Major Kinetic Traps for the Oxidative Folding of Leech Carboxypeptidase Inhibitor†

Silvia Salamanca,‡ Li Li,§ Josep Vendrell,*,‡ Francesc X. Aviles,‡ and Jui-Yoa Chang*,§

Institut de Biotecnologia i Biomedicina and Department de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain, and Research Center for Protein Chemistry, Institute of Molecular Medicine, The University of Texas, Houston, Texas 77030

Received February 24, 2003; Revised Manuscript Received April 18, 2003

ABSTRACT: The leech carboxypeptidase inhibitor (LCI) is a 66-amino acid protein, containing four disulfides that stabilize its structure. This polypeptide represents an excellent model for the study and understanding of the diversity of folding pathways in small, cysteine-rich proteins. The pathway of oxidative folding of LCI has been elucidated in this work, using structural and kinetic analysis of the folding intermediates trapped by acid quenching. Reduced and denatured LCI refolds through a rapid, sequential flow of one- and two-disulfide intermediates and reaches a rate-limiting step in which a mixture of three major three-disulfide species and a heterogeneous population of non-native four-disulfide (scrambled) isomers coexist. The three three-disulfide intermediates have been identified as major kinetic traps along the folding pathway of LCI, and their disulfide structures have been elucidated in this work. Two of them contain only native disulfide pairings, and one contains one native and two non-native disulfide bonds. The coexistence of three-disulfide kinetic traps adopting native disulfide bonds together with a significant proportion of fully oxidized scrambled isomers shows that the folding pathway of LCI features properties exhibited by both the bovine pancreatic trypsin inhibitor and hirudin, two diverse models with extreme folding characteristics. The results further demonstrate the large diversity of disulfide folding pathways.

A well-established method for analyzing the mechanism of protein folding is *oxidative folding* of disulfide-containing proteins (*1*−*9*). In this technique, proteins are initially fully reduced and denatured in the presence of a reducing agent and a denaturant. After removal of both reagents, the reduced and denatured protein is allowed to refold in the presence of a redox buffer (*6*). The disulfide folding pathway is then elucidated from the mechanism of formation of the native disulfide bonds, and characterized by the heterogeneity and the structures of disulfide isomers that accumulate along the process of oxidative folding. Folding intermediates can be trapped by several techniques. The most effective method is acid quenching of the protein solution, which allows the structural characterization of the trapped intermediates. This technique also permits the study of the influence of redox agents on the folding pathway and the efficiency of recovery of the native protein (*6*, *10*).

A considerable number of three-disulfide proteins have been studied by the technique of oxidative folding. Among them are the extensively reported BPTI[1] (bovine pancreatic trypsin inhibitor) (*1*, *2*, *7*, *8*), hirudin (*10*, *11*), potato carboxypeptidase inhibitor (*12*, *13*), thick anticoagulant peptide (*14*, *15*), epidermal growth factor (*16*, *17*), and insulin-like growth factor (*18*, *19*). However, few models aside from ribonuclease A (*20*−*23*) and α-lactalbumin (*24*−*27*) have been studied in detail among four-disulfide proteins. For these proteins, the increasing number of possible isomers (a total of 762, including the one-, two-, three-, and four-disulfide species) presents a formidable task of isolation and characterization of folding intermediates, even if only a fraction of these isomers exist along the pathway.

To date, no general consensus regarding the disulfide folding pathway has been reached from the proteins that have been investigated. The accumulated results have shown that even among small three-disulfide proteins, the folding mechanisms vary substantially. BPTI (*1*, *2*, *8*, *9*) and hirudin (*10*, *11*) represent two important cases of such divergent folding mechanisms. For BPTI, folding proceeds via a limited number of intermediates with native disulfide bonds and adopting native-like structures. Of 75 possible disulfide isomers, only five to six one- and two-disulfide intermediates were shown to populate along the folding pathway of BPTI, and all of them were found to adopt native disulfide bonds. In the case of hirudin, a highly heterogeneous population of one- and two-disulfide intermediates were identified along

† This work was supported by Grant BIO2001-2046 (MCYT, Ministerio de Ciencia y Tecnología, Spain) and by the Centre de Referencia en Biotecnologia (Generalitat de Catalunya, Spain). J.-Y.C. acknowledges the support by the endowment from the Robert Welch Foundation.

* To whom correspondence should be addressed. J.-Y.C.: Institute of Molecular Medicine, 2121 W. Holcombe Blvd., Houston, TX 77030; telephone, (713) 500-2458; e-mail, Rowen.Chang@uth.tmc.edu. J.V.: Institut de Biotecnologia i Biomedicina and Department de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain; telephone, 34-93-581-1315; e-mail, josep.vendrel@uab.es.

‡ Universitat Autònoma de Barcelona.

§ The University of Texas.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; GdmCl, guanidine hydrochloride; GSH, reduced glutathione; GSSG, oxidized glutathione; HPLC, high-performance liquid chromatography; LCI, leech carboxypeptidase inhibitor; IGF, insulin-like growth factor; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight.

10.1021/bi034308p CCC: $25.00
Published on Web 05/16/2003

the folding pathway. Most importantly, fully oxidized three-disulfide (scrambled) isomers, which have not been observed in the case of BPTI, were shown to serve as major folding intermediates of hirudin. For hirudin, folding proceeds through an initial stage of nonspecific disulfide formation (packing) which forms the scrambled isomers, which is followed by disulfide reshuffling (consolidation) of the heterogeneous scrambled population which leads to the native structure.

