Conformational Analysis of Intermediates Involved in the in Vitro Folding Pathways of Recombinant Human Macrophage Colony Stimulating Factor β by Sulfhydryl Group Trapping and Hydrogen/Deuterium Pulsed Labeling[†]

Y. Heidi Zhang,[‡] Xuguang Yan,[‡] Claudia S. Maier,[‡,§] Michael I. Schimerlik,[||] and Max L. Deinzer*[,‡]

Departments of Chemistry and Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97330

Received May 28, 2002; Revised Manuscript Received October 28, 2002

ABSTRACT: In vitro oxidative folding of reduced recombinant human macrophage colony stimulating factor β (rhm-CSFβ) involves two major events: disulfide isomerization in the monomeric intermediates and disulfide-mediated dimerization. Kinetic analysis of rhm-CSFβ folding indicated that monomer isomerization is slower than dimerization and is, in fact, the rate-determining step. A time-dependent determination of the number of free cysteines remaining was made after refolding commence. The folding intermediates revealed that rhm-CSFβ folds systematically, forming disulfide bonds via multiple pathways. Mass spectrometric evidence indicates that native as well as non-native intrasubunit disulfide bonds form in monomeric intermediates. Initial dimerization is assumed to involve formation of disulfide bonds, Cys 157/159−Cys′ 157/159. Among six intrasubunit disulfide bonds, Cys 48−Cys 139 and Cys′ 48−Cys′ 139 are assumed to be the last to form, while Cys 31−Cys′ 31 is the last intersubunit disulfide bond that forms. Conformational properties of the folding intermediates were probed by H/D exchange pulsed labeling, which showed the coexistence of noncompact dimeric and monomeric species at early stages of folding. As renaturation progresses, the noncompact dimer undergoes significant structural rearrangement, forming a native-like dimer while the monomer maintains a noncompact conformation.

Protein folding is an important process for understanding structure−function relationships. An understanding of the fundamental principles of protein folding can lead to high-level production of biologically active proteins of therapeutic and diagnostic importance (*1*, *2*). A general solution to the protein folding problem requires the delineation of folding pathways, which can be determined 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 intermediates along the folding pathways (*3*−*6*).

The protein folding process is highly complex; it involves various noncovalent interactions such as hydrogen bonding and covalent interactions such as disulfide formation. In proteins containing disulfide bonds, folding pathways can be elucidated by examining the extent of disulfide bond formation at various stages of folding using a sulfhydryl group trapping procedure in which cysteine residues not involved in disulfide bonds are alkylated (*7*, *8*). By analyzing the pattern of modified and nonmodified cysteine residues in folding intermediates, detailed folding pathways have been defined for several disulfide containing proteins, including bovine pancreatic trypsin inhibitor (BPTI) (*9*−*11*), tick anticoagulant peptide (TAP) (*12*, *13*), bovine pancreatic ribonuclease A (RNase A) (*14*−*19*), α-lactalbumin (*20*−*22*), and hen egg white lysozyme (HEWL) (*23*, *24*). These studies suggest that protein folding mechanisms are highly complex and may involve non-native disulfide isomers as intermediates.

Folding can also be studied by monitoring the formation of hydrogen bonds that occur as unstructured polypeptides attain compact conformations. It is difficult to detect the formation and breakage of all hydrogen bonds during protein folding (*25*); however, it is possible to probe the solvent accessibility of folding intermediates by deuterium labeling using amide hydrogens (NHs) as reporter groups (*26*−*28*). Amide H/D[1] exchange pulsed labeling utilizes deuterium exchange to label NHs on the polypeptide backbone that are unprotected at defined folding times (*29*). The level of deuterium trapped at different folding times can be determined by nuclear magnetic resonance (NMR) (*30*, *31*) and/

[†] This work was supported by NIEHS Grants ES 00040 and ES 00210 to M.L.D. and M.I.S.

* Corresponding author. E-mail: max.deinzer@orst.edu. Fax: (541) 737-0497. Phone: (541) 737-1773.

[‡] Department of Chemistry, Oregon State University.

[§] Present address: Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research, and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CA Utrecht, The Netherlands.

[||] Department of Biochemistry and Biophysics, Oregon State University.

[1] Abbreviations: rhm-CSFβ, recombinant human macrophage colony stimulating factor β; H/D, hydrogen/deuterium exchange; CID-MS-MS, collision-induced dissociation tandem mass spectrometry; NMR, nuclear magnetic resonance; *E. coli*, *Escherichia coli*; cDNA, complementary DNA; RP-HPLC, reversed-phase high-pressure liquid chromatography; SEC, size exclusion chromatography; CN, cyanylation; kDa, kilodaltons; aa, amino acid, ACN, acetonitrile; TFA, trifluoroacetic acid.

10.1021/bi0203841 CCC: $22.00
Published on Web 12/05/2002

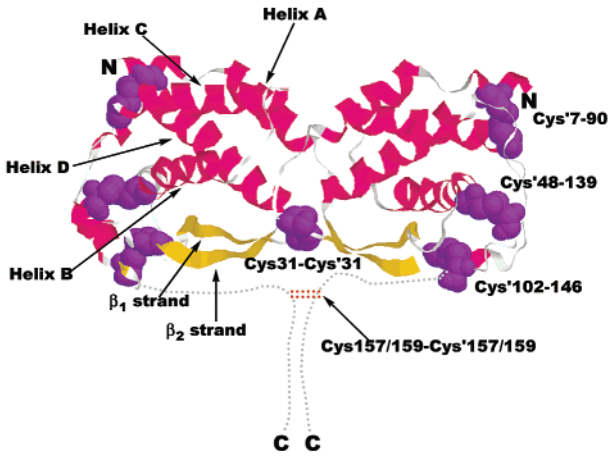

FIGURE 1: Schematic representation of rhm-CSFβ based on the X-ray crystal structure of m-CSFα (*38*). The C-terminal regions (aa 154−221) and the Cys 157/159−Cys′ 157/159 disulfide bonds are shown as dotted lines.

or electrospray ionization mass spectrometry (ESI-MS) (*32*−*34*). This approach can reveal the extent of solvent exposure of the protein backbone at various folding times, allowing the determination of the conformational changes during folding. In addition, transient folding intermediates that are undetected by other conventional spectroscopic methods may be identified by H/D pulsed labeling (*35*, *36*).

