFβ were revealed by the charge state distributions in the ESI-MS spectra of the intact proteins (Figure 13). The distribution of charge state centers on 21+ in native rhm-CSFβ and 33+ in the modified protein. The difference of +12 in the dimeric proteins may be attributed to the accessibility of six additional basic amino acid residues (K^{154}, D^{156}, Y^{161}, K^{163}, D^{169}, H^{176}) for protonation as the surface area in the C-terminal region increased resulting from the removal of Cys157/159-Cys'157/159. Hydrogen deuterium exchange experiments also indicated observable structural changes that had taken place upon the removal of Cys157/159-Cys'157/159. As a result of disulfide bonds' removal, there is an increase of 19 fast exchanging NHs ($k>10min^{-1}$), a decrease of 9 fast exchanging NHs ($k\sim1min^{-1}$) and a decrease of 5 slow exchanging NHs ($k\sim0.01min^{-1}$). The shift in population in these rate categories indicates a general increase in the amount of deuterium exchange as the disulfide bonds are removed and a less compact regional three-dimensional structure of the protein resulted. As demonstrated in this study, structural changes that are not readily detected by optical methods can be identified and localized by ESI charge state distribution in combination with hydrogen/deuterium exchange experiments.

Detailed structural features of the C-terminal region of rhm-CSFβ have not been revealed by x-ray crystallography. The present study provides significant structural descriptions of this region. The similar CD spectra indicate that the α-helical bundle core is not disturbed by the removal of Cys157/159-Cys'157/159. The fluorescence data also indicates little change as the disulfide bonds are removed. Interestingly, Trp^{191} is located on the C-side of the disulfide bond removal site. Apparently, the modification has no effect on the fluorescence property of this residue. However, significant differences in deuterium exchange behavior of this region have been revealed by hydrogen/deuterium exchange studies. These data suggest that within the C-terminal region (aa151-221) there are sub-regions that are affected to various degrees by the removal of Cys157/159-Cys'157/159. Upon the removal of Cys157/159-Cys'157/159, the sub-region that is on the N-side of the disulfide bond removal site (aa151-159) is shown to exhibit more significant exchange behavior than the sub-region on the C-side (aa160-221). It is possible that Cys157/159-Cys'157/159 are placed relatively close to the protein α-helical core (Figure 4). Upon the removal of these disulfide bonds, the sub-region (aa151-159) whose structure maintained by Cys157/159-Cys'157/159 becomes flexible and exhibits different deuterium exchange behavior. The sub-region that is already flexible (aa160-221) prior to the chemical modification is not

strongly affected by the modification as supported by the unchanged fluorescence property of Trp^{191} as well as the peptide deuterium exchange data.

Hydrogen isotope exchange kinetics for the intact protein and individual peptides combine to provide a detailed structural characterization. Fitting of global exchange data separates amide hydrogens into groups with different exchange rates. In contrast, the summation of NHs within the same exchanging group of peptides covering 90% of the backbone does not add up to the same number provided by global exchange data. This is partly because peptides covering 10% of backbone could not be identified in the peptic digest. More importantly, the N-terminal NH from every peptide was lost during the HPLC step. In some cases, depending on the sequence of the peptide, the NH from the amino acid second from the N-terminus was also lost (Engen, Gmeiner et al. 1999). In contrast, in the intact protein, only the NH at the N-terminus was lost. Irrespective of the more significant loss of labeling determined by peptide analysis, deuterium exchange rate data on the peptide level can locate the regions where the three-dimensional structure has changed. Even though this approach can only estimate the number of NHs within a given range of exchange kinetics, it is useful in the identification of structural alteration caused by chemical modification, mutation and ligand binding. A similar approach that focuses

on the CID MS/MS experiment can be used to estimate deuterium levels at individual peptide bonds (Deng 1999). In this way, more detailed structural information can be obtained for large proteins and protein complexes.

2.5 Conclusion

By cDNA cloning, another form of m-CSF, named m-CSFα can be obtained. In the membrane-bound form, m-CSFα and β are identical in their N-and C-terminal regions, but m-CSFβ contains inserts at residue 149 as a result of alternative mRNA processing (Pandit 1992). The N-terminus (aa1-149) contains the receptor-binding domain, and the C-terminus contains a transmembrane region that is believed to anchor m-CSF in an active, cell-associated form. It has been proposed that m-CSF is expressed in variable lengths from a single gene in order to provide membrane-associated m-CSF with variable length spacers to allow direct contact with receptors on neighboring cells (Cerretti 1988). In the membrane-bound form of m-CSFβ, aa 150-221 is part of the length spacer that connects the signal sequence (N-terminus) to the trans-membrane domain (C-terminus). This agrees with previous experiments which indicated that the C-terminal

region in m-CSFβ (aa150-221) is not important for the protein's biological activity (Pandit 1992).

A previous study reports the characterization of a stable dimeric intermediate isolated from the *in vitro* folding reaction of rhm-CSFβ (Maier, 1999). This refolding intermediate contains seven disulfide bonds and four carboxyamidomethylated cysteinyl residues. In addition, its charge state distribution closely resembles that of the $CN^{157,159}$-modified rhm-CSFβ indicating they may be the same intermediate that exists in both the folding and unfolding pathways.

The present study suggests that Cys157/159-Cys'157/159 play a minor role in maintaining the three-dimensional structure in the C-terminus of the protein. An important question arises: do these disulfide bonds play any role in the folding pathway of rhm-CSFβ? Most likely, these disulfide bonds help to stabilize the folded structure rather than introducing structure into the polypeptide chain as the protein folds. It is possible that they may provide additional stability for the proper protein conformation in the final stage of folding.

2.6 Materials and methods

2.6.1 Materials.

Purified, biologically active rhm-CSFβ was provided by Dr. C. Cogwill (Chiron Corp. Emeryville, CA). NFS-60 cell line was provided by Dr. J. Weaver (Chiron Corp. Emeryville, CA). Tris(2-carboxyethyl)phosphine (TCEP) and immobolized pepsin were obtained from Pierce (Rockford, IL). 1-cyano-4-dimethylaminopyridinium fluoroborate (CDAP) was obtained from Sigma Chemicals. Modified trypsin was purchased from Promega (Madison, WI). Deuterium oxide and urea-d_4 were obtained from Aldrich Chemical Co. (Milwaukee, WI). All other chemicals and reagents were of the highest grade commercially available.

2.6.2 Partial reduction of rhM-CSFβ.

The freshly dissolved rhM-CSFβ (12nmol) was solubilized in 300 μl of 20 mM ammonium acetate buffer (pH 3.5) containing 1 mM EDTA. Partial reduction of rhM-CSFβ was initiated by adding 0.05 equivalent of

TCEP of total cysteine content and followed by incubation at room temperature for 5 hrs. After TCEP was removed by centrifugal filtration using ultrafree MC filters with a 5,000 Da molecular mass cutoff (Millipore), a 25-fold molar excess of CDAP solution over the cysteine content was added to the partially reduced protein mixture. Cyanylation of the sulfhydryl groups was done at room temperature for 10 min. CDAP was then removed by centrifugal filtration. The reduced, cyanylated protein was purified by reversed phase high performance liquid chromatography (HPLC) on a Vydac C4 TP214 column (300Å, 5 μm, 4.6X250 mm, 1 ml/min flow rate) employing a binary linear gradient elution with 0.1% trifluroacetic acid (TFA) as solvent A and acetonitrile (ACN) containing 0.09% trifluroacetic acid as solvent B. The proteins were eluted using a 60 min 44-54% B gradient. Protein fractions were collected manually and lyophilized.

2.6.3 Identification of reduced disulfide bonds.

The lyophilized modified rhM-CSFβ was dissolved in 30 μl of 6M guanidine-HCl in 1 M NH_4OH (pH 11.3). Cleavage of the peptide chain was performed by adding 30 μl of 1 M NH_4OH (pH 11.5) followed by incubation at room temperature for 1 hr. Excess ammonia was removed in a vacuum

centrifuge. Truncated peptides containing residual disulfide bonds were completely reduced by reacting with 250 µl of 7.2 mM TCEP solution at 37°C for 30 min (pH 8.5). To the completely reduced protein mixture, 8 µl of iodoacetamide (20mg/ml) in 0.1M TrisCl at pH 8.5 was added and the mixture was incubated for 1 hr at room temperature. The reaction mixture was subjected to mass spectrometric analysis utilizing a Shimadzu HPLC equipped with two LC-10AD pumps coupled to a PE-Sciex API III triple quadrupole mass spectrometer equipped with a nebulizer-assisted electrospray source. A capillary column packed with Vydac C4 TP214 material (5 µm particle size, 300 Å pore size, 20 µL/min flow rate) was used. The binary gradient elution included 0.05%TFA as solvent A and ACN containing 0.05% TFA as solvent B. The proteins were eluted using a 8min 10-90% B gradient. MS spectra were analyzed using PE Sciex software (MacSpec and BioSpec).

The carboxyamidomethylated protein mixture was subjected to tryptic digestion by reacting with 2 µg of trypsin in 100 µl of 0.1 M tris buffer (pH 7) containing 1 mM $CaCl_2$. The tryptic digest mixture was analyzed using a Waters HPLC equipped with two pumps coupled to a Finnigan quadrupole ion-trap mass spectrometer. A capillary column packed with C18 Luna material (5 µm particle size, 300 Å pore size, 5 µL/min flow rate) was used. The identities of tryptic peptides were determined by collision-

induced dissociation (CID) experiments. The biological activity of the modified protein was determined using m-CSF dependent NSF-60 cell culture assay (Nakoinz 1990).

2.6.4 Circular dichroism and fluorescence spectroscopy.

CD spectra were obtained using a Jasco J720 spectropolarimeter (Jasco Inc., Easton, MD) over the range of 185-310nm in a cylindrical 1-cm path length quartz cuvette and with a bandwidth of 2nm and a scan speed of 10nm/min. Fluorescence data was acquired on a 8100C fluorescence spectrometer (SLM Instruments Inc., Urbana-Champaign, IL) with 8nm emission and excitation slit widths using a square 0.5 cm path length quartz cuvette.

Fluorescence was measured at excitation wavelengths of 280nm and 290nm and expressed as a ratio of emission intensity to reference emission intensity.

2.6.5 Deuterium exchange-in experiments.

The lyophilized $CN^{157,159}$-rhm-CSFβ (250 μg) was dissolved in 10 μl of 10 mM ammonium phosphate buffer (pH 6.8) containing 1 mM EDTA. In-exchange was initiated by diluting the protein solution 20-fold with 10 mM ammonium phosphate/D_2O buffer (pH 6.8) containing 1 mM EDTA. At each time point, 10 μg of protein sample was removed from the labeling solution and diluted 1:1 with 0.1 M ammonium phosphate buffer (pH 2.3, 0°C, 1:1 D_2O/H_2O) to quench isotopic exchange. The samples were stored in liquid N_2 until analysis. Unlabeled protein (0% reference) was prepared by dissolving the modified protein in 0.1 M ammonium phosphate buffer (pH 2.3, 0°C, 1:1 D_2O/H_2O). The completely deuterated protein (100% reference) was prepared by incubating the modified protein in 8 M urea-d_4/ D_2O at 37°C for 6h. The same exchange procedure was applied to rhm-CSFβ.

The deuterated protein samples were analyzed by LC-ESI-triple quadrupole MS (PE-Sciex API III^+) using a self-packed C4 column (5 μm particle size, 300 Å pore size, 20 μL/min flow rate) and a binary gradient elution involving 0.05% TFA as solvent A and ACN containing 0.05%TFA as solvent B. After desalting at 10%B for 2min, the protein was eluted using a 3 min 10-80% B program. The HPLC injector and the column were

submerged in ice/water slurry, and the transfer syringe was rinsed with D_2O and pre-cooled on ice before use. MS spectra were analyzed by BioSpec software.

The deuterated $CN^{157,159}$-rhm-CSFβ was digested by immobolized pepsin (protein/enzyme = ½ (w/v)) in 0.1 M ammonium phosphate buffer (pH 2.3, 0°C, 1:1 D_2O/H_2O) for 8min. The digest mixture was analyzed using LCQ-ESI-ion trap mass spectrometer using a C18 Luna column (5 μm particle size, 100 Å pore size, 15 μL/min flow rate) and 0.05%TFA as solvent A and ACN containing 0.05% TFA as solvent B. After being desalted for 2min, peptides were eluted using the following program: 10-20%B for 1min, 20-30%B for 1min, 30-45%B for 2min, 45-60%B for 1min, 60-90%B for 3min. To minimize deuterium back exchange, the injector and the column were immersed in an ice/water bath, and the transfer syringe was rinsed with D_2O and cooled on ice before use. Due to the compactness of the modified protein, this digest procedure resulted in only 57% coverage of the protein backbone. Consequently, the complete digestion of the deuterated modified protein was obtained by incubating deuterated protein with 0.1 M ammonium phosphate buffer (pH 2.3, 0°C, 1:1 D_2O/H_2O) containing 8 M urea-d_4 and 1 M TCEP at 0°C for 5min prior to pepsin digest. This procedure yielded 90% coverage of the entire backbone but did not provide cyanylated peptides because TCEP reduced

cyanylated cysteine residues. The peptides were identified by CID experiments. The deuterated rhm-CSFβ was subjected to pepsin digestion using the same procedure as described above.

2.6.6 Data analysis.

Because the HPLC elution was performed with protiated solvents, a small number of deuteriums at the amide positions are lost. Corrections were made for this deuterium loss using eq. 1 (Zhang and Smith 1993):

$$D = [(m-m_{0\%})/(m_{100\%}- m_{0\%})] \times N \qquad \text{Equation 6}$$

where D is the number of deuteriums present in a particular peptide segment or protein after incubation in deuterated solvent, and m, $m_{0\%}$, and $m_{0\%}$ represent the average molecular weight of a peptide or protein obtained for non-deuterated, partially deuterated and completely deuterated samples, respectively. N is the total number of exchangeable amide hydrogens in the protein or the peptide of interest. The present experiments yielded 10-20% back exchange in both proteins and peptides.

Amide hydrogens undergo deuterium exchange at different rates depending on whether the hydrogens participate in intramolecular hydrogen bonding, as well as on the extent to which the hydrogens are shielded from the solvent. Exchange-in data for proteins and peptides can be analyzed using eq. 2 (Engen, Smithgall et al. 1997), which distinguishes amides according to whether they are fast, intermediate, slow or non-exchanging:

$$D = N - \sum_{i=1}^{n} \exp(-k_i t) \qquad \text{Equation 7}$$

where D is the deuterium level found in a protein or peptide with N amide linkages following incubation of the intact protein in the deuterated solvent for time t and k_i are the first order rate constants for isotopic exchange at every amide linkage. Thus, higher-order structure, structural changes, and structural dynamics of the protein that control exchange rate behavior can be defined. Data fitting for deuterium exchange was accomplished using Origin6.1 (Microcal Software, Inc).

2.7 References

Bourette, R. P. R., L.R. (2000). Early events in m-CSF receptor signaling. *Growth Factors* **17**, 155-166.

Cerretti, D. P. W., J.; Anderson, D.; Cosman, D. (1988). Human Macrophage Colony Stimulating Factor: An Alternative RNA and Protein Processing from a Single Gene. *Mol. Immunol.* **25**(8), 761-770.

Deng, Y. P., H; Smith, D.L. (1999). Selective Isotope Labeling Demonstrates that Hydrogen Exchange at Individual Peptide Amide Linkages can be determined by Collision-Induced Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **121**, 1966-1967.

Engen, J. R., Gmeiner, W. H., Smithgall, T. E. & Smith, D. L. (1999). Hydrogen exchange shows peptide binding stabilizes motions in Hck SH2. *Biochemistry* **38**(28), 8926-35.

Engen, J. R., Smithgall, T. E., Gmeiner, W. H. & Smith, D. L. (1997). Identification and localization of slow, natural, cooperative unfolding in the hematopoietic cell kinase SH3 domain by amide hydrogen exchange and mass spectrometry. *Biochemistry* **36**(47), 14384-91.

Glocker, M. O., Arbogast, B., Schreurs, J. & Deinzer, M. L. (1993). Assignment of the inter- and intramolecular disulfide linkages in recombinant human macrophage colony stimulating factor using fast atom bombardment mass spectrometry. *Biochemistry* **32**(2), 482-8.

Kragelund, B. B. K., J.; Poulsen, F.M. (1995). Local Perturbations by Ligand Binding of Hydrogen Deuterium Exchange Kinetics in a Four-Helix Bundle Protein, Acyl Coenzyme A Binding Protein (ACBP). *J. Mol. Biol.* **250**, 695-706.

Maier, C.S.; Yan, X-G.; Zhang, H.Y.; Schimerlik, M.I.; Dienzer, M.L. (1999). Investigating the glutatione-dependent *in vitro* folding pathway of homodimeric cytokine rhm-CSFβ. *Proc. 47th. Conf. Mass. Spectrum. Allied Topics.* Dalles, Texas.

Nakoinz, I. L., M.T.; Weaver, J. F.; Ralph, P. (1990). Differentiation of the IL-3-Dependent NFS-60 cell line and Adaptation to Growth in Macrophage Colony Stimulating Factor. *The Journal of Immunology* **145**(3), 860-864.

Pandit, J. B., A.; Jancarik, J.; Halenbeck, R.; Koths, K.; Kim, S.H. (1992). Three-dimensional structure of dimeric human recombinant macrophage colony stimulating factor. *Science* **258**, 1358-1362.

Resing, K. A. A., N.G. (1998). Deuterium Exchange Mass Spectrometry as a Probe of Protein Kinase Activation. Analysis of Wild Type and Constitutively Active Mutants of MAP Kinase Kinase-1. *Biochemistry* **37**, 463-475.

Stanley, E. R. B., K.L.; Einsteine, D.B.; Lee, P.S.; Pixley, F.J.; Wang, Y.; Yeung, Y.G. (1997). Biology and Action of Colony-Stimulating Factor-1. *Mol. Reprod. Dev.* **46**, 4-10.

Wang, F. B., J. S.; Tang, X-J. (1997). Hydrogen exchange / electrospray ionization mass spectrometry studies of substrate and inhibitor binding and conformational changes of Ecoli dihydrogenase reductase. *Biochemistry* **36**(13), 3755-3759.

Wu, J. W., J. T. (1997). A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Science*(6), 391-398.

Zhang, Z. & Smith, D. L. (1993). Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **2**(4), 522-31.

3 Consequences of Removing Disulfide Bridges on Folding of Recombinant Human Macrophage Colony Stimulating Factor Beta

3.1 Abstract

Hydrogen exchange pulsed labeling, monitored by ESI-MS, has been employed to study the *in vitro* folding of recombinant human macrophage colony stimulating factor beta (rhm-CSFβ) and its derivative in which two of the nine disulfide bridges (Cys157/Cys159-Cys'157/Cys'159) were selectively reduced and alkylated. The overall folding kinetics of rhm-CSFβ were very similar to those of $CN^{157,159}$-modified rhm-CSFβ. For both proteins, a substantial amount of three-dimensional structure was formed within 7.5 ms after folding was initiated. Little structural change was detected when the folding time was extended to 48 hr. These data indicated that the removal of Cys157/159-Cys'157/159 did not affect the overall folding of rhm-CSFβ in the presence of seven native disulfide bonds.