Therefore, further study of folding models from novel proteins is required to better understand the underlying causes that generate such a diversity of disulfide folding pathways. In this work, we report the pathway of oxidative folding of leech carboxypeptidase inhibitor (LCI), a 66-residue polypeptide rich in cysteines (*28*). The recently reported three-dimensional structure shows that LCI folds in a compact domain consisting of a five-stranded antiparallel β-sheet and a short α-helix (*29*). The molecule is stabilized by the presence of four disulfide bridges, all of them located within secondary structure elements. LCI displays no structural similarity to other protease inhibitors, except to potato carboxypeptidase inhibitor (*30*, *31*), with which it shares C-terminal sequence homology and a substrate-like-manner inhibitory mechanism for pancreatic-like carboxypeptidases. LCI is also a potent inhibitor of plasma carboxypeptidase B, also called TAFI (thrombin activatable fibrinolysis inhibitor), an enzyme involved in the equilibrium of blood clots. LCI could thus be a possible candidate for pharmacological application in thrombosis disease. Our laboratory has recently described the unfolding pathway of LCI (*32*), showing that the rate constant of the process is low when compared to those of other disulfide-containing proteins and that the protein is highly stable. All these features confirm the suitability of LCI as a model for the study of oxidative folding pathways.

EXPERIMENTAL PROCEDURES

Materials. Recombinant LCI was obtained by heterologous expression in *Escherichia coli* following a procedure described previously (*28*). The recombinant protein that contains a construction-added glycine as the N-terminal residue was purified by ion-exchange chromatography on a TSK DEAE column (Amersham Biosciences) followed by reverse-phase HPLC. The protein was more than 99% pure, as judged by HPLC, and its molecular mass was confirmed by MALDI-TOF. Thermolysin (P-1512), dithiothreitol, reduced glutathione (GSH), oxidized glutathione (GSSG), and β-mercaptoethanol were obtained from Sigma with purities of greater than 99%.

Control Folding Experiments. Native LCI (0.5 mg) was initially reduced and denatured in 400 μL of Tris-HCl buffer (0.1 M, pH 8.4) containing 6 M GdmCl and 30 mM dithiothreitol. The reaction was carried out at 22 °C for 90 min. The sample was then passed through a NAP-5 column (Amersham Biosciences), previously equilibrated in Tris-HCl buffer (0.1 M, pH 8.4), to initiate oxidative refolding. Desalting took approximately 1 min. Reduced and unfolded LCI was recovered in 0.55 mL and immediately diluted to a final concentration of 0.5 mg/mL in the same Tris-HCl buffer, both in the absence (control −) and in the presence (control +) of 0.25 mM β-mercaptoethanol. Folding intermediates were trapped in a time course manner at selected times by mixing aliquots of the sample with an equal volume of 4% trifluoroacetic acid in water. Trapped folding intermediates were analyzed by HPLC and reverse-phase chromatography. HPLC conditions are given in the figure legends.

Folding of LCI in the Presence of Redox Agents. Reduction, denaturation, and desalting were carried out as described in the preceding paragraph. GSSG (0.5 mM) or a mixture of GSSG and GSH (0.5 and 1 mM, respectively) were added immediately after unfolded LCI was desalted with a NAP-5 column. Folding intermediates were trapped with an equal volume of 4% trifluoroacetic acid in water and analyzed by HPLC and reverse-phase chromatography.

Analysis of Folding Intermediates of LCI. Acid-trapped folding intermediates of LCI were purified from HPLC and freeze-dried. The samples were derivatized with 50 μL of vinylpyridine (0.1 M) in Tris-HCl buffer (0.1 M, pH 8.4) at 23 °C for 35 min. The reaction was quenched with 4% trifluoroacetic acid. Vinylpyridine-derivatized samples were then analyzed by mass spectrometry to characterize the number of disulfide bonds in the population of folding intermediates.

Characterization of the Disulfide Structure of the Major Kinetic Traps. The major HPLC fractions containing non-native LCI species detected in the control experiments were isolated, freeze-dried, and derivatized with vinylpyridine as reported for the analysis of folding intermediates of LCI. Vinylpyridine-derivatized samples were further purified by HPLC and freeze-dried. Isolated intermediates III-A1, III-A2, and III-B (20 μg each) were treated with 2 μg of thermolysin (Sigma, P-1512) in 30 μL of *N*-ethylmorpholine/acetate buffer (50 mM, pH 6.4). Digestions were carried out for 16 h at 37 °C. Thermolytic products were then isolated by HPLC and analyzed by amino acid sequencing and mass spectrometry to identify the disulfide-containing peptides.

Amino Acid Sequencing and Mass Spectrometry. The amino acid sequence of disulfide-containing peptides was analyzed by automatic Edman degradation using a Perkin-Elmer Life Sciences Procise sequencer (model 494) equipped with an on-line phenylthiohydantoin-derivate analyzer. The molecular masses of disulfide-containing peptides were determined by a matrix-assisted laser desorption ionization time-life-of-flight mass spectrometer (Perkin-Elmer Life Sciences Voyager-DE STR).

RESULTS

Oxidative Folding of LCI in the Absence of Redox Agents. Folding experiments of reduced and denatured LCI were first performed using simple conditions in Tris-HCl buffer alone (control −) and in Tris-HCl buffer in the presence of β-mercaptoethanol (0.25 mM). HPLC profiles of folding intermediates trapped by acid quenching at selected time points are presented in Figure 1. A high degree of heterogeneity of intermediates is observed during the early stages of folding. The chromatographic behavior of the folding intermediates clearly shows the accumulation of two fractions of major kinetic traps at the late stage of folding, designated here as III-A and III-B. Their patterns remain indistinguishable regardless of the presence of β-mercaptoethanol.

