Section II. Work 4

Role of Kinetic Intermediates in the Folding of Leech Carboxypeptidase Inhibitor

Summary

The oxidative folding and reductive unfolding pathways of leech carboxypeptidase inhibitor (LCI; four disulfides) have been characterized in this work by structural and kinetic analysis of the acid-trapped folding intermediates. The oxidative folding of reduced and denatured LCI proceeds rapidly through a sequential flow of 1-, 2-, 3-, and 4-disulfide (scrambled) species to reach the native form. Folding intermediates of LCI comprise two predominant 3-disulfide species (designated as III-A and III-B) and a heterogeneous population of scrambled isomers that consecutively accumulate along the folding reaction. Our study reveals that forms III-A and III-B exclusively contain native disulfide bonds and correspond to stable and partially structured species that inter-convert, reaching an equilibrium prior to the formation of the scrambled isomers. Given that these intermediates act as kinetic traps during the oxidative folding, their accumulation is prevented when they are destabilized, thus leading to a significant acceleration of the folding kinetics. III-A and III-B forms appear to have both native disulfides bonds and free thiols similarly protected from the solvent; major structural rearrangements through the formation of scrambled isomers are required to render native LCI. The reductive unfolding pathway of LCI undergoes an apparent all-or-none mechanism, although low amounts of intermediates III-A and III-B can be detected, suggesting differences in protection against reduction among the disulfide bonds. The characterization of III-A and III-B forms shows that the former intermediate structurally and functionally resembles native LCI, whereas the III-B form bears more resemblance to scrambled isomers.

Keywords: carboxypeptidase inhibitor; protein folding; disulfide folding pathways; oxidative folding; reductive unfolding; folding intermediates.

Introduction

The “new view” of protein folding, which has emerged in the recent years from a combination of experimental work and theoretical approximations, postulates the folding process as a parallel flow of molecules that follow multiple folding routes to reach the native state (1, 2). As folding proceeds, some semi-stable conformations corresponding to local free energy minima (intermediates) may be transiently accumulated, acting as kinetic traps (3). Thus, understanding protein folding requires identification of the intermediate(s) that form(s) along the preferential pathways leading from the unfolded state to the native form (4). Unfortunately, characterizing kinetic folding intermediates is usually a difficult issue due to their short half-life. An important part of our knowledge about the role and nature of intermediates along the folding process comes from studies of disulfide-rich proteins in which transient folding forms have been trapped and characterized (5).

Oxidative folding is one of the well-established methods used to analyze the folding of disulfide-containing proteins (6-14). For these proteins, the folding pathway is characterized and defined by the heterogeneity and structures of the disulfide isomers that accumulate along the folding process. Folding intermediates can be trapped by acidification of the protein solution and separated by reversed-phase high performance liquid chromatography (RP-HPLC), which allows their further structural characterization. Application of the oxidative folding and acid-trapping method has allowed the elucidation of folding pathways of several 3-disulfide proteins such as hirudin (15, 16), potato carboxypeptidase inhibitor (PCI) (17, 18), tick anticoagulant peptide (TAP) (19, 20), epidermal growth factor (21, 22), insulin-like growth factor (IGF-1) (23, 24), and the extensively investigated model of bovine pancreatic trypsin inhibitor (BPTI) (6, 7, 9-11). However, few models aside from ribonuclease A (RNase A) and α-lactalbumin (αLA) have been studied in detail among 4-disulfide proteins (25-32). In these cases, analysis of the folding pathway represents another level of technical challenge due to the increase in the number of possible disulfide intermediates.

The above-mentioned studies have not indicated any predominant folding scenario, and even among small 3-disulfide proteins the folding mechanism varies substantially. For proteins as BPTI, intermediates with native disulfide bonds and native-like structures prevail along the folding pathway (10, 11). The non-covalent interactions that stabilize the native BPTI play a crucial role in guiding the early folding events and hence dictate the formation of a limited number of intermediates that admit the prevalence of native

disulfides. In the case of hirudin and PCI, two other 3-disulfide proteins, folding proceeds through an initial non-specific disulfide pairing (packing) that leads to the formation of a heterogeneous population of 3-disulfide scrambled isomers; this is followed by disulfide reshuffling (consolidation) of these intermediates to finally acquire the native form (15, 17). For the latter proteins, non-covalent interactions do not seem to participate significantly in guiding protein folding during the early phase of non-specific packing. Within this context, folding studies of novel protein models are required to better understand the underlying causes of such a diversity of disulfide folding pathways.

Leech carboxypeptidase inhibitor (LCI) is a 66-residue cysteine-rich protein that folds in a compact domain consisting of a five-stranded antiparallel β-sheet and a short α-helix, as reported by our group (Fig. 1) (33). The molecule is stabilized by four disulfide bridges, which are all located within secondary structure elements (Fig. 1). LCI is a potent metallocarboxypeptidase inhibitor that binds tightly to pancreatic carboxypeptidases A1, A2, B (CPA1, CPA2, CPB) and to plasma CPB, also called thrombin-activatable fibrinolysis inhibitor (TAFI) (34). Assuming that leeches secrete LCI during feeding, LCI may participate in the elimination of blood clots by inhibiting TAFI, an enzyme shown to retard fibrinolysis (35, 36). LCI could help to maintain the liquid state of the blood during feeding and possibly block the host defense mechanisms involving mast cell proteases (33). The profibrinolytic effect of LCI has been demonstrated *in vitro*, suggesting a potential pharmacological application in thrombotic diseases (Salamanca et al., manuscript in preparation).

We have recently described both, the unfolding pathway and thermodynamic stability (37), and the oxidative folding process of this protein (38), showing that 3- and 4-disulfide intermediates act as kinetic traps along its folding pathway. In the present work, we study in depth the kinetic, thermodynamic, conformational and functional properties of several disulfide intermediates along the pathways of oxidative folding and reductive unfolding of LCI.

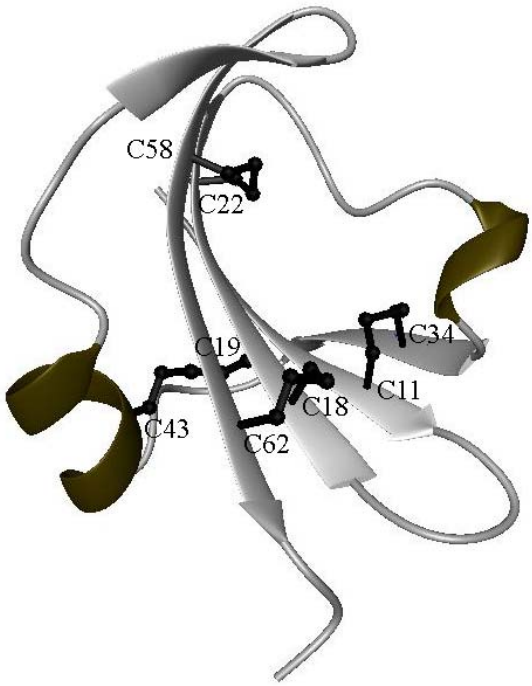


Fig. 1. Schematic view of the native three-dimensional structure of leech carboxypeptidase inhibitor. The cysteine residues are depicted in the structure. The Protein Data Bank accession number for the structure of LCI is 1DTV. The amino acid sequence of LCI and its disulfide pairing are shown at the bottom. The inhibitory site for metallocarboxypeptidases is at the C-terminus, after Cys62.

Experimental Procedures

Materials — Recombinant LCI was obtained by heterologous expression in *Escherichia coli* with an added glycine at the N-terminus. The protein was purified by ion-exchange chromatography on a TSK-DEAE column (Tosohaas), followed by RP-HPLC (34). The recombinant protein was more than 99% pure, as judged by HPLC analysis. The chromogenic substrates N-(4-methoxyphenylazoformyl)-Phe-OH and N-(4-methoxyphenylazoformyl)-Arg-OH were obtained from Bachem. Bovine CPA was purchased from Sigma. Human CPA1, CPA2 and CPB were prepared following described procedures (39). Dithiothreitol (DTT), guanidine hydrochloride (GndHCl), thermolysin (P-1512) and 2-mercaptoethanol were purchased from Sigma with purities greater than 99%.

Oxidative Folding of Fully Reduced LCI — Native LCI (1 mg) was reduced and denatured in Tris-HCl buffer (0.1 M, pH 8.4) containing 8 M GdnHCl and 50 mM DTT, at 22ºC for 2 h. To initiate folding, the sample was passed through a PD-10 column (Sephadex-25, Amersham Biosciences), previously equilibrated with Tris-HCl buffer (0.1

M, pH 8.4). Reduced and denatured LCI was recovered in 1.2 ml and immediately diluted to a final protein concentration of 0.5 mg/ml in the same Tris-HCl buffer, both in the absence (control -) and presence (control +) of 0.25 mM 2-mercaptoethanol. Folding intermediates of LCI were trapped in a time course manner at selected times by mixing aliquots of the sample with 2% trifluoroacetic acid (TFA). Trapped folding intermediates were analyzed by RP-HPLC.

Analysis of the Folding Intermediates of LCI by RP-HPLC — Analysis and isolation of folding intermediates of LCI were achieved by RP-HPLC using the following conditions. Solvent A was 0.1% TFA and solvent B acetonitrile containing 0.1% TFA. The column used was a 4.6 mm Protein C4 (Vydac). A linear 20-40% gradient of solvent B was applied during 50 min, with a flow rate of 0.75 ml/min.

Stop/Go Folding — Acid trapped intermediates were isolated by RP-HPLC, freeze-dried, and allowed to carry on the folding by dissolving the sample (0.5 mg/ml) in Tris-HCl buffer (0.1 M, pH 8.4), both in the absence and presence of 0.25 mM 2-mercaptoethanol. Folding intermediates were trapped with 2% TFA and analyzed by RP-HPLC. Scrambled isomers of LCI were separated from 3-disulfide intermediates by treatment with vinylpyridine and further isolation by RP-HPLC.

Oxidative Folding of LCI in the Presence of Denaturants — The procedures of unfolding and refolding were as described in the oxidative folding experiments. Immediately after the desalting of unfolded LCI through a PD-10 column, selected concentrations of denaturants (0.5-5 M GdnHCl, 1-8 M urea) were added. Folding intermediates were similarly trapped by acidification and analyzed by RP-HPLC.

Reductive Unfolding — Native LCI and the 3-disulfide intermediates (0.5 mg) were dissolved in 1 ml of Tris-HCl buffer (0.1 M, pH 8.4) with different concentrations of DTT (0.1-100 mM). Reduction was carried out at 22ºC. To monitor the kinetics of unfolding, time-course aliquots of the samples were trapped with 2% TFA, and analyzed by RP-HPLC. In addition, native LCI was dissolved in the above-mentioned Tris-HCl buffer containing selected concentrations of GdnHCl or urea. Aliquots of the samples were likewise removed at time intervals, trapped with acid and analyzed by RP-HPLC.

Disulfide-Pairing Analysis of the Major Intermediates in the LCI Reductive Unfolding — The acid-trapped intermediates were purified by RP-HPLC and freeze-dried. Each sample (20 μg) was derivatized with either 50 μl of vinylpyridine (0.1 M) in Tris-HCl buffer (0.1 M, pH 8.4) or with 500 μl of vinylpyridine (0.25 M) in Tris-HCl buffer

(0.1 M, pH 6.4) at 22ºC for 45 min. The vinylpyridine-derivatized samples were further purified by RP-HPLC, freeze-dried, and treated with 2 μg of thermolysin in 30 μl of N-ethylmorpholine/acetate buffer (50 mM, pH 6.4). Digestion was carried out at 37ºC for 16 h. Thermolytic products were then purified by RP-HPLC using a 4.6 mm Vydac C18 column (a linear gradient from 0 to 60% solvent B in 60 min) and analyzed by mass spectrometry (MS). Disulfide-containing peptides were further reduced with 10 mM tributylphosphin and analyzed by MS to identify their peptidic composition. The N-terminal sequence of selected peptides was also analyzed by automated Edman degradation.

Deuterium to Proton Exchange Followed by MS — Acid trapped intermediates were isolated by RP-HPLC and freeze-dried. The samples (50 μg) were resuspended in deuterated glycine buffer (20 mM, pD 2.5), incubated at 90ºC for 2.5 h to exchange all labile protons and then maintained at room temperature for 1 h to promote protein refolding. The deuterated proteins were diluted 1:4 with ammonium citrate (50 mM, pH 4.0) to start the hydrogen exchange. Aliquots were taken at different time points and analyzed by matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF MS) until an exchange plateau was reached. Samples were prepared by mixing equal volumes of the protein solution and matrix solution (sinapic acid in 30% acetonitrile with 0.1% TFA). At each exchange time, six samples were analyzed by duplicate. The average of the mass values, corresponding to the centroid of the peaks, was calculated for each exchange time and compared to an external unlabeled control, whose mass was determined by duplicate measurements.

Mass Spectrometry and Amino Acid Sequencing — The molecular masses of disulfide-containing peptides were determined by MALDI-TOF MS on a Bruker Ultraflex spectrometer. Samples for the deuterium to proton (D/H) exchange experiments were analyzed by the same spectrometer. The amino acid sequences of selected thermolytic peptides were analyzed by automatic Edman degradation using a Beckman LF3000 Protein Sequencer.

Circular Dichroism and NMR Spectroscopy — Samples for circular dichroism (CD) spectroscopy were prepared by dissolving the protein to a final concentration of 0.2 mg/ml in 0.1% TFA (pH 2.0). CD analyses were carried out in a Jasco J-715 spectrometer at 25ºC using a cell of 2 mm path length. Protein samples for ^{1}H NMR experiments were prepared

by dissolving the protein in H_2O/D_2O (9:1 ratio, v/v) with a concentration of 1 mg/ml at pH 2.0. NMR spectra were acquired on a Bruker AMX 500-MHz spectrometer at 25ºC.

CP Inhibitory Activity — The inhibitory activity of selected LCI folding intermediates was assayed by measuring the inhibition of the hydrolysis of the chromogenic substrate N-(4-methoxyphenylazoformyl)-Phe-OH by CPAs and N-(4-methoxyphenylazoformyl)-Arg-OH by CPB. The assay was performed in Tris-HCl buffer (50 mM, pH 7.5) containing 100 mM NaCl, with a substrate concentration of 100 μM. The inhibition constants (K_i) for the complexes of LCI intermediates with different carboxypeptidases were determined at the presteady-state as described for tightly binding inhibitors (40). The protein concentration of the LCI intermediates was determined from the A_{280} of the solution (LCI extinction coefficient: $E_{0.1\%}$=2.12).

Results

Accumulation of 3-Disulfide Intermediates and Scrambled Isomers along the Oxidative Folding Pathway of LCI — Oxidative folding of fully reduced LCI was carried out in the Tris-HCl buffer in the absence and presence of 2-mercaptoethanol as thiol catalyst. The RP-HPLC profiles of acid-trapped folding intermediates at selected time points are shown in Fig. 2. A high degree of heterogeneity of intermediates is observed at the beginning of the folding reaction, with identical RP-HPLC profiles in both refolding conditions (control - and control +). This initial stage is followed by the accumulation of two fractions (III-A and III-B) of major intermediates that act as kinetic traps. At this point (at 8 h), the RP-HPLC patterns are similar regardless of the presence of a reducing agent. The last stage of the folding process is characterized by an accumulation of a heterogeneous population of intermediates, which is more pronounced when the refolding is performed in the absence of a thiol catalyst (control -). Purified intermediates from the RP-HPLC analyses were derivatized with vinylpyridine, and analyzed by MALDI-TOF MS to evaluate the disulfide bond content of the folding intermediates. Folding of LCI was shown to undergo a sequential conversion through 1-, 2-, 3-, and 4-disulfide intermediates to reach the native structure (data not shown). Both 3-disulfide (III-A and III-B) and a mixture of non-native 4-disulfide (scrambled) isomers co-exist as folding intermediates and major kinetic traps of LCI folding. The folding of LCI cannot reach completion in the absence of a thiol catalyst, indicated by the fact that only ~ 30% of the protein was recovered in the native form after 48 h of refolding (Fig. 2). In the presence of 2-

mercaptoethanol, the recovery of native LCI was more than 90%, confirming the role of this redox agent in promoting the disulfide reshuffling and the conversion of scrambled forms to their native conformation.

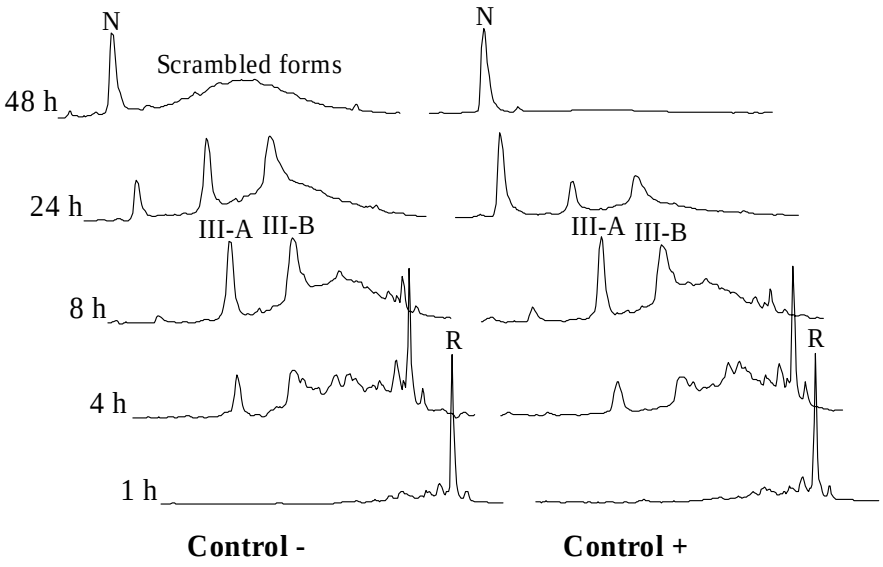


Fig. 2. RP-HPLC analysis of the acid-trapped intermediates of LCI along its oxidative folding. Folding was carried out in Tris-HCl buffer (pH 8.4) with (+) or without (-) the redox agent 2-mercaptoethanol (0.25 mM). Acid-trapped intermediates were analyzed at the noted times by RP-HPLC as detailed under "Experimental Procedures". *N* and *R* indicate the elution positions of native and fully reduced species of LCI, respectively. *III-A* and *III-B* are two major intermediates of three native-disulfides and s*crambled forms* are intermediates of four disulfides.

Evolution of the 3-Disulfide Intermediates and Scrambled Isomers along the Oxidative Folding Pathway of LCI — Our previous study on the oxidative folding of LCI revealed the presence of at least two 3-disulfide intermediates (III-A1 and III-A2) in fraction III-A and one 3-disulfide intermediate in fraction III-B after 8 h of refolding (38). Assignment of their disulfide pairings showed that isomers III-A2 and III-B contain three native disulfide bonds: Cys11-Cys34, Cys18-Cys62 and either Cys19-Cys43 or Cys22-Cys58, respectively, while isomer III-A1 contains one native and two non-native disulfide bonds: Cys11-Cys34, Cys19-Cys62, and Cys18-Cys43.

In the present work, purified fractions III-A and III-B from different refolding time points were derivatized with vinylpyridine (at pH 8.4), and analyzed by RP-HPLC in order to know their composition in disulfide isomers along the folding process. The analysis showed that fraction III-B only contains one predominant 3-disulfide intermediate along

the reaction, with the three native disulfide bonds previously described (data not shown). In contrast, within fraction III-A other 3-disulfide bonded forms were detected apart from the two previously characterized species (III-A1 and III-A2). This heterogeneity was not observed when the derivatization with vinylpyridine was performed at pH 6.4 with a lower protein concentration (data not shown). This fact suggests that such heterogeneity could be an artifact caused by the working pH (8.4), the high protein concentration and the conformation of the intermediates, since all these factors might affect the disulfide exchange rate. Structural analysis of the only species observed at pH 6.4 shows that it corresponds to the folding intermediate III-A2. Thus, both 3-disulfide kinetic traps that populate LCI folding correspond to species containing three native disulfide bridges.

To further assess the kinetic role of the 3-disulfide intermediates and scrambled 4-disulfide isomers, we performed stop/go experiments on these species. Acid-trapped intermediates III-A and III-B were isolated and allowed to resume the folding in the absence and presence of 2-mercaptoethanol. The data presented in Fig. 3 clearly show how such intermediates inter-convert along the folding reaction reaching an equilibrium that is slightly biased toward the III-A intermediate and finally form the 4-disulfide scrambled isomers. At this initial stage the RP-HPLC profiles are indistinguishable regardless of the presence of the thiol catalyst, suggesting that scrambled forms are not yet formed. The equilibrium, which is reached faster beginning from the III-B form than from the III-A form, would represent a rate-limiting step for the folding of LCI. Acid-trapped 4-disulfide (scrambled) isomers were also isolated, separated from 3-disulfide intermediates by treatment with vinylpyridine (which only modifies the latter) followed by RP-HPLC, and allowed to resume the folding in the absence and presence of 2-mercaptoethanol. The reshuffling of non-native 4-disulfide isomers into the native disulfide bonding pattern takes place directly and supposes yet another rate-limiting step for the folding of LCI (Fig. 3). As expected, in the stop/go experiments of scrambled forms the presence of 2-mercaptoethanol strongly promotes rearrangements, allowing conversion of more than 90% of the scrambled forms into native LCI, while in the absence of the redox agent only ~ 7% of the protein is recovered as native form at the end of the process.

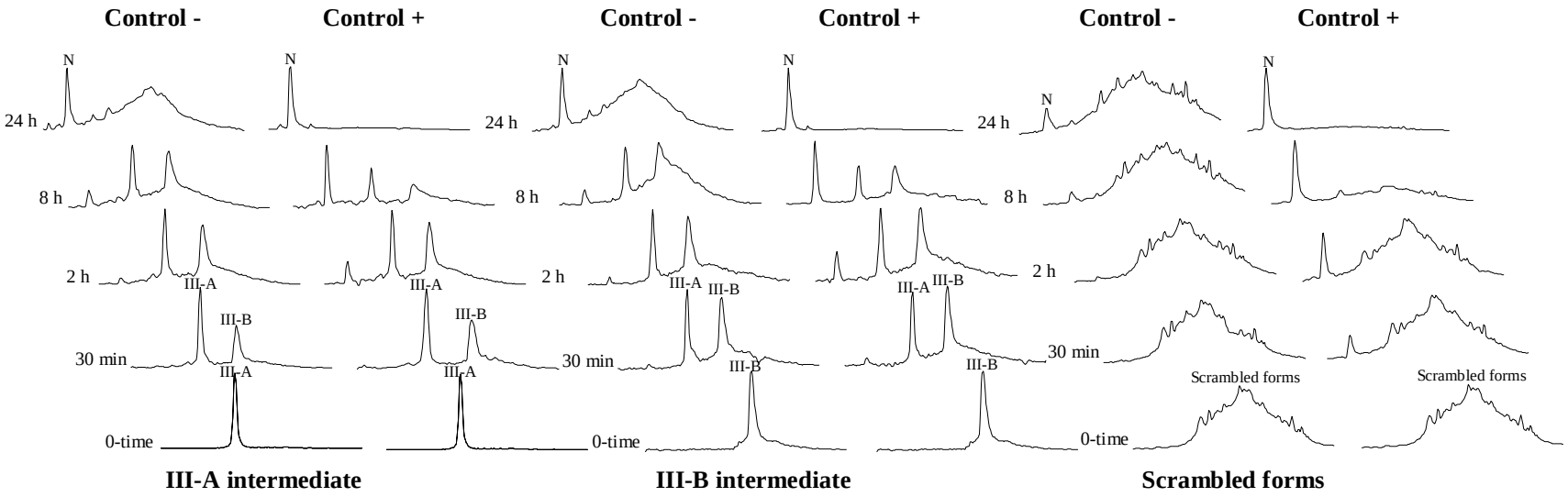


Fig. 3. Stop/Go folding of the 3-disulfide intermediates and scrambled forms of LCI. Acid-trapped intermediates were purified by RP-HPLC, freeze-dried, and dissolved in Tris-HCl buffer (pH 8.4) to allow the folding either in the absence (-) or presence (+) of 2-mercaptoethanol (0.25 mM). Folding intermediates were trapped with acid and analyzed by RP-HPLC. *N*, *III-A* and *III-B* stand for the native and the two 3-disulfide intermediates of LCI, respectively.

Oxidative Folding of LCI in the Presence of Denaturants — Oxidative folding of LCI was performed in the presence of increasing concentrations of GdnHCl or urea in order to evaluate the influence of denaturant on the prevalence of the 3-disulfide intermediates formed during the LCI folding process (Fig. 4). The comparison of these results with the control folding experiments in Fig. 2 shows that the accumulation of 3-disulfide intermediates decreases in the presence of higher denaturant concentrations. Intermediates III-A and III-B still accumulate under low denaturing conditions (up to 2 M urea concentration), being an indication of the high stability of these species (Fig. 4A). We can also observe a higher prevalence of intermediate III-A under these folding conditions. Interestingly, the higher recovery of native LCI in the presence of 0.5-1 M GdnHCl correlates with a lower accumulation of 3-disulfide intermediates at these conditions (Fig. 4B). In contrast, the refolding process performed in the presence of 1-2 M urea does not alter native LCI recovery.