KEY WORDS: hydrogen exchange pulsed labeling, quenched flow, rhm-CSFβ, protein folding

3.2 Introduction

The processes by which proteins fold from unstructured polypeptides into their unique native conformations have been an important topic for structural biology research (Wedemeyer 2000). In proteins that contain disulfide bonds, the proper formation of the disulfide bonds contributes significantly to the stability of the native conformation and help direct protein folding (Creighton 1993). The effects of disulfide bond formation on protein folding have been investigated for several protein systems. For example, Eyles et al observed that the disulfide bond Cys6-Cys127 was crucial for stabilizing the partially folded intermediates in lysozyme and that the removal of this disulfide bond led to the disappearance of an important intermediate during folding (Eyles, Radford et al. 1994). Ruoppolo et al observed that the removal of Cys58-Cys110 and Cys26-Cys84 had a dramatic effect on the folding kinetics of RNase A (Ruoppolo 2000). These authors suggested that the cysteine residues might work as internal catalysts and promote shuffling of disulfides to accelerate isomerization reactions in the establishment of the native conformation of RNase A.

Recombinant human macrophage colony stimulating factor beta is a glycoprotein that stimulates the proliferation, differentiation and survival of cells belonging to the monocyte-macrophage lineage both *in vivo* and *in*

vitro (Stanley 1994). rhm-CSFβ is a homodimer maintained by six intramolecular and three intermolecular disulfide bridges (Figure 4). Two intermolecular disulfide bonds (Cys157/159-Cys'157/159) are located in the tail regions near the C-termini. A previous study showed that the removal of these two disulfide bridges did not affect the biological activity of rhm-CSFβ nor did it strongly alter the overall three-dimensional structure of the protein (Zhang 2001). However, the regions near the disulfide removal sites became more accessible to hydrogen deuterium exchange, suggesting a more flexible conformation. Therefore, these disulfide bonds served to maintain the proper three-dimensional fold of the protein in the C-terminal regions. An interesting question arises: do these disulfide bonds play any role in the formation of the three-dimensional structure during rhm-CSFβ folding? The present study addresses this question by investigating the folding kinetics of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ monitored by deuterium/hydrogen (D/H) exchange pulsed labeling.

3.3 Results and discussion

3.3.1 Equilibrium unfolding by far-UV circular dichroism and fluorescence spectroscopy.

The thermodynamic stabilities of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ were studied by equilibrium unfolding in urea, monitored by far-UV CD spectroscopy. The degrees of unfolding for both proteins were monitored by their ellipticity at 222 nm at 25°C (Figure 21).

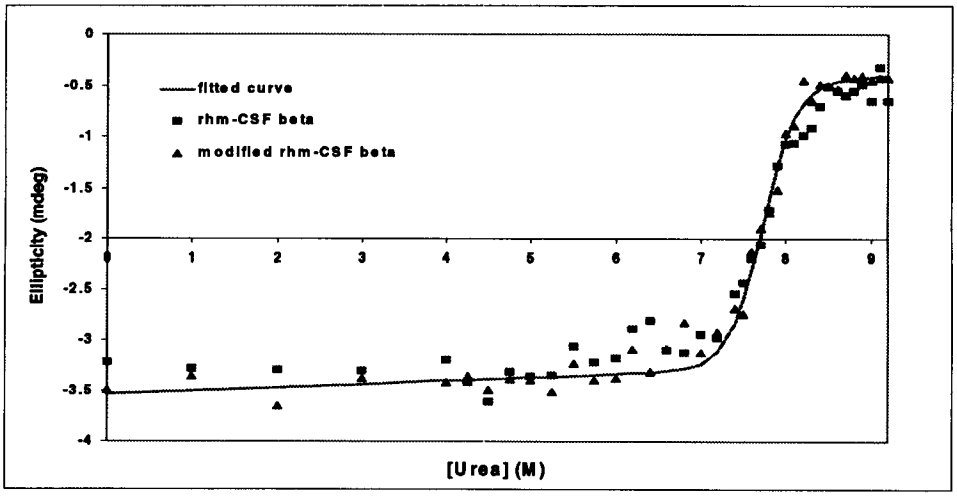


Figure 21. Equilibrium unfolding of $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ studied by circular dichroism spectroscopy at 222 nm at 25°C.

The denaturation curves for rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ were essentially superimposable, suggesting that these proteins had similar stabilities. Both proteins exhibited highly cooperative unfolding transitions with an apparent lack of equilibrium unfolding intermediates. The dependence of ΔG° on the denaturant concentration for a system in which only the folded and unfolded states were present at equilibrium was analyzed by the following equation:

$$\Delta G^{\circ}[\text{urea}] = \Delta G^{\circ}[0] + m^{*}c \qquad \text{Equation 8}$$

where ΔG°[urea] represents the free energy of unfolding at various urea concentration, ΔG°[0] represents the free energy at zero concentration urea, m measures the sensitivity of ΔG° on denaturant concentration, and c represents urea concentration. This equation was used to fit the data from the CD experiments, yielding $\Delta G^{\circ}_{rhmCSF\beta}$= 23.01±2.75 kcal/mol, $m_{rhmCSF\beta}$ = -2.97±0.37 kcal/mol/M, $\Delta G^{\circ}_{\text{modified rhmCSF}\beta}$= 23.15±1.93 kcal/mol, and $m_{\text{modified rhmCSF}\beta}$ = -3.00±0.25 kcal/mol/M. The equilibrium *m*-values for rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ were nearly equivalent, suggesting that both proteins exhibited similar sensitivities to urea denaturation.

The equilibrium unfolding processes of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ were also monitored by fluorescence spectroscopy. As

demonstrated in the CD experiments, the equilibrium unfolding curves for these proteins were essentially identical, also suggesting that rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ had similar stabilities.

3.3.2 Folding of rhm-CSFβ studied by deuterium/hydrogen (D/H) exchange pulsed labeling.

Deuterium/hydrogen exchange pulsed labeling required the use of an automated instrument for mixing to study the folding events on the msec time scale. A BioLogic quenched flow instrument was used for mechanically performing protein folding, hydrogen labeling, and quenching reactions. Oxidized rhm-CSFβ was initially unfolded in a D_2O-denaturant solution in which all exchangeable amide sites were deuterated. Folding, initiated by a rapid 11-fold dilution of the denaturant, was allowed to proceed over variable time periods between 7.5 ms and 10 s before the partially folded protein was exposed to a hydrogen pulse. Under this pulsed labeling condition (pH 10, 10 ms, 8°C), the half life for D/H exchange of a free ND is about 1 ms (Molday 1972). Therefore, the amide sites in still unstructured parts of the protein became fully protonated while those protected by prior formed structures remained deuterated. The labeling pulse was terminated by adjusting the solution pH to 2.5 and the

temperature to 0°C, conditions under which D/H exchange was slow. The protein samples were analyzed using LC-ESI-MS. To minimize further deuterium exchange loss from solvent hydrogen through back exchange during HPLC analysis, the chromatography step was optimized for speed. The frozen samples were thawed and analyzed by mass spectrometry within 5 min. ESI-MS data showed that 227 deuteriums were maintained in rhm-CSFβ after 7.5 ms of folding (Figure 22).

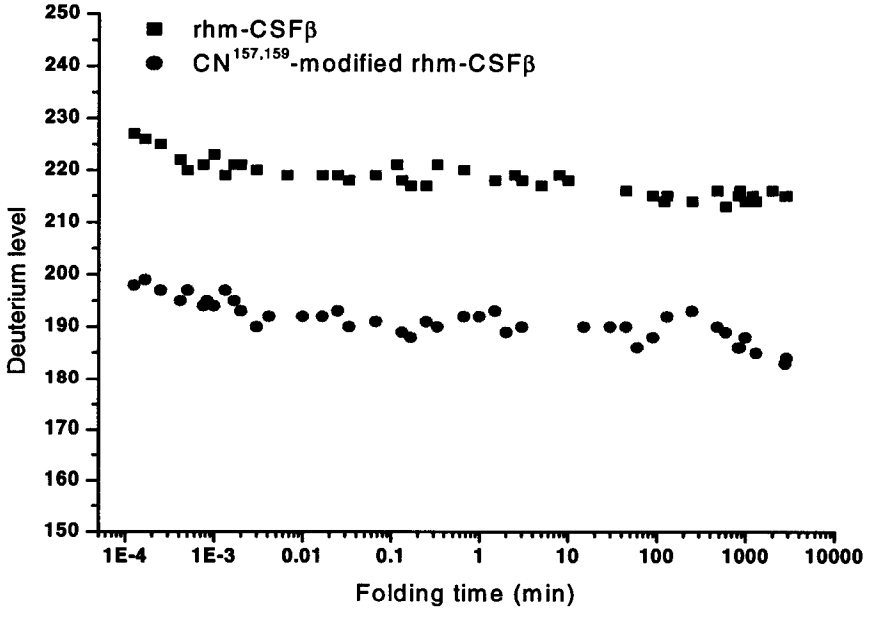


Figure 22. Deuterium levels maintained in $CN^{157,159}$-modified rhm-CSFβ and rhm-CSFβ during protein folding.

Interestingly, a loss of only 13 additional deuteriums was detected when the folding time was extended to 10 s. This suggested that little structural change took place during folding in the range of 7.5 ms to 10 s.

In light of these results, the folding reaction was allowed to proceed beyond 10 s. However, because the quenched flow instrument was incapable of mixing on the time scale of minutes or hours, a manual mixing procedure was adapted in place of the automated mixing provided with the quenched flow instrument. In this procedure, the pulsed labeling condition (pH 7, 10 s, 8°C) was carefully selected so that the strength of the labeling pulse was kept the same as in the previous experiments performed with the quenched flow instrument. Pulsed labeling experiments with rhm-CSFβ were performed at 30 different folding times between 10 s and 48 hr (Figure 22). The samples were analyzed by LC-ESI-MS. The deuterium levels in these samples were found to be similar as those found in samples obtained from much shorter folding times. In the range of 10 s to 48 hr, the deuterium levels fluctuated in the range of 215 and 225, reflecting little if any structural change. Most importantly, two samples obtained for the 10 s folding reaction, one by automated mixing and another by manual mixing, showed the same level of deuterium incorporation. Therefore, the pulsed labeling conditions for both methods indeed gave the same results.

These data indicated that rhm-CSFβ folded rapidly (within 7.5 ms) under the present folding condition (0.8 M urea at pD 8.5 at 8°C) and no significant structural change could be detected by extending the folding time. This finding was consistent with results obtained from previous stopped-flow CD experiments with rhm-CSFβ, which showed that secondary structure formation occurred rapidly (M. Schimerlik, private communication). However, other studies using fluorescence spectroscopy and size exclusion chromatography (C. Maier, private communication) indicated formation of tertiary structure on the time scale of min to hr, which was not detected in the present study.

3.3.3 Folding of $CN^{157,159}$-modified rhm-CSFβ investigated by D/H exchange pulsed labeling.

Folding studies of $CN^{157,159}$-modified rhm-CSFβ were carried out as described for rhm-CSFβ. Both automated and manual mixing methods were employed to investigate the folding kinetics of the oxidized protein in the time range of 7.5 ms to 48 hr. Mass spectrometric analyses of these samples showed that nearly 198 deuteriums were retained in the protein within 7.5 ms after the folding reaction was initiated. The loss of 10 additional deuteriums was observed as the folding time was increased to

10s. As the folding reaction was prolonged to 48 hr, the measured deuterium levels fluctuated in the range of 183 to 190 (Figure 22). These data suggested that shortly after the folding reaction was initiated $CN^{157,159}$-modified rhm-CSFβ attained a significant amount of folded structure. As the folding time was increased, the loss of a small amount of deuterium was detected, indicating little structural change had taken place. Interestingly, the deuterium levels in $CN^{157,159}$-modified rhm-CSFβ on average were lower than the deuterium levels found in rhm-CSFβ by about 30 deuteriums for the same folding time.

3.3.4 Localizing deuteriums in folded proteins by proteolysis and CID MS-MS.

In order to locate the amide sites corresponding to the deuterium differences in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ, pepsin digestion was applied to D/H exchanged proteins after various folding times. Thirteen peptides covering more than 90% of the protein backbone were identified in the pepsin digests of rhm-CSFβ as well as in $CN^{157,159}$-modified rhm-CSFβ.

The deuterium content in all peptides from rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ after 7.5 ms of folding were determined by collision induced dissociation (CID) experiments (Table 5).

Table 5. Number of amide deuteriums retained in rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ after 7.5 ms of folding.

Segment	# of NHs	rhm-CSFβ	$CN^{157,159}$-rhm-CSFβ
Intact proteins	404	227	198
4-19	16	5	4
20-27	7	7	6
37-55	17	11	11
56-62	6	5	5
63-76	12	7	7
77-82	5	4	4
83-105	11	10	11
106-113	6	4	4
114-128	14	14	13
129-135	7	5	2
136-143	8	2	1
151-189	29	9	1
190-221	27	3	1

The peptides derived from both proteins exhibited an important similarity: the peptides of the interior regions of the proteins retained far more deuteriums than those in the N-termini and the C-termini. For instance, the peptides covering aa 4-19 showed deuterium levels of 5 (31%) and 4 (25%)

for rhm-CSFβ and its derivative while the peptides covering 83-105 showed deuterium levels of 10 (91%) and 11 (100%), respectively. However, the deuterium levels in the C-terminal regions were different between the two proteins. This was especially pronounced in the peptides covering aa151-189: 9 deuteriums (31%) were retained in the peptide from rhm-CSFβ while only 1 deuterium (3.4%) was retained in the modified protein. These data indicated that both proteins achieved similar overall structures but exhibited different structural features in the C-terminal regions after 7.5 ms of folding.

The numbers of amide deuteriums from the peptides did not equal to those from the intact proteins because back exchange with the protonated solvent in peptides was more complex than that for the intact proteins. For instance, the deuteriums at the N-termini from every peptide were lost during HPLC analysis. In some cases, depending on the sequence of the peptide, the deuteriums from the amino acid second from the N-termini might be lost as well (Engen, Smithgall et al. 1997). In contrast, only two amide deuteriums at the two N-termini were lost with the intact rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ.

In order to determine the folding kinetics, the deuterium content in peptides derived from both proteins were analyzed over folding times ranging from 7.5 ms to 48 hr. However, none of the peptides analyzed

exhibited any change in deuterium levels as the folding time increased. For instance, peptides covering aa 83-105 from both proteins showed unchanged deuterium levels of 10 and 11 (Figure 23). In peptides covering aa 151-189, the deuterium levels remained the same at 9 and 1 (Figure 23) for rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ, respectively. Thus, rate constants for specific kinetic processes could not be determined. It is intriguing that the intact proteins exhibited slight deuterium loss while none of the peptides showed any decrease in deuterium levels. It is possible that most of the amide deuteriums loss occurred in sites that were closer to the N- and C-termini of the peptides.

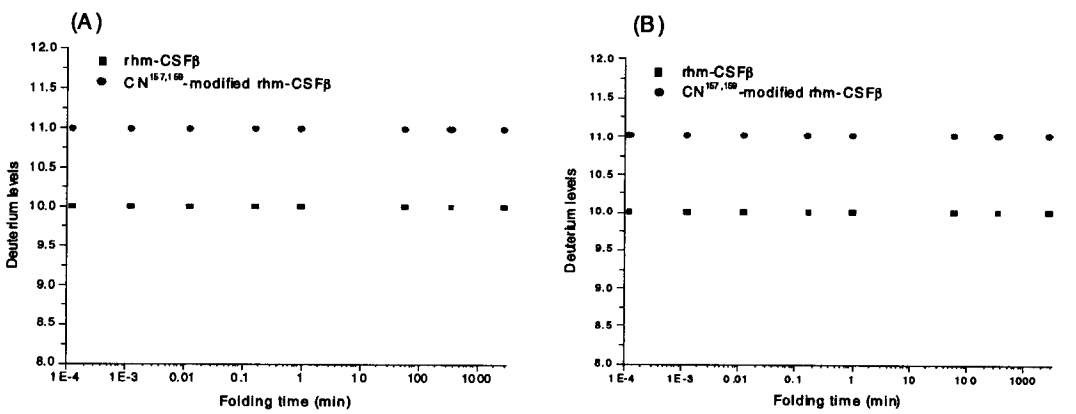


Figure 23. Deuterium levels in peptides covering aa 83-105 (A) and peptides covering aa 151-189 (B) remained unchanged during the folding of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ.

3.4 Conclusion

Deuterium/hydrogen exchange pulsed labeling was used to study the folding kinetics of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ. Both proteins exhibited fairly high levels of deuterium, which were sustained throughout folding. The fact that little additional deuterium loss was detected in the two proteins suggested that the proteins attained three-dimensional structural features characteristic of the folded proteins. In the presence of all nine disulfide bonds, unfolded rhm-CSFβ quickly folded into the native conformation. In the case of $CN^{157,159}$-modified rhm-CSFβ, the data suggest that the disulfide bonds linking Cys157/159-Cys'157/159 did not greatly affect protein folding kinetics. It may be that D/H exchange pulsed labeling was not able to detect the change in folding kinetics resulting from the absence of Cys157/159-Cys'157/159.

A comparison of the data from the native state deuterium exchange experiment (Zhang, unpublished result) and the deuterium exchange pulsed labeling experiment provided interesting results. In the native state deuterium exchange experiment, fully deuterated proteins were subjected to exchange with protic solvent for 10 s. The deuterium levels determined in these samples were nearly identical as those from the pulsed labeling experiment in which the proteins were subjected to a 10 s hydrogen pulse.

These results indicated that the folded proteins prior to the proton pulse in the pulsed labeling experiment shared similar structural features with the folded proteins prior to the 10 s D/H exchange. It has been suggested that residues with slowly exchanging amide deuteriums fold early on the kinetic folding pathways (Woodward 1993). In the case of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ folding, there were many slow exchanging amide deuteriums which must have been shielded from exchange during early folding.

3.5 Material and methods

3.5.1 Materials.

Purified rhm-CSFβ was provided by Dr. C. Cogwill (Chiron Corp. Emeryville, CA). Tris(2-carboxyethyl)phosphine (TCEP) and immobilized pepsin were obtained from Pierce (Rockford, IL). 1-cyano-4-dimethylaminopyridinium fluoroborate (CDAP) was obtained from Sigma Chemicals. Modified trypsin was purchased from Promega (Madison, WI). Deuterium oxide and urea-d_4 were obtained from Aldrich Chemical Co.

(Milwaukee, WI). All other chemicals and reagents were of the highest grade commercially available.