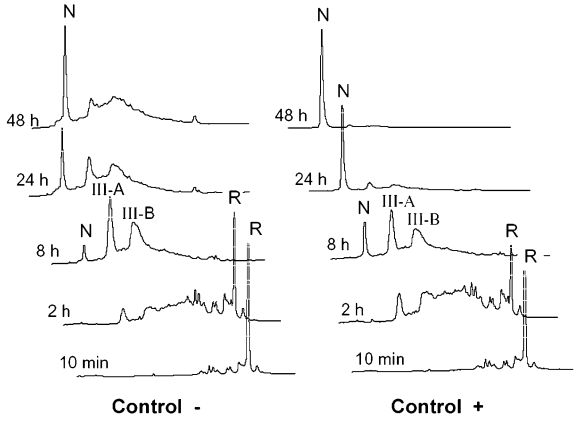

FIGURE 1: HPLC analysis of the folding intermediates of LCI trapped by acidification. Folding was carried out in Tris-HCl buffer (0.1 M, pH 8.4) in the presence (control +) or absence (control −) of 0.25 mM β-mercaptoethanol. Folding intermediates were trapped by mixing with an equal volume of 4% aqueous trifluoroacetic acid and analyzed by HPLC using the following conditions. Solvent A was 0.1% trifluoroacetic acid in water; solvent B was an acetonitrile/water mixture (9:1, v/v) containing 0.085% trifluoroacetic acid. The sample was applied to a Zorbax 300SB C-18 column (4.6 mm × 5 μm). A linear gradient was developed between 31 and 47% solvent B over the course of 45 min, with a flow rate of 0.5 mL/min. The column temperature was 23 °C. N and R indicate the elution positions of the native and fully reduced species of LCI, respectively. III-A and III-B are two major fractions of three-disulfide intermediates identified along the pathway of oxidative folding of LCI.

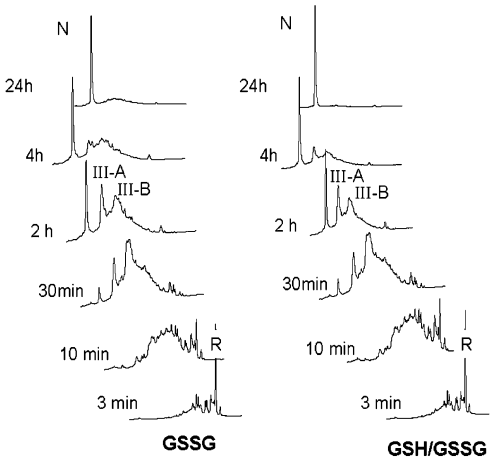

FIGURE 2: HPLC chromatograms of the intermediates of oxidative folding of LCI in the presence of redox agents. Folding was performed in Tris-HCl buffer (0.1 M, pH 8.4) in the presence of GSSG (0.5 mM) or a mixture of GSH and GSSG (0.5 and 1 mM, respectively). Folding intermediates were trapped by mixing with an equal volume of 4% aqueous trifluoroacetic acid and analyzed by HPLC using the same conditions described in the legend of Figure 1. N and R indicate the elution positions of the native and fully reduced species of LCI, respectively. III-A and III-B are two major fractions of three-disulfide intermediates.

Purified intermediates from the HPLC analyses were reacted with vinylpyridine and analyzed by MALDI mass spectrometry in an effort to evaluate the disulfide bond content of folding intermediates. The data (not shown) allow calculation of the relative concentrations of the fully reduced, one-disulfide, two-disulfide, three-disulfide, and four-disulfide ("scrambled" isoforms plus the native) species for each time course-trapped sample. The results are shown in Figure 3 (parts A and B). The recovery of native LCI can be distinguished from that of other four-disulfide species by its unique HPLC elution position (Figure 1) and is represented by bars in Figure 3.

The folding of LCI could not reach completion in the absence of β-mercaptoethanol, since only ~40% of the protein was recovered as the native structure after 48 h of folding, while the remaining 60% of LCI was trapped as folding intermediates. Among them, 35% corresponds to three-disulfide species represented by III-A and III-B and 25% to non-native four-disulfide species made up of a heterogeneous population of scrambled isomers. In the presence of β-mercaptoethanol, the recovery of native LCI is significantly higher, with more than 90% of the protein recovered as the native structure after 48 h of folding. Under these conditions, the recovery rate is also more closely related to the generation of four-disulfide LCI (Figure 3B). These results confirm the coexistence of three-disulfide and non-native four-disulfide isomers as folding intermediates and major kinetic traps of LCI folding. They also demonstrate the role of β-mercaptoethanol as a thiol catalyst for promoting the disulfide shuffling and the conversion of scrambled LCI to the native structure.

Oxidative Folding of LCI in the Presence of Redox Agents. Oxidative folding of LCI was subsequently performed in the presence of redox agents, including GSSG (0.5 mM) and a mixture of GSH and GSSG (1 and 0.5 mM, respectively). Folding intermediates were similarly analyzed by HPLC. As shown in Figure 2, HPLC patterns obtained under both conditions are almost indistinguishable, and they are also very similar to those obtained in the control experiments (Figure 1). Fractions III-A and III-B were also shown to accumulate along the folding pathway under these conditions, confirming that isoforms eluted within these two fractions are major intermediates of the oxidative folding of LCI. However, the addition of GSSG accelerates the rate of LCI folding by approximately 10−12-fold. The increase in the rate of folding can be observed both through the kinetics of disulfide oxidation that leads to the generation of native LCI and by the total recovery of the native species. Similar effects of GSSG on the efficiency of oxidative folding have been demonstrated in other proteins (*10*, *12*, *14*, *17*).

Analysis of the composition of disulfide species of time course-trapped intermediates (Figure 3C,D) revealed the rapid accumulation of three- and four-disulfide LCI during the folding. It is also apparent that the accumulation of native LCI follows that of four-disulfide species. These data again confirm that four-disulfide scrambled isomers are essential folding intermediates of LCI and that the native LCI can be attained directly through either three-disulfide species with native pairing or reorganization of scrambled four-disulfide species.