Recombinant human macrophage colony stimulating factor β (rhm-CSFβ) stimulates the proliferation, differentiation, and survival of cells belonging to the monocyte−macrophage lineage (*37*). Mature rhm-CSFβ is a homodimer with three intersubunit disulfide bonds (Cys 31−Cys′ 31, Cys 157/159−Cys′ 157/159) and three intramolecular disulfide bonds (Cys 48−139, Cys 7−90, Cys 102−146) (Figure 1) (*38*). It 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 (*39*). However, analysis of the folding pathway of the nine-disulfide protein presents a daunting technical challenge because more than 9×10^9 disulfide-bonded species are potential folding intermediates. Even if only a fraction of these isomers exist along the pathways, their separation, identification, and quantitation would be extremely difficult.

In the present study, the folding events of rhm-CSFβ were elucidated by monitoring the folding kinetics using size exclusion chromatography (SEC). The number of disulfides was deduced by alkylating the free cysteines on the basis of a corresponding experimentally observed mass shift. The folding intermediates were characterized by reversed-phase high-pressure liquid chromatography (RP-HPLC) and electrospray ionization mass spectrometry (ESI-MS). In addition, H/D pulsed labeling was performed to examine the extent of solvent accessibility of the folding intermediates and to monitor the conformational changes that took place during folding.

MATERIALS AND METHODS

Materials. Reduced and oxidized glutathiones were obtained from Sigma Chemical Co. Iodoacetamide was purchased from Fluka. Prepacked 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.

Preparation of Unfolded Protein. Denaturation and disulfide bond reduction of rhm-CSFβ was achieved by diluting the native 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 h.

In Vitro Oxidative Folding. Folding was initiated by rapid buffer exchange of the denatured, reduced protein in the folding buffer which contained 50 mM Tris (pH 8.5), 0.4 M urea, 5 mM GSH, and 0.64 mM GSSG using a prepacked Sephadex column. The 8-fold difference between reduced and oxidized glutathione was selected so that folding proceeded at a reasonable rate. A low concentration of urea was included to prevent protein aggregation.

Folding Monitored by Size Exclusion Chromatography. Folding was performed with different protein concentrations: 0.12, 0.34, and 0.64 mg/mL. Aliquots were removed at folding times ranging from 2 to 1404 min and analyzed by SEC−HPLC on a Ultrahydrogel-500 column equilibrated in H_2O/50 mM Tris and 5 mM EDTA (pH 7) (4:1) employing a Shimadzu HPLC system. Peak areas for the monomeric and dimeric species were obtained from the UV 280 nm trace and were fitted to a half-Gaussian modified Gaussian function using PeakFit (Jandel Scientific). Four data sets at three different protein concentrations were fit globally using the computer program Scientist (MicroMath, Inc.). Equations used for global analysis of models A−C were as follows:

model A

$$D(t) = \left(\frac{k_1 k_2 M_0}{2(k_2 - k_1)}\right)\left[\left(\frac{1}{k_2}\right)(e^{-k_2 t} - 1) + \left(\frac{1}{k_1}\right)(1 - e^{-k_1 t})\right] \quad (1)$$

models B and C

$$D(t) = D_0 + \frac{M_0}{2}\left\{\left(\frac{k_2 k_4}{k_4 - k_1 - k_2} + \frac{k_1 k_3}{k_3 - k_1 - k_2}\right) \times \left(\frac{1}{k_1 + k_2}\right)(1 - e^{-(k_1 + k_2)t}) + \left(\frac{k_2}{k_4 - k_1 - k_2}\right)(e^{-k_4 t} - 1) + \left(\frac{k_1}{k_3 - k_1 - k_2}\right)(e^{-k_3 t} - 1)\right\} \quad (2)$$

where D_0 is a small amount of native dimer present at time zero and M_0 is the concentration of unfolded monomer at time zero.

Folding Monitored by Sulfhydryl Group Trapping. The protein concentration in the folding reaction was 0.3 mg/mL. At various times ranging from 10 s to 7200 min, folding was quenched by adding 25 mM iodoacetamide to the reaction aliquots. Carboxamidomethylated 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% trifluoroacetic acid (TFA) as solvent A and acetonitrile (ACN) containing 0.05% TFA as solvent B. The proteins were eluted using an 8 min gradient from 10% to 90% solvent B. MS

spectra were analyzed by PE Sciex software (MacSpec and BioSpec).

Isolation of Folding Intermediates. Three folding times, 1, 120, and 4800 min, were selected for detailed analysis. The alkylated protein mixture (0.5 mg, 10 nmol) isolated for each time point was analyzed by RP-HPLC on a Vydac C4 TP214 column (300 Å, 5 μm, 4.6 $\times$ 250 mm, 1 mL/min flow rate) employing a binary linear gradient with 0.1% TFA as solvent A and ACN containing 0.09% TFA as solvent B. Proteins were eluted using a 60 min 44−54% solvent B gradient. The shallow gradient was developed to improve peak separation of folding intermediates, which exhibited similar chromatographic properties. Protein fractions were collected manually and dried in a Speed-Vac (Savant).

Structural Characterization of Folding Intermediates. Fractions collected from freeze-drying were resuspended in 100 mM phosphate buffer (pH 2.5). All fractions were then subjected to the same digestion procedures to ensure consistency in sample preparation. Each fraction was added to 8 M urea and 1 M TCEP in order to reduce the existing disulfide bonds. Denaturation and disulfide bond reduction were allowed to proceed for 30 min at room temperature to ensure complete digestion in the subsequent step. Each fraction was then 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 10−60% over a 90 min period. The peptic peptides were identified by CID-MS-MS.

In a separate experiment, D-SS9-cam0 isolated from HPLC analysis was digested by immobolized pepsin (protein/enzyme = 1:2) for 10 min at room temperature in the absence of urea and TCEP. The digest contained disulfide-bonded peptides that were analyzed by RP-HPLC and CID-MS-MS. Simultaneously, a control sample of the native rhm-CSFβ was also digested by pepsin without reduction of disulfides.

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 7200 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 fully deuterated protein (100% reference) was prepared by incubating rhm-CSFβ in 8 M urea-d_4/D_2O at 37 °C for 6 h.

Electrospray Ionization Mass Spectrometry of Deuterated Samples. The samples from pulsed labeling were 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% solvent B gradient. To prevent deuterium loss during analysis, the HPLC injector and column were immersed in ice/water slurry, and the transfer syringe was precooled on ice before use. MS spectra were analyzed by BioSpec software.