3.5.2 Preparation of CN $^{157, 159}$-modified rhm-CSFβ.

The freshly dissolved rhm-CSFβ (12nmol) was solubilized in 300 μl of 20 mM ammonium acetate buffer (pH 3.5) containing 1 mM EDTA. Partial reduction of rhm-CSFβ was initiated by adding 0.05 equivalent of TCEP of total cysteine content and followed by incubation at room temperature for 5 hrs. After TCEP was removed by centrifugal filtration using ultrafree MC filters with a 5,000 Da molecular mass cutoff (Millipore), a 25-fold molar excess of CDAP solution relative to the cysteine content was added to the partially reduced protein mixture. Cyanylation of the sulfhydryl groups was carried out at room temperature for 10 min. CDAP was then removed by centrifugal filtration. The reduced, cyanylated protein was purified by reverse phase high performance liquid chromatography (HPLC) on a Vydac C4 TP214 column (300Å, 5 μm, 4.6X250 mm, 1 ml/min flow rate) employing a binary linear gradient elution with 0.1% trifluroacetic acid (TFA) as solvent A and acetonitrile (ACN) containing 0.09%

trifluroacetic acid as solvent B. The proteins were eluted using a 60 min 44-54% B gradient. Protein fractions were collected manually and lyophilized.

3.5.3 Equilibrium unfolding studied by circular dichroism and fluorescence.

rhm-CSFβ and its derivative were dissolved to a protein concentration of 0.3ug/ul in 50 mM Tris buffer solution (pH 8.5) containing concentrations of urea ranging from 1 M to 9 M. The solutions were equilibrated overnight at room temperature. All measurements were carried out at 25°C. CD spectra were obtained using a Jasco J720 spectropolarimeter (Jasco Inc., Easton, MD) over the range of 185-310 nm in a cylindrical 1-cm path length quartz cuvette with a bandwidth of 2nm and a scan speed of 10nm/min. Fluorescence data were acquired on an SLM 8100C fluorescence spectrometer (SLM Instruments Inc., Urbana-Champaign, IL) with 8 nm emission and excitation slit widths using a square 0.5 cm path length quartz cuvette. The intrinsic tryptophan residues were excited at 280 nm, and emission spectra were recorded at a wavelength range of 300 to 420 nm and expressed as a ratio of emission intensity to reference emission intensity.

3.5.4 Folding studied by quenched-flow deuterium hydrogen exchange pulsed-labeling.

Deuterium hydrogen exchange pulsed labeling experiments were performed on a QFM-4 quenched flow module (BioLogic, Echirolles, France) at 8°C. rhm-CSFβ and its derivative (80 uM) were unfolded in a deuterated buffer containing 50 mM Tris, 8.5 M urea-d_4 at pD 8.5. Under these conditions, all backbone amide protons exchange rapidly with deuterons from the solvent. Folding was initiated by an 11X dilution of the denatured proteins with the folding buffer containing 50 mM Tris at pD 8.5. The resulting denaturant concentration (0.77 M urea-d_4, pD 8.5, 10°C) was well below the equilibrium unfolding transition for rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ. The folding times ranged from 7.5 ms to 10 s. The labeling reaction was initiated by rapidly mixing the partially folded protein solution with an equal volume of 25 mM Tris buffer in H_2O at pH 11, resulting in a jump to pH 10. The 10 ms pulse was terminated by lowering the pH to 2.5 in a final mixing step with the quenching buffer (0.5 M HCl). The samples were concentrated using a Millipore microconcentrator and stored at -80°C prior to mass spectrometric analysis.

3.5.5 Folding studied by manual mixing deuterium hydrogen exchange pulsed-labeling.

rhm-CSFβ (80 uM) and its derivative were unfolded in a deuterated buffer containing 50 mM Tris, 8.5 M urea-d_4 at pD 8.5. Folding was initiated by an 11X dilution of the denatured proteins with the folding buffer containing 50 mM Tris at pD 8.5 and 8°C. The folding times ranged from 10 s to 48 hr. The labeling reaction was initiated by rapidly mixing the partially folded protein solution with an equal volume of 25 mM Tris buffer in H_2O at pH 6, resulting in a jump to pH 7. The 10 s pulse was terminated by lowering the pH to 2.5 in a final mixing step with the quenching buffer (0.5 M HCl). The samples were concentrated using a Millipore microconcentrator and stored at -80°C prior to mass spectrometric analysis. The deuterated refolded protein samples were analyzed by LC-ESI-triple quadrupole MS (PE-Sciex API III$^+$) using a self-packed C4 column (5 μm particle size, 300 Å pore size, 20 μL/min flow rate) and a binary gradient elution involving 0.05% TFA as solvent A and ACN containing 0.05%TFA as solvent B. After desalting at 10%B for 2min, the protein was eluted using a 3 min 10-80% B program. The HPLC injector and the column were submerged in ice/water slurry, and the transfer syringe pre-cooled on ice before use. MS spectra were analyzed by BioSpec software.

3.5.6 Determination of the localized deuterium content in folded proteins by pepsin digestion and CID-MS-MS.

The samples obtained from various folding times were digested with pepsin (protein/enzyme = ½) in 0.1 M ammonium phosphate buffer (pH 2.5, 0°C, 1:1 D_2O/H_2O) containing 8 M urea-d_4 and 1 M TCEP at 0°C for 5min. This procedure yielded 90% coverage of the entire backbone. The digest mixture was analyzed using the LCQ-ESI-ion trap mass spectrometer. A C18 Luna HPLC column (5 μm particle size, 100 Å pore size, 15 μL/min flow rate) and 0.05%TFA as solvent A and ACN containing 0.05% TFA as solvent B was used for separation. After desalting for 2min, peptides were eluted with the following program: 10-20%B for 1min, 20-30%B for 1min, 30-45%B for 2min, 45-60%B for 1min, 60-90%B for 3min. To minimize deuterium back exchange, the injector and column were immersed in an ice/water bath, and the transfer syringe was rinsed with D_2O and cooled on ice before use. MS spectra were analyzed using Sequest software.

3.6 References

Creighton, T. E. (1993). "Proteins: structures and molecular properties." W.H. Freeman and Company, New York. Second Edition.: 316-323.

Engen, J. R., T. E. Smithgall, et al. (1997). "Identification and localization of slow, natural, cooperative unfolding in the hematopoietic cell kinase SH3 domain by amide hydrogen exchange and mass spectrometry." Biochemistry **36**(47): 14384-91.

Eyles, S. J., S. E. Radford, et al. (1994). "Kinetic consequences of the removal of a disulfide bridge on the folding of hen lysozyme." Biochemistry **33**(44): 13038-48.

Molday, R. S. E., S.W.; Kallen, R.G. (1972). "Primary structure effects on peptide group hydrogen exchange." Biochemistry **11**(2): 150-158.

Ruoppolo, M. V., F.; Klink, T.A.; Raines, R.T.; Marino, G. (2000). "Contribution of individual disulfide bonds to the oxidative folding of Ribonuclease A." Biochemistry **32**(12033-12042).

Stanley, E. R. B., K.L.; Einstein, D.B.; Lee, P.S.; Yeung, Y.G. (1994). "The biology and action of colony stimulating factor-1." Stem Cells **12**: 15-24.

Wedemeyer, W. J. W., E.; Narayan, M.; Scheraga, H.A. (2000). "Disulfide bonds and protein folding." Biochemistry **39**(15): 4207.

Woodward, C. (1993). Trends Biochem Sci **18**: 359-361.

Zhang, Y. H. Y., X-G.; Maier, C.L.; Schimerlik, M.I.; Dinezer, M.L. (2001). Prot. Sci. **10**: 2336-2345.

4 Structural Elucidation of Intermediates in rhm-CSFβ Folding by Disulfide Bond Quenching and H/D pulsed labeling

4.1 Abstract

In vitro folding of recombinant human macrophage colony stimulating factor beta (rhm-CSFβ) involves the systematic formation of disulfide bonds *via* multiple pathways. Evidence was obtained by mass spectrometry for the formation of native as well as non-native intramolecular disulfide bonds in the monomeric folding intermediates. Initial dimerization involved the formation of disulfide bonds Cys157/159-Cys'157/159. Among three intra-subunit disulfide bonds Cys48-Cys139 and Cys'48-Cys'139 were the last to form, while Cys31-Cys'31 formed the last inter-subunit disulfide bond. An hydrogen/deuterium (H/D) exchange pulsed labeling experiment showed the coexistence of dimeric and monomeric species of non-compact conformation at early stages of folding. As folding proceeded, the deuterium level dropped rapidly in the dimer, losing 20% deuterium label within 50 min, suggesting that the non-compact dimer underwent significant structural rearrangement. The deuterium level remained high in the monomer throughout extended folding, indicating that the monomer maintained a non-compact conformation.

Key Words: oxidative folding, disulfide bond quenching, H/D pulsed labeling, rhm-CSFβ,

4.2 Introduction

Protein folding is an important process for understanding structure-function relationship. A thorough understanding of the fundamental principles of protein folding may lead to the high level production of biologically active proteins of therapeutic and diagnostic importance (Dobson 1998; Liu 1998). A general solution to the protein folding problem requires the delineation of folding pathways, which can be constructed by identifying the structural and energetic characteristics of the folding intermediates and placing them in some kind of logical order. Extensive research efforts have been devoted to elucidating the thermodynamics and kinetics of the intermediates along the folding pathways (Lu 1992; Onda 1997; Chang 2001; Zhou 2001).

The protein folding process is highly complex; it involves various chemical interactions of individual amino acids, such as disulfide bonding and hydrogen bonding. By experimentally manipulating these chemical interactions, it is possible to trap intermediate conformations that occur

during protein folding. In proteins containing disulfide bonds, for example, the folding process can be studied by examining the extent of disulfide bond formation at various stages of folding by disulfide bond quenching, i.e. trapping of the intermediates with nascent sulfhydryl groups (Creighton 1978). Detailed folding pathways have been defined for several proteins by analyzing the pattern of modified and non-modified cysteine residues in the folding intermediates (Weissman 1991; Lu 1992; Glocker, Arbogast et al. 1994).

Folding pathways can also be elucidated by monitoring the formation of hydrogen bonding that occurs in protein folding. It is difficult, if not impossible, to manipulate and detect the formation and breakage of all hydrogen bonds throughout protein folding. However, it is possible to trap the intermediates by deuterium labeling using amide hydrogens (NHs) as reporter groups. Amide H/D exchange pulsed labeling utilizes deuterium exchange to label NHs on the polypeptide backbone that are unprotected at defined folding times (Clarke 1997). The levels of deuterium trapped at different folding times can be determined by nuclear magnetic resonance (NMR) and/or electrospray ionization mass spectrometry (ESI-MS). Transient folding intermediates that are undetected by other conventional spectroscopic methods have been identified by H/D pulsed labeling (Roder 1988; Bai, Sosnick et al. 1995; Tsui, Garcia et al. 1999).

rhm-CSFβ was expressed from *Escherichia coli* as a truncated form (aa 4-218, 49KDa) of one of the three human cDNA clones (Pandit 1992). It is a glycoprotein that stimulates the proliferation, differentiation and survival of cells belonging to the monocyte-macrophage lineage both *in vivo* and *in vitro* (Stanley 1994). The effects of rhm-CSFβ are mediated through binding to its receptor, M-CSFR, which is an integral transmembrane glycoprotein that functions as a ligand-activated protein tyrosine kinase (Bourette 2000). Mature rhm-CSFβ is a homodimer with three inter-subunit disulfide bonds (Cys31-Cys'31, Cys157/159-Cys'157/159) and three intramolecular disulfide bonds (Cys48-139, 7-90, 102-146) (Figure 4). The assembly of the monomeric subunits and the maturation of the biological activity were shown to depend on the progressive formation of the correct disulfide bonds during *in vitro* folding (Glocker, Arbogast et al. 1994). Consequently, it is important to study the structures of disulfide intermediates and their order of formation in folding. However, the number of possible disulfide bonded intermediates increases rapidly with increasing numbers of cysteine residues available for disulfide bond formation. For a protein with nine disulfide bonds and eighteen cysteine residues, such as rhm-CSFβ, the number of possible disulfide bonded species is enormous, exceeding 30,000,000 (Jaenicke 1989). Therefore, it is a daunting task to attempt isolation and analysis of all

disulfide scrambled isomers occurring in the folding pathways of rhm-CSFβ. However, isolating and characterizing the few major disulfide-bonded intermediates and placing them in a time-dependent order may help to elucidate some folding events for rhm-CSFβ. In the present study, the folding of rhm-CSFβ was investigated by disulfide bond quenching and H/D pulsed labeling.

4.3 Results

4.3.1 In vitro folding monitored by disulfide bond quenching.

In order to study the intermediates involved in rhm-CSFβ folding, a purified sample of native rhm-CSFβ was denatured and reduced. The complete conversion from dimer to monomer was verified by size exclusion chromatography and mass spectrometry (data not shown). Folding was initiated by rapid buffer exchange of the denatured, reduced protein solution into the folding buffer containing GSSG/GSH. Carboxyamidomethylation was used to trap the intermediates in the time course folding experiment.

Mixtures of trapped intermediates were analyzed by mass spectrometry without HPLC fractionation.

The mass spectra recorded of the intermediates for folding times ranging from 1 min to 7,200 min showed that the charge state distributions shifted from one centered on peaks with low m/z values to one centered on peaks with high m/z values (Figure 24). A charge state distribution centered on peaks with low m/z values represented monomeric proteins whose molecular weights (MWs) were half that for dimeric proteins (Figure 24. 1min). A distribution centered on high m/z values was indicative of compact dimeric proteins (Figure 24. 7,200 min). During this transition, there were charge state distributions centered on medium m/z values (Figure 24. 120 min). These distributions corresponded to dimeric species that had non-compact structures in which many basic sites were available for protonation during electrospray ionization. The shift in charge state distributions represented a conversion from non-compact monomers to non-compact dimers and eventually to compact dimers. The overlapping charge state distributions indicated monomeric and dimeric species coexisted at different stages of folding.

Each mass spectrum represented a time-dependent mixture of folding intermediates. For instance, the mass spectrum after 1 min folding indicated the presence of two protein forms (Figure 25. A). Deconvolution

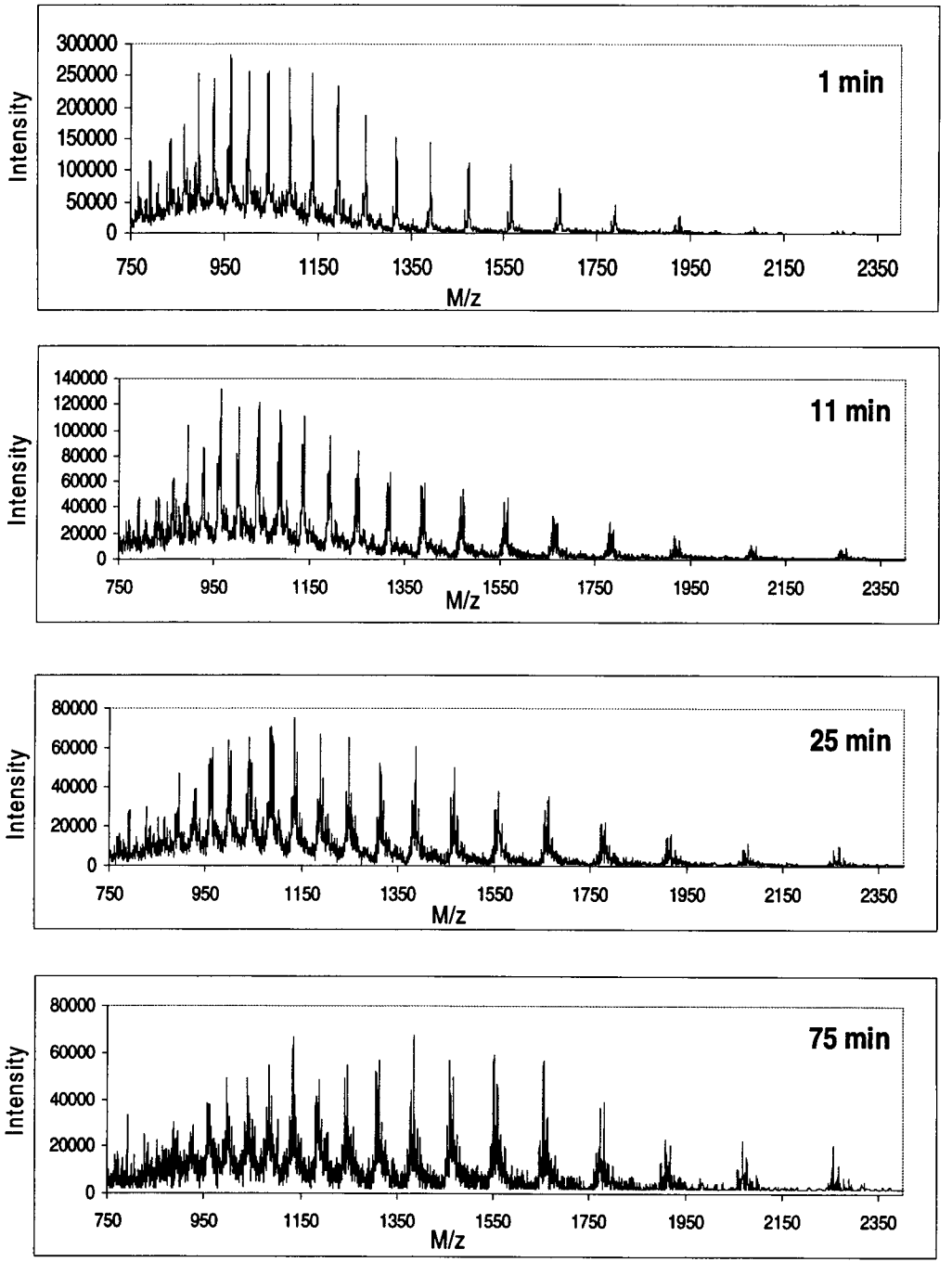


Figure 24. Carboxyamidomethylated intermediates of rhm-CSFβ after folding for different incubation periods in GSH/GSSG (7.8/1).