Structural Analysis of the Major Three-Disulfide Kinetic Traps. Fractions III-A and III-B were isolated, freeze-dried, and reacted with vinylpyridine, and the derivatized samples were further purified by HPLC using the same conditions that were employed for the separation of folding intermediates. The chromatographic analysis shows that fraction III-B contains one predominant three-disulfide species and III-A comprises two major three-disulfide species (III-A1 and III-A2) (Figure 4). III-A1, III-A2, and III-B were isolated; their molecular masses were determined by MALDI-TOF mass spectrometry, and aliquots were digested with thermolysin.

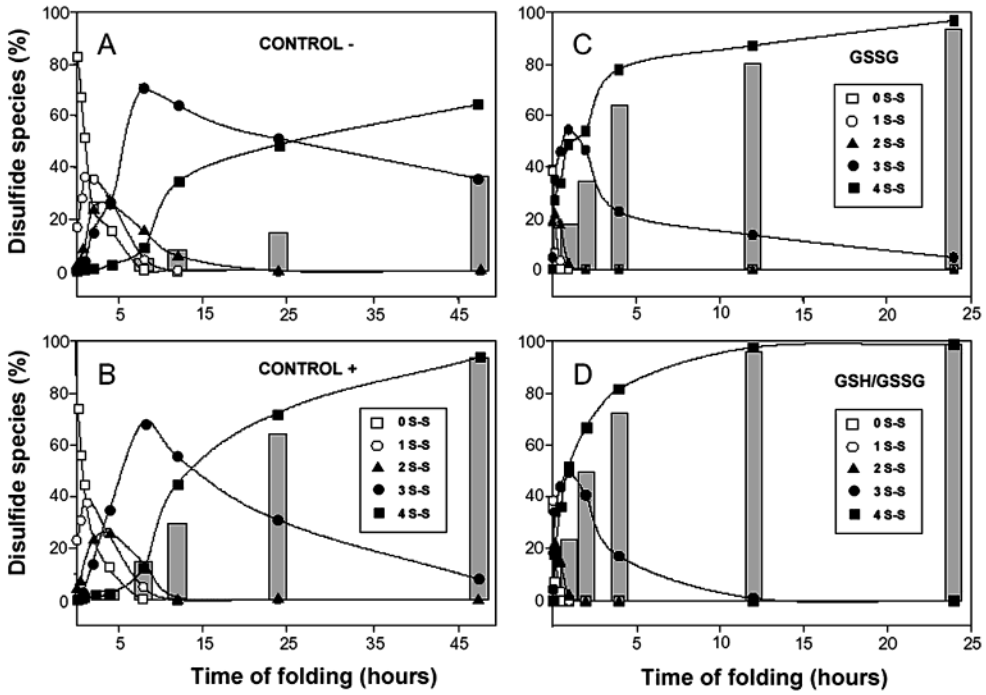

FIGURE 3: Quantitative analysis of various disulfide species along the pathway of oxidative folding of LCI. Folding was performed using the following conditions: (A) Tris-HCl buffer alone (control −), (B) Tris-HCl buffer in the presence of 0.25 mM β-mercaptoethanol (control +), (C) Tris-HCl buffer containing GSSG (0.5 mM), and (D) Tris-HCl buffer containing a mixture of GSSG and GSH (0.5 and 1 mM, respectively). 0 S−S, 1 S−S, 2 S−S, 3 S−S, and 4 S−S represent the completely reduced, one-disulfide, two-disulfide, three-disulfide, and four-disulfide species, respectively. The four-disulfide species include the native and all scrambled isoforms. The recovery of native LCI is represented by bars shown in each panel. Quantitative analysis of various disulfide species of LCI was based on the peak response of MALDI spectra (data not shown).

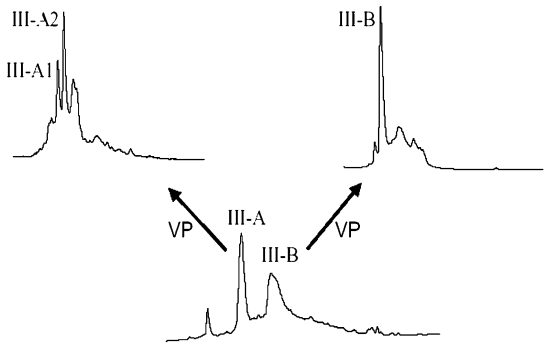

FIGURE 4: HPLC isolation of three-disulfide folding intermediates of LCI derivatized with vinylpyridine. Fractions III-A and III-B were isolated, freeze-dried, and derivatized with vinylpyridine. Derivatized III-A and III-B were further purified by HPLC using the same conditions described in the legend of Figure 1. The results reveal that fraction III-A contains two major three-disulfide intermediates (III-A1 and III-A2) and fraction III-B contains one predominant three-disulfide intermediate (III-B).

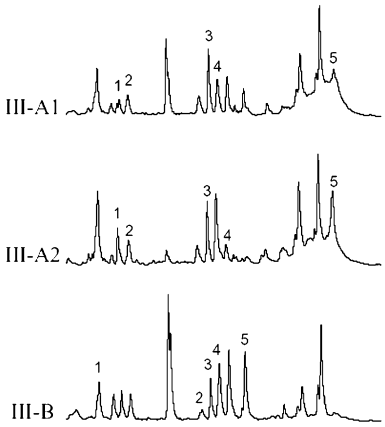

FIGURE 5: Peptide maps of thermolysin-digested III-A1, III-A2, and III-B. Thermolysin digestion product peptides were separated by HPLC using the conditions described in the legend of Figure 1, except for the linear gradient, which was 0 to 60% solvent B over the course of 60 min. Peptides were analyzed by Edman sequencing and MALDI mass spectrometry in an effort to identify the cysteine (vinylpyridine-derivatized)- and disulfide-containing peptides. The results are summarized in Table 1. The data that were obtained were used to define the disulfide structure of the intermediates depicted in Figure 6.