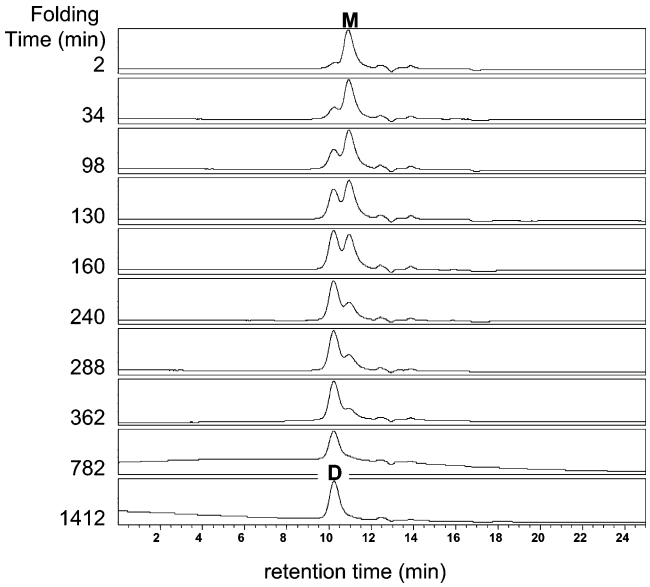

FIGURE 2: SEC−HPLC chromatograms of intermediates of rhm-CSFβ folding performed in the presence of GSH/GSSG (7.8/1) at 22 °C. M refers to monomeric species that eluted at 10.9 min, and D refers to dimeric species that eluted at 10.2 min.

Data Analysis. Because the HPLC elution was performed with protonated solvents, a small number of deuteriums at the amide positions are lost. Corrections were made for this deuterium loss according to the equation (*40*):

$$D = [(m - m_{0\%})/(m_{100\%} - m_{0\%})]N \qquad (3)$$

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

RESULTS

Kinetic Studies of in Vitro Oxidative Folding. Under SEC−HPLC conditions, monomeric species eluted with a retention time of 10.9 min while dimeric species eluted after 10.2 min (Figure 2). Given the complexity of the folding system, the monomeric and dimeric species probably represented ensembles of proteins rather than individual species. The kinetic data (Figure 3) could be adequately fit to several different models that described two major folding events: disulfide isomerization in the monomeric intermediates and disulfide-mediated dimerization.

The simplest mechanism that fit the data is shown (Figure 4A), where M, M_1, and M_2 are monomeric species and D represents dimeric species. The values of parameters k_1 (5.2 $\times$ 10^{-3} min^{-1}) and k_2 (2.1 $\times$ 10^{-2} min^{-1}) can be reversed without affecting the fit; however, the monomer isomerization step must be slow compared to dimerization (k_3). It is

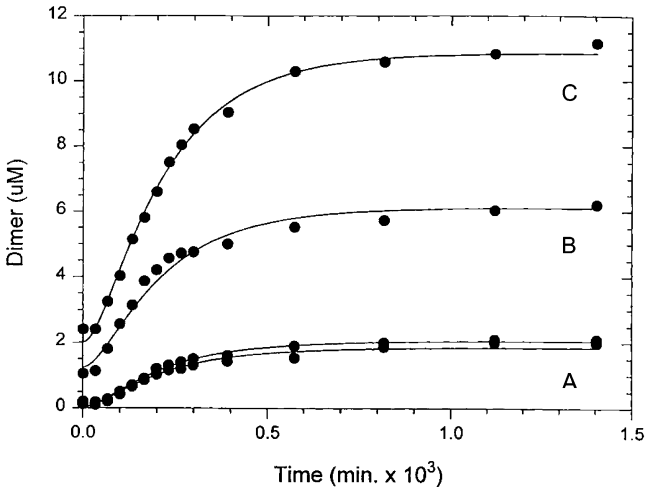

FIGURE 3: Kinetic curves showing the formation of dimeric species as a function of folding times in the presence of GSH/GSSG (7.8/1) at 22 °C at three different protein concentrations: (A) 0.12 mg/mL, (B) 0.34 mg/mL, and (C) 0.64 mg/mL.

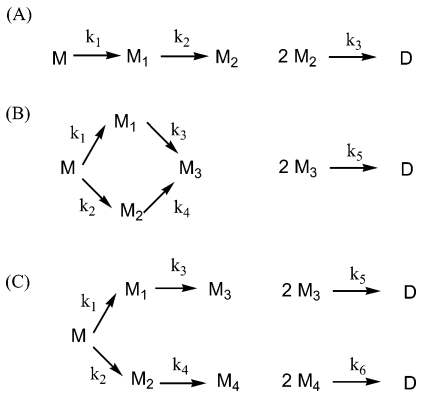

FIGURE 4: Kinetic models of rhm-CSFβ folding. M, M_1, M_2, M_3, and M_4 are monomeric species, and D represents dimeric species. Kinetic equations were derived assuming that dimerization is fast compared to kinetics steps in monomer interconversion; thus, M_2, M_3, and M_4 were assumed to be in a steady state, i.e., $d[M]/dt \approx 0$. (A) k_1 is $(5.2 \pm 0.39) \times 10^{-3}$ min^{-1} and k_2 is $(2.1 \pm 0.04) \times 10^{-2}$ min^{-1}. (B) k_1 is 0.020 ± 0.025 min^{-1}, k_2 is 0.002 ± 0.017 min^{-1}, k_3 is 0.0055 ± 0.0034 min^{-1}, and k_4 is 0.002 ± 0.0079 min^{-1}. (C) This model gives rise to the same equation as model B with the same values for the fitted parameters k_1 through k_4.

important to note that the model with more complicated pathways (Figure 4B,C) also gave satisfactory results. The inclusion of additional parameters, however, did not improve the fit in the more complicated models.

Although it was not possible to resolve individual monomeric and dimeric species by SEC−HPLC, the kinetic results (Figure 3) did require a model involving the formation of at least two monomeric intermediates whose interconversion was slow, prior to dimer formation. The kinetic data gave essentially the same results for at least three models (Figure 4) although the fitting values were not as well defined for the more complicated ones. The association of the rate-limiting step with monomeric species and not with dimeric species indicates that disulfide bond shuffling in the monomeric intermediates and proper disulfide formation are probably the important steps for productive folding of rhm-CSFβ.

In Vitro Folding Monitored by Sulfhydryl Group Trapping. The mass spectra recorded for rhm-CSFβ folding intermediates showed charge state distributions that shifted from one centered on highly charged ion peaks (low *m*/*z* values) to one centered on peaks with a lower number of charges (high *m*/*z* values) (Figure 5). Charge state distributions in mass spectra reflect the extent of protein protonation during electrospray ionization and therefore can reveal protein conformation in solution. A compactly ("compact" here refers to proteins with disulfide bonds and not necessarily a folded native state) folded protein has only limited numbers of basic sites exposed to the solvent and yields high *m*/*z* charge state distributions. In contrast, a loosely packed protein often has more basic sites available for protonation, resulting in low *m*/*z* charge state distributions.