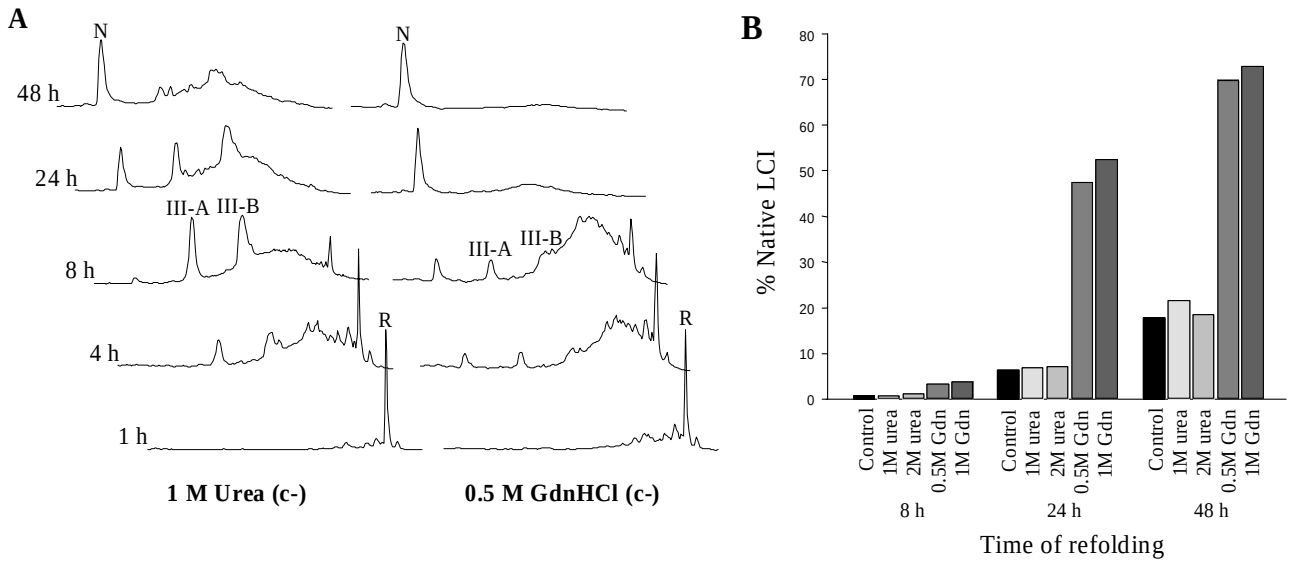


Fig. 4. Effect of denaturant on the folding intermediates of LCI. Reduced LCI was allowed to refold in Tris-HCl buffer (pH 8.4) under selected denaturing conditions. ***A.*** Folding experiments performed in the presence of 1 M urea or 0.5 M GdnHCl, without 2-mercaptoethanol (c-). Folding intermediates were trapped by acidification and analyzed by RP-HPLC. *N* and *R* stand for the native and reduced LCI, respectively. Intermediates III-A and III-B are indicated. ***B.*** The graphics shows the percentage of native form found at different denaturing conditions and time points, calculated from the peak areas in the corresponding RP-HPLC chromatograms. Maximum standard deviation is ± 6%.

The folding pathway of LCI drastically changed when the refolding experiments were carried out at high denaturing conditions (more than 2 M GdnHCl or 4 M urea). Examination of time course trapped intermediates revealed that 3-disulfide species do not longer accumulate in these conditions and that the reshuffling of the accumulated heterogeneous 4-disulfide scrambled isomers becomes the rate-limiting step of the folding reaction (Fig. 5). Important differences in native LCI recovery are observed in the absence and presence of 2-mercaptoethanol. For instance, when fully reduced LCI is allowed to refold in presence of 4 M GdnHCl and 2-mercaptoethanol, ~ 35% of the protein attains the native structure after 24 h of refolding, while less than ~ 2% is obtained in absence of the thiol catalyst (Fig. 5).

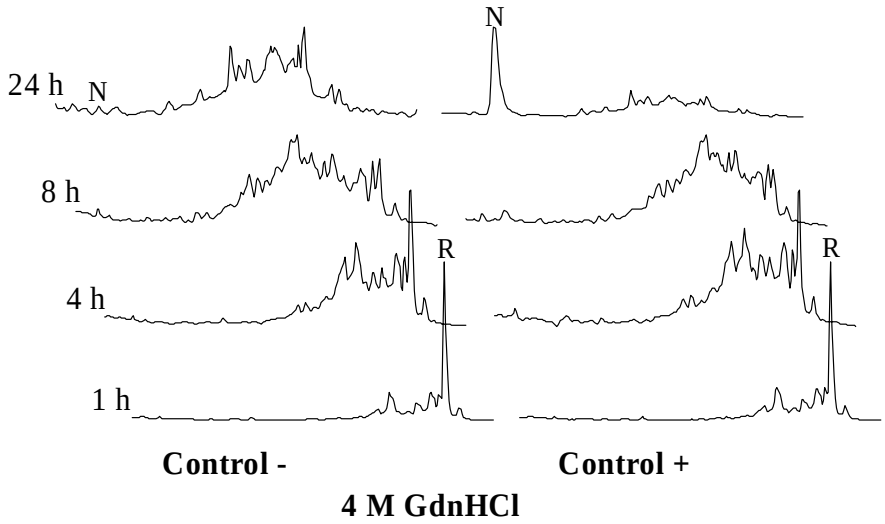


Fig. 5. Oxidative folding of LCI in presence of high denaturing conditions. Folding was carried out in Tris-HCl buffer (pH 8.4) in the presence (+) or absence (-) of 2-mercaptoethanol (0.25 mM) at high denaturant concentration (4 M GdnHCl). Acid-trapped intermediates were analyzed at the noted times by RP-HPLC. The elution positions of native (N) and reduced (R) LCI are indicated.

Reductive Unfolding of Native LCI and 3-Disulfide Intermediates — Reductive unfolding of native LCI was performed at pH 8.4 using various concentrations of DTT as reducing agent. Reduction undergoes an apparent all-or-none mechanism in which only low amounts of partially reduced intermediates accumulate (Fig. 6). The unfolding intermediates were trapped in a time course manner by acidification and were analyzed by RP-HPLC. Two different fractions of 3-disulfide intermediates (assessed by treatment with vinylpyridine and molecular mass analysis) were detected. These species subsequently convert to the fully reduced LCI (*R*) without a significant buildup of 1- or 2-disulfide

intermediates along the pathway. The same behavior was observed when the analysis was performed in the presence of different concentrations of DTT ranging from 2 to 100 mM. The two 3-disulfide intermediates fractions have a RP-HPLC elution equivalent to that of the intermediates III-A and III-B observed along the pathway of oxidative folding (Fig.2).

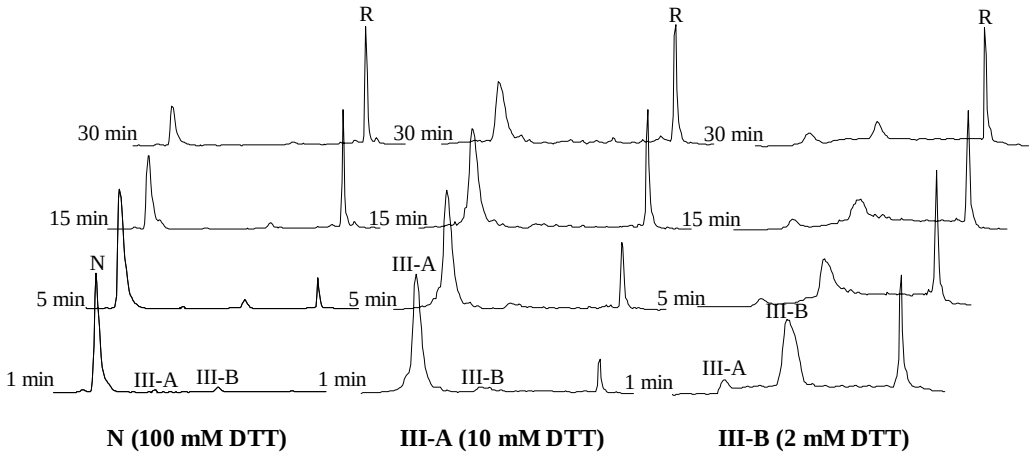


Fig. 6. Reductive unfolding of native LCI and 3-disulfide intermediates. The native LCI and the 3-disulfide intermediates were treated with various concentrations of DTT. Time course intermediates were trapped by acidification and analyzed by RP-HPLC. *N* and *R* stand for the native and reduced LCI. Intermediates III-A and III-B are indicated.

The unfolding intermediates were isolated to carry out structural analyses. They were treated with vinylpyridine, further purified by RP-HPLC, and digested with thermolysin. Thermolytic peptides were isolated by RP-HPLC and analyzed by MALDI-TOF MS and Edman sequencing to identify the structures of the disulfide-containing peptides. The results confirm that these unfolding intermediates are indeed identical to the predominant 3-disulfide oxidative folding intermediates of LCI (data not shown). These intermediates accumulate little along the reductive unfolding pathway. At early stages of the process, intermediate III-A comprises only ~ 1% of the total protein, while species III-B represents about 3-4%. When the experiment was performed in presence of a high concentration of denaturant (4 M GdnHCl), these intermediates did not accumulate at all. Reductive unfolding of purified intermediates III-A and III-B was also performed at pH 8.4 using various concentrations of DTT. In all conditions, the reduction of the three native disulfides takes place in a cooperative and concerted manner, and both forms unfold to the fully reduced LCI without further accumulation of 1-, or 2-disulfide intermediate species (Fig. 6). However, the inter-conversion between both intermediates can also be observed

along the reduction process. Since the inter-conversion process is slower than the reduction reaction and the intermediate III-A resists higher DTT concentrations than intermediate III-B, the inter-conversion cannot reach the same equilibrium as in the stop/go folding experiments; thus the amount of converted intermediate remains small along the process and with a higher prevalence of intermediate III-A.

Conformational Properties of the 3-Disulfide Intermediates and Scrambled Forms of LCI — The structural features of isolated LCI folding intermediates were analyzed by using CD spectroscopy, D/H exchange followed by MALDI-TOF MS, and ^{1}H-NMR spectroscopy. The CD spectrum of LCI is quite peculiar, with a well-defined ellipticity minimum at 210 nm and a maximum at 228 nm (Fig. 7). The former may be related to the presence of a high percentage of residues in β–structure, and the latter to both β–structures and loops or to an asymmetric environment of Tyr64 (34, 41). Previous CD spectroscopy measurements showed that the degree of LCI denaturation correlates with the decrease in ellipticity at 228-nm, and complete disappearance of this signature is observed when the protein is completely unfolded (34). The shape of CD spectrum of the III-A species is similar to that of the native protein. However, the 228-nm maximum is only about 30% of that of the native LCI and the minimum in ellipticity shifts to 205 nm (Fig. 7). The CD spectra of the III-B species and scrambled isomers exhibit clear differences to that of the native form. At 228 nm both species show negative values and the minimum is located at about 200 nm (Fig. 7), indicating that they are less structured intermediates than III-A.

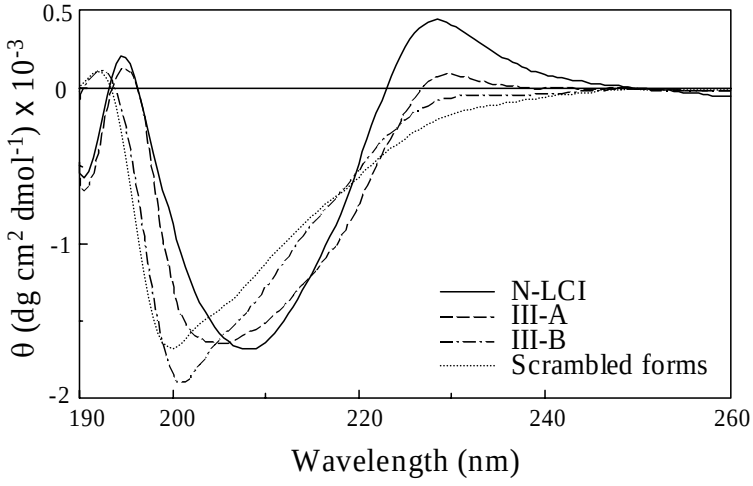


Fig. 7. Circular dichroism analyses of 3-disulfide intermediates and scrambled forms of LCI. CD analyses of LCI intermediates were carried out in 0.1% aqueous TFA (pH 2.0) at a final protein concentration of 0.2 mg/ml. Spectra were recorded in a Jasco J-715 spectrometer at 25ºC.

The conformational stability of the above-mentioned LCI folding intermediates was also investigated by D/H exchange experiments followed by MALDI-TOF MS (42). The extent of hydrogen exchange was quite different for the native form and its folding intermediates. Native LCI retains 27 deuterons at the end of the reaction, whereas the intermediates III-A and III-B, and the scrambled isomers retain 16 deuterons (standard deviation ±5%). This approximately 40% of decrease in protected deuterons probably reflects the low level of conformational packing in the intermediates. However, hydrogen exchange is far from being free and a slow exchange core exists in all these kinetic traps, as expected if they are, at least, partially structured.

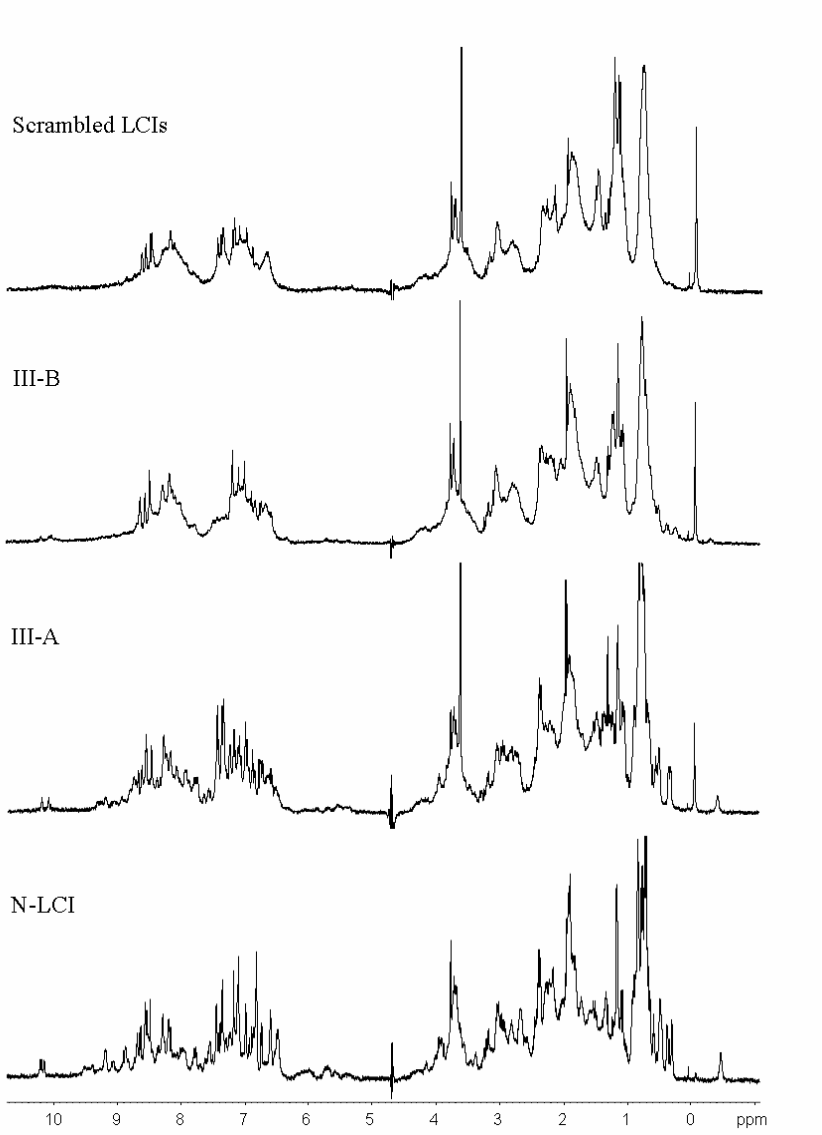


Fig. 8. ^{1}H-NMR analyses of 3-disulfide intermediates and scrambled forms of LCI. Monodimensional spectra were recorded in a Bruker AMX 500-MHz spectrometer at 25ºC. Samples were prepared in H_2O/D_2O (9:1 ratio, v/v) at pH 2.0.

Another adequate approach to assess protein conformation is NMR. The ^{1}H-NMR spectra of native LCI and intermediate III-A display very similar signal dispersion, peak sharpness, and upfield/downfield shifted resonances, a clear indication that this intermediate corresponds to a properly folded species (Fig. 8). In contrast, intermediate III-B and scrambled isomers spectra exhibit a clear band broadening and peak collapse (Fig. 8). However, chemical shift dispersion is appreciably greater than expected for a random coil conformation, an additional indication that these intermediates correspond to partially folded species.

CP Inhibitory Activity of the 3-Disulfide Intermediates and Scrambled Forms of LCI — LCI is a tightly binding, competitive inhibitor of carboxypeptidases A and B (34). Equilibrium dissociation constants for the complexes of the 3-disulfide intermediates and scrambled forms with different CPs were determined (Table 1). Surprisingly, the inhibitory activities of native LCI and intermediate III-A are practically identical, both in the nanomolar range. In contrast, the inhibitory capabilities of intermediate III-B and scrambled forms are, respectively, one and two orders of magnitude lower than that of the native form. These results correlate well with the conformational properties of the LCI folding intermediates described above and with the deduced degree of "nativeness" of them.

Table 1. Inhibition constants (K_i) of 3-disulfide intermediates and scrambled forms of LCI measured for different types of carboxypeptidases

Carboxypeptidase type	K_i LCI intermediates (nM)			
	Native	III-A	III-B	Scrambled
Bovine CPA	1.1 ± 0.2	1.2 ± 0.3	13 ± 2	110 ± 30
Human CPA1	1.3 ± 0.4	1.4 ± 0.4	38 ± 5	420 ± 45
Human CPA2	4.5 ± 0.6	4.8 ± 0.5	160 ± 35	1500 ± 400
Human CPB	1.2 ± 0.3	1.1 ± 0.4	32 ± 6	105 ± 25

Discussion

Folding Pathways among 4-Disulfide Proteins — Oxidative folding is the process by which a reduced and unfolded disulfide-containing protein gains both its native disulfide bonds and its native structure (8). The disulfide folding pathways of several model proteins have been characterized using the oxidative folding approach, exhibiting an unexpected diversity (32, 38). A protein disulfide folding pathway can be characterized by the level of

heterogeneity of its folding intermediates, the occurrence of predominant intermediates, and the accumulation of fully oxidized scrambled isomers.

In general, 4-disulfide proteins display a higher extent of secondary structure than 3-disulfide proteins. This fact affects the accessibility, proximity and reactivity of the thiols of the former proteins, drastically interfering in the rates of disulfide bond rearrangement and thus, complicating the folding landscape (43). Therefore, only few 4-disulfide models have been studied in detail. In the case of RNase A, its oxidative folding is characterized by an initial stage of sequential oxidation of the disulfide intermediates, leading to the formation of 1-, 2-, 3-, and 4-disulfide ensembles without any prevalent accumulation (27, 31, 44). Its rate-limiting step is the formation of two native-like species containing three native disulfide bonds, which are located in the protein core, protected from reduction and reshuffling. However, their thiols remain accessible to solvent and are subsequently oxidized to form the native protein (Fig. 9).

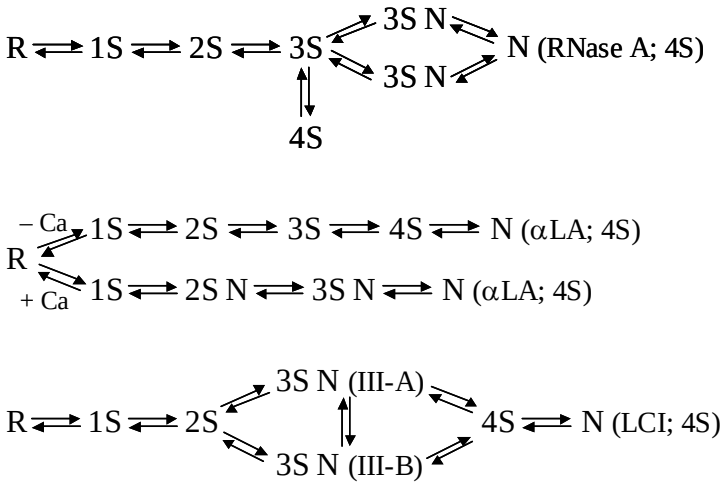


Fig. 9. Comparative disulfide folding pathways of RNase A, αLA and LCI. *R* and *N* indicate the reduced and native forms, respectively. *1S, 2S, 3S* and *4S* are ensembles of molecules with the corresponding number of disulfide bonds. *3S N* stands for 3-disulfide species with native bonds.

Another 4-disulfide protein that has been extensively characterized is α-lactalbumin, with a folding pathway dependent on the presence of calcium (29, 30, 32, 45-47). In its absence, oxidative folding proceeds through heterogeneous 1- to 4-disulfide intermediates, with a final conversion of 4-disulfide scrambled species to the native structure, which represents the major rate-determining step. No native-like conformations are predominant along the folding pathway (Fig. 9). Binding of calcium favors the formation of the β-sheet

domain of α-LA, and then only two major disulfide intermediates with two and three native bonds accumulate along the folding. The formation of the fourth bond accounts for the rate-limiting step of folding in these conditions (Fig. 9).

Oxidative Folding Pathway of LCI — In this context, LCI represents a new 4-disulfide model, which could give us further insight into the oxidative folding pathways. Previously, we have elucidated that denatured and reduced LCI folds through a heterogeneous mixture of 1- and 2-disulfide intermediates, leading to the formation of two populations of intermediates, 3-disulfide species and 4-disulfide (scrambled) isomers, which apparently act as kinetic traps (38). In the present work, we have stated that, as it happens for RNase A and α-LA, both predominant 3-disulfide intermediates (III-A and III-B) possess native disulfides, which are directly formed by oxidation from the 2-disulfide ensemble without any detectable accumulation of other 3-disulfide species (Fig. 9). This would suggest that as it happens for RNase A (48), the 2-disulfide ensemble of LCI may be enthalpically biased toward native disulfide bonds relative to the populations predicted taking into account entropic factors, allowing a faster and preferential formation of the third native disulfide bond.

The use of stop/go experiments clearly demonstrates that the rate of inter-conversion between the two 3-disulfide intermediates of LCI is much faster than their rate of conversion into scrambled forms (Fig. 9). It also shows that the rate of inter-conversion is faster from III-B to III-A intermediate, which is found at a slightly higher concentration at equilibrium, probably due to its higher thermodynamic stability and nativeness. III-A and III-B are probably metastable forms equivalent to what Scheraga and co-workers have defined as disulfide-insecure intermediates (43). Inside such kind of intermediates, thiol groups are as well protected as their disulfide bonds; therefore such thiols cannot be simply exposed and oxidized by a local unfolding process. Structural fluctuations that expose the thiol groups are also likely to expose the disulfide bonds and promote their reshuffling instead of oxidation of the free thiols to the native pairing. In the two LCI intermediates both, disulfide bonds and free thiols, are similarly protected from the solvent. The presence of the external thiol reagent does not affect the first stages of the stop/go experiments of these intermediates, showing that the protein free thiols are not solvent accessible and thus cannot interact with the external reagent. In RNase A or α-LA intermediates, local fluctuations may occur around the thiol groups of the fourth native disulfide bond allowing its oxidation without affecting the overall protein conformation. But LCI has a lower

secondary structure content than RNase A or α-LA, and the unfolding events in LCI 3-disulfide intermediates are likely to affect the whole core of the molecule leading to an overall rearrangement that also exposes the disulfide bonds.

LCI 3-disulfide intermediates probably differ from the previously described disulfide-insecure species in some aspects. First, they are able to inter-convert in a fast way and so, minor local fluctuations may possibly allow solvent-independent disulfide inter-change. This disulfide inter-change is an internal process in which all the reacting groups are protected from the exterior, since neither the rate of intermediates inter-conversion nor the concentration of the species at the equilibrium are affected by the presence of external thiols. Secondly, whereas 3-disulfide-insecure species described to date preferentially reshuffle to an unstructured 3-disulfide ensemble forming metastable dead-end pathways, LCI 3-disulfide intermediates III-A and III-B simultaneously oxidize, reshuffle and convert into a heterogeneous population of 4-disulfide scrambled isomers. Reshuffling of non-native 4-disulfide isomers into the native state is the last stage of the LCI oxidative folding and can be considered as the strongest rate-determining step (Fig. 9). Unlike that of the 3-disulfide intermediates, the disulfide bonds of unstructured scrambled forms are solvent-accessible and the addition of an external thiol group strongly accelerates the kinetics of native-disulfide formation from the scrambled population.

Effect of 3-Disulfide Intermediates Stability on the Folding Pathway of LCI — Despite the absence of a disulfide bond that might staple key secondary structural elements, both 3-disulfide intermediates display a striking stability in denaturing environment; so they are located in strong thermodynamic local minima that slow down LCI folding pathway. When LCI folding reaction is performed under mild denaturing conditions promoting partial unfolding of the intermediates, the rate and efficiency of LCI folding pathway increase. Under these conditions, the intermediates are under-stabilized, being less effective kinetic traps and accumulating to a lesser extent (49, 50); native disulfides and free cysteines are probably more solvent-accessible, and can easily convert into the scrambled forms by local unfolding events.

By adding enough denaturant to strongly destabilize III-A and III-B species, LCI oxidative folding pathway changes completely. The 3-disulfide intermediates do not longer accumulate and LCI folding proceeds through a sequential oxidation of 1-, 2-, 3-, and 4-disulfides forms, which accumulate as scrambled isomers. The disulfide reshuffling of the scrambled intermediates to finally attain the native form becomes the only rate-limiting step of the reaction. Due to the high denaturant concentrations, the relative abundance

among the scrambled isomers differs from that observed in the absence of denaturant. Probably, the scrambled isomers displaying more open and relaxed conformations, for instance the beads-form, show a higher prevalence, as observed for other proteins such as PCI, TAP and IGF-1 (51-53). These scrambled isomers display a higher difficulty to attain the native-bond pairing, hence this last stage of LCI folding becomes extremely slow under these conditions. The "simplified" LCI folding pathway observed in the presence of high concentrations of denaturants shows much resemblance to those exhibited by less structured 3-disulfide proteins (i.e. PCI, hirudin) (15, 17), suggesting that the differences among their folding processes are due to the higher extent of regular secondary structure displayed by LCI and not to the different number of disulfide bonds.

Reductive Unfolding Pathway of LCI — 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, as observed for hirudin or PCI (54, 55). On the other hand, a sequential reduction of the native disulfide bonds is generally associated with the presence of predominant folding intermediates, as in the case of BPTI or RNase A (55, 56).

Reinforcing the above-mentioned theory, transient accumulation of two intermediate species during the reductive unfolding of LCI was detected, which correspond to the 3-disulfide intermediates that act as kinetic traps in the oxidative folding pathway: the III-A and III-B forms. In LCI, these intermediates accumulate at a lesser extent than in the case of RNase A or BPTI, in agreement with the different characteristics of the 3-disulfide intermediates and folding pathways. In the oxidative folding reaction of BPTI or RNase A, one may expect a preferential protection toward reduction of those native bonds hidden in the protein core. Thus, the less stable and more solvent-accessible disulfide bonds can be preferentially reduced by local unfolding events, with the accumulation of the correspondent intermediate. Global unfolding only occurs after reduction of the covalent bonds hidden in the protein core. In the case of LCI, the disulfide bonds between Cys18-Cys62 and Cys11-Cys34 appear to be slightly more stable and protected than Cys22-Cys58 and Cys19-Cys43. This allows the detection of intermediates in which the former bonds are still formed and one of the two other disulfides is also present. It explains why the two 3-disulfide intermediates can still inter-convert prior to their complete reduction in the presence of moderate concentrations of reducing agent. However, in LCI, the differences in protection against reduction between disulfides bonds are too small to allow "locking

in" intermediate forms before the total reduction of the polypeptide and, on the overall, they are reduced almost in a concerted manner following an all-or-none mechanism.