The resultant thermolytic peptides were separated by HPLC (Figure 5) and isolated for molecular mass analysis and Edman sequencing. Table 1 summarizes the M_r and sequence data of peptides containing disulfide bonds and free cysteines (modified with vinylpyridine). Taken together, these data permit us to deduce the disulfide structures of the three major three-disulfide intermediates (Figure 6). Two major shoulder fractions of III-A1 and III-A2 (Figure 4) were also characterized and were shown to contain exclusively (>95%) four-disulfide scrambled species as determined by MALDI mass spectrometry.

All three intermediates share a common thermolytic peptide (III-A1-3, III-A2-3, and III-B-3) which was shown to have a molecular mass of 1653 Da and contain the native Cys^{11}−Cys^{34} disulfide pairing. Both III-A1 and III-A2 have free Cys^{22} and Cys^{58} derivatized with vinylpyridine, as unambiguously identified by Edman sequencing and mass spectrometry analyses. These coincidences, together with the initial difficulty of separating III-A1 and III-A2 intermediates, suggest that they have very similar isomeric structures. Analysis of thermolytic peptides III-A1-5 and III-A2-5 further establishes that III-A1 and III-A2 adopt either the

Table 1: Amino Acid Sequences and Molecular Masses of Cysteine- and Disulfide-Containing Peptides Derived from Thermolysin Digestion of Folding Intermediates of LCI[a]

peptide	sequence	residue numbers	disulfide	M_r, found (exptl) (Da)
III-A1-1	ICRG	21−24	Cys^{22}−VP	554.51 (552.65)
III-A1-2	STGQCR	54−59	Cys^{58}−VP	756.16 (755.81)
III-A1-3	LC	10 and 11	Cys^{11}−Cys^{34}	1652.62 (1650.85)
	AAPLPSEGECNPHP	25−38		
III-A1-4	LCYQPDQ	10−16	Cys^{11}−Cys^{34}	2285.21 (2282.49)
	AAPLPSEGECNPHP	25−38		
III-A1-5	YQPDQVCCFI	12−21	Cys^{18}−Cys^{43}	2717.6 (2713.13)
	TAPWC	39−43	Cys^{19}−Cys^{62}	
	TTCIPYVE	60−67		
III-A2-1	ICRG	21−24	Cys^{22}−VP	554.51 (552.65)
III-A2-2	YSTGQCR	53−59	Cys^{58}−VP	919.73 (918.98)
III-A2-3	LC	10 and 11	Cys^{11}−Cys^{34}	1653.92 (1650.85)
	AAPLPSEGECNPHP	25−38		
III-A2-4	GSHTPDESF	1−9	no Cys	977.35 (975.97)
III-A2-5	YQPDQVCCFI	12−21	Cys^{18}−Cys^{62}	2717.6 (2713.13)
	TAPWC	39−43	Cys^{19}−Cys^{43}	
	TTCIPYVE	60−67		
III-B-1	ICRG	21−24	Cys^{22}−Cys^{58}	1097.9 (1096.26)
	STGQCR	54−59		
III-B-2	PHPTAPWCREGAV	36−48	Cys^{43}−VP	1526.8 (1525.71)
III-B-3	LC	10 and 11	Cys^{11}−Cys^{34}	1653.1 (1650.85)
	AAPLPSEGECNPHP	25−38		
III-B-4	LCYQPDQ	10−16	Cys^{11}−Cys^{34}	2285.21 (2282.49)
	AAPLPSEGECNPHP	25−38		
III-B-5	VCC	17−19	Cys^{19}−VP	962.5 (960.16)
	TTCIP	60−64	Cys^{18}−Cys^{62}	

[a] The peptides that were analyzed are those marked numerically in Figure 5. The underlined sequences were confirmed by Edman degradation. Vinylpyridine (VP)-modified peptides exhibit an additional molecular mass of 106.1 Da.

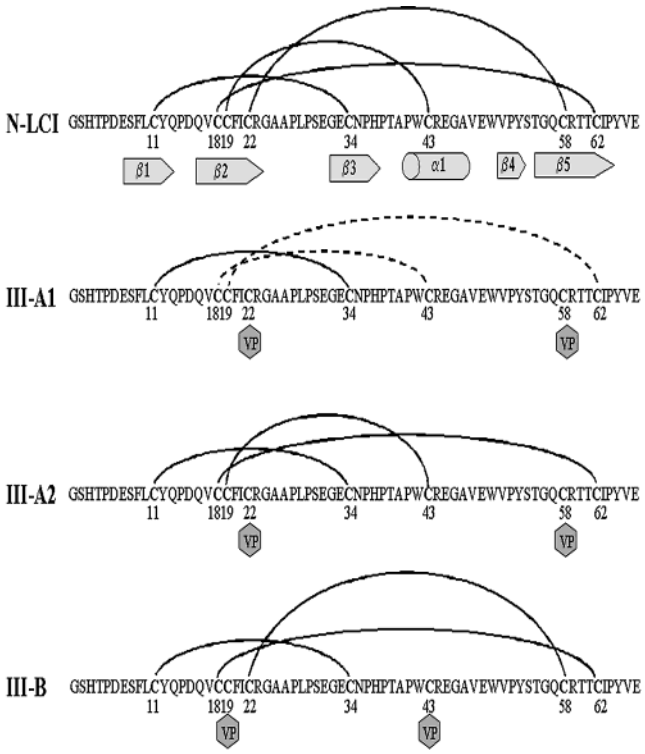

FIGURE 6: Disulfide structures of the major kinetic traps III-A1, III-A2, and III-B. The disulfide pairing of native LCI (N-LCI) is shown at the top for comparison. Solid lines represent native disulfide pairings. Dashed lines in IIIA-1 indicate non-native disulfide pairings. Free cysteines are labeled with vinylpyridine (VP).