A charge state envelope centered on peaks with low *m*/*z* values represented monomeric proteins (Figure 5, 1 min) while a distribution centered on high *m*/*z* values was found for compact dimeric proteins (Figure 5, 7200 min). During this transition, there were charge state distributions centered on medium *m*/*z* values (Figure 5, 120 min). These distributions corresponded to dimeric species that had noncompact structures in which there were more basic sites available for protonation during electrospray ionization than in the compact dimer. The shift in charge state distributions represented a conversion from noncompact monomers to noncompact dimers and eventually to compact dimers. The overlapping charge state distributions indicated the coexistence of monomeric and dimeric species 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 6A). Deconvolution of the mass spectrum provided the molecular masses for these two proteins, 25034 and 24919 Da, corresponding to monomeric proteins with nine and seven alkylated cysteine residues, respectively (Figure 6B). After 300 min folding, the folding mixture became more complex, exhibiting multiple overlapping charge state envelopes (Figure 6C). Deconvolution identified these species to be monomeric and dimeric proteins of various degrees of carboxamidomethylation (Table 1). The oxidative folding of rhm-CSFβ approached completion after 7200 min. The mass spectrum showed a distribution of charge states centered on 21+, the same as in the mass spectrum of the native protein. When the charge states were deconvoluted, the presence of a small population of a monomer containing four disulfide bonds and one 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 molecular masses. The identities of each species in the folding mixture were assigned on the basis of their molecular masses obtained by deconvoluting the mass spectra and comparing them with calculated values (Table 1). Quantitative comparisons between the monomeric and dimeric proteins cannot be made simply on the basis of their relative peak intensities in the mass spectra. This is because it is not possible to correlate the intensity of ion signals of different species directly with their fractional population present in solution. Several factors significantly influence the ESI signal of different ions. First, instrumental conditions governing the electrospray ionization process

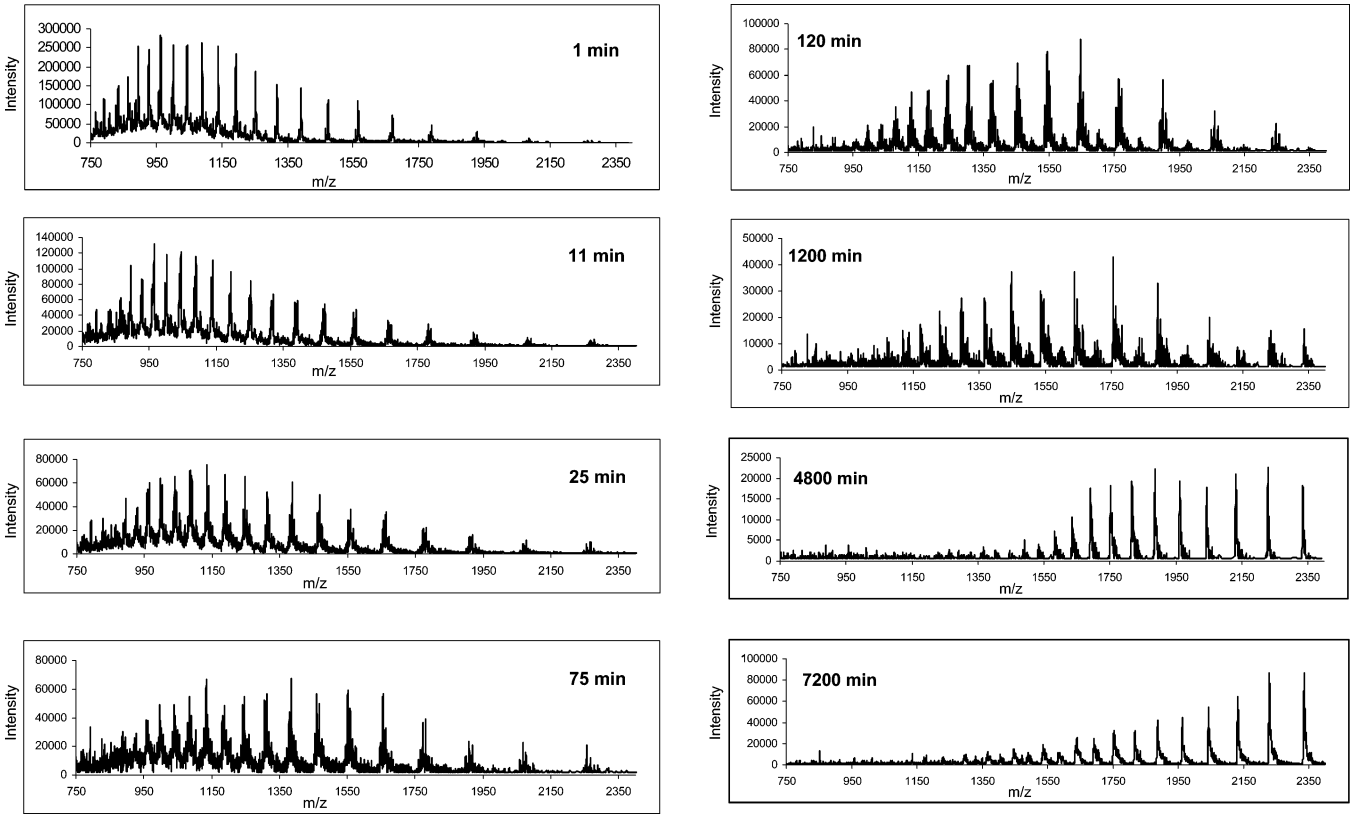

FIGURE 5: Carboxamidomethylated intermediates of rhm-CSFβ folding performed in the presence of GSH/GSSG (7.8/1) at 22 °C. The protein concentration was 0.3 mg/mL.