Conformation and Functionality of LCI Folding Intermediates — LCI three-dimensional structure consists of a five-stranded anti-parallel β-sheet and a short α-helix (Fig. 1) (33). The protein is stabilized by four disulfides, all of them located within secondary structure elements (Fig. 1) (33). The III-A intermediate has two free cysteines, Cys22 and Cys58, which in the native form connect the C-terminal end of β2 and the N-terminal end of β5. The III-B species lacks the disulfide bridge formed between Cys19 and Cys43, which links the β2 and the α-helix.

The III-B species and the scrambled population are marginally structured forms, while maintaining yet some conformational order and activity. In contrast, the III-A intermediate corresponds to a structured and properly folded species, as assessed by NMR and CD spectroscopy. In addition, it has a RP-HPLC elution time very similar to that of native LCI, indicative of similar hydrophobicity. Besides its inhibitory capability is indistinguishable from that of the native LCI for all tested carboxypeptidases. One question is why has LCI evolved to be a 4-disulfide protein instead of a protein with 3-disulfide bonds, with the same inhibitory efficiency and a less complicated and faster folding pathway? Although proteins perform in a very efficient way their role *in vivo*, it is now clear that they are no fully optimized. They only fulfill the minimum requirements in terms of stability and folding efficiency that allow them to operate properly in the cell (57-59). Thus, in the case of LCI, one may assume that a 3-disulfide bonded variant would not be stable enough to perform efficiently its functions *in vivo*. This assumption makes sense if one takes into account that LCI is a protease inhibitor from leech saliva, evolved to act in blood, a fluid very rich in proteases. Despite its nativeness, the III-A intermediate displays higher fluctuation and lower conformational stability than native LCI, as shown by the lower protection to D/H exchange. By analogy, a native 3-disulfide bonded LCI would be probably more susceptible to proteolytic attacks.

Our results, and the comparison made with others, clearly indicate that the folding pathway of disulfide-containing proteins hinge critically on the presence of localized stable structures. The different structural content of the 3-disulfide intermediates characterized in the present work suggests that the accumulation of kinetic intermediates along the disulfide folding reaction relies mainly on their ability to protect their native disulfide bridges from rearrangement in the interior of totally or partially folded protein conformations.

Acknowledgments — We are grateful to Dr. Ulf Ekström for kindly revising this manuscript.

References

1. Dill KA & Chan HS. From Levinthal to pathways to funnels. *Nat Struct Biol* 1997; 4: 10-19
2. Honig B. Protein folding: from Levinthal paradox to structure prediction. *J Mol Biol* 1999; 293: 283-293
3. Wolynes PG, Onuchic JN & Thirumalai D. Navigating the folding routes. *Science* 1995; 267: 1619-1620
4. Bai Y, Sosnick TR, Mayne L & Englander SW. Protein folding intermediates: native-state hydrogen exchange. *Science* 1995; 269: 192-197
5. Creighton TE, Darby NJ & Kemmink J. The roles of partly folded intermediates in protein folding. *Faseb J* 1996; 10: 110-118
6. Creighton TE. Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol* 1978; 33: 231-297
7. Creighton TE & Goldenberg DP. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J Mol Biol* 1984; 179: 497-526
8. Creighton TE. Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 1986; 131: 83-106
9. Creighton TE. Protein folding. *Biochem J* 1990; 270: 1-16
10. Weissman JS & Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991; 253: 1386-1393
11. Goldenberg DP. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 1992; 17: 257-261
12. Wedemeyer WJ, Welker E, Narayan M & Scheraga HA. Disulfide bonds and protein folding. *Biochemistry* 2000; 39: 4207-4216
13. Frand AR, Cuozzo JW & Kaiser CA. Pathways for protein disulphide bond formation. *Trends Cell Biol* 2000; 10: 203-210
14. Woycechowsky KJ & Raines RT. Native disulfide bond formation in proteins. *Curr Opin Chem Biol* 2000; 4: 533-539
15. Chatrenet B & Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 1993; 268: 20988-20996
16. Chang JY. Controlling the speed of hirudin folding. *Biochem J* 1994; 300: 643-650

17. Chang JY, Canals F, Schindler P, Querol E & Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 1994; 269: 22087-22094
18. Venhudova G, Canals F, Querol E & Aviles FX. Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 2001; 276: 11683-11690
19. Chang JY. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 1996; 35: 11702-11709
20. Chang JY & Ballatore A. Structure and heterogeneity of the one- and two-disulfide folding intermediates of tick anticoagulant peptide. *J Protein Chem* 2000; 19: 299-310
21. Wu J, Yang Y & Watson JT. Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci* 1998; 7: 1017-1028
22. Chang JY, Li L & Lai PH. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 2001; 276: 4845-4852
23. Miller JA, Narhi LO, Hua QX, Rosenfeld R, Arakawa T, Rohde M, Prestrelski S, Lauren S, Stoney KS, Tsai L & et al. Oxidative refolding of insulin-like growth factor 1 yields two products of similar thermodynamic stability: a bifurcating protein-folding pathway. *Biochemistry* 1993; 32: 5203-5213
24. Yang Y, Wu J & Watson JT. Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3)IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J Biol Chem* 1999; 274: 37598-37604
25. Scheraga HA, Konishi Y & Ooi T. Multiple pathways for regenerating ribonuclease A. *Adv Biophys* 1984; 18: 21-41
26. Welker E, Narayan M, Volles MJ & Scheraga HA. Two new structured intermediates in the oxidative folding of RNase A. *FEBS Lett* 1999; 460: 477-479
27. Scheraga HA, Wedemeyer WJ & Welker E. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol* 2001; 341: 189-221
28. Rao KR & Brew K. Calcium regulates folding and disulfide-bond formation in alpha-lactalbumin. *Biochem Biophys Res Commun* 1989; 163: 1390-1396
29. Ewbank JJ & Creighton TE. Pathway of disulfide-coupled unfolding and refolding of bovine alpha-lactalbumin. *Biochemistry* 1993; 32: 3677-3693
30. Ewbank JJ & Creighton TE. Structural characterization of the disulfide folding intermediates of bovine alpha-lactalbumin. *Biochemistry* 1993; 32: 3694-3707
31. Rothwarf DM, Li YJ & Scheraga HA. Regeneration of bovine pancreatic ribonuclease A: identification of two nativelike three-disulfide intermediates involved in separate pathways *Biochemistry* 1998; 37: 3760-3766

32. Chang JY & Li L. Pathway of oxidative folding of alpha-lactalbumin: a model for illustrating the diversity of disulfide folding pathways. *Biochemistry* 2002; 41: 8405-8413
33. Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000; 7: 322-328
34. Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998; 273: 32927-32933
35. Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-2718
36. Bouma BN & Meijers JC. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
37. Salamanca S, Villegas V, Vendrell J, Li L, Aviles FX & Chang JY. The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* 2002; 277: 17538-17543
38. Salamanca S, Li L, Vendrell J, Aviles FX, Chang JY. Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 2003; 42: 6754-6761
39. Reverter D, Ventura S, Villegas V, Vendrell J & Aviles FX. Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* 1998; 273: 3535-3541
40. Morrison JF. The slow-binding and slow, tight-binding inhibition of enzyne-catalysed reactions. *Trends Biochem Sci* 1982; 7: 102-105
41. Chen YH, Yang YT & Martínez HH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* 1972; 13: 3350-3359
42. Villanueva J, Canals F, Villegas V, Querol E & Aviles FX. Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. *FEBS Lett* 2000; 472: 27-33
43. Welker E, Narayan M, Wedemeyer WJ & Scheraga HA. Structural determinants of oxidative folding in proteins. *Proc Natl Acad Sci USA* 2001; 98: 2312-2316
44. Narayan M, Welker E, Wedemeyer WJ & Scheraga HA. Oxidative folding of proteins. *Acc Chem Res* 2000; 33: 805-812
45. Ikeguchi M, Fujino M, Kato M, Kuwajima K & Sugai S. Transition state in the folding of alpha-lactalbumin probed by the 6-120 disulfide bond. *Protein Sci* 1998; 7: 1564-1574

46. Chang JY. The folding pathway of alpha-lactalbumin elucidated by the technique of disulfide scrambling. Isolation of on-pathway and off-pathway intermediates. *J Biol Chem* 2002; 277: 120-126
47. Permyakov EA & Berliner LJ. alpha-Lactalbumin: structure and function. *FEBS Lett* 2000; 473: 269-274
48. Xu X, Rothwarf DM & Scheraga HA. Nonrandom distribution of the one-disulfide intermediates in the regeneration of ribonuclease A. *Biochemistry* 1996; 35: 6406-6417
49. Zhang JX & Goldenberg DP. Amino acid replacement that eliminates kinetic traps in the folding pathway of pancreatic trypsin inhibitor. *Biochemistry* 1993; 32: 14075-14081
50. Zhang JX & Goldenberg DP. Mutational analysis of the BPTI folding pathway: II. Effects of aromatic-->leucine substitutions on folding kinetics and thermodynamics. *Protein Sci* 1997; 6: 1563-1576
51. Chang JY, Li L, Canals F & Aviles FX. The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. *J Biol Chem* 2000; 275: 14205-14211
52. Chang JY. Denatured states of tick anticoagulant peptide. Compositional analysis of unfolded scrambled isomers. *J Biol Chem* 1999; 274: 123-128
53. Chang JY, Marki W & Lai PH. Analysis of the extent of unfolding of denatured insulin-like growth factor. *Protein Sci* 1999; 8: 1463-1468
54. Chang JY. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 1997; 272: 69-75
55. Chang JY, Li L & Bulychev A. The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 2000; 275: 8287-8289
56. Li YJ, Rothwarf DM & Scheraga HA. Mechanism of reductive protein unfolding. *Nat Struct Biol* 1995; 2: 489-494
57. Adams MWW & Kelly RM. Biocatalysis at extreme temperatures: enzyme systems near and above 100ºC. *ACS press Washington* 1992
58. Kim DE, Gu H & Baker D. The sequences of small proteins are not extensively optimized for rapid folding by natural selection. *Proc Natl Acad Sci USA* 1998; 95: 4982-4986
59. Demetrius L. Thermodynamics and kinetics of protein folding: an evolutionary perspective. *J Theor Biol* 2002; 217: 397-411

Section II. Work 5

NMR Structural Characterization and Computational Predictions of the Major Intermediate in the Oxidative Folding of Leech Carboxypeptidase Inhibitor

Summary

The III-A intermediate, containing three native disulfide bonds, constitutes the major rate-determining step in the oxidative folding of leech carboxypeptidase inhibitor (LCI; four disulfide bonds). In this work, III-A has been directly purified from the folding reaction and structurally characterized by NMR. The data shows that this species displays a highly native-like structure although it lacks some secondary structure elements being more flexible than native LCI. III-A represents the first structurally determined example of a *disulfide insecure* intermediate; direct oxidation of this species to the fully native protein seems to be restricted by the burial of its two free cysteine residues inside a native-like structure. We show also that theoretical approaches based on topological constrains predict with good accuracy the presence of this folding intermediate. Overall, the derived results suggest that native-like interactions between segments of secondary structure rather than the cross-linking of disulfide bonds direct the folding of LCI.

Keywords: carboxypeptidase inhibitor; folding intermediate; oxidative folding; NMR structure; folding prediction.

Introduction

Much effort has gone into identifying intermediates that are assumed to be necessary for rapid protein folding (1, 2). However, folding intermediates are usually difficult to isolate and characterize due to their short half-life and highly flexible/unfolded structure. A recent breakthrough in folding has come from the theoretical field because several groups have succeeded in developing computational approximations that are able to capture the major features of folding processes (3-6). In general these algorithms can predict with good accuracy the folding nucleus of proteins by exploiting a tight dependence between native state topology and folding mechanisms. A good qualitative agreement between predictions and experiments has been found not only for two-state folding proteins, but also for larger

proteins that exhibit transient on-pathway intermediates (7, 8). Nevertheless, much of our knowledge about folding intermediates comes from studies on the oxidative folding of disulfide-containing proteins, in which transient forms have been trapped using the disulfide acid-quenching approach, analyzed by reversed-phase high performance liquid chromatography (RP-HPLC), and further characterized (9). Folding intermediates of several proteins have been analyzed in this way; among them are those from bovine pancreatic trypsin inhibitor (BPTI) (10), bovine pancreatic ribonuclease A (RNase A) (11), and lysozyme (12).

Leech carboxypeptidase inhibitor (LCI) is a 67-residue inhibitor of A/B metallo-carboxypeptidases isolated from the medical leech *Hirudo medicinalis* (13). It folds in a compact domain consisting of a five-stranded antiparallel β–sheet and a short α–helix that are tightly stabilized by four disulfide bonds (14). LCI is a potent inhibitor of plasma carboxypeptidase B, also called thrombin-activatable fibrinolysis inhibitor (TAFI), which is a well-known attenuator of fibrinolysis (15, 16). Our group has recently tested *in vitro* the pro-fibrinolytic activity of LCI proving its possible use as an enhancer of the tissue pasminogen activator therapy of thrombosis (Salamanca et al., manuscript in preparation). Knowledge about the folding determinants of this molecule constitutes a basis for the development of variants with enhanced stability. In this regard we have described the unfolding pathway and conformational stability of LCI showing that this protein has slow kinetics of unfolding and is highly stable (17). We have also extensively characterized its oxidative folding pathway using structural and kinetic analysis of the intermediates trapped by acidification (18, 19). These studies show that reduced and denatured LCI refolds through a rapid and sequential flow of 1- and 2-disulfide intermediates to reach a rate-limiting step in which two 3-disulfide species (III-A and III-B) strongly accumulate. These two intermediates contain only native disulfide bonds and act as kinetic traps that need major structural rearrangements through the formation of a heterogeneous population of 4-disulfide (scrambled) isomers to render native LCI. III-A and III-B intermediates also populate transiently the reductive unfolding pathway of LCI. The preliminary structural characterization of these two intermediates shows that III-A is a properly folded and stable species that might have high content of native structure, while III-B is a less structured form.

In the present work, we report a detailed NMR structural analysis of the III-A folding intermediate of LCI. This intermediate has been directly purified from the oxidative folding reaction using RP-HPLC, and its structure characterized by NMR spectroscopy and

compared to that of native LCI. We integrate these structural data with the folding data obtained from a predictive algorithm to provide insights into the crucial role played by III-A in the folding of LCI. Implications of this study for understanding the underlying diversity in the folding of different disulfide-rich proteins are discussed.

Experimental Procedures

Sample Preparation — The synthetic gene for LCI (13) was cloned into the pBAT4 plasmid (20), fused in frame to the OmpA signal sequence for periplasmic expression. Recombinant ^{15}N-labeled LCI was obtained by heterologous expression in *Escherichia coli* strain TG1. Fresh transformed cells were precultured in LB media containing 0.1 mg/mL carbenicillin at 37ºC; after 5 h, 0.5 mL of culture were centrifuged at 3000 rpm for 5 min, and cells were resuspended in 25 mL of M9 media containing $^{15}NH_4Cl$ as the only nitrogen source and 0.1 mg/mL carbenicillin. This second preculture was continued overnight, and the cells contained in 10 mL were used to inoculate 1 L of the same minimal media. Protein expression was induced in late phase (OD_{600} = 1.0) by adding IPTG (1 mM final concentration). LCI was purified from supernatant as described (13). In summary, the protein was initially purified using a Sep-Pak C_{18} Cartridge (Waters), followed by anion-exchange chromatography on a TSK-DEAE 5PW column (Tosohaas), and by RP-HPLC on a 4.6 mm Protein C4 column (Vydac). The ^{15}N LCI labeling was almost heterogeneous (>99%) as deduced by MALDI-TOF MS analysis on a Bruker Ultraflex spectrometer.

The III-A folding intermediate was produced by oxidative refolding of LCI as previously reported (19). Briefly, native ^{15}N-labeled LCI was reduced and denatured in 0.1 M Tris-HCl (pH 8.4) containing 8 M guanidine hydrochloride and 50 mM dithiothreitol, at 22ºC for 2 h. To initiate folding, the sample was passed through a PD-10 column (Sephadex-25, Amersham Biosciences), previously equilibrated with 0.1 M Tris-HCl (pH 8.4). Reduced and denatured LCI was recovered and immediately diluted to a final protein concentration of 0.5 mg/ml in the same Tris-HCl buffer. Folding intermediates of LCI were trapped after approximately 8 h of refolding by mixing aliquots of the sample with 2% trifluoroacetic acid (TFA). The trapped III-A intermediate was purified by RP-HPLC using the following conditions: solvent A was water containing 0.1% TFA and solvent B acetonitrile containing 0.1% TFA. A linear 20-40% gradient of solvent B was applied over 50 min, with a flow rate of 0.75 ml/min. The column used was a 4.6 mm Protein C4 (Vydac).

NMR Experiments and Structure Calculation — Protein samples for NMR experiments were prepared by dissolving lyophilized ^{15}N LCI and ^{15}N III-A in either H_2O/D_2O (9:1 ratio by volume) or D_2O, at a concentration of 1 mM and pH 3.5. All experiments were carried out on a Bruker DRX-600 spectrometer, at 27ºC. The spectrometer was equipped with a triple resonance, triple gradient probe head. The TOCSY experiments (21) were performed with different mixing times between 20 and 40 ms, while the NOESY experiments (22) were carried out with a mixing time of 120 ms. 4096 complex data points were recorded in the time domain t_2 and 700 increments in the t_1 domain. Water suppression was achieved using the WATERGATE pulse sequence (23). The ^1H-^{15}N HSQC spectra (24) were also recorded at the same temperature with 2048 complex data points in the t_2 domain and 128 points in the t_1 domain, with 256 scans. The 3D NOESY-HSQC spectra (25) were performed with a mixing time of 100 ms, and 4096 complex data points were recorded in the t_3 domain. For the amide proton exchange experiments lyophilized samples of ^{15}N native LCI and III-A were dissolved in D_2O at pH 3.5, 27ºC. A series of consecutive 2D heteronuclear ^1H-^{15}N HSQC experiments were acquired with increased delays for up to 3 days.

The collected spectra were processed using the standard XWinNMR software package of Bruker and analyzed with the SPARKY software (26). Chemical shifts were assigned applying a combination of TOCSY/NOESY techniques (27, 28). Peaks were classified according to their intensities as weak (3.8-5 Å), medium (2.8-3.8 Å), and strong (2.0-2.8 Å). A total of 20 structures were calculated by the simulated-annealing method with the program CNS (29). Structure calculations were carried out essentially according to the basic protocol described previously (30, 31). For the final refinement, NOE tables were supplemented with constraints for several hydrogen bonds identified from the determined secondary structure.

Folding Pathway Prediction — Fold-X is a program developed by Serrano's Group at the EMBL-Heidelberg to calculate the folding pathways of proteins and the effect of a point mutation on the stability of a protein (6, 32, 33). Fold-X exploits the previously described correlation between protein topology and folding mechanisms (34). The method is based on two simple assumptions: 1) only native interactions contribute significantly to the folding process; and 2) each individual residue has two accessible states, native and disordered. As a novelty, the algorithm energy function includes terms that account for the difference in solvation energy for apolar and polar groups, the free energy difference between the formation of an intramolecular H-bond compared to intermolecular H-bond

formation, and local and loop entropy. As described by the authors, the different energy terms taken into account in Fold-X have been weighted using empirical data obtained from protein engineering experiments. Fold-X uses a full atomic description of the structure of the proteins. We used the available online version of the software at http://foldx.embl.de employing the default software settings. The protein data bank accession numbers used for the calculations were 1DTV for LCI, 1D0D for BPTI and 1H20 for PCI.

Results and Discussion

Isolation of the III-A Folding Intermediate — In the past years several folding intermediates of protein models such as BPTI, RNase A and lysozyme have been characterized. However, most of these studies have been centered on the analysis of mutant analogues of the intermediates lacking a specific disulfide bridge (35-38). These analogues can be obtained in substantial quantities and in homogeneous form, but they are not "true" intermediates in the sense that they do not posses any reactive free cysteine. Many analogues in which the same cysteine residues have been substituted by different amino acids or chemically blocked differ in stability and/or functionality among them (39-41). Only a few true intermediates have been directly purified from the folding reaction and structurally characterized (42-45).

The present work was aimed to characterize in detail the structure of the major folding intermediate of LCI, the III-A. For this task, LCI was expressed in *E. coli* as uniformly ^{15}N-labeled protein and purified to homogeneity (purity >98%). Oxidative refolding of fully reduced ^{15}N LCI was carried out to generate III-A. The RP-HPLC profiles of acid-trapped intermediates at selected time points during folding in the absence (control -) and presence (control +) of 2-mercaptoethanol as thiol catalyst are shown in Figure 1. A high degree of heterogeneity of 1- and 2-disulfide intermediates is observed at the beginning of the folding reaction, with identical RP-HPLC profiles in both refolding conditions (control - and control +). This initial stage of folding is followed by the accumulation of two major intermediates (III-A and III-B) that contain three native disulfide bonds and act as strong kinetic traps. At this point (8 h), the RP-HPLC patterns are still similar regardless of the presence of the reducing agent. The last stage of the folding process is characterized by accumulation of a heterogeneous population of 4-disulfide (scrambled) isomers that is more pronounced when the refolding is performed in

the absence of 2-mercaptoethanol. This reducing agent strongly promotes disulfide reshuffling and conversion of scrambled forms to the native conformation.

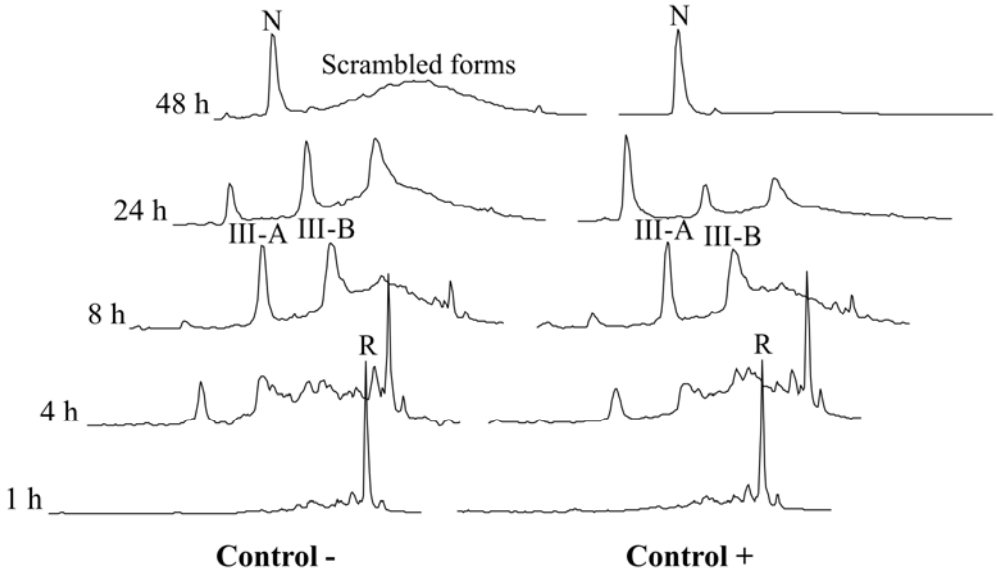


Fig. 1. RP-HPLC analysis of the folding intermediates of LCI trapped by acidification. Oxidative folding was carried out at 22ºC in Tris-HCl buffer (0.1 M, pH 8.4) with (Control +) or without (Control -) the redox agent 2-mercaptoethanol (0.25 mM). The protein concentration was 0.5 mg/mL. Folding intermediates were trapped at the noted times by acidification and analyzed by RP-HPLC using the conditions described in "Experimental Procedures". *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 intermediates containing three native disulfides and s*crambled forms* are isomers with four disulfides.

To isolate III-A, the reaction was acid-quenched approximately after 8 h of refolding in the absence of 2-mercaptoethanol to avoid the presence of possible traces of scrambled isomers. The intermediate was separated by RP-HPLC and freeze-dried after verification of its purity by treatment with vinylpyridine and analysis by matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF MS). The ^{1}H NMR spectrum of III-A also confirmed the homogeneity of the preparation (see Fig. 1S in "Supplementary Information"). The chemical shift dispersion of the III-A ^{1}H spectrum at pH 3.5 is typical for globular proteins and suggests a properly folded conformation at this pH. A low pH is a necessary condition to maintain the III-A folding intermediate in a "quenched" state.

Resonance Assignments — ^{1}H and ^{15}N NMR assignments were performed by the standard sequence-specific method on the basis of heteronuclear 2D and 3D spectra of the uniformly ^{15}N-labeled native LCI and the III-A folding intermediate recorded in H_2O at 27ºC and pH 3.5. The 2D ^{1}H-^{15}N HSQC spectra of both native LCI and III-A show excellent chemical shift dispersion with few overlaps (Fig. 2). Excluding the N-terminal residues 1-4, which are highly flexible, and the nine Pro residues, all 54 backbone NH cross-peaks could be observed in the ^{1}H-^{15}N HSQC spectrum of the native LCI. For III-A, there are 53 cross-peaks and only residue Val51 is not observed, probably due to an increased flexibility in this region (see next sections). All backbone cross-peaks were readily assigned to their corresponding spin system after sequential assignment. The list of these assignments is given in the Table 1S of "Supplementary Information".

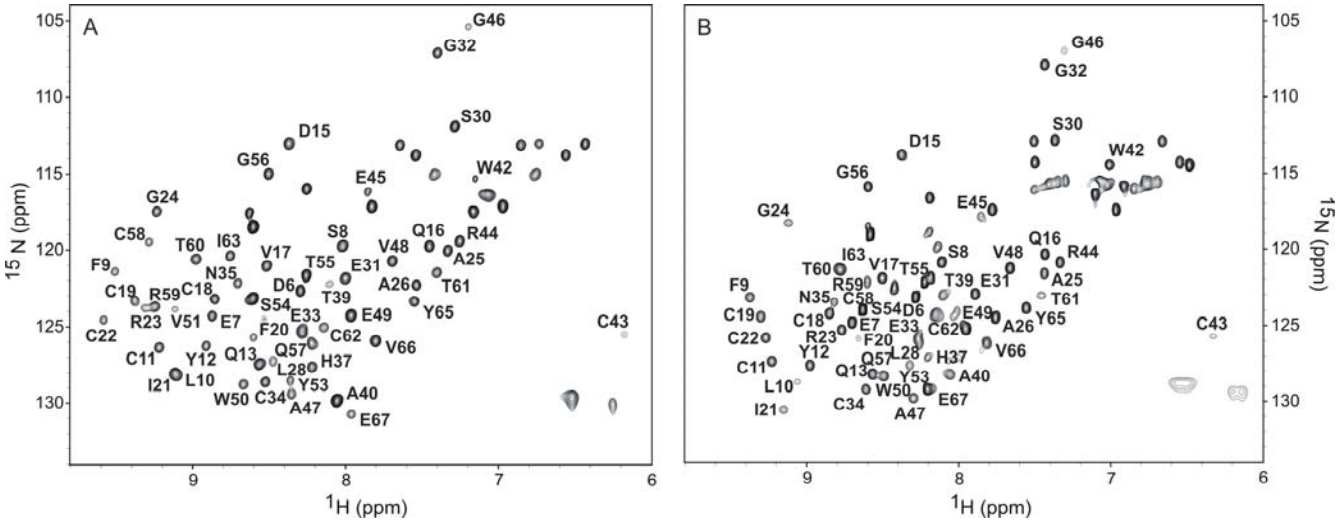


Fig. 2. 2D ^{1}H-^{15}N-HSQC spectra of uniformly ^{15}N-labeled native LCI and III-A intermediate. Backbone cross-peak assignments are indicated for both native LCI (A) and III-A (B). The experiments were carried out with 1 mM protein concentration in H_2O/D_2O (9:1) at pH 3.5 and 27ºC.