additional disulfide pairings of [Cys^{18}−Cys^{43}; Cys^{19}−Cys^{62}] and [Cys^{18}−Cys^{62}; Cys^{19}−Cys^{43}] or [Cys^{18}−Cys^{62}; Cys^{19}−Cys^{43}] and [Cys^{18}−Cys^{43}; Cys^{19}−Cys^{62}], respectively. Conventional methods of peptide mapping and Edman sequencing were unable to distinguish these two isomeric disulfide bond patterns in III-A1-5 or III-A2-5 because enzymatic cleavage between Cys^{18} and Cys^{19} was not possible. It is important to point out that the Cys^{18}−Cys^{62} and Cys^{19}−Cys^{43} disulfides are native disulfide bonds whereas the Cys^{18}−Cys^{43} and Cys^{19}−Cys^{62} disulfides are non-native disulfide bonds. Putative disulfide structures of III-A1 and III-A2 are presented in Figure 6, in which one of them adopts three native disulfide bonds and the other contains one native and two non-native disulfide bonds. Since both species are detected along the folding pathway of LCI with comparable and significant concentrations, this assignment should not affect the global understanding and interpretation of the data.

Assignment of disulfide pairings is unmistakable in the case of III-B. Aside from the native Cys^{11}−Cys^{34} disulfide bond verified by peptide III-B-3, Cys^{19} and Cys^{43} were shown to be free cysteines modified with vinylpyridine. In addition, two native disulfide bonds, Cys^{22}−Cys^{58} and Cys^{18}−Cys^{62}, were identified by analysis of peptides III-B-1 and III-B-5 (Table 1). The disulfide structure III-B is also illustrated in Figure 6, in which solid lines and dotted lines represent native and non-native disulfide pairings, respectively. All three isoforms share the native Cys^{11}−Cys^{34} pairing. Among the nine disulfide bonds present in these three kinetic traps, seven are native and two are non-native.

DISCUSSION

Diversity of the Disulfide Folding Pathway. The disulfide folding pathways of several model proteins have been elucidated by various laboratories (*1*, *2*, *8*−*27*). The results have shown that the mechanism of oxidative folding varies significantly from protein to protein. These differences are illustrated by (a) the extent of the heterogeneity of folding intermediates, (b) the predominance of intermediates contain-

ing native disulfide bonds, and (c) the accumulation of fully oxidized scrambled isomers as intermediates. Among numerous proteins that have been investigated in detail, BPTI and hirudin represent two remarkable models with extreme folding characteristics, despite their comparable size and identical disulfide content (both contain three disulfide bonds). The original model of BPTI established by Creighton was based mainly on the analysis of folding intermediates trapped by iodoacetate and separated by ion-exchange chromatography (*1*, *2*, *7*). A subsequent study, using the method of acid trapping and HPLC analysis, revealed a somewhat different pattern of folding intermediates of BPTI (*8*). Specifically, intermediates with non-native disulfides are detected at a much lower concentration than those found previously. Despite few discrepancies, the folding pathway of BPTI is characterized by the predominance of a limited number of folding intermediates that adopt native disulfide bonds and native-like structures. Most importantly, three-disulfide scrambled BPTI was absent and only one- and two-disulfide intermediates were observed during the folding of BPTI. By contrast, the folding mechanism of hirudin (*10*, *11*) differs from that of BPTI in three important aspects. (a) Folding intermediates of hirudin are far more heterogeneous. At least 30 fractions of one- and two-disulfide intermediates were identified along the folding pathways of hirudin. (b) Predominant folding intermediates adopting native disulfides are absent in this case. (c) Scrambled three-disulfide isomers, not observed in the case of BPTI, were shown to serve as folding intermediates of hirudin. Among the 14 possible scrambled isomers of hirudin, 11 have been identified as folding intermediates, and their disulfide structures have been determined (*33*, *34*). Moreover, accumulation of scrambled hirudin as folding intermediates can be greatly enhanced by the presence of oxidized glutathione or cysteine (*10*). For instance, when folding of hirudin was performed in the presence of 2 mM GSSG, a condition routinely applied for the oxidative folding of BPTI, the scrambled species accounted for more than 95% of the intermediates observed 5 min after the onset of the folding reaction. This accumulation took place even before any significant amount of native hirudin had appeared (*10*). For hirudin, folding apparently proceeds through an initial stage of nonspecific disulfide formation (packing) followed by disulfide reshuffling (consolidation) of the heterogeneous scrambled intermediates to attain the native structure. The mechanism of oxidative folding of potato carboxypeptidase inhibitor (*12*, *13*) is nearly indistinguishable from that of hirudin.

Underlying Chemistry of the Diversity of Disulfide Folding Pathways. To identify the underlying causes of these diversities, we have previously conducted kinetic analysis of reductive unfolding of five different proteins, including BPTI, hirudin, potato carboxypeptidase inhibitor, tick anticoagulant peptide, and epidermal growth factor (*35*). The experiment of reductive unfolding was intended to evaluate the relative stability and interdependency of disulfide bonds in the native protein. A striking correlation was observed between the mechanisms of reductive unfolding and oxidative folding. Those proteins with their native disulfide bonds reduced collectively in an all-or-none mechanism without detectable partially reduced species display both a high degree of heterogeneity of folding intermediates and the accumulation of scrambled isomers along the pathway of oxidative folding, as observed with hirudin. On the other hand, a sequential reduction of the native disulfide bonds is associated with the presence of predominant intermediates with native-like structures, as in the case of BPTI. These studies lead to the conclusion that the pathway and mechanism of oxidative folding hinge critically on the presence of localized stable domain structures. An outstanding example to illustrate this hypothesis is the mechanism of oxidative folding of α-lactalbumin observed in the absence or presence of calcium (*27*). The structure of α-lactalbumin consists of an α-helical domain that is primarily stabilized by hydrophobic force and a β-sheet domain that is stabilized upon binding to calcium (*36*). In the absence of calcium, the folding pathway of α-lactalbumin resembles that of hirudin and potato carboxypeptidase inhibitor, which is depicted by heterogeneous one-, two-, and three-disulfide folding intermediates and the accumulation of four-disulfide scrambled isomers along the folding pathway. Binding of calcium serves to generate a stable β-sheet domain. Consequently, the folding pathway of α-lactalbumin in the presence of calcium involves only two predominant intermediates that adopt native-like structures of the β-sheet domain, thus bearing a resemblance to what is observed in the case of BPTI (*7*, *8*).