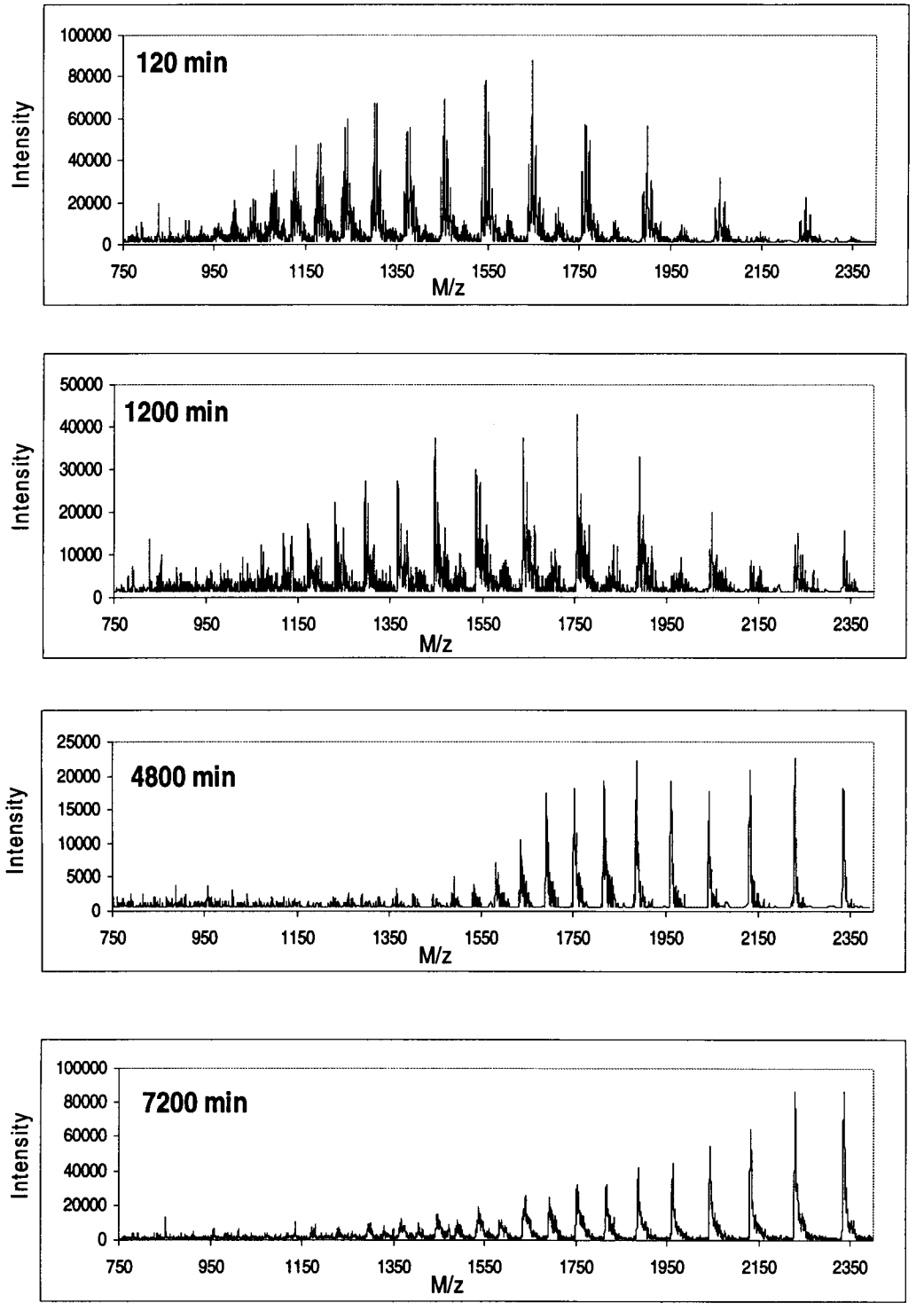
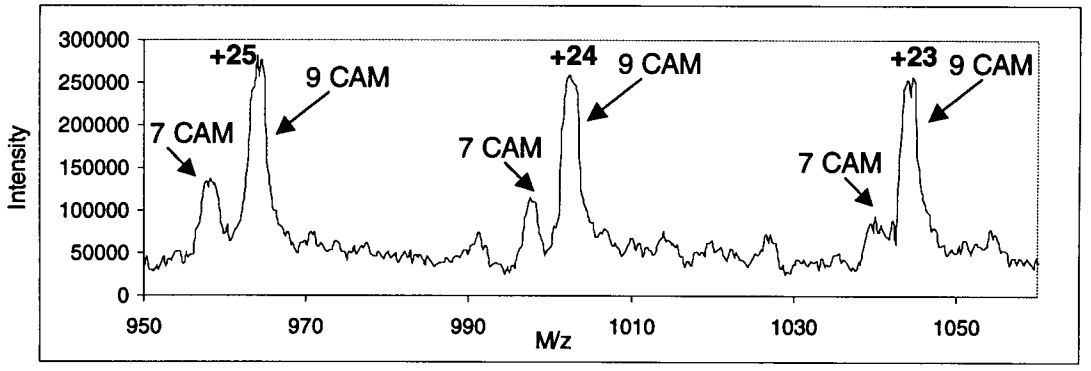


Figure 24. (continued)

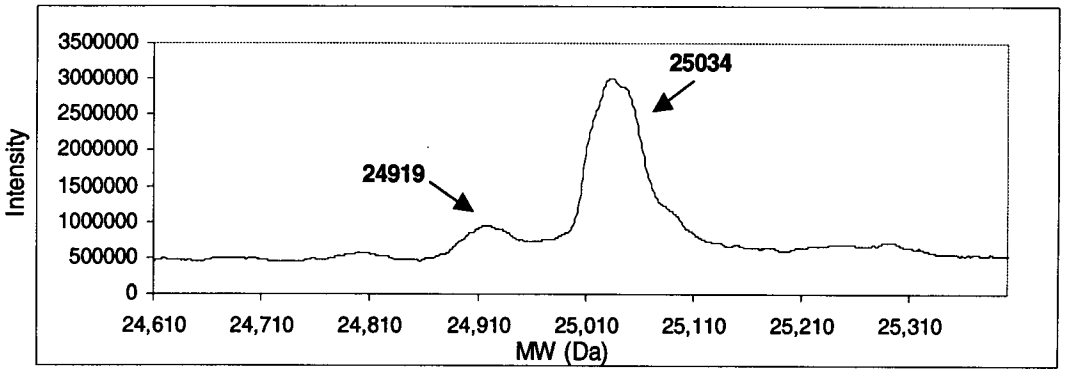
of the mass spectrum provided the MWs for these two proteins: 25,034 Da and 24,929 Da, corresponding to monomeric proteins with 9 and 7 alkylated cysteine residues, respectively (Figure 25. B). After 300 min folding, the folding mixture became more complex, exhibiting multiple overlapping charge state distributions (Figure 25. C). Deconvolution identified these species to be monomeric and dimeric proteins of various degrees of carboxyamidomethylation. The oxidative folding of rhm-CSFβ approached completion after 7,200 min. The mass spectrum showed a distribution of charge states centered on 21+, the same as in the mass spectrum of the native protein, indicating that nearly all species were converted to the native protein by this time. When the charge states were deconvoluted, the presence of a small population of a monomer containing 4 disulfide bonds and 1 alkylated cysteine residue became apparent even though the charge state distribution characteristic of a monomer was not detected.

The samples collected after various folding times consisted of complex mixtures of monomeric and dimeric species that were differentiated by their MWs. The identities of each species in the folding mixture were assigned based on their MWs obtained by deconvoluting the mass spectra and comparing them with calculated values (Table 6). It is important to note that quantitative or qualitative comparisons between the

(A) 1 min



(B) 1 min



(C) 300 min

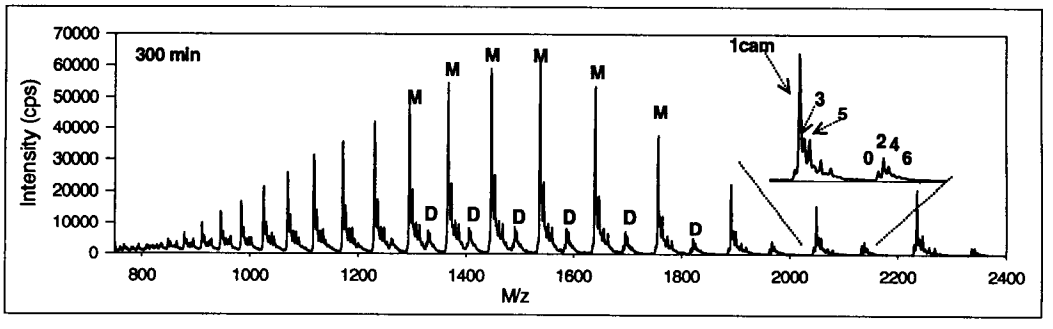


Figure 25. (A) Charge state distributions corresponding to carboxyamidomethylated monomers. (B) Molecular weights obtained by deconvolution. (C) Heterogeneous monomeric and dimeric proteins of various degrees of carboxyamidomethylation.

Table 6. Disulfide bonds in rhm-CSFβ folding intermediates.*

Time (min)	M-SS0-cam9	M-SS1-cam7	M-SS2-cam5	M-SS3-cam3	M-SS4-cam1	D-SS6-cam6	D-SS7-cam4	D-SS8-cam2	D-SS9-cam0
1	1	2							
11	1	2	3						
15	1	2	3	4					
20	2	1	3	4					
25	2	1	3	4					
45	3	2	1	4		1			
75			1	2	3	2	1		
103			2	1	3	3	1	2	
120			3	1	2	3	1	2	4
180			3	1	2	4	1	2	3
305			3	2	1	4	2	1	3
480				2	1	4	2	1	3
970				2	1		2	1	3
1,200				2	1		3	1	2
1,279				2	1		2	1	3
1,513				2	1		3	2	1
2,565					1		3	2	1
4,800					1			2	1
7,200					1				1
MW^1 (Da)	25,045	24,929	24,810	24,697	24,578	49,268	49,377	49,144	49,036
MW^2 (Da)	25,037	24,921	24,805	24,689	24,573	49,262	49,378	49,146	49,030

* cam refers to carboxyamidomethylation, SS refers to disulfide bond. MW^1 is the measured molecular weight and MW^2 is the calculated molecular weight. Numbers were assigned based on the abundance of the protein form in the mass spectra: 1 being the most populous and 4 being the least abundant.

monomeric and dimeric proteins cannot be made simply based on their relative peak intensities in the mass spectra. This is because monomeric and dimeric proteins exhibit different molar sensitivities in the mass spectrometer, perhaps because of different electrospray ionization. Monomeric proteins are less structured and are ionized more efficiently than the compact dimeric proteins. However, amongst monomeric or dimeric proteins, it is possible to make qualitative comparisons of the abundance of these species based on their relative peak intensities in the mass spectra. Numbers were assigned to each folding intermediate based on their relative intensities in the mass spectra, 1 being the most populous and 4 being the least abundant (Table 6).

Overall monomeric proteins were converted to dimeric species and eventually to the native rhm-CSFβ (Table 6). The folding process was characterized by an increasing number of disulfide bonds and a decreasing number of alkylated cysteine residues in both monomeric and dimeric folding intermediates. At early folding times, the monomer with 9 carboxyamidomethylated cysteinyl residues, M-SS0-cam9, was the most abundant. As folding proceeded, other monomers with fewer alkylated cysteinyl residues became more abundant, indicating that more disulfide bonds had formed. A dimer containing 4 carboxyamidomethylated cysteinyl residues, D-SS7-cam4, was detected in the 45 min folding mixture which

marked the beginning of the transition from monomeric to dimeric proteins. Subsequently, other dimeric proteins appeared exhibiting a lesser degree of carboxyamidomethylation and a higher degree of disulfide bond formation. Folding of rhm-CSFβ approached completion after 7,200 min when the folding sample contained mainly D-SS9-cam0, a dimeric protein with 9 disulfide bonds. In the same sample, a monomer containing 4 disulfide bonds and 1 carboxyamidomethylated cysteinyl residue, M-SS4-cam1, was also detected, representing species trapped in non-native disulfide bonded conformations and was, therefore, less capable of productive folding.

4.3.2 Isolation of folding intermediates.

To further characterize the folding intermediates, the folding reaction was performed on a larger amount of denatured and reduced rhm-CSFβ (0.5 mg, 10 nmol) for 1, 120, and 4,800 min. The folding intermediates were separated on an analytical Vydac C4 column by RP-HPLC. All fractions were collected, lyophilized and subjected to mass spectrometric analysis. Under the present HPLC conditions, the native dimer (D) had a retention time of 12.5 min and the fully alkylated monomeric protein (M) had

a retention time of 13.4 min. As expected, reduced and denatured rhm-CSFβ monomer migrated with the longest retention time due to its hydrophobic nature. The measured molecular weights for the native dimer and the fully alkylated monomer were 49,036 Da and 25,034 Da, respectively.

The HPLC trace for the folding mixture obtained after 1 min showed a single peak at 13.4 min (Figure 26). Mass spectrometric analysis of the isolated protein indicated that the protein had an average MW 25,038 Da and was a fully alkylated monomer M-SS0-cam9 (Table 7). The presence of a monomeric protein containing 7 alkylated cysteine residues, M-SS1-cam7, was detected in the oxidative folding sample that was not HPLC fractionated. However, this protein was not detected in any HPLC fraction, perhaps due to its low abundance in the folding mixture. HPLC analysis of the folding mixture quenched after 120 min folding provided four fractions of intermediates, identified with retention times of 12.4, 12.8, 13.1, and 13.4 min (Figure 26). All fractions were analyzed by mass spectrometry (Table 7). The 12.4-min fraction represented a single protein with an average MW 49,030 Da, the same as D-SS9-cam0. The 12.8-min fraction contained two species with MWs 49,144 Da and 49,377 Da, corresponding to dimeric proteins with 2 and 6 carboxyamidomethylated cysteinyl residues, namely D-SS8-cam2 and D-SS6-cam6. The 13.1-min fraction contained two

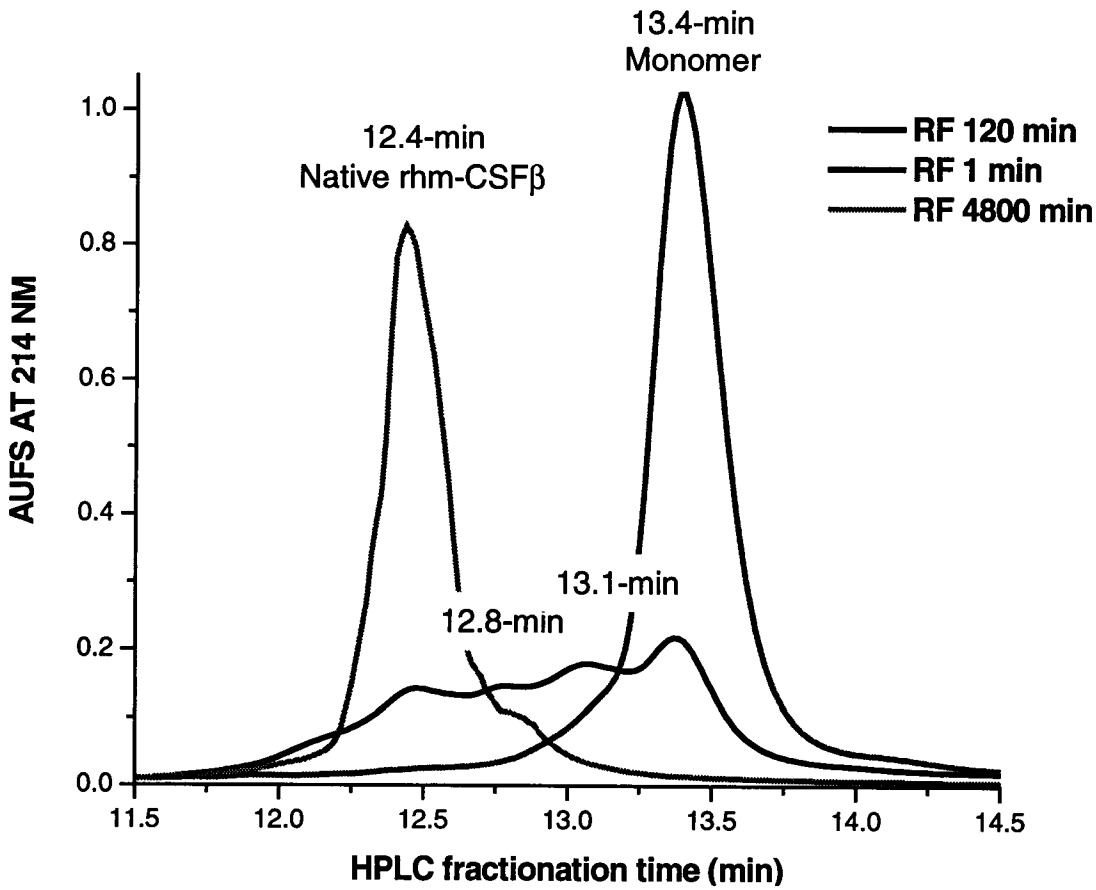


Figure 26. Disulfide bond quenched folding intermediates separated by RP-HPLC.

Table 7. Mass spectrometric analysis of HPLC fractionated folding mixtures. *

Time (min)	HPLC pools	MW (Da)	Identification	Cys-cam	Free Cys	Cys-SG
1	13.4 min	25,038	M-SS0-cam9	7, 31, 48, 90, 102, 139, 146, 157, 159	none	none
120	12.5 min	49,030	D-SS9-cam0	none	7, 31, 48, 90, 102 139, 146, 157, 159	none
	12.8 min	49,144 49,377	D-SS8-cam2 D-SS6-cam6	31, 48, 139	7, 48, 90, 102, 139, 146, 157, 159	none
	13.1 min	25,054 24,805	M-SS2-cam4/GS1 M-SS2-cam5	31, 48, 146, 157, 159	7, 90, 102, 139	31
	13.4 min	24,692	M-SS3-cam3	31, 157, 159	7, 48, 90, 102, 139, 146	none
4800	12.4 min	49,031	D-SS9-cam0	none	31, 48, 90, 139,146, 157, 159	none
	12.9min	49,146	D-SS8-cam2	31	48, 90, 139, 146, 157, 159	none

* RF refers to refolding time, cam refers to carboxyamidomethylation and MW refers to molecular weight.

monomeric species. One with MW 24,805 Da corresponding to a monomer with 5 carboxyamido-methylated cysteinyl residues M-SS2-cam5. The other had MW 25,045 Da representing a monomer with 1 glutathionylated and 4 carboxyamido-methylated cysteinyl residues M-SS2-cam4-SG1. The 13.4-min fraction contained a single protein with MW 24,692 Da that was consistent with a monomer with 3 carboxyamidomethylated cysteinyl residues M-SS3-cam3. The 12.8-min and the 13.1-min HPLC pools contained protein mixtures and were subjected to further separation by reversed phase and ion exchange chromatography. However, no improved separation was achieved, probably due to the similarities of the hydrodynamic and ionic properties of the co-eluting proteins.

The mixture collected after 4,800 min folding was separated by RP-HPLC, producing a major peak with a retention time of 12.4-min and a minor peak with a retention time of 12.9-min. The mass spectra of these two fractions provided the MWs: 49,031 Da corresponding to the native rhm-CSFβ, 49,146 Da consistent with a dimeric protein with 2 carboxyamidomethylated cysteinyl residues D-SS8-cam2. The monomeric protein with 4 disulfide bonds and 1 carboxyamidomethylated cysteinyl residue, M-SS4-cam1, was not identified in any HPLC fraction even though this protein was detected in the non-separated mixture.