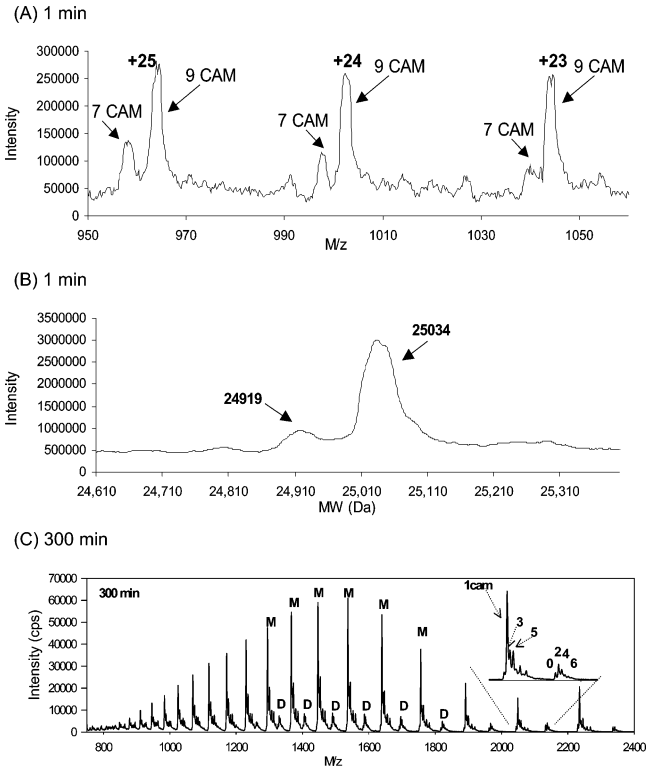

FIGURE 6: Sample mass spectra of folding intermediates. (A) Charge state distributions for two carboxamidomethylated monomers after 1 min folding. (B) Molecular masses obtained by deconvoluting the mass spectrum shown in (A). (C) Charge state distributions for a mixture of monomeric and dimeric proteins of various degrees of carboxamidomethylation after 300 min folding.

influence ion abundances. Second, there seems to be a positive discrimination for ion species with the highest affinity for the droplet surface; i.e., the ionization efficiency seems to be relatively higher for hydrophobic than for more hydrophilic species. Third, the transmission and detection efficiency for high mass ions carrying fewer charges are significantly reduced compared to the highly charged ions. Thus, monomeric and dimeric proteins exhibit different molar sensitivities in the mass spectrometer. Moreover, monomeric proteins are less structured and are ionized more efficiently than compact dimeric proteins.

Despite its lack of quantitative capability, mass spectrometry provided an important advantage to studying oxidative folding by allowing for simultaneous analysis of complex folding mixtures. Among monomeric or dimeric proteins, it was 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 on the basis of its relative intensities in the mass spectrum, 1 being the most populous and 4 being the least abundant (Table 1).

rhm-CSFβ folding was characterized by an increasing number of disulfide bonds and a decreasing number of alkylated cysteine residues in both monomeric and dimeric intermediates. At early folding times, the monomer with nine carboxamidomethylated 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 six carboxamidomethylated cysteinyl residues, D-SS6-cam6, 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 carboxami-

Table 1: Qualitative Analysis of Disulfide Bond Trapped Intermediates Found during Oxidative Folding of rhm-CSFβ in the Presence of GSH/GSSG (7.8/1) at 22 °C[a]

	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
RF time (min)									
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	1	2		
103			2	1	3	1	2	3	
120			3	1	2	3	1	2	4
180			3	1	2	3	1	2	4
305			3	2	1	3	2	1	4
480				2	1	4	2	1	3
970				2	1		2	1	3
1200				2	1		3	1	2
1279				2	1		2	1	3
1513				2	1		3	2	1
2565					1		3	2	1
4800					1			2	1
7200					1				1
mass (Da)									
measured	25045 ± 5	24929 ± 5	24810 ± 5	24697 ± 5	24578 ± 5	49377 ± 5	49268 ± 5	49144 ± 5	49036 ± 5
expected	25037 ± 5	24921 ± 5	24805 ± 5	24689 ± 5	24573 ± 5	49262 ± 5	49378 ± 5	49146 ± 5	49030 ± 5

[a] cam refers to carboxamidomethylation, and SS refers to disulfide bond.

domethylation and a higher degree of disulfide bond formation. Folding of rhm-CSFβ approached completion after 7200 min when the folding sample contained mainly D-SS9-cam0, a dimeric protein with nine disulfide bonds. In the same sample, a monomer containing four disulfide bonds and one carboxamidomethylated cysteinyl residue, M-SS4-cam1, was also detected. This species was probably trapped in a non-native disulfide-bonded conformation that was less capable of productive folding. As described above, it was not considered useful to try to determine the relative amounts of D-SS9-cam0 and M-SS4-cam1 present because of their different ionization behavior during ESI. However, it is estimated that M-SS4-cam1 is much less than 10% of the amount of D-SS9-cam0 present.

Isolation of Folding Intermediates. Under RP-HPLC conditions, the native dimer (D) eluted at 12.5 min and the fully alkylated monomeric protein (M) eluted at 13.4 min. The monomer was more hydrophobic than the dimer because the structure was lost upon denaturation and disulfide bond reduction. The measured molecular masses for the native dimer and the fully alkylated monomer were 49036 and 25034 Da, respectively. RP-HPLC analysis of the disulfide bond trapped folding samples indicated that the conversion of the denatured and reduced rhm-CSFβ to the compactly folded dimer was 87% complete within 7200 min.

The HPLC trace for the folding mixture obtained after 1 min showed a single peak at 13.4 min (Figure 7). Mass spectrometric analysis of the isolated protein indicated that the protein had an average molecular mass of 25038 Da and was a fully alkylated monomer, M-SS0-cam9 (Table 2). The presence of a monomeric protein containing seven 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 with

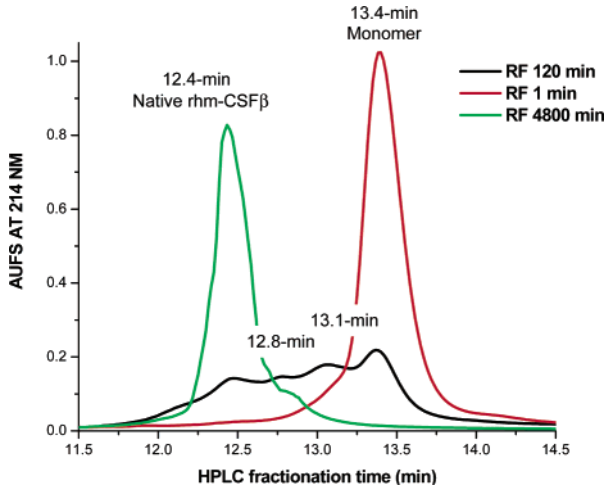

FIGURE 7: RP-HPLC chromatograms of disulfide bond trapped intermediates of rhm-CSFβ folding performed in the presence of GSH/GSSG (7.8/1) at 22 °C.

retention times of 12.4, 12.8, 13.1, and 13.4 min (Figure 7). All fractions were analyzed by mass spectrometry (Table 2). The 12.4 min fraction represented a single protein with an average molecular mass of 49030 Da, the same as D-SS9-cam0. The 12.8 min fraction contained two species with molecular masses of 49144 and 49377 Da, corresponding to dimeric proteins with two and six carboxamidomethylated cysteinyl residues, namely, D-SS8-cam2 and D-SS6-cam6. The 13.1 min fraction contained two monomeric species. One had molecular mass 24805 Da corresponding to a monomer with five carboxamidomethylated cysteinyl residues, M-SS2-cam5. The other had molecular mass 25045 Da representing a monomer with one glutathionylated and four carboxamidomethylated cysteinyl residues, M-SS2-cam4-SG1. The 13.4 min fraction contained a single protein with molecular mass 24692 Da that was consistent with a monomer with three carboxamidomethylated cysteinyl residues, M-SS3-cam3.