Pathway of Oxidative Folding of LCI. The pathway of oxidative folding of LCI elucidated here exhibits mechanisms observed in both BPTI and hirudin. The folding of LCI proceeds through a heterogeneous mixture of one- and two-disulfide intermediates, guiding the formation of two different populations of intermediates that act as kinetic traps. One population consists of a complex mixture of scrambled four-disulfide species. The other consists of three predominant three-disulfide isomers. Among them, two contain exclusively native disulfide bonds. Like in the case of hirudin, a sequential flow of heterogeneous one- and two-disulfide intermediates leads to the formation of scrambled (four-disulfide) species which are able to effectively convert to the native structure only in the presence of thiol catalysts. On the other hand, and like BPTI, intermediates adopting native disulfide bonds populate the late stages of folding and serve as rate-limiting steps in the attainment of the native structure. The pathway of oxidative folding of LCI thus further broadens the diversity spectrum of disulfide folding pathways.

The effect of the GGGS/GSH system in promoting the efficiency of oxidative folding has also been demonstrated in this study. Specifically, the kinetics of disulfide formation (i.e., the flow of intermediates) is accelerated by the addition of GSSG to the folding buffer, and the kinetics of disulfide shuffling (scrambling) are promoted by the presence of GSH (or β-mercaptoethanol). In the case of LCI, addition of GSSG (0.5 mM) increases the rate of disulfide formation by ~12-fold. In the absence of any thiol reagent ("control −" experiment), scrambled four-disulfide intermediates become trapped and unable to reshuffle their non-native disulfide bonds and convert to the native structure during the folding. However, the population and pattern of folding intermediates of LCI remain largely unchanged regardless of the presence of redox agents. This unique effect of the GSSG/GSH redox system on the mechanism of oxidative folding has been likewise demonstrated with numerous disulfide proteins,

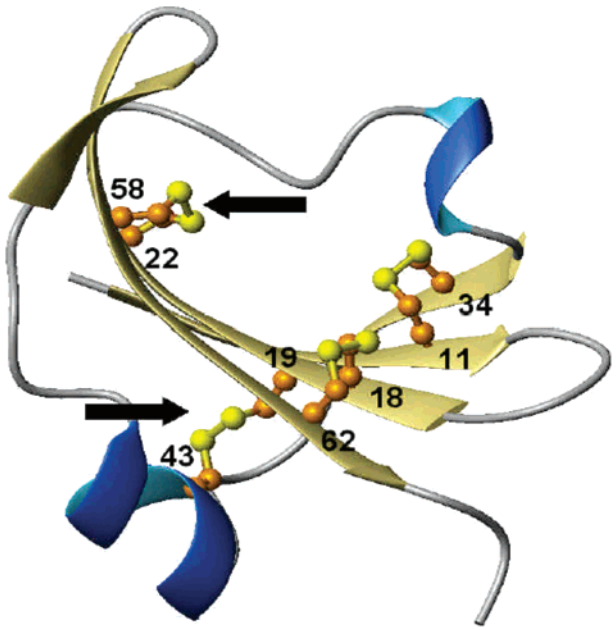

FIGURE 7: Schematic view of the native structure of leech carboxypeptidase inhibitor. The four native disulfide bonds (Cys^{11}−Cys^{34}, Cys^{18}−Cys^{62}, Cys^{19}−Cys^{43}, and Cys^{22}−Cys^{58}) are numbered. The arrows indicate the two disulfide bonds (Cys^{22}−Cys^{58} and Cys^{19}−Cys^{43}) absent in intermediates III-A and III-B, respectively. The Protein Data Bank entry for the structure of LCI is 1DTV.

including hirudin (*10*), potato carboxypeptidase inhibitor (*12*), tick anticoagulant peptide (*14*), and apo-α-lactalbumin (*27*).

Molecular Basis for the Stability of Three-Disulfide Kinetic Traps of LCI. The structure of LCI consists of a five-stranded antiparallel β-sheet and one short α-helix (*29*), as seen in Figure 7 and schematically displayed in Figure 6. All four native disulfide bridges are located within regular secondary structure elements. The III-B intermediate has two free cysteines, Cys^{19} and Cys^{43}, which connect α-helix 1 and β-sheet 2 in the native structure. The absence of this disulfide bridge suggests that the structure of III-B would display a disconnected α-helical 1 section and a main core of five antiparallel β-sheets stabilized by the remaining three native disulfide bonds. The stability of the β-sheet core domain may account for the sluggish formation of the last native disulfide (Cys^{19}−Cys^{43}) and allow III-B to serve as one of the major kinetic traps. Alternatively, a kinetic barrier required for the cooperative formation of the α-helix 1 and the Cys^{19}−Cys^{43} disulfide may explain the slow conversion of III-B to the native LCI. For III-A2, the absence of the Cys^{22}−Cys^{58} disulfide implies that the linkage between the C-terminal end of β-sheet 2 and the N-terminal end of β-sheet 5 and the subsequent stabilization of the complete β-sheet core domain represents a rate-limiting step in reaching the native structure. In the case of III-A1, conversion to the native LCI would require disruption of the interaction between β-strands 2 and 5, as well as a displacement of α-helix 1. All three kinetic traps, despite their native disulfide bonds and presumably native-like structure, must somehow twist their conformation and overcome an energy barrier to form the native LCI.