4.3.3 Identification of folding intermediates by peptide mapping.

To further characterize the structures of the various folding intermediates isolated after 1, 120 and 4,800 min folding, these proteins were analyzed by peptide mapping using pepsin proteolysis and collision induced dissociation tandem mass spectrometry (CID MS-MS). Each HPLC fractionated pool was subjected to reduction to remove existing disulfide bonds in order to achieve complete enzymatic digestion. Consequently, those cysteine residues that did not participate in disulfide bridges at the time of disulfide bond quenching were carboxyamidomethylated and those residues that were members of disulfide bonds were free of alkylation.

In the HPLC fraction collected after 1 min folding, a single protein, M-SS0-cam9, was detected. A CID MS-MS experiment of the pepsin-digested sample verified that all 9 cysteine residues were carboxyamidomethylated in this monomer (Table 7), indicating that no disulfide bond had formed after 1 min folding.

Four HPLC fractions were obtained after 120 min folding. Mass spectrometric studies of these fractions indicated that two of the fractions contained mixtures of proteins that could not be further separated from each other. Therefore, the peptic digests of each HPLC pool contained

peptides derived from co-eluted proteins (Table 7). In the 12.5-min fraction, a single dimeric protein, D-SS9-cam0, was isolated and it contained no modified cysteinyl residues, indicating that all cysteinyl residues were involved in disulfide bonding. In the 12.8-min fraction that consisted of dimeric proteins D-SS8-cam2 and D-SS6-cam6, peptide mapping showed that several cysteine residues were carboxyamidomethylated, including Cys 31, 48 and 139. In addition, cysteine residues 7, 48, 90, 102, 139, 146, 157 and 159 were found to be free of alkylation. These non-modified residues were among the disulfide bonds that had formed after 120 min folding in proteins that eluted at 12.8-min. In these proteins, the intermolecular disulfide bond linking Cys 31 and Cys' 31 had not yet formed. Interestingly, Cys 48 and Cys 139 were found in both the modified and non-modified forms, indicating that some protein molecules contained an open Cys48-Cys139 disulfide bridge while others had the intact Cys48-Cys139 disulfide bond.

In the 13.1-min fraction collected after 120 min folding, two monomeric proteins, M-SS2-cam4-SG1 and M-SS2-cam5, were identified by ESI-MS analysis (Table 7). CID MS-MS experiments of this fraction showed a highly complex mixture of alkylated and non-modified cysteinyl residues. Carboxyamidomethylated cysteinyl residues included Cys 31, 48, 146, 157 and 159. The same CID MS-MS experiment also indicated that

Cys 7, 90, 102 and 139 were free of alkylation. Cysteine 31 was also found to be glutathionylated. Because M-SS2-cam4-SG1 and M-SS2-cam5 were eluted with the same HPLC retention time and they both contained 5 modified cysteine residues (Cys31, 48, 146, 157 and 159), it is likely that they might contain similar disulfide structure in which case Cys 31 was carboxyamidomethylated in M-SS2-cam5 and Cys 31 was glutathionylated in M-SS2-cam4-SG1. Residues Cys48, 146, 157 and 159 were carboxyamidomethylated in both monomers. These data verified that in these monomeric proteins, Cys 31, 157 and 159, which form the intermolecular disulfide bonds in native rhm-CSFβ, were not involved in any disulfide bridges. Importantly, some cysteine residues that were involved in disulfide bonds at the time of disulfide bond quenching did not have the same disulfide partners as in the native protein. For instance, in native rhm-CSFβ, Cys 102 forms a disulfide bridge with Cys 146, and Cys 139 bonds with Cys 48. However, Cys 48 and Cys 146 were modified with carboxyamidomethylation while Cys 139 and Cys 102 were free of alkylation. These data signified that the disulfide bonds formed in the monomeric proteins after 120 min were not the native disulfides as formed in rhm-CSFβ. The presence of non-native disulfide bonds suggested that the folding of rhm-CSFβ involved disulfide bond shuffling, which represented an important isomerization process catalyzed by GSH/GSSG.

The 13.4-min fraction was shown to contain a single monomeric protein, M-SS3-cam3, in which three cysteine residues were carboxyamidomethylated, including Cys 31, 157 and 159.

After folding had proceeded for 4,800 min, two HPLC pools were collected. One fraction contained D-SS9-cam0 and the other contained D-SS8-cam2. Analysis of the peptic peptides of the first fraction indicated that all 9 cysteine residues were free of alkylation, which confirmed their involvement in disulfide bonds. The pepsin digests of the later fraction showed that all cysteine residues, other than Cys 31, were free of alkylation. Thus, D-SS8-cam2 contained two intermolecular disulfide bonds linking Cys 157/159 and Cys' 157/159. Interestingly, this protein had similar molecular weight and HPLC retention time as a protein found in the 12.8-min fraction after 120 min folding. These are likely the same protein intermediate present in the folding mixture over an extended time. In this case, D-SS8-cam2 in the 12.8-min fraction after 120 min folding must contain unmodified Cys31 and Cys'31. Therefore, D-SS6-cam6 isolated from the same HPLC pool must contain unmodified Cys31, Cys157, Cys159 and Cys'31, Cys'157 and Cys'159 (Table 7).

4.3.4 In vitro folding monitored by deuterium exchange pulsed labeling.

The progress of rhm-CSFβ folding was also monitored by deuterium exchange pulsed labeling. Denaturation, reduction, and folding of rhm-CSFβ were performed in protonated solvents as in the disulfide bond quenching experiment. After each folding time, an aliquot of sample was removed and subjected to 10-fold dilution into deuterated solvent for 10 s. During the labeling time, all accessible amide hydrogens (NHs) became deuterated, while those that remained protonated were assumed to be either involved in hydrogen bonding or buried deep within the folded protein. The labeling pulse was quenched by lowering the solution pH to 2.5 and the temperature to 0°C. In this experiment, disulfide bond quenching was not performed because disulfide bond formation was assumed to slow significantly under the H/D exchange quenching conditions (acid pH and low temperature). The protein samples were analyzed directly by ESI-MS.

After 1 min folding, a monomer was present and nearly 80% of all exchangeable NHs had exchanged to deuteriums, indicating that the protein had a non-compact overall conformation (Figure 27). After 40 min, a dimer appeared along with the monomer. The dimer also exhibited a high level of deuterium, incorporated by ~80% of the available amide sites. The

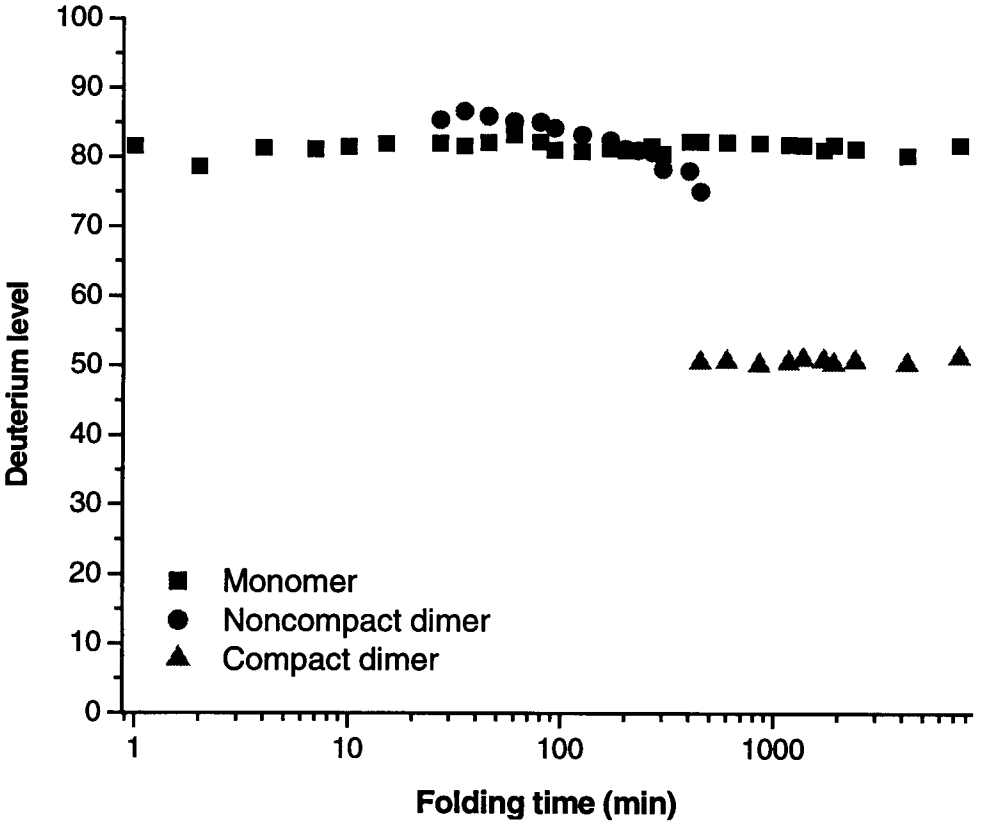


Figure 27. Time-dependence of deuterium levels in refolded rhm-CSFβ after H/D pulsed labeling.

high levels of deuterium incorporation indicated there was easy access of the deuterated solvent to NHs in the dimer; therefore the dimer must also have a non-compact overall conformation. The monomer continued to be present in the folding mixture and the level of deuterium incorporation remained high at ~80%. The deuterium level in the non-compact dimer decreased steadily from 80% to 75% between 40 min and 400 min folding. After 400 min, deuterium incorporation in the dimer rapidly decreased from 75% to ~50% of all amides within 50 min. The deuterium level in the compact dimer remained at ~50% throughout the remainder of the folding time course. Deuterium incorporation for the native rhm-CSFβ was 46%. These results indicate that the native dimer structure was more compact and less susceptible to H/D exchange than the compact dimer present in later stages of folding.

4.4 Conclusion

HPLC analysis indicated that the transition from the denatured and reduced polypeptides to the compactly folded rhm-CSFβ was 87% complete within 7,200 min (120 hr) under the present folding condition. Complex mixtures of folding intermediates were present throughout the

course of folding. The time-dependent folding events revealed that rhm-CSFβ molecules passed through both monomeric and dimeric intermediate states, suggesting that the denatured and disulfide bond reduced rhm-CSFβ folded *via* multiple pathways (Figure 28). Intermediates isolated after 1, 120, and 4,800 min folding were subjected to detailed structural analysis by HPLC fractionation and peptide analysis. Monomeric intermediates were predominant during early folding. Several cysteine residues were shown to participate in non-native disulfide bonds in monomeric intermediates, suggesting that disulfide bond shuffling represented an important step during early folding. Initial dimerization was achieved by forming intermolecular disulfide bonds Cys157/159-Cys157/159. The presence of D-SS8-cam2 in both 120 and 4,800 min folding mixture indicated that it might be a thermodynamically stable intermediate whose conversion to the native protein was slow. The conversion of D-SS8-cam2 to the native state only required the formation the last intermolecular disulfide bond Cys31-Cys'31. This process was relatively slow perhaps because D-SS8-cam2 required major structural rearrangement to facilitate the formation of Cys31-Cys'31.

H/D pulsed labeling data indicated the coexistence of dimeric and monomeric species of non-compact conformations, which is in agreement with multiple pathways for rhm-CSFβ folding. The deuterium level in the

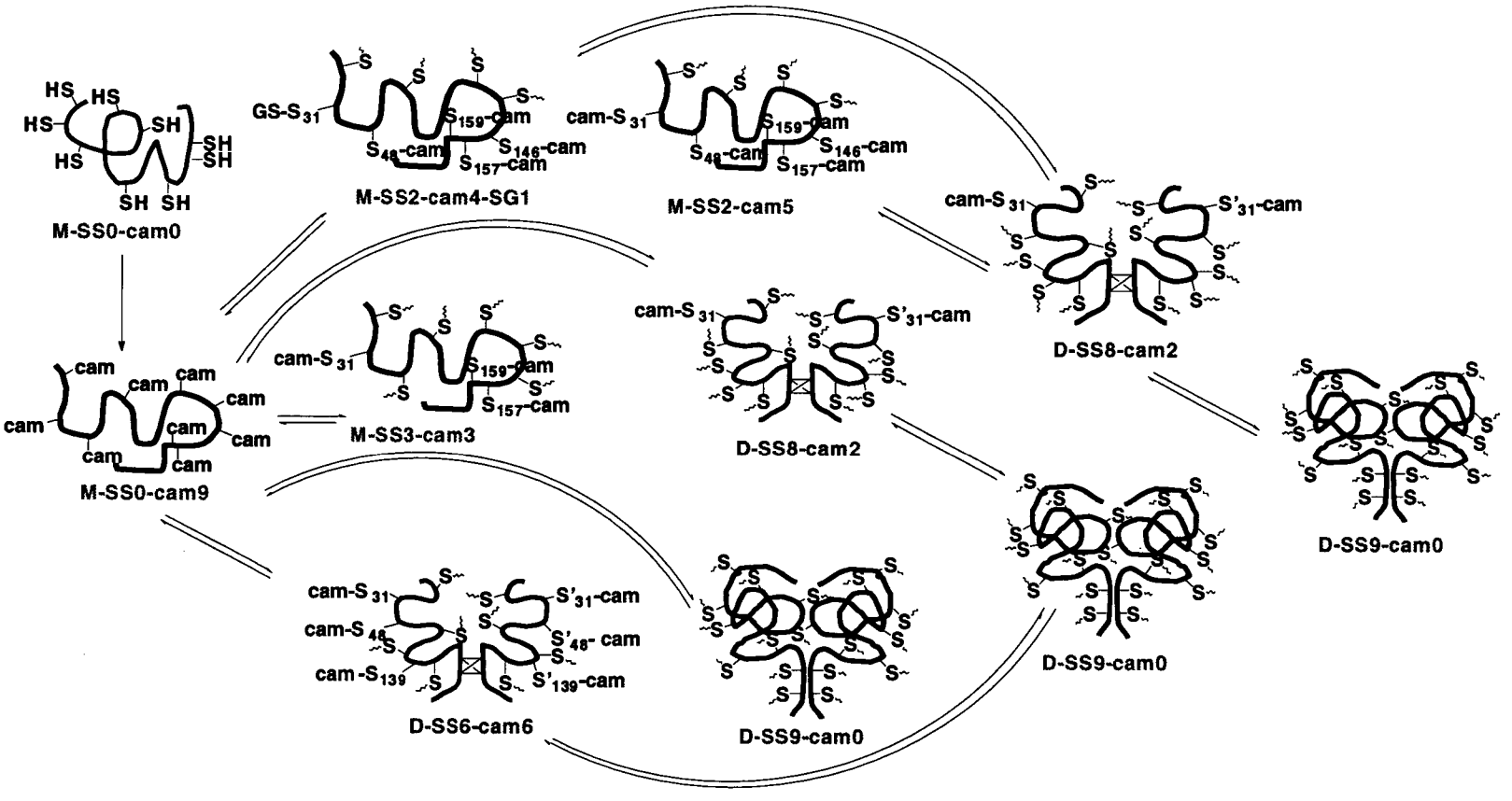


Figure 28. Heterogeneous mixtures of monomeric and dimeric intermediates in rhm-CSFβ folding.

non-compact dimer decreased steadily from 80% to 75% between 40 min and 400 min (Figure 27), indicating minor conformational changes associated with protein folding. After 400 min, the non-compact dimer started to collapse to a more compact form, losing 20% deuterium label within 50 min. The rapid reduction of the deuterium level in the non-compact dimer suggested that the protein underwent a rapid structural contraction, which significantly limited the accessibility of NHs to solvent deuterium. Clearly, the transformation of the non-compact to the compact dimer involved slow conformational changes followed by a fast structural change. The deuterium level remained at 80% for the monomer throughout the folding tie course suggesting that deuterium accessibility in the monomer was not affected even after intramolecular disulfide bonds had formed.

Folding of rhm-CSFβ monitored by disulfide bond quenching and H/D pulsed labeling was conducted under the same conditions, including the same protein concentration, oxidant concentration, i.e. GSH/GSSG ratio, reaction temperature, pH, and folding times. Therefore, these two experiments should provide results regarding the same intermediates on the folding pathways. As demonstrated in the disulfide bond quenching experiment, rhm-CSFβ folding was characterized by complex mixtures of monomeric and dimeric intermediates. Consequently, the monomer, the

non-compact dimer and the compact dimer detected in the H/D pulsed labeling experiment must represent ensembles, which contain proteins with similar degrees of deuterium accessibility. The contents of the ensembles can be approximated by correlating the data obtained from disulfide bond quenching and H/D pulsed labeling studies. The monomer ensemble contained all monomeric species. The non-compact dimer ensemble included D-SS6-cam6 and D-SS7-cam4. The compact dimer ensemble consisted of D-SS8-cam2 and D-SS9-cam0. Therefore, the transition from the non-compact to the compact dimer involved an ensemble switch from species with fewer disulfide bonds to those with a high degree of disulfide bonds. The compact dimer obtained after 7,200 min folding exhibited 5% higher deuterium incorporation than the native rhm-CSFβ. This difference indicates that D-SS9-cam0 did not fully attain the same compact conformation as the native protein even though all native disulfide bonds had formed.

A previous folding study performed with rhm-CSFβ was performed under reducing conditions (0.4 mM residual DTT SH plus 0.4 mM GSH plus 0.5 mM protein SH / 0.4 mM GS^- = 3.3) and provided evidence for the following model (Glocker, Arbogast et al. 1994):

$$2M \rightarrow 2M' \leftrightarrow D' \rightarrow D \qquad \text{Equation 9}$$

At early folding times, monomers (M) formed intermediates (M') that involved no disulfide bonds. After 17 hr folding, Cys 157/159 and Cys' 157/159 formed the first disulfide bonds and facilitated the initial dimerization, resulting in the formation of a dimeric intermediate (D'). Further oxidation converted D' to the fully biologically dimer (D) after 72 hr folding.