Spin systems were sequentially connected throughout $d_{\alpha N(i,i+1)}$, $d_{NN(i,i+1)}$, and/or $d_{\beta N(i,i+1)}$ nuclear Overhauser enhancements (NOEs) observed in the 2D NOESY and 3D NOESY-HSQC. The whole stretch of sequential connectivities at the spectra was only broken at the nine Pro residues for native LCI and also at residue Val51 for III-A. The sequential assignment paths of a region of the 2D NOESY spectra of native LCI and III-A are shown in Figure 2S of the "Supplementary Information". It is worth mentioning that for native LCI most of the proton resonances are coincident with the previous assignment

reported by our group at pH 6.5 (14). The ^{1}H and ^{1}H-^{15}N HSQC spectra of the III-A intermediate did not change along the experiment (data not shown), indicating that, as expected, intramolecular disulfide rearrangements did not occur to a significant extent at the working pH.

Position of most cross-peaks in the ^{1}H-^{15}N HSQC and 2D NOESY spectra of native LCI and III-A was similar (Fig. 2 and 2S). However, on the overall smaller cross-peak dispersion was observed for the folding intermediate. This indicates a less compact conformation as compared to the native protein. Individual protein chemical shifts depend strongly on local chemical and electronic environments. Therefore, direct comparisons of chemical shift data between native LCI and III-A can elucidate local areas within the intermediate where differences are present. Accordingly, a comparison of assigned backbone NH and $C^{\alpha}H$ was carried out (see Fig. 3S in "Supplementary Information"). As might be expected, the largest differences are centered in the two regions encompassing Cys22-Arg23 and Gln57-Thr60. This can be unequivocally attributed to the absence of the Cys22-Cys58 disulfide bond in the intermediate.

Secondary Structure Analysis — The elements of secondary structure in native LCI and III-A were first delineated using information provided by the NOE data (27, 28). For native LCI, a five-stranded antiparallel β-sheet was identified by strong long-range interstrand $d_{\alpha\alpha(i,j)}$, $d_{\alpha N(i,j)}$ NOEs, medium intensity interstrand $d_{NN(i,j)}$ NOEs, and strong sequential $d_{\alpha N(i,i+1)}$ NOEs with weak or absent $d_{NN(i,i+1)}$ NOEs. This antiparallel β-sheet involves residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), Val51-Tyr53 (β4) and Gly56-Ile63 (β5) (Fig. 3). A short stretch of strong sequential $d_{NN(i,i+1)}$ NOEs, weak or absent $d_{\alpha N(i,i+1)}$ NOEs, and medium range $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+3)}$, and $d_{\alpha N(i,i+4)}$ NOEs indicated the presence of a short α-helix (α1) located between residues Pro41 and Gly46.

For the III-A intermediate, a four-stranded antiparallel β-sheet was identified by strong sequential $d_{\alpha N(i,i+1)}$ NOEs with weak or absent $d_{NN(i,i+1)}$ NOEs involving residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), and Cys58-Ile63 (β5') (Fig. 3). Strong long-range interstrand $d_{\alpha\alpha(i,j)}$ contacts were found between β3-β1 (Cys34-Cys11), β1-β2 (Ser8-Ile21, Leu10-Cys19), and β2-β5' (Cys18-Cys62). Strong long-range interstrand $d_{\alpha N(i,j)}$ NOEs were detected between β3-β1 (His37-Phe9), β1-β2 (Ser8-Cys22, Cys11-Cys19), and β2-β5' (Gln16-Tyr65, Cys18-Ile63). Finally, medium intensity interstrand $d_{NN(i,j)}$ NOEs were identified between β3-β1 (Glu33-Tyr12, Asn35-Leu10), β1-β2 (Glu7-Cys22, Phe9-Phe20, Cys11-Cys18, Gln13-Gln16), and β2-β5' (Val17-Ile63,

Cys19-Thr61, Ile21-Arg59). All these contacts were also present in the spectra of the native LCI defining the antiparallel strands β3-β1-β2-β5.

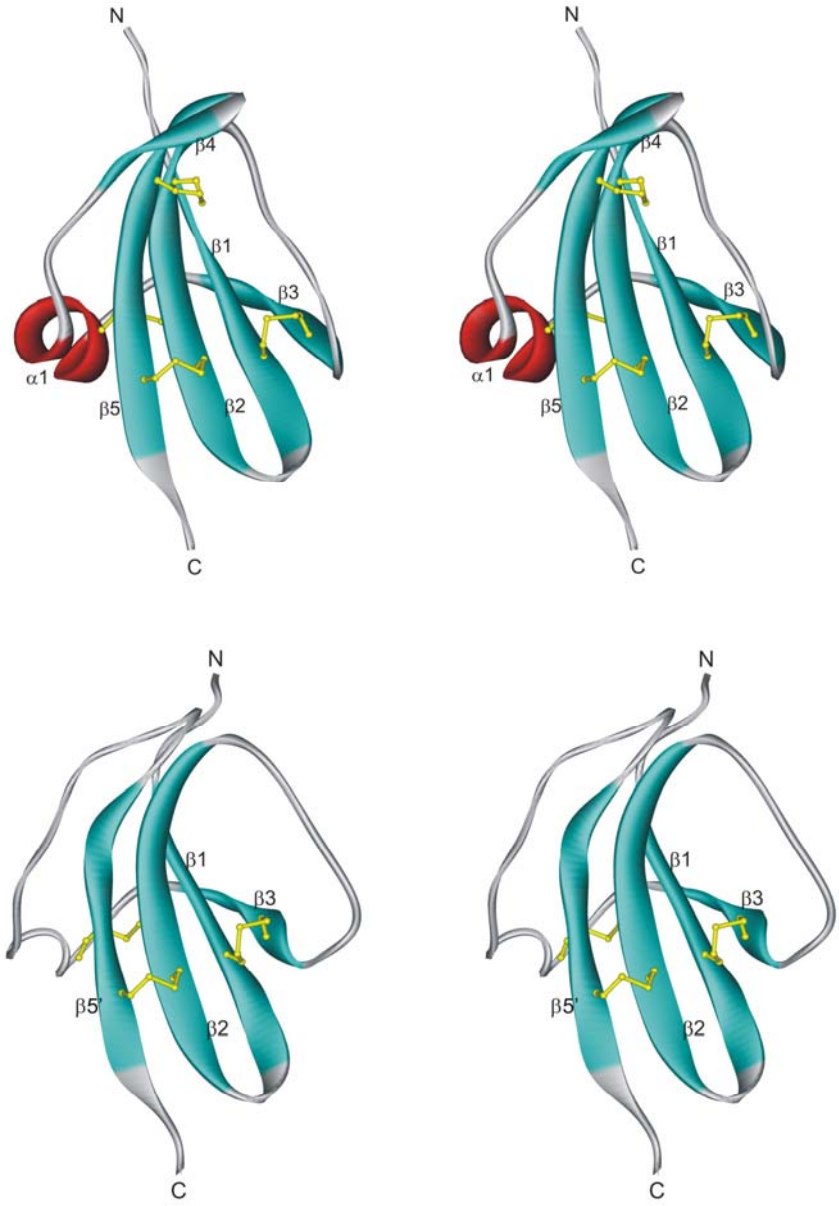


Fig. 3. Stereoview of the structure of native LCI and III-A intermediate. Ribbon representation of the calculated structure for native LCI (*upper panel*) and III-A (*lower panel*). β-strands are shown in blue and the α-helix in red. *N* and *C* indicate the location of the N- and C-terminal tails of both proteins. The disulfide bonds are shown in yellow.

Three main differences in terms of secondary structure could be observed between native LCI and the III-A intermediate. First, an increased flexibility is present in the region between residues Glu49-Thr55 as compared to that in the native protein. Proton resonances

of residue Val51 cannot be observed in the spectra of III-A. In addition, NH-NH contacts observed in native LCI between β4-β5 are lost in the intermediate (e.g. Val51-Cys58 and Tyr53-Gly56). This probably precludes the formation in III-A of the short strand present in the native form (β4). Second, the strand β5' of III-A (located in the C-terminus) is significantly shorter than the corresponding strand in the native LCI (β5); in the intermediate some contacts are lost between β2 and β5' (e.g. the NH-NH contact between Arg23-Gln57). Despite these missing contacts, the C-terminus of the intermediate is far from being free; the β5' strand in the intermediate is large and stable enough to keep the C-terminus tail in a proper orientation, providing the conformational rigidity needed for the binding to the carboxypeptidase. This explains why the inhibitory capability of III-A is indistinguishable from that of the native LCI for all tested carboxypeptidases (19). The other remarkable difference is the absence of a well-defined α-helix in the intermediate. However, the region between residues Pro41-Gly46 is very similar to a short α-helix. Some strong $d_{NN(i,i+1)}$ and medium $d_{NN(i,i+2)}$ NOEs are found here (e.g. Cys43-Arg44, Arg44-Glu45, Trp42-Arg44), together with weak or absent $d_{\alpha N(i,i+1)}$ NOEs.

Three-Dimensional Structure Calculation — The backbone structures of native LCI and III-A intermediate were calculated using the simulated annealing method with the program CNS (29). The results of these calculations are given in the Table 2S of "Supplementary Information". For native LCI, with the exception of the five N-terminal and the three C-terminal residues, the ensemble of 20 calculated structures is well defined (Fig. 4), with an average backbone root mean square deviation (rmsd) of 1.40 Å (residues 6-64). This rmsd value is representative of a good-quality backbone structure determination. Figure 5 indicates the rmsd of each residue in the bundle of 20 structures, thus showing the most flexible regions in native LCI. The three-dimensional structure of this molecule at pH 3.5 is similar to that calculated previously at pH 6.5 (14): a five-stranded antiparallel β-sheet with a β3-β1-β2-β5-β4 topology, and a short α-helix that packs onto the most compact part of the β-structure interacting with the end and the beginning of the β1 and β2 strands, respectively (Fig. 3 and 4). A high percentage of residues belong to regular secondary structure elements (nearly 45%), which are cross connected and stabilized by the presence of four disulfide bridges: Cys11-Cys34 (β1-β3), Cys18-Cys62 (β2-β5), Cys19-Cys43 (β2-α1), and Cys22-Cys58 (β2-β5). They provide high stability and compactness to the protein.

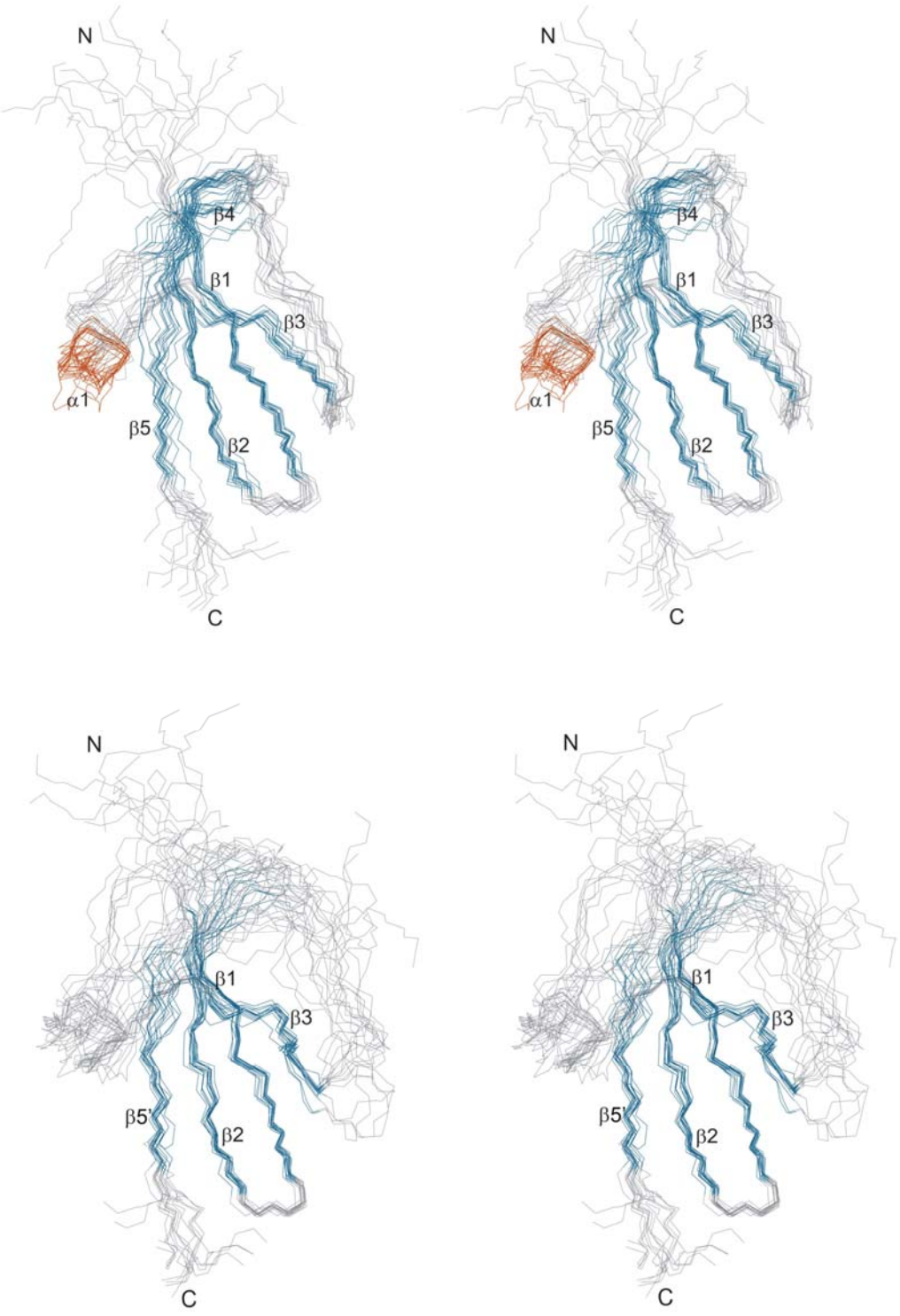


Fig. 4. Stereoview with the superposition of the calculated structures for native LCI and III-A intermediate. The pictures show the backbone atoms of the bundle of lowest energy structures calculated for native LCI (*upper panel*) and III-A (*lower panel*). β-strands are indicated in blue and the α-helix in red. The N- and C-terminal tails are labeled.

The calculation of the structure of the III-A intermediate confirms that it also possesses a well-defined globular conformation that includes a four-stranded antiparallel β-sheet with a β3-β1-β2-β5' topology (Fig. 3). However, some parts of this molecule are more flexible than in native LCI (Fig. 4). This is shown by the higher average backbone rmsd value for residues 6-64 of the 20 calculated structures: 2.47 Å. The rmsd of each residue is shown in Figure 5 and provides evidence for the presence of highly variable flexible regions, mainly between residues Arg23-Gly32 and Arg44-Gln57. The disulfide pairings of III-A were unambiguously determined during three-dimensional structure calculations and were in complete agreement with a previous assignment carried out by digestion of the vinylpyridine-derivatized intermediate with thermolysin and analysis of the resulting disulfide-containing peptides by MALDI-TOF MS and automated Edman degradation (19): Cys11-Cys34 (β1-β3), Cys18-Cys62 (β2-β5'), and Cys19-Cys43 (β2-"α"). The missing disulfide bond present in the native form between Cys22 and Cys58 seems to account for the lower compactness of such intermediate as compared to that of the native protein. However, the structural similarities between the native protein and the III-A intermediate are striking.

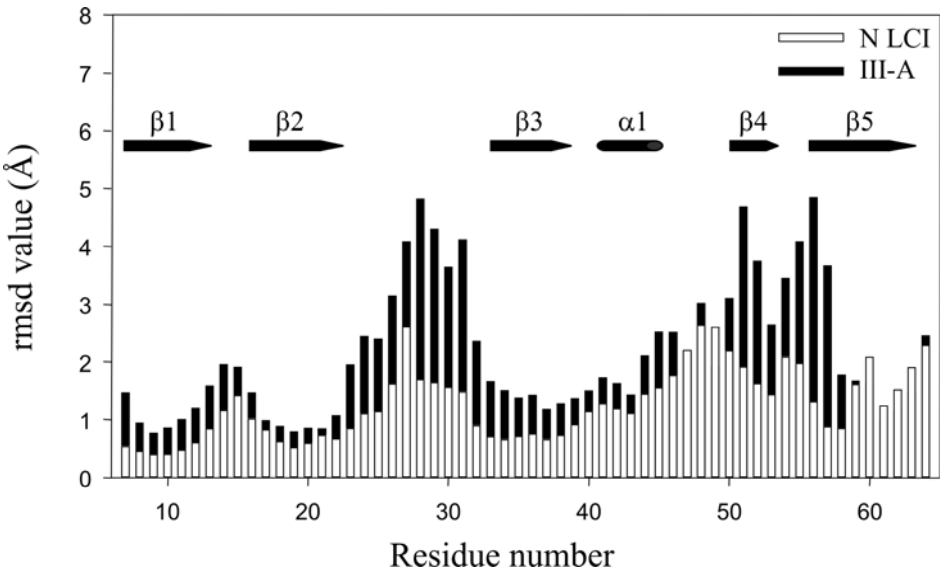


Fig. 5. Comparison of local rmsd values for backbone atoms of native LCI and III-A. The rms deviations of the backbone atoms from the 20 calculated structures are plotted vs the residue number (residues 6-64). Secondary structure elements for native LCI are indicated inside the graphic.

Understanding the Role of the III-A Intermediate in the Folding of LCI — In a recent study we hypothesized that the way the oxidative folding pathway of LCI proceeds depends on the ability of its secondary structure elements to protect progressively native disulfide bonds from rearrangement in the interior of a totally or partially folded structure (19). This would result in a final native structure in which disulfide bonds should be highly protected. Our view is strongly reinforced here by the results of amide proton exchange experiments on native LCI and III-A intermediate. Both forms were dissolved in D_2O at pH 3.5 and several ^{1}H-^{15}N HSQC spectra were recorded over time. For native LCI, maximum exchange was achieved after approximately 10 h and the spectra did not significantly change after that time point. The following residues were found protected: Phe9-Gln13 (β1), Gln16-Arg23 (β2), Gly32, Asn35 and His37 (β3), Thr39, Cys43 (α1), Val51 (β4), Gln57-Arg59 and Thr60-Ile63 (β5), and Tyr65 (see Fig. 4S in "Supplementary Information"). Thus, the protected residues are located within all secondary structure elements around the cysteine residues and this fact clearly indicates that in native LCI the four disulfide bonds are buried and not solvent-accessible.

For the III-A intermediate, after 15 min of exchange 35 residues from all secondary structure elements were found protected (in the native form were 36 residues). As expected, residues located in the highly flexible regions were already exchanged at this early time point (see Fig. 4S in "Supplementary Information"). Similarly to the native form, in the intermediate maximum exchange was achieved after about 10 h. The protected residues were Phe9-Gln13 (β1), Gln16-Cys19 and Ile21-Cys22 (β2), Asn35 (β3), Thr60 and Ile63 and Tyr65 (β5'). Thus, in III-A both free cysteines (Cys 22 and 58) and the three disulfide bonds are located in protected regions or close to them and therefore are not solvent-accessible or have limited accessibility. The final number of protected residues for native LCI and the intermediate (27 and 15 residues, respectively) is in good agreement with those previously found measuring the global D/H exchange by MALDI-TOF MS (26 and 16 residues, respectively) (19). This approximately 40% of decrease in protected residues between native LCI and III-A, mainly localized in the "missing" secondary structure elements of the latter (α1, β4 and β5'), is a reflection of its lower level of conformational packing.

Analogs of BPTI and RNase A intermediates that lack only one native disulfide bond and act as major kinetic traps along the folding process display structures similar to the native ones (40, 46-48). Here we demonstrate that this is also the case of the III-A

intermediate of LCI. However, critical differences among the folding pathways of these proteins are seen. For both BPTI and RNase A, the last step in their folding reactions consists of a direct oxidation of two free cysteines to form the native structure. In LCI, however, a major disorganization of the acquired tertiary structure (in III-A) to form a heterogeneous ensemble of non-native 4-disulfide (scrambled) isomers is necessary before the native structure can be gained. This was shown by stop/go experiments using the III-A intermediate and scrambled isomers (19). We have previously postulated that III-A is a metastable form equivalent to what Scheraga and co-workers had defined as *disulfide-insecure* intermediates (49). In these forms, the free thiol groups are as well protected as their disulfide bonds; therefore such thiols cannot be simply exposed and oxidized by a local unfolding process as it happens with BPTI or RNase A on-pathway intermediates. Structural fluctuations that expose the thiol groups are also likely to expose the disulfide bonds and promote their reshuffling instead of oxidation of the free thiols to the native pairings. The data from this work, together with information from the stop/go assays confirm that indeed III-A acts as a *disulfide-insecure* intermediate. Unlike in III-A, the four disulfide bonds of the unstructured scrambled isomers are solvent-accessible, therefore the addition of an external thiol strongly accelerates their kinetics of reshuffling to attain native disulfide pairings and the native structure.

In BPTI or RNase A analogs, deletion of the disulfide bond introduces only small and localized structural perturbations with respect to the native form, in the vicinity of the eliminated bond (40, 46-48). In contrast, long-range effects are detected in III-A: the α-helix and β4 strand present in the native form disappear in the intermediate and the loop connecting strands β2 and β3 is distorted. In addition, a higher degree of flexibility is detectable in some protein regions when compared to those of native LCI. One of these flexible stretches is the region around Cys43, even if in the intermediate this residue is bonded to Cys19. Interestingly, the other major kinetic trap of LCI folding, III-B, lacks this disulfide bond. We have previously shown by stop/go experiments that there is an inter-conversion of intermediates III-A and III-B to each other, reaching equilibrium prior to the formation of scrambled isomers (19). This disulfide interchange is an internal process in which all reaction groups are protected from the exterior, since neither the rate of inter-conversion nor the concentration of the species at the equilibrium are affected by the presence of external thiols. Minor local fluctuations due to an increase of flexibility in the

backbone of the involved regions, such as the one observed here, may be well behind this solvent-independent disulfide interchange.

Predicting the Folding Pathway of LCI — As concluded above, the III-A intermediate, representing the major kinetic trap in the oxidative folding process of LCI, is a highly structured molecule. Several groups have hypothesized that the folding of disulfide-rich proteins hinges critically on their secondary structure organization (19, 50). In order to provide additional evidences for this view, we tested whether a folding predicting algorithm, which only uses the native structure of the protein as an input and whose calculations are mainly based on topological considerations, may be able to predict the presence and the structural properties of the highest energy barrier in the folding of LCI. For this task we used the Fold-X software developed by Serrano's group at the EMBL (6, 32, 33). This algorithm, which takes into account also sequence features underlying protein topology, has been shown to predict even subtle differences within protein folding pathways. In Figure 6A we show the probability for LCI residues to be ordered during the folding process. The first β-hairpin, formed by β1, β2 and the turn connecting them, exhibits the highest probability to be ordered at the beginning of the folding reaction. When the number of ordered residues increases (up to 55%) folding of β3 and β5 also reaches high probability. The association of these two regions corresponds sharply to the native-like regions in the III-A intermediate. Conversely, the fourth β-strand as well as the α-helix exhibit very low propensity to fold. Thus, the hierarchy of structure formation predicted using the Fold-X algorithm provides a description of the folding process of LCI consistent with the experimental data and predicts the main structural features of the highest energy barrier in the folding reaction. Since topology is characterized by the way secondary structure elements are connected in the sequence and are organized together, and the Fold-X algorithm does not use the disulfide bonds as an input, this successful prediction indicates that secondary structure plays a crucial role in the folding of LCI.

Predicting Folding Diversity in Disulfide-Rich Proteins — Application of the oxidative folding technique has allowed elucidation of folding pathways of several disulfide-rich proteins. Of special interest is the group of 3-disulfide proteins that includes the extensively investigated model BPTI (10), as well as potato carboxypeptidase inhibitor (PCI) (51), epidermal growth factor (EGF) (52), etc. The studies have shown that even among these small 3-disulfide proteins the folding mechanisms may vary substantially. These differences are illustrated by (a) the extent of heterogeneity of folding intermediates;

(b) the predominance of intermediates containing native disulfide bonds; and (c) the accumulation of 3-disulfide scrambled isomers as intermediates. The underlying causes of this diversity have been postulated to be associated with the fashion of how local structural elements of disulfide proteins are being stabilized (50).