The structural properties of intermediates III-A1 and III-A2 are reminiscent of those of two major isoforms described for insulin-like factor 1 (IGF-1) (*37*−*40*). In both IGF and LCI, the two isoforms are products of isomerization *via* two adjacent cysteines (Figures 6 and 7). At physiological pH, the native IGF (Cys^{6}−Cys^{48}, Cys^{18}−Cys^{61}, and Cys^{47}−Cys^{52}) exists in equilibrium with a scrambled species, designated IGF-s (swap) (Cys^{6}−Cys^{47}, Cys^{18}−Cys^{61}, and Cys^{48}−Cys^{52}) with a molar ratio of 60:40 (*38*, *40*). The same molar ratio is found between IIIA-2 and III-A1 of LCI (Figure 4). If the structural property of IGF-1 constitutes a valid reference, we have indeed correctly assigned the disulfide structures of III-A2 and III-A1, in which the predominant species (III-A2) adopts all native disulfide bonds.

REFERENCES

1. Creighton, T. E. (1978) *Prog. Biophys. Mol. Biol. 33*, 231−297.
2. Creighton, T. E., and Goldenberg, D. P. (1984) *J. Mol. Biol. 179*, 497−526.
3. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) *Biochemistry 39*, 4207−4216.
4. Frand, A. R., Cuozzo, J. W., and Kaiser, C. A. (2000) *Trends Cell Biol. 10*, 203−210.
5. Woycechowsky, K. J., and Raines, R. T. (2000) *Curr. Opin. Chem. Biol. 4*, 533−539.
6. Creighton, T. E. (1986) *Methods Enzymol. 131*, 83−106.
7. Creighton, T. E. (1990) *Biochem. J. 270*, 1−16.
8. Weissman, J. S., and Kim, P. S. (1991) *Science 253*, 1386−1393.
9. Goldengerg, D. P. (1992) *Trends Biochem. Sci. 17*, 247−261.
10. Chang, J.-Y. (1994) *Biochem. J. 300*, 643−650.
11. Chatrenet, B., and Chang, J.-Y. (1993) *J. Biol. Chem. 268*, 20988−20996.
12. Chang, J.-Y., Cannals, F., Schindler, P., Querol, E., and Aviles, F. X. (1994) *J. Biol. Chem. 269*, 22087−22094.
13. Venhudova, G., Canals, F., Querol, E., and Aviles, F. X. (2001) *J. Biol. Chem. 276*, 11683−11690.
14. Chang, J.-Y. (1996) *Biochemistry 35*, 11702−11709.
15. Chang, J.-Y., and Ballatore, A. (2000) *J. Protein Chem. 19*, 299−310.
16. Wu, J., Yang, Y., and Watson, J. T. (1998) *Protein Sci. 7*, 1017−1028.
17. Chang, J.-Y., Li, L., and Lai, P.-S. (2001) *J. Biol. Chem. 276*, 4845−4852.
18. Miller, J. A., Narhi, L. O., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A. (1993) *Biochemistry 32*, 5203−5213.
19. Yang, Y., Wu, J., and Watson, J. T. (1999) *J. Biol. Chem. 274*, 37598−37604.
20. Scheraga, H. A., Konishi, Y., and Ooi, T. (1984) *Adv. Biophys. 18*, 21−41.
21. Scheraga, H. A., Wedemeyer, W. J., and Welker, E. (2001 *Methods Enzymol. 341*, 189−221.
22. Welker, E., Narayan, M., Volles, M. J., and Scheraga, H. A. (1999) *FEBS Lett. 460*, 477−479.
23. Rothwarf, D. M., Li, Y.-J., and Scheraga, H. A. (1998) *Biochemistry 37*, 3760−3766.
24. Rao, K. R., and Brew, K. (1989) *Biochem. Biophys. Res. Commun. 163*, 1390−1396.
25. Ewbank, J. J., and Creighton, T. E. (1993) *Biochemistry 32*, 3677−3693.
26. Ewbank, J. J., and Creighton, T. E. (1993) *Biochemistry 32*, 3694−3707.
27. Chang, J.-Y., and Li, L. (2002) *Biochemistry 41*, 8405−8413.
28. Reverter, D., Vendrell, J., Canals, F., Hosrtmann, J., Aviles, F. X., Fritz, H., and Sommerhoff, C. P. (1998) *J. Biol. Chem. 273*, 32927−32933.
29. Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A., and Aviles, F. X. (2000) *Nat. Struct. Biol. 7*, 322−328.
30. Molina, M. A., Avilés, F. X., and Querol, E. (1992) *Gene 116*, 129−138.
31. Molina, M. A., Marino, C., Oliva, B., Avilés, F. X., and Querol, E. (1994) *J. Biol. Chem. 269*, 21467−21472.
32. Salamanca, S., Villegas, V., Vendrell, J., Li, L., Aviles, F. X., and Chang, J.-Y. (2002) *J. Biol. Chem. 277*, 17538−17543.
33. Chang, J.-Y., Schindler, P., and Chatrenet, B. (1995) *J. Biol. Chem. 270*, 11992−11997.
34. Chang, J.-Y. (1995) *J. Biol. Chem. 270*, 25661−25666.
35. Chang, J.-Y., Li, L., and Bulychev, A. (2000) *J. Biol. Chem. 275*, 8287−8289.

36. Permyakov, E. A., and Berliner, L. J. (2000) *FEBS Lett. 473*, 269−274.
37. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) *Biochemistry 31*, 1749−1756.
38. Miller, J. A., Narhi, L. O., Hua, Q. X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K. S., Tsai, L., and Weiss, M. A. (1993) *Biochemistry 32*, 5203−5213.
39. Narhi, L. O., Hua, Q. X., Arakawa, T., Fox, G. M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J. A., and Weiss, M. A. (1993) *Biochemistry 32*, 5214−5221.
40. Chang, J.-Y., Maerki, W., and Lai, P. S. (1999) *Protein Sci. 8*, 1463−1468.

BI034308P