The present study examined the intermediates on the folding pathways of rhm-CSFβ under slightly stronger reducing conditions (0.5 mM GSH plus 0.11 μM protein SH / 0.64 mM GS^- = 7.8). Even though the present condition was more reducing than Glocker's study, oxidation of disulfide bonds proceeded more quickly than reported by Glocker: the first dimeric proteins were observed after 45 min compared to 17 hr. This discrepancy may be due to the superior detection sensitivity provided by electrospray ionization mass spectrometry (ESI-MS) used in the present study in contrast to fast atom bombardment mass spectrometry (FAB-MS) used by Glocker. In other words, dimeric proteins might be present during early folding in Glocker's study but were not detected due to the low detection sensitivity of FAB-MS. Another difference between the two folding studies is that the monomeric intermediates isolated in the present study contained disulfide bonds which Glocker failed to observe. This difference is puzzling because a much higher molar ratio of iodoacetamide

to free cysteine residues (20:1) was employed in the present study than in Glocker's study (1:1). Consequently, the presence of disulfide bonds detected in the monomeric intermediates in the present study, which was deduced from the lesser degree of carboxyamidomethylation, was not caused by underalkylation. Finally, the data presented in the present study indicated that the denatured and disulfide bond reduced rhm-CSFβ folded *via* multiple pathways involving both monomeric and dimeric intermediates. Consequently, rhm-CSFβ folding is not a simple sequential formation of native disulfide bonds as proposed by Glocker. Instead, rhm-CSFβ folding is a diffusion-like process as described by the "folding funnel", which involves bumpy energy landscapes with kinetics traps and energy barriers where the protein folds to the native state *via* multiple pathways in the conformational space (Dill 1997; Dobson 1998).

m-CSFα (aa4-149) is a shorter form of m-CSFβ and contains all disulfide bonds found in mCSFβ except Cys157/159-Cys'157/159. A study of folding intermediates involved in the folding of m-CSFα reported the presence of both monomeric and dimeric intermediates (Wilkins 1993). The folding of m-CSFα involved a rapid formation of a small amount of non-native dimeric intermediates followed by a slow progression *via* both monomeric and dimeric intermediates. After 45 min folding, three intermediates were isolated and structurally characterized. One monomer

contained two native disulfide bonds (Cys48-139 and Cys7-90), another monomer contained three native disulfide bonds (Cys7-90, Cys48-139, and Cys102-146), and a dimer had all native disulfide bonds except Cys102-146 and Cys'102-'146. It is unclear whether this dimer was the same one referred to as the non-native dimeric intermediate detected in early folding. However, it is evident that m-CSFα folded *via* multiple pathways. Several similarities exist between the folding processes of m-CSFα and m-CSFβ. First, the folding of both proteins involve parallel pathways and supported the folding funnel model. Second, disulfide shuffling occurred in both folding processes and represented an important step in the conversion of unfolded polypeptides to folded proteins.

Despite a few general similar folding characteristics between m-CSFα and m-CSFβ, there is evidence for significant differences in the folding processes for the two proteins. The folding of m-CSFα was initiated by diluting the denatured and reduced monomer 10-fold without the addition of any chemical oxidant. The transition from monomer to fully folded dimer was complete within 25 hr at room temperature. In contrast, the folding of m-CSFβ approached completion after 120 hr in the presence of GSSG/GSH at room temperature. When glutathione was excluded from the folding reaction, the rate of folding slowed significantly, requiring at least six days of folding to approach completion. Clearly, the presence of four

cysteine residues, Cys157, Cys159, Cys'157 and Cys'159, greatly complicated the folding process in m-CSFβ, requiring much more time to complete than for m-CSFα under similar conditions.

Disulfide bonds Cys157/159-Cys'157/159 were found not be important for the biological activity or thermodynamic stability of rhm-CSFβ (Zhang 2001). Consequently, the roles of Cys157/159-Cys'157/159 were thought to involve the maintenance of some structural order in the C-terminal regions of the protein. However, the present study showed that Cys157/159 and Cys'157/159 formed early in folding, much sooner than Cys31-Cys'31. Because Cys157/159-Cys'157/159 were involved in the initial dimerization, they must play an important role in subunit association. Residues Cys157, Cys159, Cys'157 and Cys'159 were located in the C-termini where little structural organization existed. It is possible that they were more available for intermolecular interaction than Cys31 and Cys'31, which could be made unavailable by the conformational restrictions presented by nearby secondary structural units formed early in folding.

4.5 Material and methods

4.5.1 Materials.

Reduced and oxidized glutathione were obtained from Sigma Chemical Co.. Iodoacetamide was purchased from Fluka. Pre-packed Sephadex G-25 PD10 columns were acquired from Pharmacia Biotech. Pepsin was purchased from Pierce. All other reagents were of the highest grade commercially available.

4.5.2 In vitro folding monitored by disulfide bond quenching.

Denaturation and disulfide bond reduction of the native rhm-CSFβ was achieved by diluting the protein 10-fold into 50 mM Tris buffer (pH 8.5) containing 9 M urea, 0.25 M DTT and 5 mM EDTA. The sample was incubated at room temperature for 3 hr. Utilizing a HiTrap column, *in vitro* folding was initiated by rapid buffer exchange of the denatured, reduced protein solution into the folding buffer that contained 50 mM Tris (pH 8.5), 0.4 M urea, 5 mM GSH and 0.64 mM GSSG. The final protein

concentration in the folding reaction was 0.3 mg/ml. At various times ranging from 10 s to 7,200 min, the folding reaction was quenched by adding 25 mM iodoacetamide to reaction aliquots. Carboxyamidomethylated protein mixtures were analyzed using a LC-10AD double-pump system coupled to a PE-Sciex API III+ mass spectrometer. A capillary column packed with Vydac C4 TP214 material (5 μm particle size, 300 Å pore size, 5 μL/min flow rate) was used. The binary gradient elution included 0.05% trifluroacetic acid (TFA) as solvent A and ACN containing 0.05% TFA as solvent B. The proteins were eluted using an 8-min gradient from 10% to 90% B. MS spectra were analyzed by PE Sciex software (MacSpec and BioSpec).

4.5.3 Isolation of folding intermediates.

Denatured and disulfide bond reduced rhm-CSFβ was prepared as described above. After various folding times, the folding reaction was stopped by adding 25 mM iodoacetamide to the folding solution. Carboxyamidomethylation was performed at room temperature for 5 min. The alkylated protein mixture (0.5 mg, 10 nmol) was analyzed by reversed phase high performance liquid chromatography (RP-HPLC) on a Vydac C4

TP214 column (300Å, 5 μm, 4.6X250 mm, 1 ml/min flow rate) employing a binary linear gradient with 0.1% TFA as solvent A and acetonitrile (ACN) containing 0.09% TFA as solvent B. The proteins were eluted using a 60 min 44-54%B gradient. Protein fractions were collected manually and dried in a Speed-Vac (Savant).

4.5.4 Identification of folding intermediates.

Fractions collected from freeze-drying were resuspended in 100 mM phosphate buffer (pH 2.5). Each fraction was added to 8 M urea and 1 M TCEP in order to remove the existing disulfide bonds. Denaturation and disulfide bond reduction were allowed to proceed for 30 min at room temperature. Subsequently, the proteins were incubated with immobolized pepsin (protein/enzyme=1:2) for 10 min at room temperature. The peptic peptides were analyzed using a Waters HPLC equipped with two pumps coupled to a Finnigan quadrupole ion-trap mass spectrometer. A capillary column, packed with C18 Luna material (5 μm particle size, 300 Å pore size, 5 μL/min flow rate), was equilibrated in H_2O containing 0.03% TFA and eluted with a linear gradient of acetonitrile containing 0.03% TFA of

10% to 60% over a 90 min period. The peptic peptides were identified by CID-MS-MS.

4.5.5 In vitro oxidative folding monitored by H/D pulsed labeling.

Denaturation, disulfide bond reduction and oxidation of rhm-CSFβ were performed as described above. At folding times ranging from 10 s to 7,200 min, aliquots were removed from the folding reaction and subjected to a 10 s deuterium pulse by diluting the protein sample 10-fold into 50 mM Tris buffer (pD 8.5). Deuterium pulsed labeling was quenched by reducing the final pH to 2.5 and the temperature to 0°C. Samples were stored in liquid N_2 prior to mass spectrometric analysis. Unlabeled protein (0% reference) was prepared by dissolving native rhm-CSFβ in 0.1 M ammonium phosphate buffer (pH 2.5, 0°C, 1:1 D_2O/H_2O). The completely deuterated protein (100% reference) was prepared by incubating rhm-CSFβ in 8 M urea-d_4/D_2O at 37°C for 6 hrs.

4.5.6 Electrospray ionization mass spectrometry of deuterated samples.

The samples from pulsed labeling experiments were thoroughly desalted by washing with 0.03% TFA/H_2O using a C4 MicroTrap (Michrom BioResources, Inc) installed in the injector. Protein samples were analyzed by LC-ESI-triple quadrupole MS (PE-Sciex API III$^+$) using a custom-packed C4 column (5 μm particle size, 300 Å pore size, 20 μL/min flow rate) and a binary gradient elution involving 0.03% TFA as solvent A and ACN containing 0.03%TFA as solvent B. The protein was eluted using a 3 min 10-80% B gradient. The HPLC injector and column were submerged in ice/water slurry, and the transfer syringe was pre-cooled on ice before use. MS spectra were analyzed by BioSpec software.

4.5.7 Data analysis

Because the HPLC elution was performed with protiated solvents, a small number of deuteriums at the amide positions are lost. Corrections were made for this deuterium loss according to eq. 6 (Zhang & Smith, 1993):

$$D = [(m-m_{0\%})/(m_{100\%} - m_{0\%})] \times N \qquad \text{Equation 6}$$

where D is the number of deuteriums present in a particular peptide segment or protein after incubation in deuterated solvent, and m, $m_{0\%}$, and $m_{100\%}$ represent the average molecular weight of a peptide or protein obtained for non-deuterated, partially deuterated and completely deuterated samples, respectively. N is the total number of exchangeable amide hydrogens in the protein or the peptide of interest.

4.6 References

Bai, Y., T. R. Sosnick, et al. (1995). "Protein folding intermediates: native-state hydrogen exchange." Science **269**(5221): 192-7.

Bourette, R. P. R., L.R. (2000). "Early events in m-CSF Receptor Signaling." Growth Factors **17**: 155-166.

Chang, J. Y. L., L.; Lai, P-H. (2001). "A major kinetic trap for the oxidative folding of human epidermal growth factor." J Biol Chem **276**(7): 4845-4852.

Clarke, A. R. W., J.P. (1997). "Protein folding pathways and intermediates." Curr. Opin. Biotech. **8**: 400-410.

Creighton, T. E. (1978). "Experimental studies of protein folding and unfolding." Prog Biophys Mol Biol **33**: 231-297.

Dill, K. A. C., H.S. (1997). "From Levinthal to pathways to funnels." Nature Structural Biology **4**(1): 10-19.

Dobson, C. M. S., A.; Karplus, M. (1998). "Protein folding: a perspective from theory and experiment." Angew Chem Int Ed Engl **37**: 868-893.

Glocker, M. O., B. Arbogast, et al. (1994). "Disulfide linkages in the in vitro refolded intermediates of recombinant human macrophage-colony-stimulating factor: analysis of the sulfhydryl alkylation of free cysteine residues by fast-atom bombardment mass spectrometry." Proc Nat'l Acad Sci USA **91**(13): 5868-72.

Jaenicke, R. R., R. (1989). "Folding proteins." Protein structure, a practical approach: Edited by T.E. Creighton. IRL press.

Liu, Z. R., J.; Gierash, L.M. (1998). "Protein folding." Bioorganic chemistry peptides and proteins Sidney, M. Hecht. Oxford University Press. New York.

Lu, H. S. C., C.L.; Narhi, L.O.; Merewether, L.A.; Pearl, W.R.; Boone, T.C. (1992). "Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in E coli." J Biol Chem **267**(13): 8770-8777.

Lu, H. S. C., C.L.; Narhi, L.O.; Merewether, L.A.; Pearl, W.R.; Boone, T.C. (1992). "Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in E. Coli." J. Biol. Chem. **267**(13): 8770-8777.

Onda, M. T., E.; Takahashi, N.; Hirose, M. (1997). "Refolding process of ovalbumin from urea-reduced state." J Biol Chem **272**(7): 3973-3979.

Pandit, J. B., A.; Jancarik, J.; Halenbeck, R.; Koths, K.; Kim, S.H. (1992). "Three-dimensional structure of dimeric human recombinant macrophage colony stimulating factor." Science **258**: 1358-1362.

Roder, H. E., G.A.; Englander, S.W. (1988). "Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and NMR." Nature **335**(20): 700-704.

Stanley, E. R. B., K.L.; Einstein, D.B.; Lee, P.S.; Yeung, Y.G. (1994). "The biology and action of colony stimulating factor-1." Stem Cells **12**: 15-24.

Tsui, V., C. Garcia, et al. (1999). "Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through and obligatory intermediate." Protein Sci **8**(1): 45-9.

Weissman, J. S. K., P.S. (1991). "Reexamination of the folding of BPTI: predominance of native intermediates." Science **253**: 1386-1393.

Wilkins, J. A. C., J.; Randahawa, Z.I.; Wood, D.; Warren, M.K.; Witkowska, H. E. (1993). "A study of intermediates involved in the folding pathway for recombinant human macrophage colony-stimulating factor (M-CSF): Evidence for two distinct folding pathways." Protein Science **2**: 244-254.

Zhang, Y. H. Y., X-G.; Maier, C.L.; Schimerlik, M.I.; Dinezer, M.L. (2001). "Structural Comparison of Recombinant Human Macrophage Colony Stimulating Factor b (rhm-CSFb) and Its Derivative Using Hydrogen Deuterium Exchange and Electrospray Ionization Mass Spectrometry." Prot. Sci. **10**: 2336-2345.

Zhou, H.-X. (2001). "Single chain versus dimeric protein folding: thermodynamic and kinetic consequences of covalent linkage." J Am Chem Soc **123**: 6730-6731.

5 Summary

Elucidation of the pathways of protein folding and unfolding has remained one of the most difficult tasks in protein chemistry. For proteins containing disulfide bonds, unfolding and folding can be followed by reduction and oxidation of the native disulfide bonds. The breakage and the formation of disulfide bonds can be chemically trapped and characterized. Therefore, the disulfide bond coupled folding and unfolding pathways can be constructed based on the heterogeneity and structures of the trapped intermediates. Unfolding of a protein containing disulfide bonds can be achieved by reducing disulfide bonds in the absence of denaturant, a method referred to as reductive unfolding. In oxidative folding, a disulfide bond reduced protein is oxidized to recover its native disulfide bonds and native conformation as well as biological activity. Under the same conditions, reductive unfolding and oxidative folding exist in equilibrium, which can be shifted by adjusting the experimental condition.

In rhm-CSFβ, the assembly of its monomeric subunits and its maturation of the biological activity depend on the progressive formation of the correct disulfide bonds during *in vitro* folding (Glocker, Arbogast et al. 1994). Clearly, the formation of native disulfide bonds is intimately linked to

attaining the native three-dimensional structure. Knowledge about the conversion from the reduced to the native rhm-CSFβ can be obtained by studying the intermediates isolated from the reductive unfolding pathway as well as those from the oxidative folding pathway. The goal is to obtain disulfide-bonded intermediates and determine their thermodynamic and kinetic properties. In this project, intermediates from both reductive unfolding and oxidative folding pathways of rhm-CSFβ were isolated and structurally characterized.

Reductive unfolding was initiated by incubating rhm-CSFβ with a reducing agent in the absence of any chemical denaturant. One intermediate was isolated and identified as $CN^{157,159}$-modified rhm-CSFβ. The removal of disulfide bonds Cys157/159-Cys'157/159 did not affect the biological activity of the protein nor its stability. However, a comparison of H/D exchange kinetics of rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ indicated that the C-terminal regions of the protein became more exposed to solvent while structures in other regions of the protein remained the same. These results suggested that disulfide bonds Cys157/159-Cys'157/159 served to maintain some structural order in the C-termini.

To assess the roles of disulfide bonds Cys157/159-Cys'157/159 in folding, D/H pulsed labeling was employed to monitor the folding processes of urea denatured rhm-CSFβ and $CN^{157,159}$-modified rhm-CSFβ.

Denaturation was performed in the absence of any reducing agent. Therefore, the unfolded proteins retained the intact native disulfide structures: rhm-CSFβ maintained all nine native disulfide bonds and $CN^{157,159}$-modified rhm-CSFβ contained seven native disulfide bonds. Deuterium exchange data indicated that both proteins attained secondary structures within 7.5 ms and did not experience further structural change as folding proceeded to 48 hr. These results suggested that the folding kinetics of oxidized rhm-CSFβ were not affected by the absence of Cys157/159-Cys'157/159. However, it is possible that the folding of the denatured protein was affected but the effects were not readily detected by D/H pulsed labeling.

Extensive effort was devoted to isolating other disulfide-bonded intermediates in order to construct the unfolding pathways for rhm-CSFβ. Various experimental parameters for the disulfide bond reduction were carefully adjusted, including TCEP/SH stoichiometry, reaction times, and/or reaction temperature (Appendix 2). However, these procedures did not produce any homogeneous, partially reduced intermediates whose structures and stabilities could be characterized. Consequently, the reductive unfolding study was not fully successful in constructing the unfolding pathways for rhm-CSFβ.

Oxidative folding of rhm-CSFβ was investigated by monitoring the structural changes that occurred as the denatured and reduced rhm-CSFβ folded into the native state with the aid of a chemical oxidant. rhm-CSFβ folding, monitored by disulfide bond quenching and H/D pulsed labeling, was conducted under the same conditions, including the same protein concentration, oxidant concentration, i.e. GSH/GSSG ratio, reaction temperature, pH, and folding times. Therefore, these two experiments provided results regarding the same intermediates in the folding pathways. rhm-CSFβ folding was characterized by a heterogeneous mixture of monomeric and dimeric intermediates, indicating that rhm-CSFβ folded via multiple pathways. The disulfide structures of seven intermediates were identified. These data indicated that disulfide bond shuffling catalyzed by GSH/GSSG represented an important isomerization process during folding. A dimeric intermediate, D-SS8-cam2, represented a kinetic trap whose conversion to the native protein was slow, perhaps because significant structural rearrangement was necessary for D-SS8-cam2 to convert to the native protein. H/D pulsed labeling data indicated that many intermediates exhibited similar solvent accessibility making the structural characterization of individual species impossible.

The roles of disulfide bonds Cys157/159-Cys'157/159 in folding must be examined in the contexts of folding (from the protein containing intact

disulfide bonds) and oxidative folding (from the denatured and disulfide bond reduced protein). In the presence of seven native disulfide bonds, rhm-CSFβ folding proceeded very quickly and Cys157/159-Cys'157/159 did not seem to play any important role. By contrast, when folded from the reduced protein, folding required much more time to complete and Cys157/159-Cys'157/159 facilitated the initial subunit association. Clearly, Cys157/159-Cys'157/159 demonstrated different activities in different folding systems. In the presence of the seven native disulfide bonds, non-covalent forces were quickly restored and Cys157/159-Cys'157/159 only assisted the local folding in the C-termini. In the absence of other native disulfide bonds, Cys157/159 and Cys'157/159 served to link the subunits, perhaps because they were located in less structurally confined regions and therefore could easily form inter-subunit contacts.

Traditionally, protein folding was understood as a sequential process involving obligatory intermediates. However, in the context of a "new view" on protein folding, the concept of discrete folding pathways has been superseded by three dimensional energy landscapes or folding funnels, which depict the energetic search to the native state (Dill 1997). Such a view does not preclude the concept that folding might still occur *via* preferred routes and defined intermediate states can still be envisioned (Lazaridis 1997). Information taken from the oxidative folding studies

indicated that the folding of rhm-CSFβ was not a simple sequential formation of native disulfide bonds. The time-dependent folding events revealed that rhm-CSFβ molecules passed through both monomeric and dimeric intermediate states, suggesting that the denatured and disulfide bond reduced rhm-CSFβ folded *via* multiple pathways. These data clearly depicted rhm-CSFβ folding as a complex energy surface.