Table 2: Partial Structural Identification of Disulfide Bond Trapped Folding Intermediates by RP-HPLC and Peptide Mapping[a]

time (min)	HPLC pools (min)	mass (Da)	identification	Cys-cam	free Cys	Cys-SG
1	13.4	25038 ± 5	M-SS0-cam9	7, 31, 48, 90, 102, 139, 146, 157, 159	none	none
120	12.5	49030 ± 5	D-SS9-cam0	none	7, 31, 48, 90, 102, 139, 146, 157, 159	none
	12.8	49144 ± 5	D-SS8-cam2	31, 48, 139	7, 48, 90, 102, 139, 146, 157, 159	none
		49377 ± 5	D-SS6-cam6			
	13.1	25054 ± 5	M-SS2-cam4/GS1	31, 48, 146, 157, 159	7, 90, 102, 139	31
		24805 ± 5	M-SS2-cam5			
	13.4	24692 ± 5	M-SS3-cam3	31, 157, 159	7, 48, 90, 102, 139, 146	none
4800	12.4	49031 ± 5	D-SS9-cam0	none	31, 48, 90, 139, 146, 157, 159	none
	12.9	49146 ± 5	D-SS8-cam2	31	48, 90, 139, 146, 157, 159	none

[a] RF refers to refolding time, and cam refers to carboxamidomethylation.

The 12.8 and 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 could be achieved, due to similarities of the chromatographic properties of the coeluting proteins.

The mixture collected after 4800 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 molecular masses: 49031 Da corresponding to the native rhm-CSFβ and 49146 Da consistent with a dimeric protein with two carboxamidomethylated cysteinyl residues, D-SS8-cam2. The monomeric protein with four disulfide bonds and one carboxamidomethylated cysteinyl residue, M-SS4-cam1, was not identified in any HPLC fraction even though this protein was detected in the nonseparated mixture, but probably lost during collection.

Identification of Folding Intermediates by Peptide Mapping. To further characterize the structures of the folding intermediates isolated after 1, 120, and 4800 min folding, these proteins were analyzed by peptide mapping using pepsin proteolysis and 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 sulfhydryl group trapping were carboxamidomethylated, 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 nine cysteine residues were carboxamidomethylated in this monomer (Table 2), 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 resolved. Therefore, the peptic digests of each HPLC pool contained peptides derived from coeluted proteins (Table 2). 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 carboxamidomethylated, including Cys 31, Cys 48, and Cys 139. In addition, cysteine residues 7, 48, 90, 102, 139, 146, 157, and 159 were found to be free of alkylation. These nonmodified 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 nonmodified forms, indicating that some protein molecules contained an open Cys 48−Cys 139 disulfide bridge while others had the intact Cys 48−Cys 139 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 2). CID-MS-MS experiments of this fraction showed a highly complex mixture of alkylated and nonmodified cysteinyl residues. Carboxamidomethylated cysteinyl residues included Cys residues 31, 48, 146, 157, and 159. The same CID-MS-MS experiment also indicated that Cys residues 7, 90, 102, and 139 were free of alkylation. Cysteine 31 was also found to be glutathionylated. The two forms, M-SS2-cam4-SG1 and M-SS2-cam5, were eluted with the same HPLC retention time, and they both contained five modified cysteine residues (Cys 31, Cys 48, Cys 146, Cys 157, and Cys 159). It is likely that they might contain a similar disulfide structure, in which case Cys 31 was carboxamidomethylated in M-SS2-cam5 and Cys 31 was glutathionylated in M-SS2-cam4-SG1. Cysteine residues 48, 146, 157, and 159 were carboxamidomethylated in both monomers. These data verified that, in these monomeric proteins, Cys 31, Cys 157, and Cys 159, which participate in the intermolecular disulfide bonds in native rhm-CSFβ, existed as free thiols. Importantly, some cysteine residues that were involved in disulfide bonds at the time of sulfhydryl group trapping did not have the same disulfide partners as in the native protein. For example, in native rhm-CSFβ, Cys 102 forms a disulfide bridge with Cys 146, and Cys 139 forms a disulfide with Cys 48 in our folding experiments. However, Cys 48 and Cys 146 were modified with carboxamidomethylation while Cys 139 and Cys 102 were free of alkylation. These data showed 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 Cys 31, Cys 157, and Cys 159 residues were carboxamidomethylated.

After folding had continued for 4800 min, two HPLC pools were collected. One fraction contained D-SS9-cam0

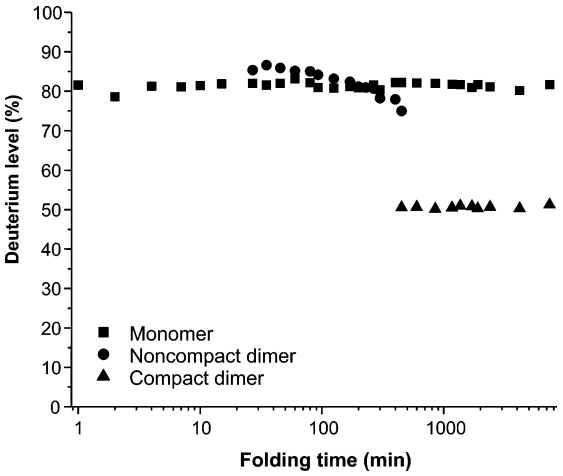

FIGURE 8: Deuterium levels in refolded rhm-CSFβ after H/D pulsed labeling. Folding was performed in the presence of GSH/GSSG (7.8/1) at 22 °C. Deuterium labeling was achieved by 10× dilution of refolded protein into the deuterated buffer for 10 s and was quenched by reducing the final pH to 2.5 and the temperature to 0 °C.

and the other contained D-SS8-cam2. Analysis of the peptic peptides of the first fraction indicated that all nine cysteine residues were free of alkylation, confirming 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 a molecular mass and HPLC retention time similar to those of a protein found in the 12.8 min fraction after 120 min folding. These may be 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 Cys 31 and Cys′ 31. Therefore, D-SS6-cam6 isolated from the same HPLC pool must contain unmodified Cys 31, Cys 157, Cys 159 and Cys′ 31, Cys′ 157, and Cys′ 159 (Table 2). The presence of D-SS8-cam2 in both 120 and 4800 min folding mixtures indicated that it might be a thermodynamically stable intermediate whose conversion to the native protein was slow. The conversion of D-SS8-cam2 to D-SS9-cam0 only required the formation of the last intermolecular disulfide bond Cys 31−Cys′ 31. This process was relatively slow perhaps because D-SS8-cam2 required major structural rearrangement to facilitate the formation of Cys 31−Cys′ 31.