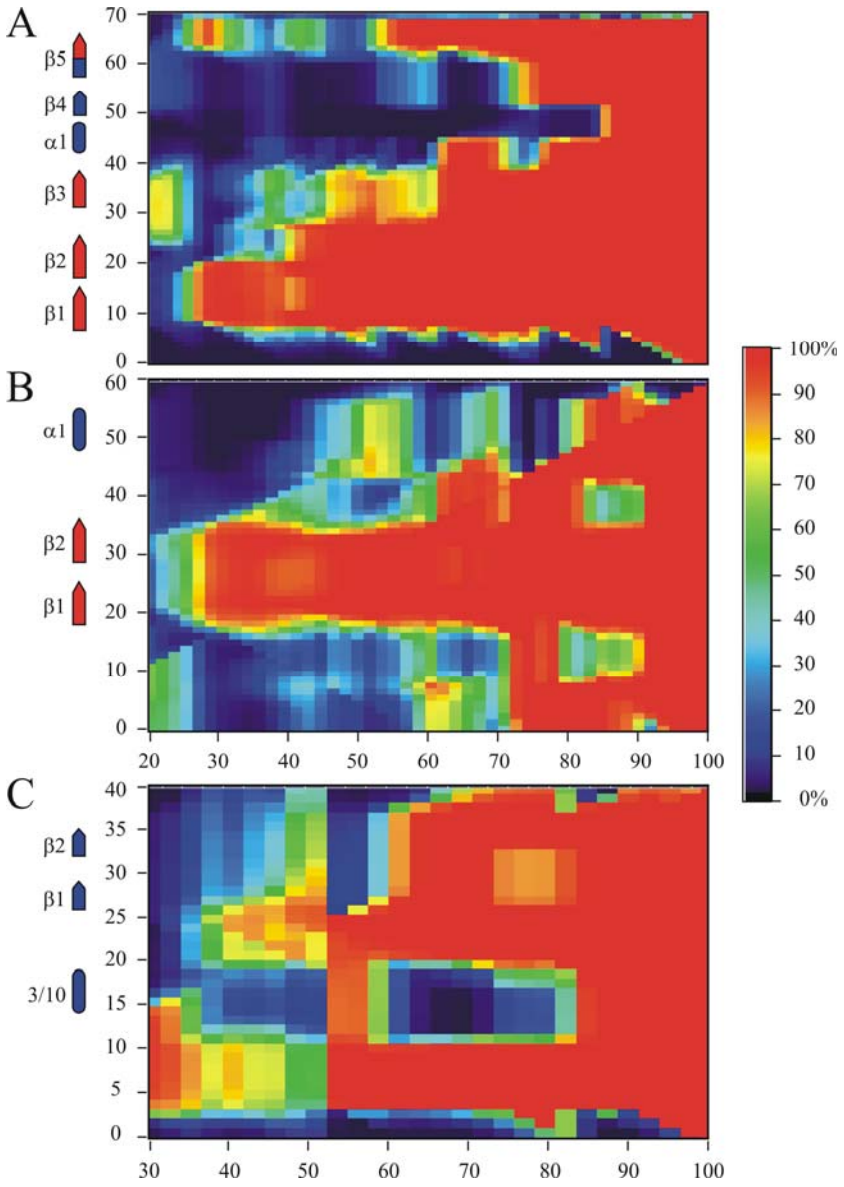


Fig. 6. Folding pathways of LCI, BPTI and PCI predicted with Fold-X. Hierarchy of structure formation for LCI (A), BPTI (B) and PCI (A) computed with Fold-X software. The residue number is reported in the vertical axis. The horizontal axis corresponds to percentage of ordered residues. The secondary structure elements of the native proteins are shown on the left side. The color code reflects the probability for a residue to be ordered during the folding process (dark blue corresponds to 0% and red to 100%).

In this work, we have determined the theoretical folding pathways of two 3-disulfide proteins exhibiting extremely different folding mechanisms to see whether the Fold-X algorithm could capture distinctive features lying beneath this folding diversity. BPTI is an example of a disulfide-rich protein in which a limited number of 1- and 2-disulfide intermediates are observed during its folding reaction. Most importantly, only intermediates adopting native disulfides and native-like structures dominate the folding of BPTI. In Figure 6B we show the predicted probability for BPTI residues to be ordered during the folding process. The β-hairpin comprising residues 19-34 exhibits the highest probability to be ordered at the beginning of the folding process; only after more than 65% of the residues are ordered the rest of the secondary elements consolidates the structure. Thus, the β-hairpin clearly nucleates the folding in the predicted pathway. To the best of our knowledge BPTI is the only disulfide-rich protein for which the folding nucleus has been determined experimentally. In excellent agreement with the theoretical data reported here, this was shown to be located in the β-hairpin of BPTI folding intermediates containing either one or two native disulfide bonds (53). In addition, in stable analogs of these species prepared by chemical trapping or protein engineering and characterized by NMR, this β–hairpin appears to be already formed (54).

In contrast to BPTI, PCI is a 3-disulfide protein with little secondary structure whose folding proceeds through an initial stage of non-specific disulfide formation (packing), followed by disulfide reshuffling (consolidation) of partially packed 3-disulfide scrambled isomers to acquire the native structure (51). No predominant intermediates populate the folding pathway of this small protein. The Fold-X analysis reveals that no specific protein region displays a high probability to be ordered at the beginning of the folding reaction (Fig. 6C). Only after more than 60% of the protein is ordered the protein structure begins to consolidate in a concerted manner. The absence of a protein region able to nucleate the folding process agrees with the experimental data in which an absence of major folded intermediates is observed, and provides an explanation for the accumulation of a population of scrambled species with similar probabilities to fold into the native structure in the last phase of the folding reaction. Thus, a simple energy function like the one incorporated in Fold-X appears to be enough to capture the main properties of the dissimilar folding processes displayed by LCI, BPTI and PCI. These results reinforce the hypothesis that, as it occurs with small proteins without disulfide bonds, the way a small disulfide-rich protein folds is critically influenced by its secondary structure organization.

It also suggests that a large component of the observed heterogeneity during a particular folding event is determined topologically.

Conclusions

The III-A intermediate constitutes the major kinetic trap in the oxidative folding of LCI. In this work, we have determined its structure by NMR, which represents one of the few available structures of folding intermediates isolated directly from a folding reaction. Despite the fact that this intermediate lacks a native disulfide bond, we show that it is a highly structured molecule with striking structural similarity to the native state. Comparison of native and intermediate structures allows deciphering why III-A accumulates along the folding reaction: it acts as a *disulfide insecure* intermediate, which protects both their native disulfide bonds and free cysteine residues from rearrangement and direct oxidation, respectively, in the interior of a highly folded protein conformation. Although III-A is a fully functional form that is formed quickly and efficiently in the LCI folding pathway, a conformational search for the formation of the last disulfide bond takes place while losing most of the tertiary structure already gained in this intermediate. The results of this study, together with previous stability data (19) demonstrate that the fourth disulfide bond restricts conformational flexibility allowing a net gain in stability and structural specificity to the native form. This assumption makes sense taking into account that LCI is a protease inhibitor from leech saliva evolved to act in blood. In addition, here we show that theoretical approximations based on topological constrains predict accurately the main characteristics of the folding pathway of LCI and those of other proteins such as BPTI or PCI. The overall data provide direct evidence for the importance of native-like interactions between elements of secondary structure in directing the folding of disulfide-rich proteins.

Acknowledgments — We thank Dr. R. Guerois for kindly revising this manuscript and Dr. S. Bronsoms for helpful discussions.

The atomic coordinates of the resulted 20 energy-minimized conformers for the native LCI and the III-A intermediate have been deposited with the Protein Data Bank, accession codes YXXX and YXXX, respectively.

References

1. Creighton TE, Darby NJ & Kemmink J. The roles of partly folded intermediates in protein folding. *Faseb J* 1996; 10: 110-118
2. Creighton TE. How important is the molten globule for correct protein folding? *Trends Biochem Sci* 1997; 22: 6-10
3. Alm E & Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA* 1999; 96: 11305-11310
4. Galzitskaya OV & Finkelstein AV. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci USA* 1999; 96: 11299-11304
5. Munoz V & Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 1999; 96: 11311-11316
6. Guerois R & Serrano L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J Mol Biol* 2000; 304: 967-982
7. Clementi C, Jennings PA & Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1beta. *Proc Natl Acad Sci USA* 2000; 97: 5871-5876
8. Clementi C, Nymeyer H & Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 2000; 298: 937-953
9. Creighton TE. Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 1986; 131: 83-106
10. Weissman JS & Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991; 253: 1386-1393
11. Scheraga HA, Wedemeyer WJ & Welker E. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol* 2001; 341: 189-221
12. Radford SE, Dobson CM & Evans PA. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 1992; 358: 302-307
13. Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998; 273: 32927-32933
14. Reverter D, Fernandez-Catalan C, Baumgartner R, Pfänder R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000; 7: 322-328
15. Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-27181

16. Bouma BN & Meijers JCM. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
17. Salamanca S, Villegas V, Vendrell J, Li L, Aviles FX & Chang JY. The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* 2002; 277: 17538-17543
18. Salamanca S, Li L, Vendrell J, Aviles FX & Chang JY. Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 2003; 42: 6754-6761
19. Arolas JL, Bronsoms S, Lorenzo J, Aviles FX, Chang JY & Ventura S. Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. *J Biol Chem* 2004; 279: 37261-37270
20. Peranen J, Rikkonen M, Hyvonen M & Kaariainen L. T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem* 1996; 236: 371-373
21. Rance M. Improved techniques for homonuclear rotating frame and isotropic mixing experiments. *J Magn Reson* 1987; 74: 557-564
22. Kumar A, Ernst RR & Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun* 1980; 95: 1-6
23. Piotto M, Saudek V & Sklenar V. Gradient-tailored excitation for single quantum NMR spectroscopy in aqueous solution. *J Biomol NMR* 1992; 6: 661-665
24. Bodenhausen G & Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 1980; 69: 185-189
25. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM & Clore GM. Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1 β. *Biochemistry* 1989; 28: 6150-6156
26. Goddard TD & Kneller DG. "SPARKY 3". University of California, San Francisco 2000
27. Wüthrich K, Billeter M & Braun W. Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *J Mol Biol* 1984; 180: 715-740
28. Wüthrich K. NMR of Proteins & Nucleic Acids. Wiley, New York 1986
29. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T & Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog D* 1998; 54: 905-921
30. Holak TA, Gondol D, Otlewski J & Wilusz T. Determination of the complete 3-dimensional structure of the trypsin-inhibitor from squash seeds in aqueous-solution by nuclear magnetic-resonance and a combination of distance geometry and dynamical simulated annealing. *J Mol Biol* 1989; 210: 635-648

31. Schnuchel A, Wiltscheck R, Czisch M, Herrler M, Willimsky G, Graumann P, Marahiel MA & Holak TA. Structure in solution of the major cold shock protein from *Bacillus Subtilis*. *Nature* 1993; 364: 169-171

32. Guerois R, Nielsen JE & Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 2002; 320: 369-387

33. Kiel C, Serrano L & Herrmann C. A detailed thermodynamic analysis of ras/effector complex interfaces. *J Mol Biol* 2004; 340: 1039-1058

34. Plaxco KW, Simons KT & Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 1998; 277: 985-994

35. van Mierlo CPM, Darby NJ, Neuhaus D & Creighton TE. Two-dimensional ^{1}H nuclear magnetic resonance study of the (5-55) single-disulfde folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol* 1991; 222: 373-390

36. van Mierlo CPM, Darby NJ, Keeler J, Neuhaus D & Creighton TE. Partially folded conformation of the (30-51) intermediate in the disulfde folding pathway of bovine pancreatic trypsin inhibitor, ^{1}H and ^{15}N resonance assignments and determination of backbone dynamics from ^{15}N relaxation measurements. *J Mol Biol* 1993; 229: 1125-1146

37. Noda Y, Yokota A, Horii D, Tominaga T, Tanisaka Y, Tachibana H & Segawa S. NMR structural study of two-disulfide variant of hen lysozyme: 2SS [6-127, 30-115] – A disulfide intermediate with a partly unfolded structure. *Biochemistry* 2002; 41: 2130-2139

38. Yokota A, Hirai K, Miyauchi H, Iimura S, Noda Y, Inoue K, Akasaka K, Tachibana H & Segawa S. NMR characterization of three-disulfide variants of lyzozyme, C64A/C80A, C76A/C94A, and C30A/C115A – A marginally stable state in folded proteins. *Biochemistry* 2004; 43: 6663-6669

39. Talluri S, Rothwarf DM & Scheraga HA. Structural characterization of a three-disulfide intermediate of ribonuclease A involved in both the folding and unfolding pathways. *Biochemistry* 1994; 33: 10437-10449

40. Shimotakahara S, Rios CB, Laity JH, Zimmerman DE, Scheraga HA & Montelione GT. NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine pancreatic ribonuclease A: automated analysis of 1H, 13C, and 15N resonance assignments for wild-type and [C65S, C72S] mutant forms. *Biochemistry* 1997; 36: 6915-6929

41. Iwaoka M, Juminaga D & Scheraga HA. Regeneration of three-disulfide mutants of bovine pancreatic ribonuclease A missing the 65-72 disulfide bond: characterization of a minor folding pathway of ribonuclease A and kinetic roles of Cys65 and Cys72. *Biochemistry* 1998; 37: 4490-4501

42. Gehrmann J, Alewood PF & Craik DJ. Structure determination of the three disulfide bond isomers of alpha-conotoxin GI: a model for the role of disulfide bonds in structural stability. *J Mol Biol* 1998; 278: 401-415

43. van den Berg B, Chung EW, Robinson CV & Dobson CM. Characterisation of the dominant oxidative folding intermediate of hen lysozyme. *J Mol Biol* 1999; 290: 781-796

44. Eliezer D, Chung J, Dyson HJ & Wright PE. Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* 2000; 39: 2894-2901

45. Sato A, Koyama S, Yamada H, Suzuki S, Tamura K, Kobayashi M, Niwa M, Yasuda T, Kyogoku Y & Kobayashi Y. Three-dimensional solution structure of a disulfide bond isomer of the human insulin-like growth factor-I. *J Pept Res* 2000; 56: 218-230

46. Hurle MR, Eads CD, Pearlman DA, Seibel GL, Thomason J, Kosen PA, Kollman P, Anderson S & Kuntz ID. Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance. *Protein Sci* 1992; 1: 91-106

47. van Mierlo CP, Kemmink J, Neuhaus D, Darby NJ & Creighton TE. 1H NMR analysis of the partly-folded non-native two-disulphide intermediates (30-51,5-14) and (30-51,5-38) in the folding pathway of bovine pancreatic trypsin inhibitor. *J Mol Biol* 1994; 235: 1044-1061

48. Laity JH, Lester CC, Shimotakahara S, Zimmerman DE, Montelione GT & Scheraga HA. Structural characterization of an analog of the major rate-determining disulfide folding intermediate of bovine pancreatic ribonuclease A. *Biochemistry* 1997; 36: 12683-12699

49. Welker E, Narayan M, Wedemeyer WJ & Scheraga HA. Structural determinants of oxidative folding in proteins. *Proc Natl Acad Sci USA* 2001; 98: 2312-2316

50. Chang JY. Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry* 2004; 43: 4522-4529

51. Chang JY, Canals F, Schindler P, Querol E & Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 1994; 269: 22087-22094

52. Chang JY, Li L & Lai PH. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 2001; 276: 4845-4852

53. Bulaj G & Goldenberg DP. Phi-values for BPTI folding intermediates and implications for transition state analysis. *Nat Struct Biol* 2001; 8: 326-330

54. Staley JP & Kim PS. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *Proc Natl Acad Sci USA* 1992; 89: 1519-1523

Supplementary Information

Table 1S. ^{15}N, HN and αH backbone assignments (in ppm) for the native form of LCI and the III-A intermediate

	N LCI			III-A		
Residue	^{15}N	HN	αH1/2	^{15}N	HN	αH1/2
P5	-	-	4.59	-	-	4.46
D6	122.7	8.30	5.00	122.4	8.29	5.06
E7	124.4	8.88	4.61	124.2	8.71	4.65
S8	119.6	8.02	5.75	120.0	8.13	5.61
F9	121.4	9.52	5.29	122.5	9.38	5.27
L10	128.5	9.11	5.02	128.4	9.08	5.03
C11	126.6	9.22	5.47	126.9	9.24	5.53
Y12	126.5	8.92	4.59	127.2	8.99	4.57
Q13	127.8	8.57	4.91	127.9	8.58	4.92
D15	112.5	8.38	5.09	112.4	8.39	5.11
Q16	119.6	7.46	4.62	119.4	7.45	4.62
V17	121.0	8.53	4.79	121.1	8.52	4.79
C18	123.3	8.86	5.64	123.6	8.86	5.62
C19	123.4	9.38	5.82	123.8	9.31	5.66
F20	124.7	8.53	5.10	125.5	8.67	4.93
I21	128.5	9.13	4.75	130.4	9.17	4.70
C22	124.7	9.59	6.09	125.3	9.28	5.34
R23	123.9	9.31	4.91	124.7	8.78	4.86
G24	117.2	9.24	4.02/3.83	117.2	9.14	4.03/3.81
A25	120.0	7.34	4.37	120.8	7.45	4.41
A26	122.4	7.55	4.20	123.9	7.77	4.46
L28	128.9	8.37	4.58	126.9	8.34	4.57
S30	111.4	7.29	4.18	111.4	7.38	4.20
E31	121.9	8.01	4.45	122.2	7.91	4.47
G32	106.3	7.40	3.66/4.45	106.1	7.45	3.68/4.44
E33	125.5	8.30	4.47	125.4	8.28	4.62
C34	129.0	8.53	6.04	129.0	8.62	5.97
N35	122.2	8.71	5.44	122.8	8.83	5.51
H37	128.0	8.23	3.88	126.6	8.20	3.92
T39	122.2	8.12	4.03	122.3	8.12	3.90
A40	130.3	8.06	3.75	127.9	8.07	3.66
W42	114.9	7.16	4.22	113.1	7.04	4.28
C43	125.5	6.16	3.81	125.2	6.35	3.91
R44	119.3	7.26	3.64	120.0	7.35	3.76
E45	115.9	7.85	3.98	116.8	7.86	4.02
G46	104.4	7.20	3.77/4.33	105.2	7.35	3.79/4.21
A47	129.9	8.37	4.35	129.6	8.32	4.40
V48	120.7	7.70	4.03	120.4	7.68	4.02
E49	124.4	7.97	4.56	124.6	7.97	4.54
W50	129.2	8.68	4.38	128.0	8.51	4.44
V51	123.9	9.12	4.89	-	-	-
Y53	129.0	8.37	4.19	127.2	8.33	4.30
S54	123.2	8.61	3.76	123.3	8.66	3.76
T55	121.7	8.26	4.10	121.1	8.20	4.31
G56	114.6	8.51	3.89/4.31	114.7	8.62	3.92/3.79
Q57	127.6	8.48	5.53	127.7	8.50	4.76
C58	119.4	9.30	6.17	123.2	8.64	4.43
R59	123.8	9.26	4.75	121.4	8.61	4.01
T60	120.5	8.99	4.87	120.5	8.80	4.90
T61	121.5	7.40	4.60	122.3	7.47	4.63
C62	125.2	8.14	5.79	123.6	8.03	5.76
I63	120.3	8.76	4.60	120.6	8.79	4.61
Y65	123.4	7.55	4.47	123.2	7.57	4.48
V66	126.2	7.81	4.03	125.7	7.83	4.05
E67	131.3	7.97	4.01	129.0	8.22	4.23

Table 2S. Parameters for the NMR structure calculation of native LCI and III-A intermediate

	Native LCI	III-A intermediate
rms deviation from ideal geometry		
Bond lengths [Å]	0.0063	0.0060
Bond angles [°]	0.610	0.590
rms deviation from the mean structure [Å]		
Backbone atoms (residues 6-64)	1.40	2.47
Ramachandran analysis [%]		
Residues in favored regions	60.5	48.5
Residues in allowed regions	33.2	40.2
Residues in generously allowed regions	5.4	9.5
Residues in disallowed regions	0.9	1.8
Distance restraints		
Total NOE distance restraints	316	325
Short range	133	128
Medium range	168	184
Long range	15	13
Hydrogen bond restraints	30	18
Number violations > 0.5 Å	0	0

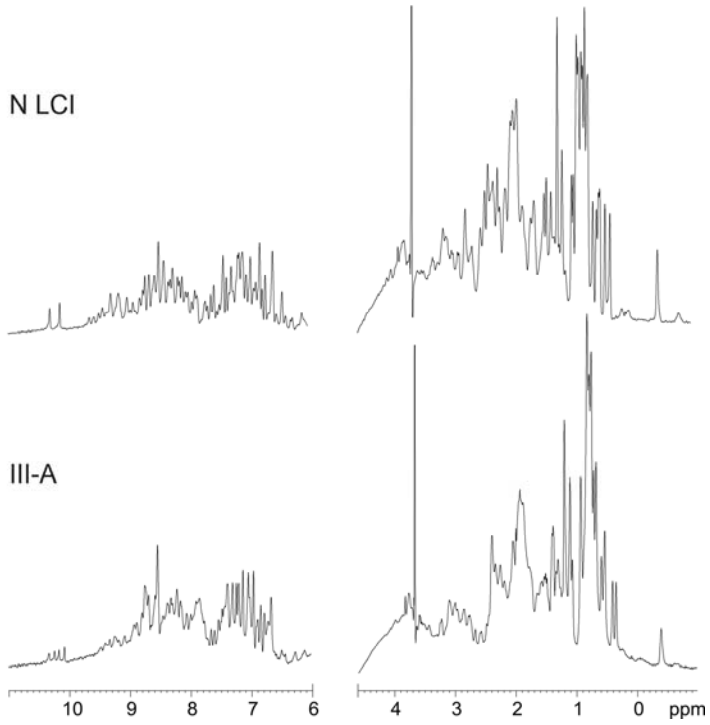
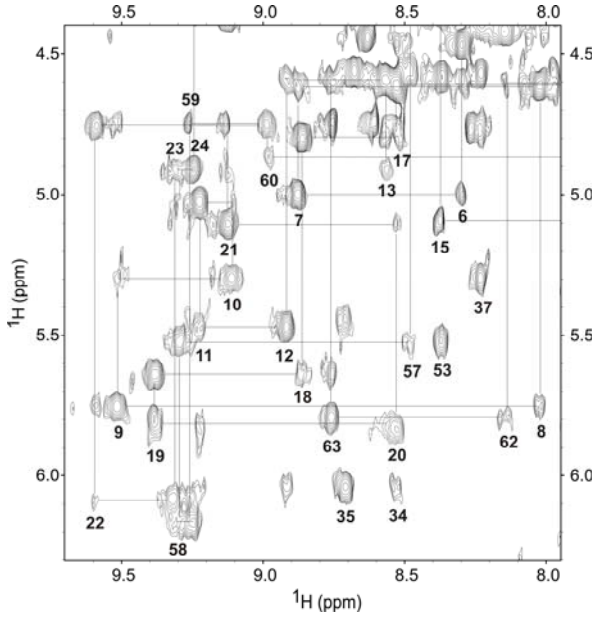


Fig. 1S. ^{1}H-NMR analysis of native LCI and III-A intermediate. Monodimensional spectra were recorded in a Bruker DRX 600-MHz spectrometer at 27ºC and at a protein concentration of 1 mM. Samples were prepared in H_2O/D_2O (9:1) at pH 3.5.



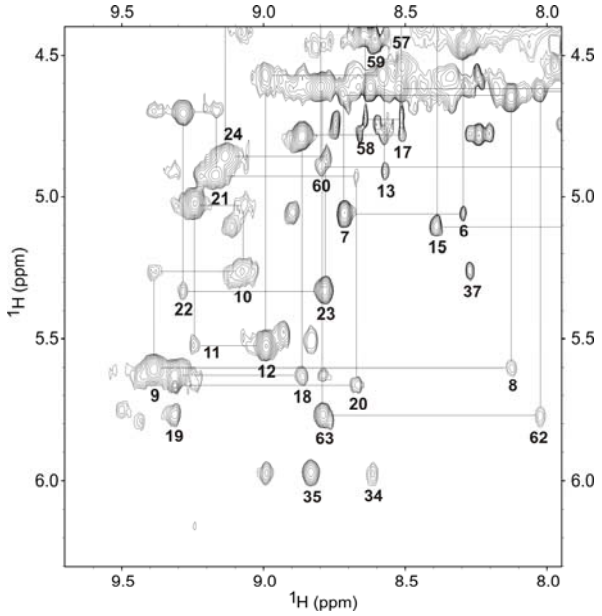


Fig. 2S. 2D NOESY spectra of native LCI and III-A intermediate. Sequential assignment pathways for native LCI (*upper panel*) and III-A (low*er panel*) are indicated between αH-NH cross-peaks in a slice of the fingerprint region. Solution conditions: 1 mM protein concentration, in H_2O/D_2O (9:1) at pH 3.5 and 27ºC.

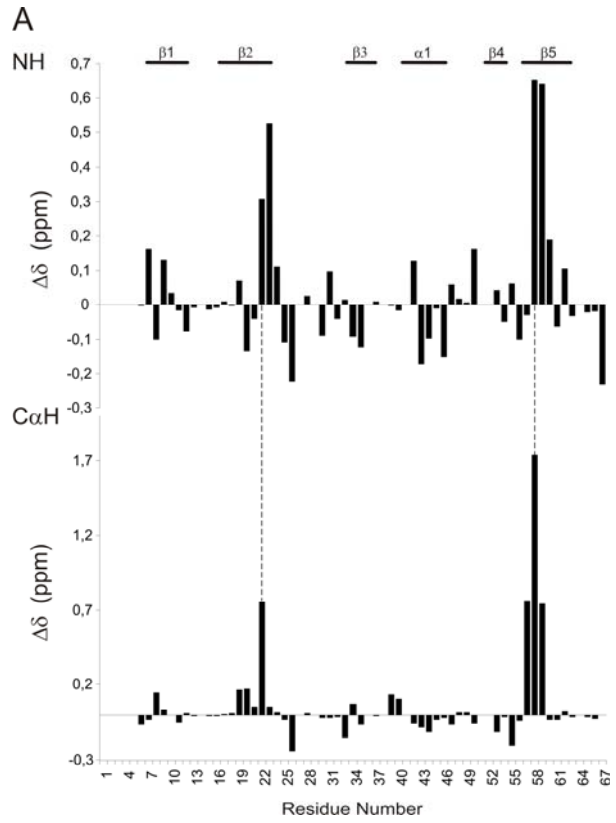


Fig. 3S. Chemical shift differences (Δδ) of backbone CαH and NH resonances between native LCI and III-A intermediate. $\Delta\delta = \delta_{native} - \delta_{III\text{-}A}$ as a function of the residue number. Backbone β-sheet and α-helix structures indicated above the plots are based on the structure of native LCI.