In order to describe the energy landscape in detail, the thermodynamic and kinetic properties of the intermediates must be determined in addition to their structural characterization. In the present study, the stability of $CN^{157,159}$-modified rhm-CSFβ was obtained by equilibrium unfolding experiments, which require large quantities of protein material (mg). Because it is difficult to prepare large quantities of homogeneous material, the stabilities for the intermediates isolated in oxidative folding were not determined. The relative stabilities of folding intermediates can also be assessed by the computation approach. The goal of a successful computational analysis is to carry out a well-defined procedure that results in the *ab initio* folding of polypeptide and obtain the free energies for various protein conformations. Energies calculated for the folding intermediates can help assign their positions on the energy landscape.

A major challenge in studying rhm-CSFβ folding results from the difficulty of isolating intermediates. As indicated in the oxidative folding study, many more disulfide-bonded intermediates were present in folding (900,000,000 possible disulfide bonded isomers). However, only a few were structurally analyzed because other intermediates were transient and did not accumulate in folding. To study the structural and energetic properties of intermediates with other disulfide structures, rhm-CSFβ folding should be investigated with mutants in which some cysteine residues have been substituted (Horovitz 1990; Roder 1997). In such studies, the effects of replacing certain cysteine residues on the relative stabilities of the intermediates can be assessed. Structural characterization of various rhm-CSFβ mutants should allow additional details of the conformation of the intermediates to be inferred. Replacements at different cysteine sites should yield distinct effects on both conformation and stability of the intermediates at different stages in the folding pathways. Both genetic (Zavodszky 2001) and chemical (Drakopoulou 1998; Pegoraro 1999) methods can be utilized to prepare mutant proteins for structural studies. However, the results from experiments on mutants must be carefully interpreted because amino acid replacements may alter different properties of the native, unfolded, and intermediate states of rhm-CSFβ.

Bibliography

Anderegg, R. J. W., D.S. (1995). "Mass spectrometric characterization of a protein ligand interaction." J Am Chem Soc **117**(4): 1374-1377.

Anderegg, R. J. W., D.S.; Stevenson, C.L.; Borchardt, R.T. (1994). J Am Soc Mass Spectrom **5**: 425-433.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**: 223-230.

Anfinsen, C. B. H., E.; Sela, M.; White, F.H.Jr. (1961). "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain." Proc. Nat. Acad. Sci. USA. **47**: 1309-1314.

Arrington, C. T., O.; Robertson, A. (1999). "Defining protein ensembles with native state NH exchange: kinetics of interconversion and cooperative units from combined NMR and MS analysis." J. Mol. Biol. **285**: 1265-1275.

Bai, Y., T. R. Sosnick, et al. (1995). "Protein folding intermediates: native-state hydrogen exchange." Science **269**(5221): 192-7.

Bai, Y. M., J.S.; Mayne, L.; Englander, S. W. (1993). "Primary structure effects on peptide group hydrogen exchange." Proteins **17**(1): 75-86.

Bourette, R. P. R., L.R. (2000). "Early events in m-CSF Receptor Signaling." Growth Factors **17**: 155-166.

Cerretti, D. P. W., J.; Anderson, D.; Tushinski, R.J.; Cosman, D.; (1988). "Human macrophage colony stimulating factor: an alternative RNA and protein processing from a single gene." Mol. Immunology. **25**(8): 761-770.

Chakravarty, S. V., R. (1999). "Residue depth: a novel parameter for the analysis of protein structure and stability." Structure **7**(7): 723.

Chang, J. Y. L., L.; Lai, P-H. (2001). "A major kinetic trap for the oxidative folding of human epidermal growth factor." J Biol Chem **276**(7): 4845-4852.

Chorush, R. A. L., D.P.; Beu, S.C.; Wood, T.D.; McLafferty, F.W. (1995). Anal Chem **67**: 1042-1046.

Chowdhury, S. K. K., V.; Chait, B.T. (1990). J Am Chem Soc **112**: 9012.

Clarke, A. R. W., J.P. (1997). "Protein folding pathways and intermediates." Curr. Opin. Biotech. **8**: 400-410.

Coyle, J. E. T., F.L.; Ashcroft, A.E.; Masselos, D.; Robinson, C.V.; Radford, S.E. (1999). "GroEL accelerates the refolding of hen lyzozyme without changing its folding mechanism." Nature structural biology **6**(7): 683-690.

Creighton, T. E. (1977). "Conformational restrictions on the pathway of folding and unfolding of the bovine pancreatic trypsin inhibitor." J Mol Biol **113**: 275-293.

Creighton, T. E. (1978). "Experimental studies of protein folding and unfolding." Prog Biophys Mol Biol **33**: 231-297.

Creighton, T. E. (1992). "Protein folding." W.H. Freeman and Company, New York.: xi-xii.

Creighton, T. E. (1993). "Proteins: structures and molecular properties." W.H. Freeman and Company, New York. Second Edition.: 316-323.

Csar, X. F. W., N.J.; McMahon, K.A.; Hamilton, J.A. (2001). "Proteomic analysis of macrophage differentiation." J Biol Chem **276**(28): 26211-24217.

Deng, Y. and D. L. Smith (1998). "Identification of unfolding domains in large proteins by their unfolding rates." Biochemistry **37**(18): 6256-62.

Deng, Y., Z. Zhang, et al. (1999). "Comparison of continuous and pulsed labeling amide hydrogen exchange/mass spectrometry for studies of protein dynamics." J Am Soc Mass Spectrom **10**(8): 675-84.

Deng, Y. P., H; Smith, D.L (1999). "Selective Isotope Labeling Demonstrates that Hydrogen Exchange at Individual Peptide Amide Linkages can be determined by Collision-Induced Dissociation Mass Spectrometry." J Am Chem Soc **121**: 1966-1967.

Dill, K. A. C., H.S. (1997). "From Levinthal to pathways to funnels." Nature Structural Biology **4**(1): 10-19.

Dobson, C. M. S., A.; Karplus, M. (1998). "Protein folding: a perspective from theory and experiment." Angew Chem Int Ed Engl **37**: 868-893.

Drakopoulou, E. V., J.; Neyton, J.; Aniort, V.; Bouet, F.; Virelizier, H.; Menez, A.; Vita, C. (1998). "Consequence of the removal of evolutional conserved disulfide bridges on the structure and function of Charbdotoxin and evidence that particular cysteine spacings govern specific disulfide formation." Biochem **37**: 1292-1301.

Engen, J. R., W. H. Gmeiner, et al. (1999). "Hydrogen exchange shows peptide binding stabilizes motions in Hck SH2." Biochemistry **38**(28): 8926-35.

Engen, J. R., T. E. Smithgall, et al. (1997). "Identification and localization of slow, natural, cooperative unfolding in the hematopoietic cell kinase SH3 domain by amide hydrogen exchange and mass spectrometry." Biochemistry **36**(47): 14384-91.

Englander, S. W. K., N. R. (1984). "Hydrogen exchange and structural dynamics of proteins and nucleic acids." Quarterly Review of Biophysics **16**(4): 521-653.

Eyles, S. J., S. E. Radford, et al. (1994). "Kinetic consequences of the removal of a disulfide bridge on the folding of hen lysozyme." Biochemistry **33**(44): 13038-48.

Gale, D. C. S., R.D. (1995). "Characterization of noncovalent complexes formed between minor groove binding molecules and duplex DNA by electrospray ionization mass spectrometry." J Am Soc Mass Spectrom **6**: 1154-1164.

Glocker, M. O., B. Arbogast, et al. (1994). "Disulfide linkages in the in vitro refolded intermediates of recombinant human macrophage-colony-stimulating factor: analysis of the sulfhydryl alkylation of free cysteine residues by fast-atom bombardment mass spectrometry." Proc Nat'l Acad Sci USA **91**(13): 5868-72.

Glocker, M. O., B. Arbogast, et al. (1993). "Assignment of the inter- and intramolecular disulfide linkages in recombinant human macrophage colony stimulating factor using fast atom bombardment mass spectrometry." Biochemistry **32**(2): 482-8.

Gregory, R. B. R., A. (1986). Methods Enzymol. **131**: 448-517.

Horovitz, A. F., A.R. (1990). "Strategy for analyzing the cooperativity of intramolecular interactions in peptides and proteins." J Mol Biol **214**: 613-617.

Jaenicke, R. R., R. (1989). "Folding proteins." Protein structure, a practical approach: Edited by T.E. Creighton. IRL press.

Kato, J. S., C.J. (1990). Blood **75**: 1780-1787.

Kawasaki, E. S. L., M.B. (1990). "Molecular biology of macrophage colony stimulating factor." Immonol. Ser. **49**: 155-176.

Kawasaki, E. S. L., M.B.; Wang, A.M.; Van Arsdell, J.; Warren, M.K.; Stanley, E.R.; Ralph, P.; Mark, D.F. (1985). "Molecular cloning of a complementary DNA encoding human macrophage colony stimulating factor." Science **230**: 291-296.

Kenny, P. T. N., K.; Orlando, R. (1992). "Fragmentation studies of peptides: the formation of y ions." Rapid Comm Mass Spec **6**: 95-97.

Kim, M. Y. M., C.L.; Reed, D.; Deinzer, M.L. (2001). "Site-Specific Amide Hydrogen/Deuterium Exchange in E. coli Thioredoxins Measured by Electrospray Ionization Mass Spectrometry." J Am Chem Soc **123**(40): 9860.

Kim, P. S. B., R.L. (1990). "Intermediates in the folding reactions of small proteins." Annu Rev Biochem **59**: 631-660.

Koyama, N. K., M.; Amano, N.; Nishimura, O. (1994). "A novel procedure for the preparation of biologically active recombinant peptides using a cyanylation reaction." Journal of biotechnology **32**: 273-281.

Kragelund, B. B. K., J.; Poulsen, F.M. (1995). "Local Perturbations by Ligand Binding of Hydrogen Deuterium Exchange Kinetics in a Four-Helix Bundle Protein, Acyl Coenzyme A Binding Protein (ACBP)." J. Mol. Biol. **250**: 695-706.

Lazaridis, T. K., M. (1997). ""New view" of protein folding reconciled with the old through multiple unfolding simulations." Science **278**: 1928-1931.

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

Li, R.-H. W., C. (1999). "The Hydrogen Exchange Core and Protein Folding." Protein Science **8**: 1571-1591.

Little, D. P. S., J.P.; Senko, M.W.; O'Conner, P.B.; McLafferty, F.W. (1994). Anal Chem **66**: 2809-2815.

Liu, Z. R., J.; Gierash, L.M. (1998). "Protein folding." Bioroganic chemistry peptides and proteins Sidney, M. Hecht. Oxford University Press. New York.

Loo, J. A. E., C.G.; Udseth, H.R.; Smith, R.D. (1990). Anal Chem **62**: 693-698.

Loo, J. A. O. L., R.R.; Udseth, H.R.; Edmonds, C.G.; Smith, R.D. (1991). Rapid Comm Mass Spec **5**: 101.

Loo, J. A. U., H.R.; Smith, R.D. (1988). Rapid Comm Mass Spec **2**: 207-210.

Lu, H. S. C., C.L.; Narhi, L.O.; Merewether, L.A.; Pearl, W.R.; Boone, T.C. (1992). "Folding and oxidation of recombinant human granulocyte

colony stimulating factor produced in E coli." J Biol Chem **267**(13): 8770-8777.

Lu, H. S. C., C.L.; Narhi, L.O.; Merewether, L.A.; Pearl, W.R.; Boone, T.C. (1992). "Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in E. Coli." J. Biol. Chem. **267**(13): 8770-8777.

Mandell, J. G. B.-O., A.; Akashi, S.; Takio, K.; Komives, E.A. (2001). "Solvent accessibility of the thrombin-thrombomodulin interface." J Mol Biol **306**: 575-589.

Mandell, J. G. F., A.M.; Komives, E.A. (1998). "Identification of Protein-Protein Interfaces by Decreased Amide Proton Solvent Accessibility." Proc. Nat. Acad. Sci. USA. **95**: 14705-14710.

Mandell, J. G. F., A.M.; Komives, E.A. (1998). "Measurement of Amide Hydrogen Exchange by MALDI-TOF Mass Spectrometry." Anal Chem **70**: 3987-3995.

McLafferty, F. W., ; Guan, Z.; Haupts, U.; Wood, T.; Kelleher, N.L. (1998). "Gaseous Conformational Structures of Cytochrome c." J Am Chem Soc **120**: 4732-4740.

Miranker, A., C. V. Robinson, et al. (1993). "Detection of transient protein folding populations by mass spectrometry [see comments]." Science **262**(5135): 896-900.

Moarefi, I. L.-B., M.; Sicheri, F.; Huse, M.; Lee, C.H.; Kuriyan, J.; Miller, W.T. (1997). "Activation of the SRC-family tyrosince kinase Hck by SH3 domain displacement." Nature **385**: 650-653.

Molday, R. S. E., S.W.; Kallen, R.G. (1972). "Primary structure effects on peptide group hydrogen exchange." Biochemistry **11**(2): 150-158.

Murphy, R. C. F., J.; Hevko, J. (2001). "Analysis of nonvolatile lipids by mass spectrometry." Chem Rev **101**: 479-526.

Nakoinz, I. L., M.T.; Weaver, J. F.; Ralph, P. (1990). "Differentiation of the IL-3-Dependent NFS-60 cell line and Adaptation to Growth in Macrophage Colony Stimulating Factor." the Journal of Immunology **145**(3): 860-864.

Ogorzalek Loo, R. R. S., R.D. (1994). J Am Soc Mass Spectrom **5**: 207.

Onda, M. T., E.; Takahashi, N.; Hirose, M. (1997). "Refolding process of ovalbumin from urea-reduced state." J Biol Chem **272**(7): 3973-3979.

Pandit, J. B., A.; Jancarik, J.; Halenbeck, R.; Koths, K.; Kim, S.H. (1992). "Three-dimensional structure of dimeric human recombinant macrophage colony stimulating factor." Science **258**: 1358-1362.

Pawlak, G. V.-D., C.; Revol, V.; Bourette, R.P.; Blanchet, J.P.; Mouchiroud, G. (1999). Exp. Hematol. **27**: 797-805.

Pegoraro, S. R., S.; Cramer, J.; Rudolph-Bohner, S.; Moroder, L. (1999). "The disulfide-coupled folding pathway of apamin as derived from diselenide-quenched analogs and intermediates." Prot. Sci. **8**: 1605-1613.

Perrin, C. A., G.M.L. (1982). "Chemical mechanisms of amide hydrogen exchange." J Amer Chem Soc **104**: 1982.

Perrin, C. L. (1989). Acc Chem Res **22**: 268.

Pinheiro, T. J. T. E., G.A.; Watts, A.; Roder, H. (1997). "Structural and kinetic description of cytochrome c unfolding induced by the interaction of lipid vesicles." Biochemistry **36**(42): 13122-13132.

Price, W. D. S., P.D.; Williams, E.R. (1996). Anal Chem **68**: 859-866.

Przybyski, M. (1995). Adv. Mass. Spectrom. **13**: 257.

Przybyski, M. G., M.O. (1996). "Electrospray mass spectrometry of biomolecular complexes with noncovalent interactions - new analytical perspectives for supramolecular chemistry and molecular recognition processes." Angew Chem Int Ed Engl **35**: 806-826.

Resing, K. A. A., N.G (1998). "Deuterium Exchange Mass Spectrometry as a Probe of Protein Kinase Activation. Analysis of Wild Type and Constitutively Active Mutants of MAP Kinase Kinase-1." Biochemistry **37**: 463-475.

Rettenmier, C. W. R., M.F.; Ashmun, R.A.; Ralph, P.; Price, K.; Sherr, C.J. (1987). "Synthesis of membrane bound colony-stimulating factor and down regulation of CSF receptor in NIH 3T3 cells transformed by cotransfection of the human CSF and CSF receptor." Mol. Cell. Biol. **7**: 2378-2387.

Roder, H. C., W. (1997). "Kinetic role of early intermediates in protein folding." Curr. Opin. Struc. Biol. **7**: 15-28.

Roder, H. E., G.A.; Englander, S.W. (1988). "Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and NMR." Nature **335**(20): 700-704.

Rohschneider, L. R. M., D. (1989). Mol. Cell. Biol. **9**: 5081-5092.

Rothwarf, D. M. S., H.A. (1991). "Regeneration and reduction of native bovine pancreatic ribonuclease A with oxidized and reduced dithiothreitol." J Am Chem Soc **113**: 6293-6294.

Rothwarf, D. M. S., H.A. (1993). "Regeneration of bovine pancreatic ribonuclease A. Steady state distribution." Biochemistry **32**: 2671-2679.

Roussel, M. F. S., C.J. (1989). Proc. Natl. Aca. Sci. USA **86**: 7924-7927.

Ruoppolo, M. V., F.; Klink, T.A.; Raines, R.T.; Marino, G. (2000). "Contribution of individual disulfide bonds to the oxidative folding of Ribonuclease A." Biochemistry **32**(12033-12042).

Smith, D. L. (1998). "Local structure and dynamics in proteins characterized by hydrogen exchange and mass spectrometry." Biochemistry (Mosc) **63**(3): 285-93.

Smith, D. L. Z., Z. (1994). "Local structure and dynamics in proteins characterized by hydrogen exchange and mass spectrometry." Mass Spectrom Rev **13**: 411.

Stanley, E. R. B., K.L.; Einstein, D.B.; Lee, P.S.; Yeung, Y.G. (1994). "The biology and action of colony stimulating factor-1." Stem Cells **12**: 15-24.

Stanley, E. R. B., K.L.; Einsteine, D.B.; Lee, P.S.; Pixley, F.J.; Wang, Y.; Yeung, Y.G. (1997). "Biology and Action of Colony-Stimulating Factor-1." Mol. Reprod. Dev. **46**: 4-10.

Stein, J. B., G.V.; Rettenmier, C.W. (1990). "Direct stimulation of cells expressing receptors for macrophage stimulating factor by a plasma membrane bound precursor of human CSF-1." Blood **76**(7): 1308-1314.

Suckau, D. S., Y.; Beu, S.C.; Senko, M.W.; Quinn, J.P.; Wamper, F.M.; McLafferty, F.W. (1993). Proc Nat'l Acad Sci USA **90**: 790.

Torella, C. R., M.; Marino, G.; Pucci, P. (1994). "Analysis of RNase A refolding intermediates by electrospray mass spectrometry." FEBS Letters **325**: 301-306.

Treisman, R. (1996). "Regulation of transcription by MAP kinase cascades." Curr. Opin. Cell. Biol. **8**: 205-215.

Tsui, V., C. Garcia, et al. (1999). "Quench-flow experiments combined with mass spectrometry show apomyoglobin folds through and obligatory intermediate." Protein Sci **8**(1): 45-9.