Peptide analysis of the isolated D-SS9-cam0 and the native rhm-CSFβ indicated that the two proteins had the same disulfide bond structure. The proteins had essentially the same HPLC profile, which included several peptides containing disulfide linkages. The CID experiment also identified the same disulfide-linked peptides that were present in both peptic digests. Together, these data indicated that D-SS9-cam0 contained the native disulfide structure.

In Vitro Folding Monitored by H/D Pulsed Labeling. 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 noncompact overall conformation (Figure 8). After 40 min, a dimer appeared along with the monomer. The dimer also exhibited a high level of deuterium incorporated at ~80% of the available amide sites. The high levels of deuterium incorporation indicated a high degree of solvent exposure in the dimer; therefore, the dimer must also have a noncompact overall conformation. The monomer was present in the folding mixture, and the level of deuterium incorporation remained high at ~80% throughout the folding time course, indicating that deuterium accessibility in the monomer was not affected even after intramolecular disulfide bonds had formed. The deuterium level in the noncompact dimer decreased steadily from 80% to 75% between 40 and 400 min folding, suggesting minor conformational changes associated with early folding. After 400 min, deuterium incorporation in the noncompact dimer rapidly decreased from 75% to ~50% of all NHs within 50 min. The rapid reduction in deuterium level indicated that the noncompact dimer underwent a rapid structural contraction, which significantly limited the accessibility of NHs to solvent deuterium. The deuterium level in the compact dimer remained at ~50% throughout the remainder of the folding time course whereas deuterium incorporation for rhm-CSFβ in the native state was 46%. These results indicate that the native dimer structure was slightly more compact and less susceptible to H/D exchange than the compact dimer present in later stages of folding.

Oxidative folding of rhm-CSFβ monitored by trapping of unreacted sulfhydryl groups and H/D pulsed labeling were 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 pathway. As demonstrated in the sulfhydryl group trapping experiment, rhm-CSFβ folding was characterized by complex mixtures of monomeric and dimeric intermediates. Consequently, the monomer, the noncompact 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 sulfhydryl group trapping and H/D pulsed labeling studies. The monomer ensemble contained all monomeric species. The noncompact dimer ensemble included D-SS6-cam6, D-SS7-cam4, and D-SS8-cam2. The compact dimer ensemble consisted of mainly D-SS9-cam0. Therefore, the transition from the noncompact 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 7200 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.

DISCUSSION

Traditionally, protein folding was viewed 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 (*41*). Such a view does not preclude the concept that folding might still occur via preferred routes and defined intermediate states can still be envisioned (*42*, *43*).

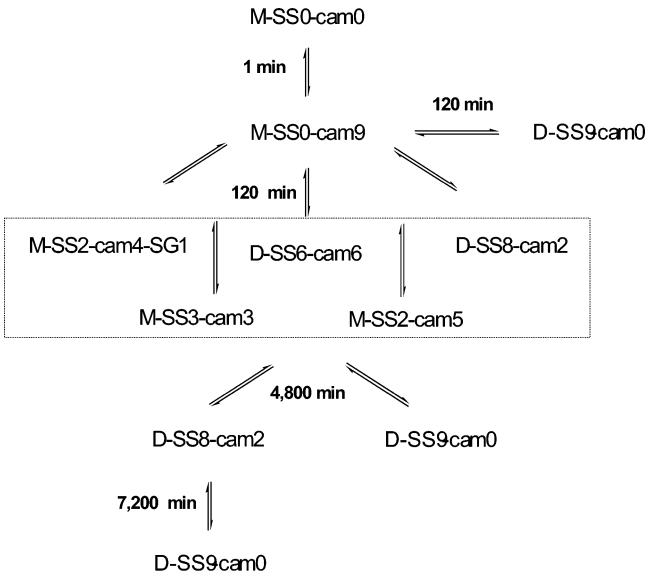

FIGURE 9: Heterogeneous mixtures of monomeric and dimeric intermediates after various folding times. M stands for monomer, D stands for dimer, SS refers to disulfide bond, and cam refers to carboxamidomethylation.

A previous folding study suggested that m-CSFβ folded according to a sequential pathway (*39*):

$$2M \rightarrow 2M' \rightarrow D' \rightarrow D \qquad (4)$$

Information taken from the present study provided a more detailed description for the folding events of rhm-CSFβ. Within the context of an overall folding process involving monomer disulfide bond isomerization and disulfide-mediated dimerization, 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 9). These data clearly depicted rhm-CSFβ folding as occurring via a complex energy surface, involving a rugged energy landscape with kinetic traps and multiple barriers.

It is interesting that a shorter form of m-CSFβ, m-CSFα (aa 4−149), a C-terminally truncated form in which the disulfide bonds, Cys 157/159−Cys′ 157/159, were missing, was also shown to fold via multiple pathways. A study of folding intermediates involved in the folding of m-CSFα reported the presence of both monomeric and dimeric intermediates (*44*). 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. Several similarities exist between the folding processes of m-CSFα and m-CSFβ. First, the folding of both proteins involved multiple pathways and followed the folding funnel model. Second, disulfide bond 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 between the two processes. m-CSFα folding 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 h at room temperature. In contrast, the folding of m-CSFβ approached completion after 120 h in the presence of GSSG at room temperature. When glutathione was excluded from the folding reaction, the folding rate slowed significantly, requiring at least 6 days of incubation to approach completion of folding. Clearly, the presence of four additional cysteine residues, Cys 157, Cys 159, 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 Cys 157/159−Cys′ 157/159 were shown to be unimportant for the biological activity or thermodynamic stability of rhm-CSFβ (*45*). Consequently, the roles of Cys 157 and Cys 159, probably as Cys 157/159−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 Cys 157/159 and Cys′ 157/159 formed early in folding, much sooner than Cys 31−Cys′ 31. Indeed, Glocker and co-workers showed by both oxidative refolding (*46*) and reductive unfolding (*46*, *47*) studies that certain sulfhydryls including Cys 157/Cys 159 are bridged in the unfolded state by melarsen oxide [*p*-[(4,5-diamino-1,3,5-triazin-2-yl)amino]phenylarsonous acid], which is a bifunctional trap for sulfhydryls in close proximity. Thus, the involvement of the disulfide pairs Cys 157/159−Cys′ 157/159 located in the C-termini where little structural organization exists in the initial stages may play an important role in subunit association. It is possible that they were more available for intermolecular interaction than Cys 31 and Cys′ 31, which could be made unavailable by the conformational restrictions presented by nearby secondary structural units formed early in folding.