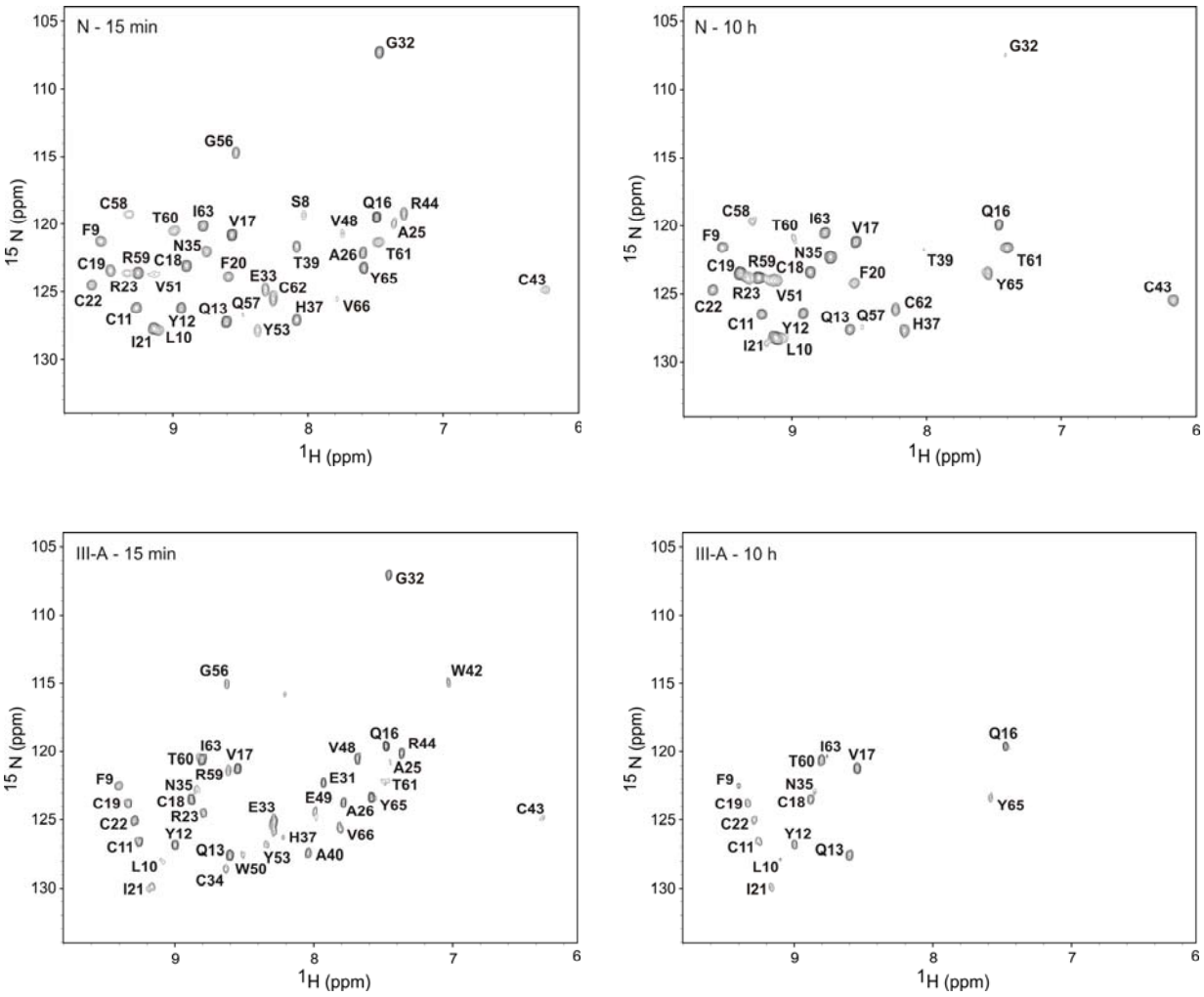


Fig. 4S. Amide exchange experiments on native LCI and III-A intermediate. The protected residues are shown in the 2D ^{1}H-^{15}N-HSQC spectra of uniformly ^{15}N-labeled native LCI and III-A intermediate recorded after 15 min and 10 h of exchange. The experiments were carried out with 1 mM protein concentration in D_2O at pH 3.5 and 27ºC.

Section II. Work 6

Characterization of an Analog of a Major Intermediate in the Oxidative Folding of Leech Carboxypeptidase Inhibitor

Summary

The oxidative folding pathway of leech carboxypeptidase inhibitor (LCI; four disulfide bonds) proceeds through the formation of two major intermediates (III-A and III-B) that contain three native disulfide bonds and act as strong kinetic traps in the folding reaction. The III-B intermediate lacks the Cys19-Cys43 disulfide bond which links the β-sheet core with the α-helix in wild-type LCI. In this work, an analog of this intermediate was constructed by replacing Cys19 and Cys43 with alanine residues. Its oxidative folding proceeds rapidly through a sequential flow of 1-, 2-, and 3-disulfide species to reach the native form in an efficient way due to the low accumulation of 2-disulfide intermediates and 3-disulfide (scrambled) isomers. The three-dimensional structure of this analog, alone and in complex with carboxypeptidase A (CPA) was determined by X-ray crystallography at 2.2 Å resolution. Its overall structure is very similar to that of wild-type LCI, although the residues in the region adjacent to the mutation sites show an increased flexibility. This lack of static order observed in the uncomplexed form is strongly reduced upon binding to the CPA. The structure of the complex also demonstrates that the analog and the wild-type LCI bind to the enzyme in the same manner, as expected by their inhibitory capabilities, which were similar for all enzymes tested. Equilibrium unfolding shows that this mutant is destabilized in ~ 1.5 kcal·mol^{-1} (~ 40%) relative to the wild-type protein. Taken together the data indicates that the fourth disulfide bond provides LCI with both high stability and structural specificity.

Keywords: carboxypeptidase inhibitor; oxidative folding; folding intermediates; crystal structure; protein stability.

Introduction

Protein folding proceeds through a series of intermediates that define the folding landscape from the unfolded polypeptide to the native structure (1, 2). Understanding the fundamental relationship between the amino acid sequence and the three-dimensional

structure of the native protein requires structural analysis of those folding intermediates. Although characterizing the intermediates is usually a difficult task due to their short half-life, studies of proteins stabilized by disulfide bonds have provided considerable insight into the field of protein folding (3). In disulfide-rich proteins, the coupling of the covalent chemistry of disulfide bond formation to the non-covalent folding processes makes it feasible to trap, isolate and characterize their intermediates (4). In this way, oxidative folding processes of disulfide-containing proteins such as bovine pancreatic trypsin inhibitor (BPTI) (5, 6), ribonuclease A (RNase A) (7, 8), and lysozyme (9, 10) have been investigated in great detail. However, the role and significance of many structured intermediates that accumulate along folding processes are still largely unknown.

Leech carboxypeptidase inhibitor (LCI) is a cysteine-rich polypeptide of 67 residues that behaves as a tight binding inhibitor of different metallocarboxypeptidases (11). Assuming that leeches secrete LCI during feeding, this inhibitor seems to maintain blood in the fluid state by inhibiting plasma carboxypeptidase B, also known as thrombin-activatable fibrinolysis inhibitor (TAFI), which acts as a potent attenuator of fibrinolysis (12, 13). Indeed, LCI significantly enhances the *in vitro* clot lysis induced by tissue plasminogen activator (Salamanca et al., manuscript in preparation), suggesting a potential use in the prevention or treatment of thrombotic disorders (14, 15). The three-dimensional structure of LCI shows that it folds in a compact domain consisting of a five-stranded antiparallel β-sheet and a short α-helix (16), with the occurrence of four disulfide bridges between cysteines 11-34, 19-43, 22-58, and 18-62, all of them located within regular secondary structure elements. The folding, unfolding and thermodynamic studies performed in our laboratory (17-19) have shown that LCI is a highly resistant molecule with slow rate constants of unfolding. Its reduced and denatured form refolds rapidly through a sequential flow of 1-, 2-, and 3-disulfide intermediates to reach a rate-limiting step in which two major 3-disulfide intermediates (designated as III-A and III-B) act as kinetic traps freezing the folding reaction. Both intermediates exclusively contain native disulfide pairings and inter-convert between them, reaching equilibrium prior to the formation of a heterogeneous population of 4-disulfide (scrambled) isomers (19). III-A and III-B are metastable forms in which both, native disulfide bonds and free thiols, appear to be similarly protected from the solvent; major structural rearrangements through the formation of scrambled isomers are required to render native LCI.

We have recently determined the NMR structure of the III-A intermediate (work 5 of this thesis) showing that it contains a native-like structure, which is responsible for its

strong accumulation along the LCI folding pathway. It displays, however, higher flexibility when compared to the native form. In the present work we intend to characterize the physical properties of the III-B intermediate. Due to the impossibility to obtain this species directly from the folding reaction, we decided to characterize a III-B analog in which the native Cys19 and Cys43 residues were replaced with alanine residues (C19A/C43A LCI). The differences in kinetic, thermodynamic, structural and functional properties resulting from the removal of the Cys19-Cys43 disulfide bond were analyzed for a better understanding of the folding process, conformational stability and functionality of LCI.

Experimental Procedures

Mutagenesis and Protein Expression — The synthetic gene for LCI (11) was cloned into the pBAT4 plasmid (20), fused in frame to the OmpA signal sequence for extracellular expression. The C19A/C43A analog was constructed using a two steps PCR method for site-directed mutagenesis. All constructs were verified by DNA sequencing. Wild-type and C19A/C43A LCI were obtained by heterologous expression in *Escherichia coli* strain BL21(DE3) using M9CAS medium containing 0.5% glycerol. Proteins were purified from the culture medium using a Sep Pak C_{18} cartridge (Waters), followed by anion-exchange chromatography on a TSK-DEAE 5PW column (Tosohaas), and by RP-HPLC on a 4.6 mm Protein C4 column (Vydac). Protein identity and purity (>98%) were confirmed by MALDI-TOF MS on a Bruker Ultraflex spectrometer and automatic Edman degradation using a Procise 492 Protein Sequencer (Applied Biosystems), respectively.

Oxidative Folding — The native protein (1 mg) was dissolved in Tris-HCl buffer (0.1 M, pH 8.4) containing 7 M GdnHCl and 50 mM DTT and kept at 23ºC for 2 h. The reduced and denatured sample was then passed through a PD-10 gel filtration column (Amersham Biosciences), previously equilibrated with Tris-HCl buffer (0.1 M, pH 8.4). The protein was recovered in 1.2 ml and immediately diluted to a final protein concentration of 0.5 mg/ml in the same Tris-HCl buffer, both in the absence (control -) and presence (control +) of 0.25 mM 2-mercaptoethanol. Folding intermediates were trapped in a time course manner at selected times by mixing aliquots from the different solutions with an equal volume of 2% TFA, and analyzed by RP-HPLC on a Protein C4 column. Using water containing 0.1% TFA as solvent A and acetonitrile containing 0.1% TFA as solvent B, a linear gradient from 20 to 40% B over 50 min was applied, at a flow rate of 0.75 ml/min.

Oxidative Folding in the Presence of Redox Agents or Denaturants — The procedures of unfolding and refolding were as described in "Oxidative Folding". Immediately after desalting the unfolded protein through a PD-10 column, GSSG (0.5 mM), GSSG and GSH (0.5 and 1 mM, respectively), or selected concentrations of denaturants (0.5-5 M GdnHCl, 1-8 M urea) were added. Folding intermediates were similarly trapped by acidification and analyzed by RP-HPLC.

Disulfide Analysis — Acid-trapped intermediates were purified from RP-HPLC and freeze-dried. The samples were derivatized with vinylpyridine (0.2 M) in Tris-HCl buffer (0.1 M, pH 8.4) at 22ºC for 45 min. The reaction was quenched with 2 % TFA. The derivatized samples were analyzed by MALDI-TOF MS to characterize the number of disulfide bonds of the folding intermediates. To determinate the disulfide pairing of native C19A/C43A LCI, the protein was purified by RP-HPLC and freeze-dried. The sample (20 μg) was derivatized with vinylpyridine (0.2 M) in Tris-HCl buffer (0.1 M, pH 6.4) at 22ºC for 45 min. The derivatized sample was further purified by RP-HPLC, freeze-dried, and digested with 2 μg of thermolysin (Sigma, P-1512) in 30 μl of N-ethylmorpholine/acetate buffer (50 mM, pH 6.4) at 37ºC for 16 h. Thermolytic products were purified by RP-HPLC using a 4.6 mm C18 column (Vydac); a linear gradient from 0 to 60% solvent B over 60 min was applied. The molecular masses of disulfide-containing peptides were determined by MALDI-TOF MS. The amino acid sequences of selected thermolytic peptides were analyzed by automatic Edman degradation.

Reductive Unfolding — Native proteins (0.5 mg) were dissolved in 1 ml of Tris-HCl buffer (0.1 M, pH 8.4) with different concentrations of DTT (0.1-100 mM). Reduction was carried out at 23ºC. To monitor the kinetics of unfolding, aliquots of the sample were removed at various time intervals, quenched with 2% TFA, and analyzed by RP-HPLC as detailed under "Oxidative Folding".

Denaturation and Unfolding in the Presence of Denaturant — Native proteins were dissolved to a final concentration of 0.5 mg/ml in Tris-HCl buffer (0.1 M, pH 8.4) containing 0.25 mM 2-mercaptoethanol and selected concentrations of denaturants (0-8 M GdnHCl, 0-8 M urea). The reaction was allowed to reach equilibrium and was typically performed at 23ºC for 20 h. The samples were then quenched with 2% TFA and analyzed by RP-HPLC using the same conditions described in "Oxidative Folding".

Circular Dichroism and Fluorescence Spectroscopy — Samples were prepared dissolving the protein in Tris-HCl buffer (0.1 M, pH 8.4) containing 0.25 mM 2-

mercaptoethanol and selected concentrations of GdnHCl (0-8 M). The final protein concentration was 27 μm for CD and 10 μm for fluorescence studies. The reaction was allowed to reach equilibrium and was performed at 23ºC for 20 h. CD analyses were carried out in a Jasco J-715 spectrometer at 25ºC using a cell of 2 mm path length. Fluorescence spectra were recorded at 25ºC on a Perkin-Elmer 650-40 spectrofluorometer using 280 nm as the excitation wavelength. Excitation and emission slits were set at 5 and 7 nm, respectively.

Thermodynamic Parameters — Thermodynamic parameters were calculated by fitting experimental data to Equation 1:

$$y = \frac{(a + b[denaturant]) - (c + d[denaturant])}{1 + \exp\{-(\Delta G_U^{H_2O} - m[denaturant])/RT\}} + (c + d[denaturant])$$

where y is the observed fractional population of native species, $\Delta G_U^{H_2O}$ is the free energy change of unfolding in the absence of denaturant, m is a parameter for cooperativity of unfolding, and $(a + b[denaturant])$ and $(c + d[denaturant])$ are terms for the base-line dependence on denaturant concentration. The midpoint concentration of unfolding, c_M, is calculated by Equation 2:

$$c_M = \Delta G_U^{H_2O} / m$$

The fitting was performed using the nonlinear least-squares algorithm provided with the software KaleidaGraph (Abelbeck Software) assuming a two-state unfolding mechanism.

CP Inhibitory Assays — The inhibitory activity of C19A/C43A LCI was assayed by measuring its ability to inhibit the hydrolysis of the chromogenic substrates N-(4-methoxyphenylazoformyl)-Phe-OH and N-(4-methoxyphenylazoformyl)-Arg-OH (Bachem) by carboxypeptidases type A and type B, respectively. The assay was performed using a substrate concentration of 100 μM in Tris-HCl buffer (50 mM, pH 7.5) containing NaCl (100 mM). The inhibition constants (K_i) were determined at the presteady-state as described for tight binding inhibitors (21). Bovine CPA and porcine CPB were purchased from Sigma and human CPA1 and CPB were prepared following described procedures (22, 23). The concentration of the purified solutions of wt and C19A/C43A LCI was determined from the A_{280} of the final solution (extinction coefficient: $E_{0.1\%}$=2.12).

Crystallization and Structure Determination — C19A/C43A LCI was crystallized in complex with bovine CPA. The complex was obtained by mixing both proteins in Tris-HCl buffer (50 mM, pH 7.5) containing NaCl (100 mM). The analog was added until complete

inhibition of the enzyme was reached. The complex was maintained for 2 hours at 20ºC and purified by gel-filtration chromatography on a Superdex 75 HiLoad 26/60 column (Amersham Biosciences) using the same buffer, and further concentrated to about 10-12 mg/ml. Suitable crystals for data collection were obtained at 20ºC using the sitting-drop vapor-diffusion method in a few weeks growing period. Drops were obtained by mixing equal volumes of protein solution (C19A/C43A LCI-CPA) and reservoir buffer containing Lithium Sulfate monohydrate (1.5 M) and Tris (100 mM, pH 8.5).

A dataset up to 2.8 Å was collected on the MPG/GBF beamline BW6 at DESY, Hamburg (Germany). Diffraction data for the structure refinement was collected at 90K. Diffraction images were taken on MARCCD Detectors. The summary of the data collection is shown in Table 1. The collected data were integrated, scaled and merged by XDS and XSCALE programs (24). The structure was determined by molecular replacement using the Molrep program from the CCP4 suite (25). The structure of the LCI-CPA2 complex taken from the PDB entry 1DTD (16) was used as a probe after removing the inhibitor part. The initial R-factor of the model was 0.46. The model was then refined by Refmac5 (25) and rebuilt by XtalViev/Xfit (26) and by a subsequent Refmac5 refinement. Waters were added by Arp/warp (27). The final R crystallographic factor was 0.19 and R_{free} 0.23. The asymmetric unit contains two complexes of C19A/C43A LCI-CPA and two additional C19A/C43A LCI molecules imprisoned in crystal lattice. The unbound inhibitor molecules display significant flexibility in the loop regions and therefore their models are not completely built because of the lack of interpretable electron density. Most of the other molecules had a clear and interpretable electron density. However, the loop region between Val132-Ser136 is missing on the map in both CPA molecules. There are also solvent-exposed side chains with missing density; these parts were omitted in the final model.

Table 1. Data collection and refinement statistics

Data collection	
Space group	$P4_32_12$
Cell constants (Å)	a=124.93 b=124.93 c=154.90
Resolution range (Å)	36-2.2
Wavelength (Å)	1.05
Observed reflections	493543
Unique reflections	98449
Whole resolution range:	
Completeness (%)	97.7
R_{merge}	2.9
I/σ(I)	31.99
Last resolution shell:	
Resolution range (Å)	2.2-2.3
Completeness (%)	63.9
R_{merge}	21.2
I/σ(I)	5.28
Refinement	
No. of reflections	66228
Resolution (Å)	30-2.2
R-factor (%)	18.9
R_{free} (%)	23.3
Average B (Å^2)	28.2
R.m.s.d. bond length (Å)	0.012
R.m.s.d. angles (°)	1.48
Content of asymmetric unit	
R.m.s.d. of complexes (Å)	0.43
No. of protein molecules	6
No. of protein residues/atoms	835/6966
No. of solvent atoms	538
No. of Zn atoms	2

Results

Oxidative Folding of C19A/C43A LCI in the Absence of Redox Agents — Oxidative folding studies of C19A/C43A LCI were carried out in Tris-HCl buffer in the absence (control -) and presence (control +) of the reducing agent 2-mercaptoethanol. Folding intermediates were trapped in a time course manner by either acid quenching with trifluoroacetic acid (TFA) or reaction with the alkylating reagent vinylpyridine. The heterogeneity and chromatographic behavior of the folding intermediates was characterized by RP-HPLC, and the disulfide content was determined by matrix-assisted laser desorption/ionization - time of flight mass spectrometry (MALDI-TOF MS) analysis.

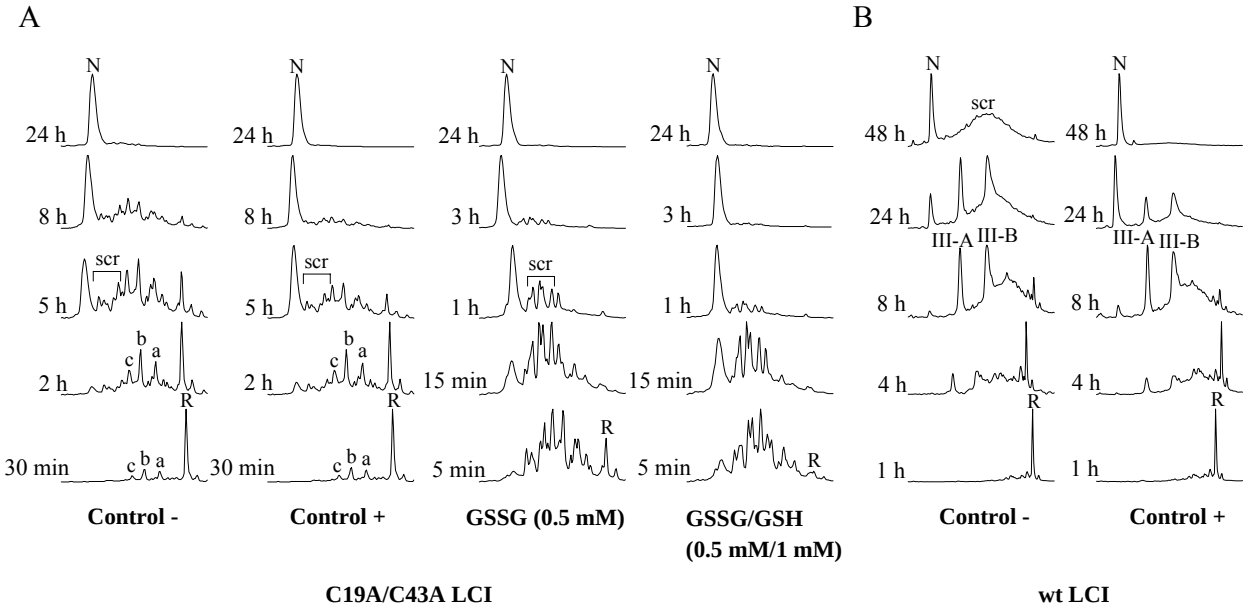


Fig. 1. RP-HPLC analysis of the folding intermediates of C19A/C43A (A) and wt (B) LCI trapped by acidification. Folding was carried out in Tris-HCl buffer (pH 8.4) in the absence (control -) and presence of selected redox agents: 2-mercaptoethanol (0.25 mM) (control +), GSSG (0.5 mM) or GSSG/GSH (0.5 mM/1 mM). Acid-trapped intermediates were analyzed at the noted times by RP-HPLC as detailed under "Experimental Procedures". *N* and *R* indicate the elution positions of native and fully reduced species, respectively. *a*, *b* and *c* are three major fractions of intermediates of C19A/C43A LCI containing mainly two disulfide bonds. *III-A* and *III-B* are two intermediates of wt LCI containing three native disulfides. *scr* indicates the elution positions of scrambled isomers.

For this analog, few intermediates populate the initial phase of the folding process. The RP-HPLC profiles clearly show the accumulation of three fractions of major intermediates in the early stages of the reaction (up to 2h), designated as *a*, *b*, and *c* (Fig.

1A). At this stage the RP-HPLC patterns remain similar regardless of the presence of reducing agent and the MS analysis of disulfide species after derivatization with vinylpyridine shows the same percentage of 1- and 2-disulfide species (Fig. 2, control - and +) indicating that the presence of the reducing agent has no apparent effect on these early stages. The analysis of disulfide species of fractions *a*, *b*, and *c* after their purification and derivatization showed that they mainly contain 2-disulfide species. Along the reaction only a low percentage of species correspond to 3-disulfide non-native (scrambled) isomers, mainly located in the fraction designated as *scr* (Fig. 1A). Therefore, the presence of 2-mercaptoethanol, which promotes disulfide reshuffling and conversion of scrambled forms into the final native structure, slightly affects the last stages of the folding pathway of C19A/C43A LCI. It becomes apparent from the quantitative analysis of disulfide species (Fig. 2) that no intermediates are able to act as strong energy barriers trapping the folding reaction. Consequently, the folding process is highly efficient, allowing more than a 90% recovery of the protein as native form after 24 h of refolding (Fig. 2). In contrast, the folding of wild-type (wt) LCI cannot reach completion in the absence of a thiol catalyst due to the strong accumulation of III-A and III-B intermediates as well as scrambled species along its folding (Fig. 1B), and only ~30% of the protein is recovered in the native form after 48 h of refolding.

Native C19A/C43A LCI can be clearly distinguished from the folding intermediates by its RP-HPLC elution position (Fig. 1A). To determine its disulfide pairing, the native analog was digested with thermolysin as described in "Experimental Procedures" and the resultant peptides were separated by RP-HPLC and analyzed by mass spectrometry and Edman sequencing. Taken together, the data permitted us to deduce the disulfide pairings: Cys11-Cys34; Cys18-Cys62; Cys22-Cys58, which are the three disulfide bonds displayed by III-B intermediate of wt LCI (19).

Oxidative Folding of C19A/C43A LCI in the Presence of Redox Agents or Denaturants — Oxidative folding of C19A/C43A LCI was performed in the presence of either oxidized glutathione (GSSG) (0.5 mM) or a mixture of reduced and oxidized glutathione (GSH and GSSG; 1 and 0.5 mM, respectively) and the acid-trapped intermediates were analyzed by RP-HPLC. Although the addition of GSSG does not have an effect on the folding efficiency of C19A/C43A LCI, it strongly accelerates its folding rate (Fig. 1A). Analysis of the composition of time course-trapped intermediates reveals that this acceleration is caused by a rapid disulfide oxidation that leads to the generation of 3-disulfide species (Fig. 2). Indeed, the fraction corresponding to scrambled isomers (*scr*)

accumulates much faster when compared with the control - experiment. The RP-HPLC patterns obtained in the presence of GSSG/GSH were similar to those obtained in the presence of GSSG alone (Fig. 1A) indicating that the addition of GSH does not strongly affect the folding process of C19A/C43A LCI. Since the redox potential of GSH has been shown to promote the reduction of 3-disulfide scrambled isomers to obtain native form (28), the low accumulation of scrambled forms along C19A/C43A LCI folding would account for its minimum effect.