Tushinski, R. S., E.R. (1982). "Survival of mononuclear phagocytes depends on lineage-specific growth factor that the differentiated cells selectively destroy." Cell **28**: 71-81.

Wakselman, M. G.-J., E.; Raoult, A.; Busse, W.D. (1976). "1-cyano-4-dimethylamino-pyridinium salts: new water-soluble reagents for the cyanylation of protein sulphydryl groups." J.C.S. Chem. Comm.: 21-22.

Wang, B. H. M., F.W. (1990). Org. Mass Spectrum. **25**: 554-556.

Wang, F. B., J. S.; Tang, X-J. (1997). "Hydrogen exchange / electrospray ionization mass spectrometry studies of substrate and inhibitor binding and conformational changes of Ecoli dihydrogenase reductase." Biochemistry **36**(13): 3755-3759.

Wedemeyer, W. J. W., E.; Narayan, M.; Scheraga, H.A. (2000). "Disulfide bonds and protein folding." Biochemistry **39**(15): 4207.

Weissman, J. S. K., P.S. (1991). "Reexamination of the folding of BPTI: predominance of native intermediates." Science **253**: 1386-1393.

Weissman, J. S. K., P.S. (1992). "Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor." Proc. Nat. Acad. Sci. USA. **89**: 9900-9904.

Wiktor-Jedrejczak, W. U., E.; Aukerman, S.L.; Pollard, J.W.; Stanley, E.R.; Ralph, P.; Ansari, A.A.; Sell, K.W.; Szperl, M. (1991). "Correction of CSF-1 of defects in the osteopetric op/op mouse suggests local, developmental, and humoral requirements for this growth factor." Exp. Hematol. **19**: 1049-1054.

Wilkins, J. A. C., J.; Randahawa, Z.I.; Wood, D.; Warren, M.K.; Witkowska, H. E. (1993). "A study of intermediates involved in the folding pathway for recombinant human macrophage colony-stimulating factor (M-CSF): Evidence for two distinct folding pathways." Protein Science **2**: 244-254.

Winger, B. E. L.-W., K.J.; Rockwood, A.L.; Smith, R.D. (1992). J Am Chem Soc **114**: 5897.

Wong, G. G. T., P.A.; Leary, A.C.; Witek-Giannotti, J.S.; Yang, Y-C.; Clark, S.C.; Saito, M.; Motoyoshi, K.; Takaku, F. (1987). "Human CSF-1: molecular cloning and expression of a 4 kb cDNA encoding the human urinary protein." Science **235**: 1504-1508.

Woodward, C. (1993). Trends Biochem Sci **18**: 359-361.

Woodward, C. (1999). "Advances in protein hydrogen exchange by mass spectrometry." J Am Soc Mass Spectrom **10**: 672-674.

Wu, J., Y. Yang, et al. (1998). "Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by

cyanylation, and subsequent structural elucidation by mass spectrometry." Protein Sci **7**(4): 1017-28.

Wu, J. W., J. T. (1997). "A novel methodology for assignment of disulfide bond pairings in proteins." Protein Science(6): 391-398.

Yan, X. G. Z., Y.H.; Schimerlik, M.I.; Dinezer, M.L. (2002). "Hydrogen deuterium exchange and mass spectrometric analysis of a protein containing multiple disulfide bonds: solution dynamics of recombinant macrophage colony stimulating factor beta." Prot. Sci. (submitted).

Zavodszky, M. C., C-W.; Huang, J-K.; Zolkiewski, M.; Wen, L.; Krishnamoorthi, R. (2001). "Disulfide bond effects on protein stability: designed variants of *Cucurbita maxima* trypsin inhibitor-V." Prot. Sci. **10**: 149-160.

Zhang, Y. H. Y., X-G.; Maier, C.L.; Schimerlik, M.I.; Dinezer, M.L. (2001). "Structural Comparison of Recombinant Human Macrophage Colony Stimulating Factor b (rhm-CSFb) and Its Derivative Using Hydrogen Deuterium Exchange and Electrospray Ionization Mass Spectrometry." Prot. Sci. **10**: 2336-2345.

Zhang, Z. and D. L. Smith (1993). "Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation." Protein Sci **2**(4): 522-31.

Zhou, H.-X. (2001). "Single chain versus dimeric protein folding: thermodynamic and kinetic consequences of covalent linkage." J Am Chem Soc **123**: 6730-6731.

Zubarev, R. A. H., D.M.; Fridriksson, E.K.; Lewis, M.A.; Carpenter, B.K.; McLafferty, F.W. (2000). "Electron capture dissociation for structural characterization of multiply charged protein cations." Anal Chem **72**: 563-573.

Appendices

A.1. The BioLogic QFM-400 instrument for quenched flow experiments.

Chemical quenched flow experiments involve fast mixing of two reactants followed by quenching of the reaction with a chemical agent. These experiments are facilitated by quenched flow instruments in which drive motors are used to force the reactants into a mixer. The mixed reactants pass into a reaction delay line where the reaction is allowed to take place for a desired time period. The duration of the reaction is determined by the volume of the reaction delay line and the flow rate of the reaction mixture through the delay line. In practice, the reaction time can be varied by changing the delay line and/or the flow rate through the delay line. The reaction mixture then passes through another mixer where the quench-solution is added to terminate the reaction. The reaction product is collected by purging the contents through the exit purge port via a syringe.

The BioLogic QFM-400 instrument consists of four syringes (S1, S2, S3 and S4) which are driven by four motors (Figure A.1).

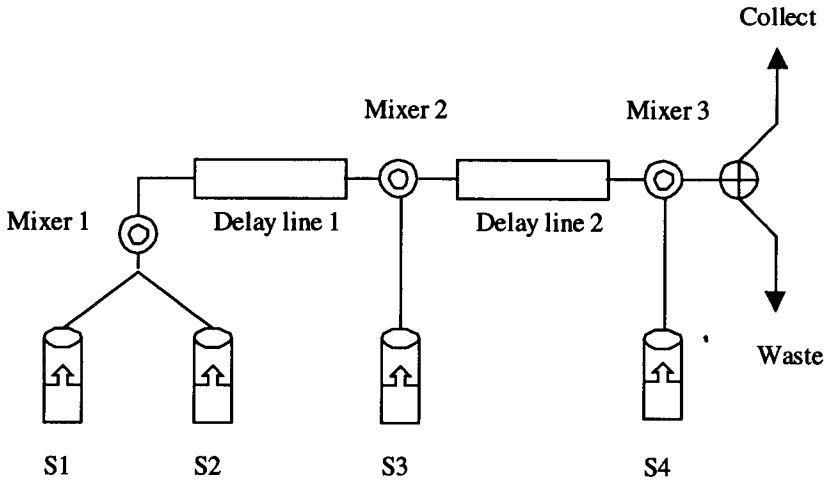


Figure A.1. The BioLogic 4-syringe mixer in the quenched flow mode.

Because the drive rates of the motors are independent of each other, different mixing ratios can be obtained by selecting syringes of different sizes. The BioLogic QFM-400 instrument contains three mixers and two delay lines that allow maximally three reactions to take place, including a final quenching reaction. At mixers 1 and 2, the reaction mixture is allowed to proceed in delay line1 and delay line 2. Volumes of the delay lines are fixed by the delay line discs (17, 40, 90, 190 ul) which can be placed at mixer 1 and 2 in any combination. However, the actual delay line volumes include the volumes of the lines connecting the discs. Therefore, the actual delay line volume for the 17 ul disc is 43 ul at mixer 1 and 53 ul at mixer 2.

The actual delay line volume for each delay line disc at each mixer position can be obtained from the MPS32 software set-up window or by calculation as described in the manual. The reaction time at each mixer is calculated by dividing the actual delay line volume at the mixer by the total speed, which is the sum of the speeds of all preceding syringes. The flow rate at each syringe can be adjusted individually from 0 ml/s to 5 ml/s by choosing the syringe of an appropriate size. The 20-ml syringe allows a syringe speed of 0-6 ml/sec (flow rates above 5 ml/sec require a preceding ramping phase).

The BioLogic QFM-400 module is operated by the MPS32 software program, which can perform automated movement of the syringes and create a drive sequence. Writing a drive sequence is central to attaining success with the BioLogic instrument. When creating a drive sequence, it is important to realize that only a portion of the reaction mixture is collected at the exit port. Therefore, the entire system must be cleaned prior to initiating a reaction by washing with a buffer three times the volume of the delay line volume at mixer 1. For a disc of 190 ul at mixer 1, approximately 600 ul of buffer is necessary to ensure that the reaction chamber is free of contamination from the previous experiment. The MPS32 allows for up to 20 phases to be included in a drive sequence. The first phase in a drive sequence often represents a washing phase. There are several important

considerations for constructing a drive sequence. First, the total speed through each mixer must be greater than 1 ml/sec and less than 12 ml/sec and must remain the same from one phase to another. Second, the reaction time must remain the same from one phase to another. To construct a drive sequence that allows for a wide range of mixing times, the easiest way is to write a sequence that allows the shortest reaction time in the continuous mode and lengthen the reaction time by incorporating the interrupt mode. To ensure that the drive sequence is correctly constructed, the hydrolysis of DNAP can be used to verify proper mixing at each mixer in the sequence as described in the manual.

A drive sequence for deuterium/hydrogen (D/H) exchange pulsed labeling is provided as an example (Chapter III experiments). In this experiment, an unfolded, fully deuterated protein is refolded in a deuterated buffer over variable time periods (11X dilution), then subjected to a 10 ms labeling pulse with a protiated solvent (2X dilution), and finally quenched by 1 M HCl. In the QFM-400 instrument, S1 contains the deuterated, refolding buffer, S2 contains the fully deuterated and unfolded protein solution, S3 contains the labeling buffer, and S4 contains 1 M HCl. The protein folding reaction takes place at mixer 1; the D/H labeling reaction takes place at mixer 2; and the quenching reaction occurs in the exit line.

In order to achieve the 11X dilution at mixing stage 1, a 5 ml syringe must be installed in S2 with syringes of 20 ml installed in S1, S3 and S4. Because the addition of an interrupt phase immediately after the mixing phase can extend the reaction time at a mixer, a drive sequence is written to achieve the smallest reaction time. In this case, the 17 ul disc is used at delay line 1 and a 40 ul disc is used at delay line 2. Thus, the actual reaction volumes at delay lines 1 and 2 are 43 ul and 60 ul, respectively. The first phase is the wash phase, which requires 3X the volume of delay line 1 (Table A.1)

Table A.1. A driving sequence for a deuterium pulsed labeling experiment.

Phase	1	2	3	4	5	6
Duration (ms)	20	20	Variable times	6	5	10
S1 (ul)	120	110		36		60
S2 (ul)		10				
S3 (ul)				36		60
S4 (ul)						48
Valve	waste	waste		waste	waste	collect

Therefore, the volume in S1, containing the less precious between the deuterated solvent in S1 and the deuterated protein in S2, is 120 ul and the

duration is 20 ms providing a speed of 6 ml/s at mixer 1. In phase 2, S1 (110 ul) and S2 (10 ul) are mixed for a time period of 21 ms, which provides a total speed of 6 ml/s at mixer 1. Thus, the actual reaction time is 7.2 ms (43ul/(6ml/s)). For reaction times longer than 7.2 ms, a third phase extending the reaction time can be added. In phase 4, the refolded protein mixture (product formed at mixer 1) is reacted with the labeling pulse in S3 in a 1/1 ratio. Delay line 2 has an actual volume of 60 ul. As a result of diffusion, 1/3 of the solution at the front-end experiences improper mixing and must be removed. Consequently, the duration time is 6 ms, S1 pushes 36 ul yielding 6 ml/s at mixer 1, S3 pushes 36 ul yielding 6 ml/s at mixer 2. The actual delay line volume is 60 ul yielding an actual reaction time of 5 ms (60ul/(12ml/s)). The designated reaction time is 10 s, therefore a duration of 5 ms can be added in phase 5. The pulsed labeling reaction of the refolded protein is quenched in phase 6 by adding 1 M HCl in phase 6. The duration time is 10 ms, S1 pushes 60 ul to maintain 6 ml/s at mixer 1, S3 pushes 60 ul to maintain 6 ml/s at mixer 2, and S4 pushes 48 ul to achieve a desirable final pH of 2.5.

A.2. Sequential disulfide bond removal

The sequential disulfide bond removal approach makes use of the partial reduction of a protein under controlled conditions and produces a protein in which reduction has opened some, but not all disulfide bridges. Every disulfide bond has the same redox potential and should be reduced to a similar extent based on thermodynamic considerations. However, in the solution of a native protein the reduction kinetics of various disulfide bonds are quite different depending on their accessibility to the reducing agents and on the conformational stability of the disulfide bonds. Theoretically, it is possible to achieve sequential reduction of disulfide bonds by controlling the reduction conditions.

Dimeric rhm-CSFβ is linked by two four-helix bundles, forming an extremely flat, elongated structure (Figure 4). There are three intra-subunit disulfide bonds in each monomer including: Cys7-Cys90, Cys48-Cys139, Cys102-Cys146. All intra-subunit disulfide bonds are at the end distal to the dimer interface. One inter-subunit disulfide bond, Cys31-Cys'31, is located at the dimer interface. The total surface buried by dimerization is ~850Å^2 from each monomer which constitutes about 9% of the total exposed area of a monomer. Two inter-subunit disulfide bonds linking Cys157/159 and Cys'157/159 are located in the C-terminal regions. It is

unclear whether Cys157 forms a disulfide bond with Cys'157 or Cys'159. Nonetheless, it is evident that these cysteine residues form symmetrical disulfide bonds that are in close proximity in the three-dimensional space.

The sequential disulfide removal strategy has been applied to study the reductive unfolding of rhm-CSFβ. The objective was to remove disulfide bonds pair by pair in order to isolate various homogeneous, partially reduced intermediates whose structures and stability could be characterized. The partial reduction condition did not include any denaturant. Reduction was carried out with TCEP at pH 3.5 where disulfide bond scrambling was minimal. The reduced cysteine residues were irreversibly alkylated by CDAP at pH 3.5. The mixture contained residual intact protein and partially reduced, cyanylated isomers that were separated by reversed phase HPLC. The purity of the isolated fraction was determined by ESI-MS.

The mixture of partially reduced rhm-CSFβ often contained intermediates with different reduction states. Therefore, experimental parameters were varied to (1) increase the yield of the predominant product in a given mixture (2) and to produce different reduction products (Figure A.2).

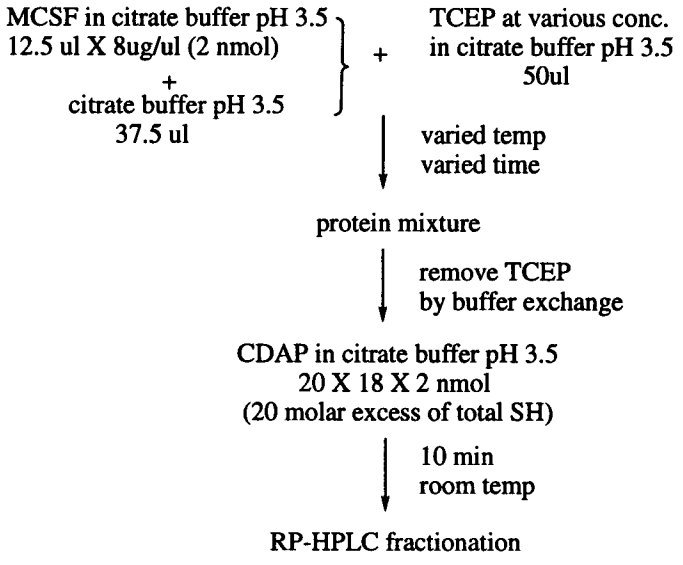


Figure A.2. Scheme for partial reduction of rhm-CSFβ.

The results obtained from these experiments are summarized in Table A.2. Under strong reducing conditions, i.e. under high TCEP/SH stoichiometry, extended reaction times, and/or high reaction temperature, more disulfide bonds were reduced and more cysteine residues alkylated. Monomeric proteins were produced in some cases. Most protein species were not homogeneous and contained various disulfide bridges.

Table A.2. Individual parameters adjusted in the partial reduction of disulfide bonds in rhm-CSFβ. Bolded molecular weights represent the predominant species in the HPLC fractionations.

TCEP/SH (mol/mol)	Rxn time (hr)	Temp (°C)	MW (Da)	# CN
50X18 / 1	2	28	24,628, **49,030**, 49,168	4.2, 0, 5.52
50X18 / 1	3	28	24,627, **49,030**, 49,217	4.2, 0, 7.48
50X18 / 1	4	28	**24,629**, 49,030	4.2, 0
50X18 / 1	5	28	**24,627**, 49,030	4.2, 0
50X18 / 1	6	28	**24,734**, 49,030	8.5
5X18 / 1	1	28	**49,030**, 49,107	0, 3.08
5X18 / 1	3	28	**49,030**, 49,145	0, 4.6
5X18 / 1	5	28	49,030, **49,136**	0, 4.24
5X18 / 1	7	28	49,030, **49,148**	0, 4.72
5X18 / 1	10	28	24,695	6.84
1X18 / 1	1	28	**49,030**, 49,146	0, 4.64
1X18 / 1	3	28	**49,030**, 49,146	0, 4.64
1X18 / 1	5	28	**49,030**, 49,148, 49,168	0, 4.65, 5.52
1X18 / 1	7	28	49,030, **49,156**, 49,235	0, 5.04, 8.2
1X18 / 1	10	28	**24,695**, 24,730, 49,030	6.84, 8.24, 0
0.5X18 / 1	1	28	**49,030**, 49,150	0, 4.8
0.5X18 / 1	2	28	**49,030**, 49,146, 49,156	0, 4.64, 5.04
0.5X18 / 1	3	28	49,030, **49,146**, 49,151	0, 4.64, 4.84
0.5X18 / 1	4	28	49,030, **49,152**	0, 4.88
0.5X18 / 1	5	28	49,030, 49,136, **49,140**	0, 4.24, 4.4
0.5X18 / 1	7	28	49,030, **49,170**	0, 5.6
0.5X18 / 1	9	28	**24,696**, 49,030	6.88, 0
0.05X18 / 1	1	28	**49,030**, 49,070	0, 1.6
0.05X18 / 1	3	28	49,030, **49,089**, 49,131	0, 2.36, 4.04
0.05X18 / 1	5	28	49,030, **49,131**	0, 4.04
0.05X18 / 1	7	28	49,030, **49,130**, 49,156	0, 4, 5.04
0.05X18 / 1	10	28	**24,695**, 49,030, 49,241	6.84, 0, 8.44
0.05X18 / 1	5	35	49,030, **49,131**	0, 4.04
0.05X18 / 1	5	45	49,030, **49,130**, 49,156	0, 4, 5.04
0.05X18 / 1	5	55	49,030, **49,130**, 49,156	0, 4, 5.04
0.05X18 / 1	5	65	**24,695**, 49,030, 49,156	6.84, 0, 5.04
0.05X18 / 1	5	72	**24,695**, 49,030	6.84, 0