ACKNOWLEDGMENT

We thank Dr. C. Cowgwill (Chiron Corp., Emeryville, CA) for providing rhm-CSFβ and the Mass Spectrometry and Nucleic Acids and Proteins core units or the Oregon State University Environmental Health Science Center for assistance.

REFERENCES

1. Jaenicke, R R. (1989) in *Protein structure, a practical approach* (Creighton, T. E., Ed.) IRL Press, New York.
2. Liu, Z. R. J., and Gierasch, L. M. (1998) *Bio-organic chemistry peptides and proteins* (Hecht, M., Ed.) Oxford University Press, New York.
3. Lu, H. S., Clogston, C. L., Merewether, L. A., Pearl, W. R., and Boone, T. C. (1992) *J. Biol. Chem. 267*, 8770−8777.
4. Onda, M. T., Takahashi, N., and Hirose, M. (1997) *J. Biol. Chem. 272*, 3973−3979.
5. Chang, J.-Y., Li, L., and Lai, P.-H. (2001) *J. Biol. Chem. 276*, 4845−4852.
6. Zhou, H.-X. (2001) *J. Am. Chem. Soc. 123*, 6730−6731.
7. Creighton, T. E. (1977) *J. Mol. Biol. 113*, 275−293.
8. Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol. 33*, 231−297.
9. Creighton, T. E. (1990) *Biochem. J. 270*, 1−16.
10. Weissman, J. S., and Kim, P. S. (1991) *Science 253*, 1386−1393.
11. Goldenberg, D. P. (1992) *Trends Biochem. Sci. 17*, 247−261.
12. Chang, J.-Y. (1996) *Biochemistry 35*, 11702−11709.
13. Chang, J.-Y., and Ballatore, A. (2000) *J. Protein Chem. 19*, 299−310.
14. Konishi, Y., Ooi, T., and Scheraga, H. A. (1982) *Biochemistry 21*, 4734−4740.
15. Konishi, Y., Ooi, T., and Scheraga, H. A. (1982) *Biochemistry 21*, 4741−4748.

16. Rothwarf, D. M., and Scheraga, H. A. (1993) *Biochemistry 32*, 2680−2689.
17. Rothwarf, D. M., and Scheraga, H. A. (1993) *Biochemistry 32*, 2690−2697.
18. Rothwarf, D. M., Li, Y.-J., and Scheraga, H. A. (1998) *Biochemistry 37*, 3760−3766.
19. Rothwarf, D. M., Li, Y.-J., and Scheraga, H. A. (1998) *Biochemistry 37*, 3767−3776.
20. Rao, K. R., and Brew, K. (1989) *Biochem. Biophys. Res. Commun. 163*, 1390−1396.
21. Ewbank, J. J., and Creighton, T. E. (1993) *Biochemistry 32*, 3677−3693.
22. Ewbank, J. J., and Creighton, T. E. (1993) *Biochemistry 32*, 3694−3707.
23. Chang, J.-Y., and Li, L. (2002) *Biochemistry 41*, 8405−8414.
24. Roux, P., Puoppolo, M., Chaffotte, A.-F., and Goldenberg, M. E. (1999) *Protein Sci. 8*, 2751−2760.
25. Creighton, T. E. (1992) *Protein Folding* (Creighton, T. E., Ed.) pp 301−315, W. H. Freeman, New York.
26. Udgaonker, J. B., and Baldwin, R. L. (1988) *Nature 335*, 694−699.
27. Roder, H., Elöve, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
28. Englander, S. W., and Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 243−265.
29. Clarke, A. R., and Waltho, J. P. (1997) *Curr. Opin. Biotechnol. 8*, 400−410.
30. Roder, H. (1989) *Methods Enzymol. 176*, 446−473.
31. Elöve, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
32. Katta, V., and Chait, B. T. (1991) *Rapid Commun. Mass Spectrum. 5*, 214−217.
33. Miranker, A., Robinson, C. V., Redford, S. E., Aplin, R. T., and Dobson, C. M. (1993) *Science 262*, 896−900.
34. Zhang, Z., and Smith, D. L. (1993) *Protein Sci. 2*, 522−531.
35. Jennings, P. A., and Wright, P. E. (1993) *Science 262*, 892−896.
36. Tsui, V., Garcia, C., Cavagnero, S., Siuzdak, H., Dyson, J., and Wright, P. E. (1999) *Protein Sci. 8*, 45−49.
37. Stanley, E. R., Einstein, D. B., Lee, P. S., and Yeung, Y. G. (1994) *Stem Cells 12*, 15−24.
38. Pandit, J., Jancarik, J., Halenbeck, R., Koths, K., and Kim, S. H. (1992) *Science 258*, 1358−1362.
39. Glocker, M. O., Arbogast, B., Schreus, J., and Deinzer, M. L. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5868−5872.
40. Zhang, Z., and Smith, D. L. (1993) *Protein Sci. 2*, 522−531.
41. Dill, K. A., and Chan, H. S. (1997) *Nat. Struct. Biol. 4*, 10−19.
42. Lazaridis, T., and Karplus, M. (1997) *Science 278*, 5345.
43. Dobson, C. M., and Karplus, M. (1998) *Angew. Chem., Int. Ed. Engl. 37*, 868−893.
44. Wilkins, J. A., Randahawa, Z. I., Wood, D., Warren, M. K., and Witkowska, H. E. (1993) *Protein Sci. 2*, 244−254.
45. Zhang, Y. H., Yan, X.-G, Maier, C. L., Schimerlik, M. I., and Deinzer, M. L. (2001) *Protein Sci. 10*, 2336−2345.
46. Happersberger, H. P. Stapleton, J., Cowgill, C., and Glocker, M. O. (1998) *Proteins: Struct., Funct., Genet., Suppl. 2*, 50−62.
47. Happersberger, H. P., Cowgill, C., and Glocker, M. O. (2002) *J. Chromatogr., B 78211−2*, 393−404.

BI0203841