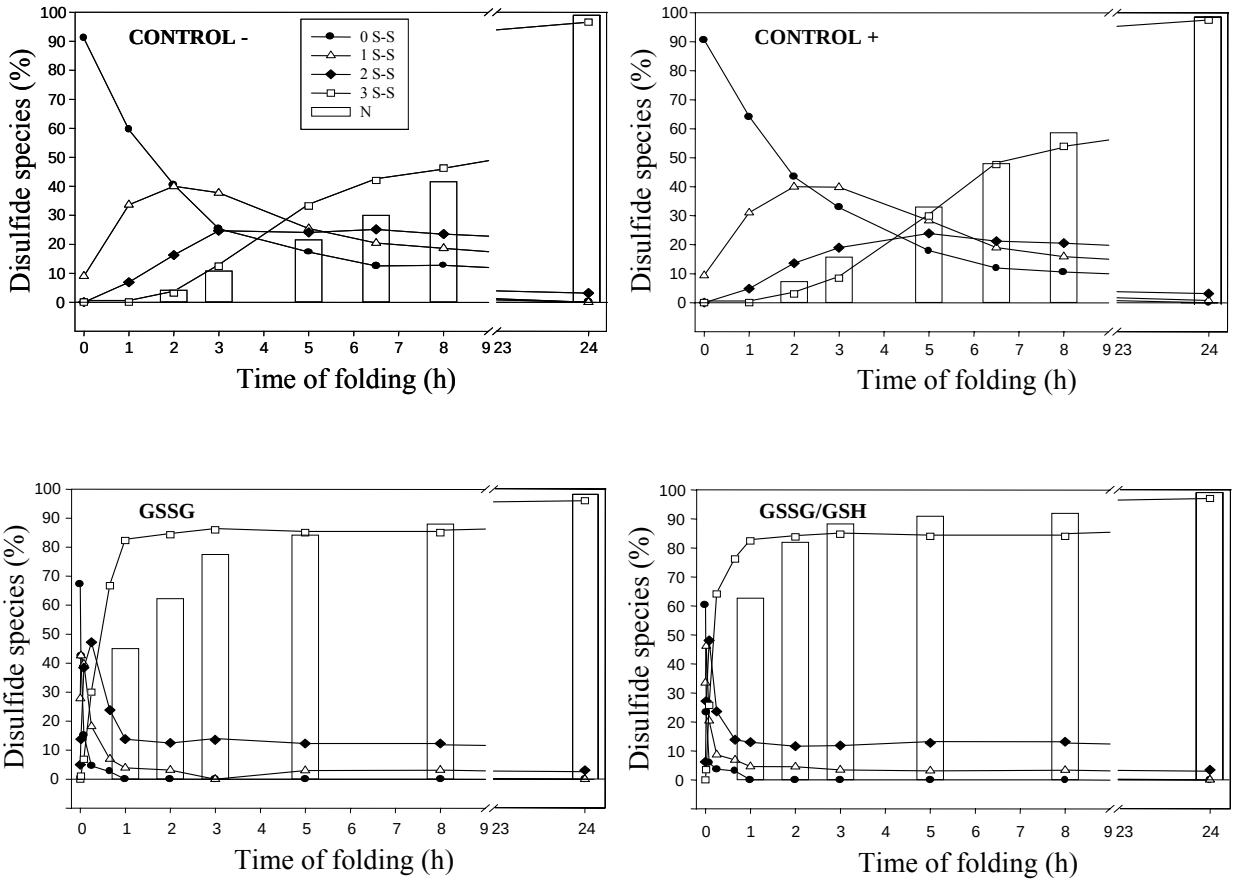


Fig. 2. Quantitative analysis of various disulfide species along the oxidative folding pathway of C19A/C43A LCI. Folding was performed in Tris-HCl buffer (pH 8.4) in the absence (control -) and presence of redox agents: 2-mercaptoethanol (0.25 mM) (control +), GSSG (0.5 mM) or GSSG/GSH (0.5 mM/1 mM). 0 S-S, 1 S-S, 2 S-S, and 3 S-S represent the completely reduced, 1-disulfide, 2-disulfide, and 3-disulfide species, respectively. The 3-disulfide species include the native form and all scrambled isomers. Quantitative analysis of disulfide species was based on the peak response of MALDI-TOF spectra (data not shown). The recovery of native form is represented by bars and was calculated from the peak areas in the corresponding RP-HPLC chromatograms.

Oxidative folding of C19A/C43A analog was subsequently performed in the Tris-HCl buffer containing different concentrations of guanidine hydrochloride (GdnHCl) or urea, both in the absence (control -) and presence (control +) of thiol catalyst. The prevalence of 2-disulfide intermediates under low denaturing conditions (up to 2 M urea concentration) when compared with the control experiments is an indication of the stability of these species (Fig. 3A). The formation of native C19A/C43A LCI is strongly accelerated by the presence of 0.5-1 M GdnHCl similarly to the effect obtained on wt LCI (Fig. 3A and 3B). However, the quicker folding acceleration and lower accumulation of these metastable forms of C19A/C43A LCI, when compared to wt LCI folding under these conditions, reflects the lower stability of C19A/C43A intermediates relative to those of wt LCI and prevents them to act as strong kinetic traps (Fig. 3A and 3B). The folding pathway of this mutant drastically changed when the refolding experiments were carried out at high denaturing conditions (e.g. 5 M GdnHCl), with a low recovery of the native form (less than 5% after 24 h of folding) (data not shown).

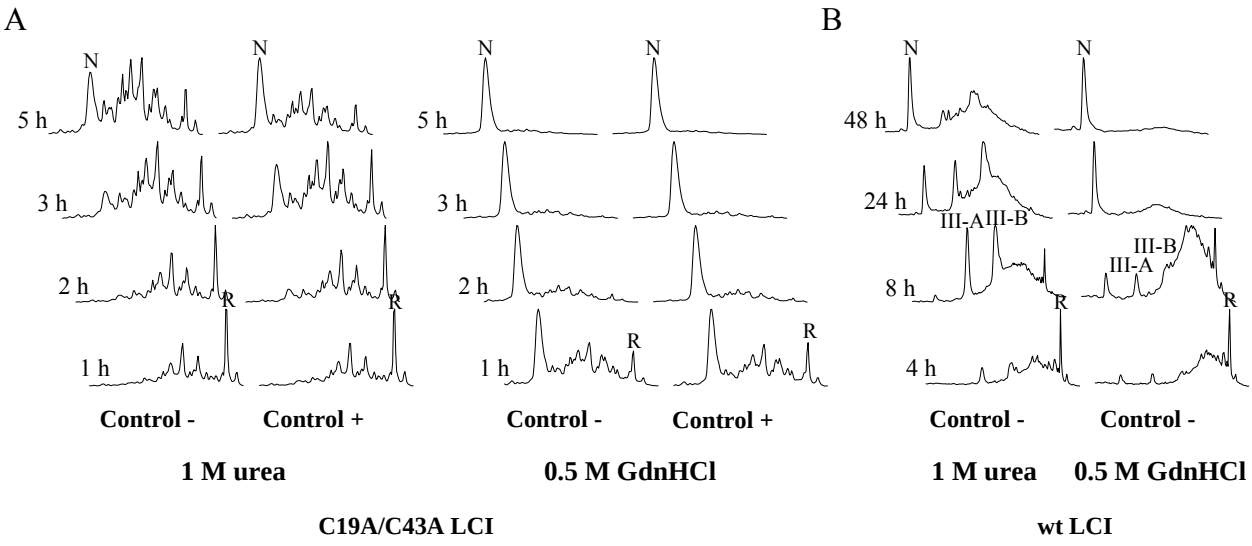


Fig. 3. Effect of denaturant on the folding intermediates of C19A/C43A (A) and wt (B) LCI. The reduced protein was allowed to refold in Tris-HCl buffer (pH 8.4) under the presence of 1 M urea or 0.5 M GdnHCl with (control +) or without (control -) 2-mercaptoethanol. Folding intermediates were trapped by acidification and analyzed by RP-HPLC. *N* and *R* indicate the elution positions of native and fully reduced species, respectively. *III-A* and *III-B* are two intermediates of wt LCI containing three native disulfides.

Conformational Stability of C19A/C43A LCI — The reductive unfolding of C19A/C43A LCI was studied at pH 8.4 using increasing concentrations of dithiothreitol (DTT) as reducing agent to test the stability of its disulfide bonds. The unfolding intermediates were trapped in a time course manner by acidification and were analyzed by RP-HPLC. Reduction of this analog undergoes an apparent all-or-none mechanism in which all three disulfides are reduced in a cooperative and concerted manner, with no accumulation of partially reduced intermediates (data not shown). This mechanism of disulfide reduction has been previously shown in numerous small single domain proteins (29, 30). The analysis also reveals that C19A/C43A LCI resists much lower concentrations of DTT than wt LCI: similar percentages of unfolded protein are obtained using 100 and 10 mM DTT for wt and C19A/C43A LCI, respectively (data not shown).

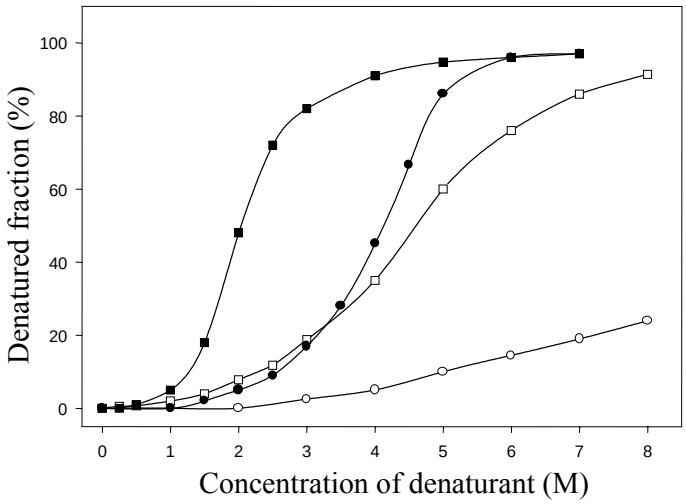


Fig. 4. Denaturation curves of the native form of C19A/C43A and wt LCI determined by RP-HPLC. The denatured fraction is calculated as the percentage of each protein converted into scrambled isomers. Wt LCI (●) and C19A/C43A LCI (■) were denatured in the presence of GdnHCl (filled symbols) and urea (open symbols) at 23ºC for 20 h in Tris-HCl buffer (0.1 M, pH 8.4) containing 2-mercaptoethanol (0.25 mM).

In the presence of denaturant and thiol initiator, unfolding of a native disulfide-containing protein is accompanied by reshuffling of its native disulfides (disulfide scrambling) (31). Denaturation of the native form of wt and C19A/C43A LCI was analyzed in the presence of increasing concentrations of denaturant and monitored by RP-HPLC, intrinsic fluorescence and CD. All the transition curves of the three experiments were sigmoid and the data were fit to a two-state mechanism (see "Experimental Procedures"). The denaturation curves obtained by RP-HPLC analysis of disulfide

scrambling in the presence of urea and GdnHCl show that the extent of unfolding, and hence the equilibrium constant between scrambled isomers and the native protein, is clearly dependent upon the strength of denaturant (Fig. 4). Urea requires higher concentrations than GdnHCl to fully denature C19A/C43A LCI, with midpoint denaturant concentrations of 1.9 and 4.2 M for GdnHCl and urea, respectively. As expected, the value of $\Delta G_U^{H_2O}$ for C19A/C43A LCI is independent of the denaturant used in the assay (Table 2). The stability data obtained from GdnHCl transition curves indicate that the mutant is destabilized in about 1.5 kcal·mol^{-1} with respect to the wt form and the c_M values for the mutant and wt forms are 1.9 and 4.2 M, respectively.

Table 2. Thermodynamic properties for unfolding of C19A/C43A and wt LCI[1]

	ΔG (kcal·mol^{-1})	m (kcal·mol^{-1}·M^{-1})	c_M (M)	$\Delta\Delta G_{\text{mut - wt}}$ (kcal·mol^{-1})
C19A/C43A LCI Urea RP-HPLC	3.0 ± 0.31	0.71 ± 0.08	4.2	
C19A/C43A LCI Gdn RP-HPLC	3.0 ± 0.30	1.61 ± 0.13	1.9	
wt LCI Gdn RP-HPLC	4.5 ± 0.39	1.08 ± 0.10	4.2	-1.5
C19A/C43A LCI Gdn Fluor	2.9 ± 0.38	1.63 ± 0.16	1.8	
wt LCI Gdn Fluor	4.3 ± 0.24	1.07 ± 0.06	4.0	-1.4
C19A/C43A LCI Gdn CD	3.7 ± 0.07	2.04 ± 0.03	1.8	
wt LCI Gdn CD	5.3 ± 0.16	1.34 ± 0.04	4.0	-1.6

[1]The experiment conditions and analysis are described in "Experimental Procedures". The errors shown correspond to the fitting errors.

Fluorescence spectroscopy provides a convenient and sensitive means for probing the environment of Trp residues and has been widely used in the equilibrium investigation of the folding/unfolding of proteins (32, 33). The fluorescence spectrum of LCI is dominated by the contribution of two Trp residues (sequence positions 42 and 50) buried in the interior of the protein that give an emission maximum at 350 nm (Fig. 5A). In the presence of denaturant, the formation of scrambled isomers exposes Trp residues to solvent causing both, a red shift to 355 nm and an increase of intensity, to its fluorescence emission. These differences between the fluorescence emission of native and scrambled forms allow to easily monitoring the unfolding reaction. The denaturation curves for wt and C19A/C43A LCI monitored by Trp fluorescence are very similar to those obtained by RP-HPLC analysis and similar thermodynamic parameters are also obtained (Table 2): $\Delta\Delta G_{\text{mut – wt}}$ = -1.4 kcal·mol^{-1} upon disulfide elimination, and 1.8 and 4.0 M concentrations

of GdnHCl are needed to denature 50% of C19A/C43A LCI and wt form, respectively (Fig. 5B). This experiment clearly indicates that the scrambling process observed upon unfolding in both, the mutant and the wt form, results in a progressive exposition of the Trp residues to solvent.

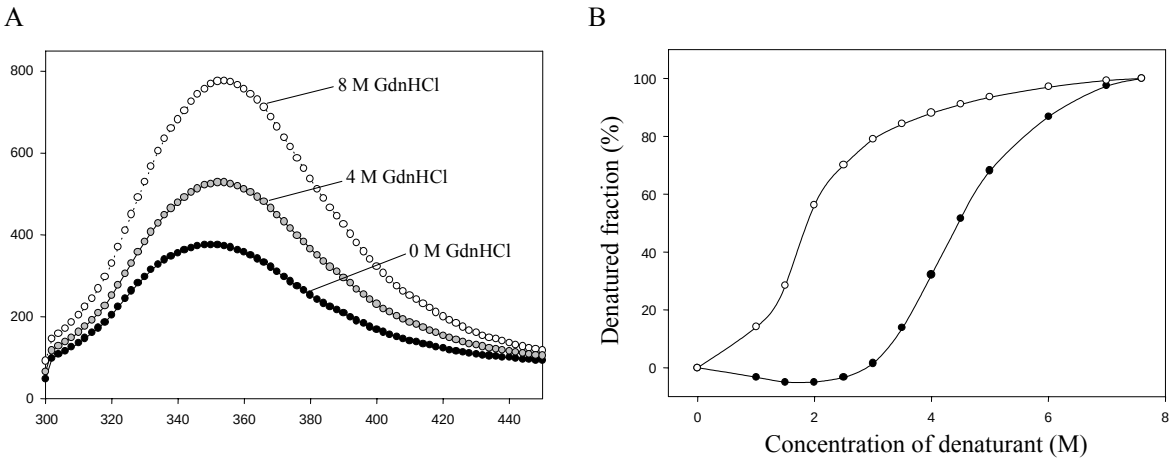


Fig. 5. Fluorescence spectroscopy. A. Fluorescence analysis of the native form of wt LCI. The protein was dissolved at a concentration of 10 μm in Tris-HCl buffer (0.1 M, pH 8.4) containing 2-mercaptoethanol (0.25 mM) and selected concentrations of GdnHCl. B. Denaturation curves of the native form of wt (●) and C19A/C43A (○) LCI determined by fluorescence. The denatured fraction is calculated as the percentage of each protein converted into scrambled isomers. Denaturation was carried out at 23ºC, for 20 h, in Tris-HCl buffer (0.1 M, pH 8.4) containing 2-mercaptoethanol (0.25 mM).

The CD spectrum of native LCI displays a well-defined minimum in ellipticity at 210 nm and a maximum at 228 nm, whose disappearance correlate with the unfolding of the protein (11). The shape of the CD spectrum of C19A/C43A is very similar to that of wt LCI (Fig. 6A), except for a slight decrease in the intensity of the 228-nm maximum. Denaturation of the native form of wt and C19A/C43A LCI in the presence of increasing concentrations of GdnHCl was followed monitoring the CD signal at 228 nm. The transition GndHCl concentration was again very similar to that obtained in the previous assays: 1.8 and 4.0 M for C19A/C43A and wt LCI, respectively (Fig. 6B). Although the calculated free energy of unfolding for both, the wt and mutant forms, is higher than those calculated in the other experiments, the global destabilization upon mutation (1.6 kcal· mol^{-1}) is in agreement with the previous results.

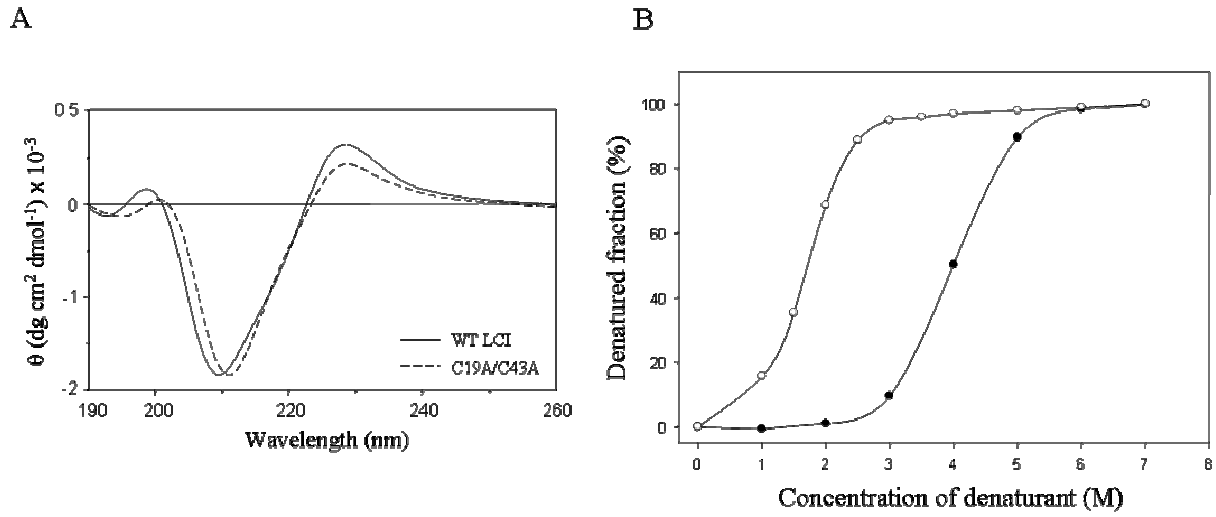


Fig. 6. Circular dichroism spectroscopy. A. CD analysis of the native form of wt and C19A/C43A LCI. CD analysis was carried out at 25ºC in sodium phosphate buffer (20 mM, pH 7.0) at a final protein concentration of 0.2 mg/ml. B. Denaturation curves of the native form of wt (●) and C19A/C43A (○) LCI determined by CD. The denatured fraction is calculated as the percentage of each protein converted into scrambled isomers. Denaturation was carried out at 23ºC, for 20 h, in Tris-HCl buffer (0.1 M, pH 8.4) containing 2-mercaptoethanol (0.25 mM).

It is worth to mentioning that, independently of the technique used to monitor equilibrium unfolding, in all cases the *m* value is higher for C19A/C43A than for wt LCI (Table 2). Since the *m* values correlate with the changes in protein accessible surface area upon unfolding (34), this increase would indicate that the mutant displays a less compact unfolded state, with a higher accessibility of the unfolded polypeptide to solvent.

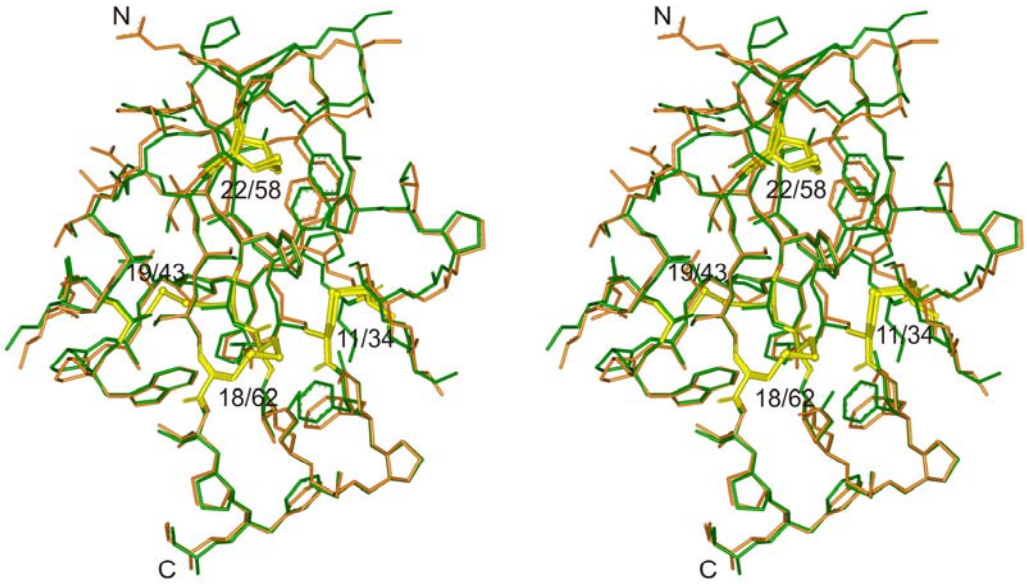
Table 3. Inhibition constants (K_i) of C19A/C43A LCI measured for different types of carboxypeptidases

Carboxypeptidase type	wt LCI	C19A/C43A LCI
	K_i (nM)	
Bovine CPA	1.1 ± 0.2	2.8 ± 0.3
Human CPA1	1.3 ± 0.4	3.4 ± 0.4
Porcine CPB	2.4 ± 0.5	5.5 ± 0.6
Human CPB	1.2 ± 0.3	3.5 ± 0.3

Function of C19A/C43A LCI — LCI is a tight binding competitive inhibitor of pancreatic-like metallocarboxypeptidases (subfamily A/B) that possess an inhibitory

activity in the nanomolar range (11). Inhibition constants for the complexes of C19A/C43A LCI with different carboxypeptidases were determined. As shown in Table 3, the inhibitory potential of this mutant is similar to that of wild-type LCI for both carboxypeptidases type A and B.

Crystal structure of C19A/C43A LCI — To understand the structural features underlying the folding, stability and functionality of C19A/C43A LCI we determined the X-ray structure of the free form and its complex with bovine CPA. Each asymmetric unit in the crystal contains two complexes of C19A/C43A LCI-CPA and two free C19A/C43A LCI molecules. The C19A/C43A LCI structure determined at 2.2 Å resolution shows that this analog is very similar to the wt form (Fig. 7A) (16). Its structure consists of a five-stranded antiparallel β-sheet with a β3-β1-β2-β5-β4 topology, and a short α-helix that packs into the most compact part of the β-structure (Fig. 7B). This antiparallel β-sheet involves residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), Val51-Tyr53 (β4) and Gly 56-Ile63 (β5), while the short α-helix is located between residues Pro41 and Gly46. The disulfide pairings of this mutant, as determined from its three-dimensional structure calculations, are Cys11-Cys34, Cys18-Cys62, and Cys22-Cys58, and completely agree with the results previously obtained by digestion with thermolysin.



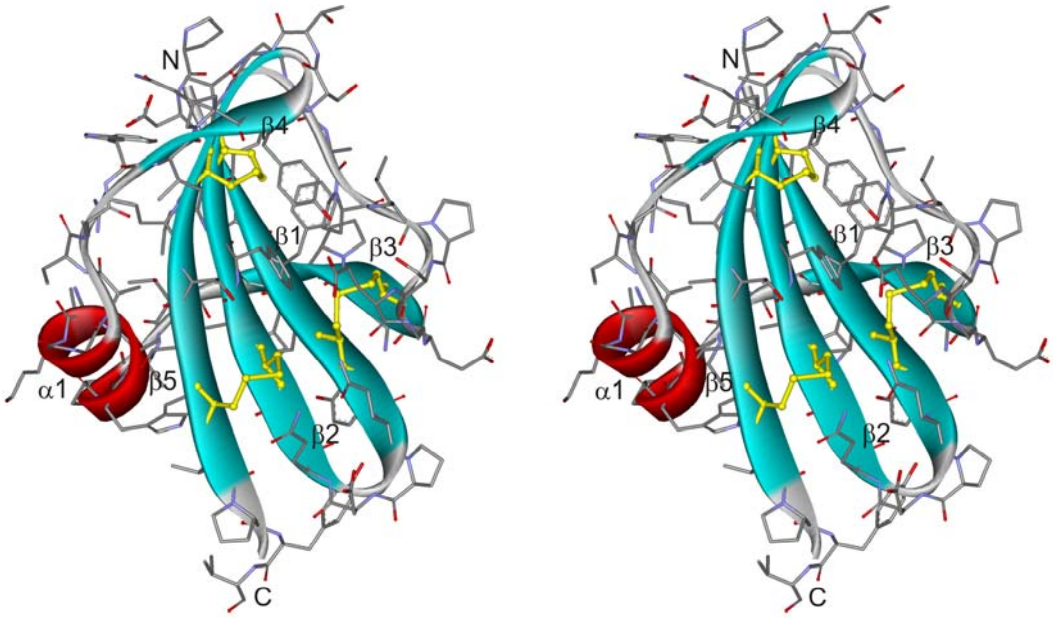


Fig. 7. Three-dimensional structure of C19A/C43A LCI. *Upper panel.* Stereo view of the overlapping between the backbone atoms from C19A/C43A (green) and wt LCI (orange). The disulfide pairings of each protein are shown yellow in the structure. *Lower panel.* Stereo representation of the C19A/C43A LCI ribbon. The helix and β-strands (β1-β5) are represented in red and light blue, respectively. The three disulfide bridges of this mutant (Cys11-Cys34, Cys22-Cys58 and Cys18-Cys62) are represented in yellow. *N* and *C* indicate the location of N- and C-terminal tails of C19A/C43A LCI.

The structure of the analog upon interaction with CPA greatly resembles that of the wt form, with a root mean square deviation (r.m.s.d.) between the bound analog and bound wt form of 0.42 Å for backbone, lower than the r.m.s.d. between the bound and free forms of the analog (0.91 Å for backbone). The free form of the mutant shows a higher flexibility in the Pro38-Trp50 region and displays a shifted α-helix (1.8 Å) out of the molecule center when compared to the wt form. Both deviations are corrected in the structure of the bound analog, indicating that the interaction with CPA induces the acquisition of the wt structure. In addition, direct comparison of temperature factors (B-factors) also shows that the analog is stabilized upon complex formation, with average B-value of 29 $Å^2$ for the bound form and 40 $Å^2$ for the free molecule.

C19A/C43A LCI interacts extensively with the carboxypeptidase (Fig. 8). The occlusion of the C-terminus of this analog in the active site groove of the carboxypeptidase constitutes the "primary binding region". The interactions of the C-terminal residues of the inhibitor (Val66, Tyr65, Pro64) with this region of the enzyme define the location of the

binding subsites S1, S2 and S3 (Fig. 8). These interactions are nearly indistinguishable from those found in the wt LCI-CPA2 complex (16) and only differ with the behavior of the residue Glu67, which in the case of C19A/C43A LCI-CPA complex is cleaved off and does not remain in the active site of the enzyme. The "secondary binding region", established by interactions between the carboxypeptidase surface and the neighboring residues of the mutant, is also highly similar to that of the wt LCI-CPA2 complex.

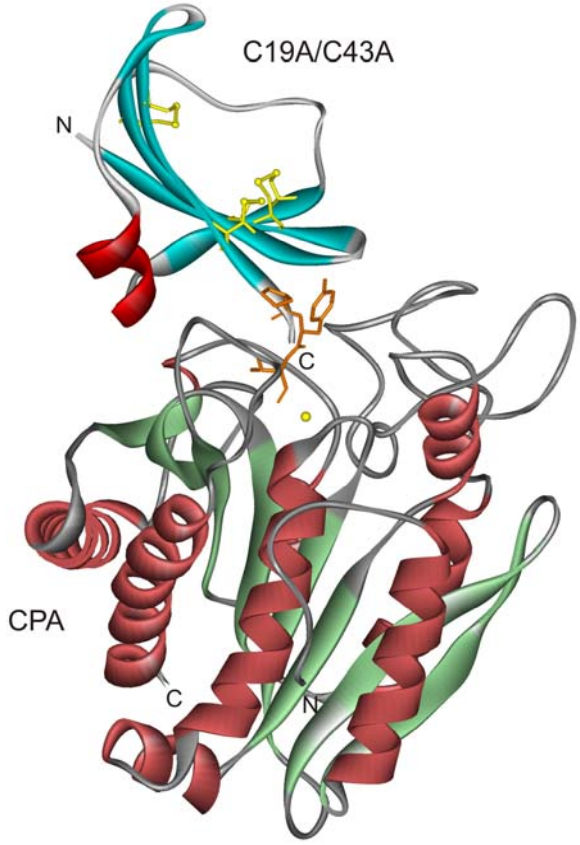


Fig. 8. Ribbon plot of the complex formed between C19A/C43A LCI and CPA. The helix and β-strands of C19A/C43A LCI are shown in red and light blue, respectively, and the disulfide bridges are in yellow. The helices and β-strands of CPA are in dark red and dark green, respectively. The side chains of C19A/C43A LCI residues involved in the interaction with CPA are explicitly shown in orange. The zinc atom of CPA is represented by a yellow sphere. The N-terminal and C-terminal domains of the mutant and the enzyme are depicted.

Discussion

C19A/C43A LCI is an analog of one of the major rate-limiting intermediates in LCI oxidative folding. The in-depth study of its folding pathway and its functional,

thermodynamic and structural properties has provided complementary data for a better understanding of wt LCI nature and the relationship between its folding and structure.

The analysis of the acid-trapped folding intermediates of this analog showed that its folding process significantly differs from that of wt LCI. While the oxidative folding of the wt protein does not show preferential accumulation of 1- or 2-disulfide species, the absence of the Cys19 and Cys43 causes the accumulation of 2-disulfide intermediates in the early stages of the analog folding. In addition, no strong kinetic traps like those displayed by wt LCI are able to freeze the mutant folding reaction, resulting in a fast and efficient process as compared to that of the wt form. The strong acceleration of the analog folding in the presence of GSSG and the lack of effect of the addition of reducing agents suggests that the native form is preferably attained by direct oxidation of 2-disulfide species rather than reshuffling of scrambled forms. In contrast to this behavior, the reshuffling of scrambled isomers constitutes the rate-limiting step of many 3-disulfide proteins such as hirudin or potato carboxypeptidase inhibitor (PCI) (35, 36). In these proteins the scrambled isomers act as strong kinetic traps, accumulating in the late stages of the folding process and therefore the addition of thiol compounds greatly facilitates the efficiency of their folding process.

When C19A/C43A LCI folding is performed under mild denaturing conditions (e.g. 0.5-1 M GdnHCl) the 2-disulfide intermediates accumulate to a lesser extent, indicating that the partial destabilization of these species prevents them from acting as transient intermediates and thus, smoothes the folding funnel to the native protein, resulting in an acceleration of the folding rate. In contrast, the complete destabilization of these intermediates observed in concentrations of GdnHCl higher than 1 M negatively affects the folding rate and efficiency. In these conditions secondary structure propensities no longer direct protein folding through preferential pathways in which predominantly productive disulfide bonding occurs. Thus, the folding process proceeds through a sequential oxidation of 1-, 2-, and 3-disulfide species that finally accumulate as scrambled forms. The lower folding rate in strong denaturing conditions is caused by the difficulty of these scrambled isomers to attain the native form. The addition of thiol agents accelerates the folding process, confirming that in these conditions scrambled isomers have become a new kinetic trap, resembling now the folding process to that of other 3-disulfide proteins such as PCI or hirudin. The differences in the relative abundances among the scrambled forms when compared to those observed in the absence of denaturant agents suggests that scrambled isomers displaying more open and relaxed conformations, for instance the beads

form, show a higher prevalence, as also observed for other proteins such as tick anticoagulant peptide (TAP) or PCI (31, 37).

The structure of wt LCI consists of a five-stranded antiparallel β-sheet and one short α-helix (16). The intermediate III-B has two free cysteines, Cys19 and Cys43, which connect the α-helix and the β-sheet in the native structure of LCI, suggesting that its structure might display a disconnected α-helical section and a main core of a five antiparallel β-sheet stabilized by the remaining three native disulfide bonds. The fact that this intermediate elutes in the RP-HPLC very close to other species precludes its homogeneous purification and further precise structural characterization; therefore an analog was constructed by replacing the two free Cys with Ala (C19A/C43A LCI). The X-ray structure of C19A/C43A LCI alone and in complex with its target, carboxypeptidase A, shows that its overall chain fold is very similar to that of wt LCI. The comparison between wt LCI-CPA2 complex and C19A/C43A LCI-CPA complex reveals that despite the missing disulfide bond in the analog both, backbone and side chains adopt a conformation that closely resembles that in the wt form. This includes those residues close to the carboxypeptidase or near the mutation point, with a r.m.s.d. between both molecules of 0.42 Å for the backbone. Despite the similarity of both crystal structures, close examination reveals that the B-factors of the atoms in the side chains of the analog near the place of mutation are clearly higher than those in the wt structure. The B-factor characterizes not only thermal motion but also the lack of static ordering in the crystal, and could be indicative of net gain in flexibility in the regions spatially adjacent to the mutation sites. This increase in flexibility around the mutation points becomes more obvious in the crystal structure of the uncomplexed analog form. Although the backbone of the bound wt LCI and unbound analog are very similar, the electronic densities of the latter are absent in the residues conforming the last turn of the α-helix (residues 45-47), clearly indicating a higher flexibility in this region. This result is sharply coincident with our previous observation that in the solution structure of III-A intermediate, the other major kinetic intermediate in LCI folding reaction, the absence of the fourth disulfide bond results in a native-like structure with an increase of backbone flexibility around the free cysteines (Cys22 and Cys58) (see “work 5” of this thesis). According to our data, the docking of the analog to the carboxypeptidase turns into a reduction of conformational flexibility and thus in a net gain of structural specificity.

As in the case of LCI, analogs of BPTI and RNase intermediates that lack only one native disulfide bond and act as major kinetic traps along the folding process display structures similar to the native one (38-41). However, there are critical differences among the folding pathways of these proteins and that of LCI. For BPTI and RNase A the native form can be obtained by direct oxidation of the two free Cys displayed by the kinetic traps (6, 8). On the contrary, to attain the native LCI structure its kinetic traps need to undergo a general rearrangement and form a heterogeneous ensemble of non-native 4-disulfide (scrambled) isomers. We have previously postulated III-B intermediate to be a metastable form equivalent to what Scheraga and co-workers have defined as *disulfide-insecure* intermediates (42). In these intermediates the free thiol groups are as well protected as their disulfide bonds by the secondary structure elements; therefore such thiols cannot be simply exposed and oxidized by a local unfolding process. Structural fluctuations that expose the thiol groups are also likely to expose the disulfide bonds and promote their reshuffling instead of direct oxidation of the free thiols to obtain the native pairing. The crystal structure of the analog shows that residues Ala19 and Ala43 expose less than 7% of their surface to solvent, similar or even lower values than those of the rest of cysteine residues involved in native disulfide bonds (Cys11, Cys18, Cys22, Cys34, Cys58 and Cys62). Thus, by analogy we could assume that the reactive thiol groups in the free cysteines of the III-B intermediate (Cys19 and Cys43) are also well protected. This fact, together with a previously reported experiment where we demonstrated that the presence of an external thiol reagent did not affect the rate of conversion of III-B into scrambled forms (19), would confirm that this species behaves as a disulfide-insecure intermediate and highlights the critical role of accessibility in oxidative folding (42). Two other factors with a general influence on oxidative folding are the proximity and reactivity of thiols and disulfide bonds. Both are responsible for the quasi-equilibrium of different disulfide forms in the pre-folding stages of oxidative folding, but obviously they are not specific enough to neither strongly favor native disulfide bonding over nonnative ensembles nor to direct the protein toward its native conformation. Thus, the present study provides a mechanistic explanation for the role of III-B as a strong kinetic trap in LCI folding, suggesting that its accumulation relies mainly on its ability to protect from rearrangement both, its native disulfide bridges and free cysteines.

Previously reported conformational data of the III-B intermediate indicated that this molecule corresponded to a rather stable but less structured form than native LCI (19). In C19A/C43A LCI the lack of the two free Cys makes impossible the equilibrium between

the III-A and III-B forms that is established in wt LCI folding and modifies the redox potential of the interior of the protein, which results in the consolidation of a more native-like conformation. This analog shows the same overall structure, a similar activity and a faster and more efficient folding process than wt LCI. The question is then, why LCI has evolved to be a 4-disulfide protein. The data from the equilibrium unfolding experiments herein provide an unequivocally answer to this query: the disulfide bond established between Cys19-Cys43 provides additional stability to the protein. C19A/C43A LCI shows a global destabilization about 1.5 kcal·mol^{-1} relative to the wt form, as shown by all methods assessed. In other words, wt LCI is about 40% more stable than the 3-disulfide analog. Probably, the primary effect of the Cys19-Cys43 disulfide bonding in LCI is the inherent destabilization of the unfolded state of the protein relative to the native state, arising from the decreased entropy of the denatured state. Accordingly, disruption of this disulfide cross-link destabilizes the native state relative to the denatured state because of the increased entropy of the denatured protein. However, entropic contributions arising from an increase in global/local flexibility upon disulfide-disruption resulting in a more disordered folded state in the disulfide-deleted mutant protein than in the native structure of wt LCI cannot be discarded; such effects have been reported for folding intermediate analogs of RNase A (41).

Overall, the fourth disulfide provides LCI with conformational stability as well as reduced flexibility, that is, increased structural specificity. These qualities appear to be highly desirable for its function *in vivo* since LCI is a protease inhibitor from leech saliva evolved to act in blood. Lower conformational stability or higher backbone fluctuation in a 3-disulfide version of this molecule would probably render a protein more susceptible to proteolytic attacks. The advances made in understanding the stability and folding behavior of LCI provide a greater insight into the nature of this protein and constitute a basis for the development of variants of this molecule with enhanced activity and/or stability. This is of great interest given that carboxypeptidase inhibitors such as LCI or PCI strongly accelerate the tissue plasminogen activator-induced lysis of human plasma clots by modulating TAFI activity (Salamanca et al., manuscript in preparation) (43, 44) and may be used as lead compounds for the optimization of thrombolytic therapies.

The coordinates of the structure reported in this work have been deposited with the Protein Data Bank under the code YXXX.

References

1. Dill KA & Chan HS. From Levinthal to pathways to funnels. *Nat Struct Biol* 1997; 4: 10-19
2. Honig B. Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol* 1999; 293: 283-293
3. Creighton TE, Darby NJ & Kemmink J. The roles of partly folded intermediates in protein folding. *Faseb J* 1996; 10: 110-118
4. Creighton TE. Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 1986; 131: 83-106
5. Weissman JS & Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991; 253: 1386-1393
6. Goldenberg DP. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 1992; 17: 257-261
7. Wedemeyer WJ, Welker E, Narayan M & Scheraga HA. Disulfide bonds and protein folding. *Biochemistry* 2000; 39: 4207-4216
8. Scheraga HA, Wedemeyer WJ & Welker E. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol* 2001; 341: 189-221
9. van den Berg B, Chung EW, Robinson CV, Mateo PL & Dobson CM. The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *EMBO J* 1999; 18: 4794-4803
10. van den Berg B, Chung EW, Robinson CV & Dobson CM. Characterisation of the dominant oxidative folding intermediate of hen lysozyme. *J Mol Biol* 1999; 290: 781-796
11. Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998; 273: 32927-32933
12. Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-27181
13. Bouma BN & Meijers JCM. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
14. Silveira A, Schatteman K, Goossens F, Moor E, Scharpe S, Stromqvist M, Hendriks D & Hamsten A. Plasma procarboxypeptidase U in men with symptomatic coronary artery disease. *Thromb Haemost* 2000; 84: 364-368

15. Eichinger S, Schönauer V, Weltermann A, Minar E, Bialonczyk C, Hirschi M, Schneider B, Quehenberger P & Kyrle PA. Thrombin-activatable fibrinolysis inhibitor and the risk for recurrent venous thromboembolism. *Blood* 2004; 103: 3773-3776
16. Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000; 7: 322-328
17. Salamanca S, Villegas V, Vendrell J, Li L, Aviles FX & Chang JY. The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* 2002; 277: 17538-17543
18. Salamanca S, Li L, Vendrell J, Aviles FX & Chang JY. Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 2003; 42: 6754-6761
19. Arolas JL, Bronsoms S, Lorenzo J, Aviles FX, Chang JY & Ventura S. Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. *J Biol Chem* 2004; 279: 37261-37270
20. Peranen J, Rikkonen M, Hyvonen M & Kaariainen L. T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. *Anal Biochem* 1996; 236: 371-373
21. Morrison JF. The slow-binding and slow, tight-binding inhibition of enzyne-catalysed reactions. *Trends Biochem Sci* 1982; 7: 102-105
22. Reverter D, Ventura S, Villegas V, Vendrell J & Aviles FX. Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* 1998; 273: 3535-3541
23. Barbosa-Pereira PJ, Segura-Martín S, Oliva B, Ferrer-Orta C, Aviles FX, Coll M, Gomis-Rüth FX & Vendrell J. Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol* 2002; 321: 537-547
24. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 1993; 26: 795-800
25. Collaborative Computational Project Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr D* 1994; 50: 760-763
26. McRee DE. XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J Struc Biol* 1999; 125: 156-165
27. Lamzin VS & Wilson KS. Automated refinement of protein models. *Acta Crystallogr D* 1993; 49: 129-149
28. Chang JY, Li L & Lai PH. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 2001; 276: 4845-4852
29. Chang JY. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 1997; 272: 69-75

30. Chang JY, Li L & Bulychev A. The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 2000; 275: 8287-8289
31. Chang JY. Denatured states of tick anticoagulant peptide. Compositional analysis of unfolded scrambled isomers. *J Biol Chem* 1999; 274: 123-128
32. Matouschek A, Kellis JT Jr, Serrano L, Bycroft M & Fersht AR. Transient folding intermediates characterized by protein engineering. *Nature* 1990; 346: 440-445
33. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L & Serrano L. Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat Struct Biol* 2002; 9: 485-493
34. Myers JK, Pace CN & Scholtz JM. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 1995; 4: 2138-2148
35. Chatrenet B & Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 1993; 268: 20988-20996
36. Chang JY, Canals F, Schindler P, Querol E & Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 1994; 269: 22087-22094
37. Chang JY, Li L, Canals F & Aviles FX. The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. *J Biol Chem* 2000; 275: 14205-14211
38. Hurle MR, Eads CD, Pearlman DA, Seibel GL, Thomason J, Kosen PA, Kollman P, Anderson S & Kuntz ID. Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance. *Protein Sci* 1992; 1: 91-106
39. van Mierlo CP, Kemmink J, Neuhaus D, Darby NJ & Creighton TE. 1H NMR analysis of the partly-folded non-native two-disulphide intermediates (30-51,5-14) and (30-51,5-38) in the folding pathway of bovine pancreatic trypsin inhibitor. *J Mol Biol* 1994; 235: 1044-1061
40. Shimotakahara S, Rios CB, Laity JH, Zimmerman DE, Scheraga HA & Montelione GT. NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine pancreatic ribonuclease A: automated analysis of 1H, 13C, and 15N resonance assignments for wild-type and [C65S, C72S] mutant forms. *Biochemistry* 1997; 36: 6915-6929
41. Laity JH, Lester CC, Shimotakahara S, Zimmerman DE, Montelione GT & Scheraga HA. Structural characterization of an analog of the major rate-determining disulfide folding intermediate of bovine pancreatic ribonuclease A. *Biochemistry* 1997; 36: 12683-12699
42. Welker E, Narayan M, Wedemeyer WJ & Scheraga HA. Structural determinants of oxidative folding in proteins. *Proc Natl Acad Sci USA* 2001; 98: 2312-2316
43. Nagashima M, Werner M, Wang M, Zhao L, Light DR, Pagila R, Morser J & Verhallen P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* 2000; 98: 333-342

44. Walker JB, Hughes B, James I, Haddock P, Kluft C & Bajzar L. Stabilization versus inhibition of TAFIa by competitive inhibitors in vitro. *J Biol Chem* 2003; 278: 8913-8921

GENERAL DISCUSSION

Metallocarboxypeptidases are enzymes that participate in alimentary protein degradation and in selective regulatory processes like blood coagulation/fibrinolysis, inflammation, prohormone/neuropeptide maturation, insect/vegetal attack/defense strategies, etc (Vendrell et al., 2004). Within this context, the modulation of their activity using specific inhibitors is of great biotechnological and biomedical interest. However, only a few metallocarboxypeptidase inhibitors have been described from natural sources and among them, only the potato and leech inhibitors (PCI and LCI) have been structurally and functionally characterized in detail (Vendrell et al., 2000; Vendrell et al., 2004).

In the work 1 of this thesis we describe the isolation of a new metallocarboxypeptidase inhibitor from the blood-sucking tick *Rhipicephalus bursa*. This protein, named TCI, was purified to homogeneity using different chromatographic strategies and its N-terminus was sequenced. The complete cDNA encoding TCI was cloned from tick mRNA by a combination of 3' and 5' RACE techniques using degenerated oligonucleotides based on the N-terminal protein sequence. It contains an open reading frame coding for a protein of 97 residues that consists of a hydrophobic signal peptide (22 residues) that precedes mature TCI. Periplasmic and intracellular expression systems were tested in *Escherichia coli* for the recombinant production of TCI. However, expression was only achieved by the former system, and its yield was moderate (2 mg/L culture), probably due to the large number of disulfide bonds of the protein (12 Cys residues). Further optimization of this expression system would be required for the production of large quantities of this protein and for isotope-labeling.

Recombinant TCI (rTCI) was shown to be correctly folded and functionally equivalent to the inhibitor isolated from ticks; natural and recombinant forms of TCI inhibit carboxypeptidases of the A/B subfamily with equilibrium dissociation constants in the nanomolar range. As a tight binding inhibitor of plasma carboxypeptidase B, also known as TAFIa, rTCI enhances the plasma fibrinolysis induced *in vitro* by tissue plasminogen activator. Previous works had showed that PCI and LCI also act as pro-fibrinolytics (Nagashima et al., 2000; Salamanca, 2003a; Walker et al., 2003). CD and NMR analysis demonstrated that rTCI is a protein strongly constrained by its disulfide bonds, unusually stable over a wide range of pH and highly resistant to denaturing conditions. These properties are desirable in a protein that has potential in human therapies, e.g. to prevent or treat thrombotic disorders. Actually, its pro-fibrinolytic action

is likely to be its biological function in ticks, since these animals need to maintain blood liquidity during feeding and digestion. Also, TCI could modulate inflammation and host defense mechanisms by affecting TAFI and/or mast-cell carboxypeptidase A (Bajzar et al., 2004).

Although TCI is a novel molecule that does not show sequence similarity with any known protein, its C-terminus resembles those of other carboxypeptidase inhibitors, suggesting a common mechanism of action. In this basis, the work 2 was carried out to determine the three-dimensional structure of TCI so that its mechanism of inhibition toward carboxypeptidases could be elucidated. The structure of TCI defines a protein motif consisting of two small domains (about 35 residues long) that are structurally equivalent despite their low sequence homology. Each domain comprises a short N-terminal α-helix and a small twisted antiparallel triple-stranded β-sheet that are stabilized by the presence of three disulfide bonds. The structure and disulfide pattern of each domain are very similar to those of proteins included within the β-defensin-fold family although they do not show sequence homology. Proteins presenting this fold display numerous pharmacological effects including ion-channel inhibition, and analgesic and myonecrotic activities (Torres & Kuchel, 2004). Thus, TCI could have another function out of fibrinolysis participating in the modulation of host defenses, e.g. as an analgesic compound.

The complexes of TCI bound to either bovine carboxypeptidase A or human carboxypeptidase B showed that both domains of TCI anchor on the surface of mammalian carboxypeptidases in a novel double-headed manner not previously described for carboxypeptidase inhibitors. Unlike PCI or LCI, which are proteins with a unique globular structure, the presence of a second domain in TCI allows the establishment of numerous secondary interactions with enzyme's residues located far away from the active site groove. Thus, the N-terminal domain of TCI binds to an exosite located in the enzyme surface. Similarly to PCI and LCI, TCI interacts with the active site groove of the enzyme via its C-terminal tail in a way that mimics substrate binding. This was already predicted in the work 1 from the resemblance of the TCI C-tail with that of PCI and LCI; such a behavior represents a clear example of convergent evolution dictated by the structures of the target carboxypeptidases. It is worth mentioning that TCI displays indistinguishable mechanisms of inhibition toward forms A and B of carboxypeptidases. However, the position of TCI domains is slightly different when bound to the carboxypeptidase in the two complexes. This fact, together with the absence of inter-domain interactions within

TCI suggests that they can adopt different orientations when binding to different carboxypeptidases, a potential "adaptability" not found in other equivalent proteinaceous inhibitors.

The structure of the TCI-hCPB complex determined in this work may be useful to better understand the biological action of TAFI, since human CPB is the enzyme most closely related to plasma CPB, also called TAFI (48% sequence identity; Barbosa-Pereira et al., 2002). In addition, the data derived from works 1 and 2 (section I), together with the information obtained in the work 3 may be used in the design of improved and/or minimized drugs to modulate TAFI activity. In the work 3, structural and enzymatic studies performed on a series of PCI mutants showed that the polar residues in the secondary binding region of PCI are not likely to be involved in the stabilization of the PCI-CPA complex, whereas aromatic residues appear to play an important role (~ 30% of the binding free energy is attributable to them). This study demonstrates that the proper positioning of the primary binding site of PCI to bind and inhibit carboxypeptidase A is significantly dependent of the secondary binding region. As such, the secondary region is involved in the maintenance of the C-terminal tail in a proper orientation, providing the conformational rigidity needed for the binding to the enzyme. These observations should be taken into account for the rational improvement of PCI and/or TCI inhibitory properties toward carboxypeptidases.

Carboxypeptidase inhibitors are small disulfide-rich proteins suitable to be used as models in the study of folding and unfolding pathways. For disulfide-containing proteins, transient forms that accumulate along the folding reaction can be trapped and characterized, a fact that cannot be accomplished for "normal" (i.e. disulfide-free) proteins that present intermediates with a very short half-life (Creighton, 1986). The folding and unfolding processes of several 3- and 4-disulfide proteins have been elucidated using different techniques showing a high diversity of mechanisms among them (Chang, 2004). For proteins like BPTI, the formation of a limited number of intermediates with native disulfide bonds and native-like structures prevail along the folding pathway, while in the case of proteins like hirudin, folding proceeds through the formation of a heterogeneous population of non-native (scrambled) intermediates that is followed by their disulfide reshuffling to finally acquire the native form. Unlike BPTI, for the latter proteins non-covalent interactions do not seem to significantly participate in guiding the acquisition of the native structure, at least during the early phases of the folding reaction (Chang, 1997).

In the work 3 we also intended to study in depth the determinants underlying the folding and unfolding of PCI, a protein with a folding mechanism very similar to that of hirudin. For this task, we used an alanine-scanning mutagenesis approach and compared the structural and folding properties of the generated mutants with those of the wild-type form. The data clearly indicated that non-covalent forces are apparently insufficient to drive the unfolded protein across the folding energy barrier, and a reduction of the system entropy and a conformational search by disulfide bond formation (through scrambled isomers) are needed. On the other hand, non-covalent interactions drive the folding of PCI at the reshuffling stage, the final and rate-limiting step of the process, given that the introduction of new non-covalent intramolecular contacts in PCI results in variants with improved folding efficiency. Thus, the low folding efficiency exhibited by PCI when compared to that of other topologically related proteins such as hirudin or TAP is probably due to the scarce non-covalent contacts displayed by this molecule. In addition, this work revealed a correlation between the conformational stability of native PCI and its structural uniqueness. PCI variants more stable than the wt form display lower amounts of scrambled species in equilibrium, whereas in less stable PCI variants the equilibrium is more displaced toward the non-native conformations. This last observation is in total agreement with those made in proteins causing conformational diseases, where stabilization of the native structure has been used as a tool to restrict conformational fluctuations that may drive to protein aggregation (Chiti et al., 2000).

Despite the low content of secondary structure displayed by PCI, the work 3 shows that its non-covalent interactions play an important role during its folding reaction. In order to test the role of non-covalent contacts in a protein displaying a high degree of secondary structure, we characterized in detail the folding process of LCI (works 4-6), which displays a much higher secondary order and one additional disulfide bond when compared to PCI. In the work 4 we show that reduced and denatured LCI refolds through a sequential flow of 1-, 2-, 3-, and 4-disulfide (scrambled) species to reach the native form. Folding intermediates of LCI comprise two major 3-disulfide species (III-A and III-B) that exclusively contain native disulfide bonds and inter-convert between them reaching equilibrium prior to the formation a heterogeneous population of scrambled isomers. From the results obtained in the stop/go experiments of this work, III-A and III-B forms appear to have both, native disulfides bonds and free thiols, similarly protected from the solvent, therefore major structural rearrangements through the formation of scrambled isomers are required to render native LCI.

These intermediates correspond to strong local minima in the energy landscape thus acting as kinetic traps during the oxidative folding. Accordingly, their accumulation is prevented when they are destabilized under mild denaturing conditions, which leads to a significant acceleration of the pathway; native disulfides and free cysteines probably become more solvent-accessible due to local unfolding events and can easily convert into the scrambled forms. By adding enough denaturant to strongly destabilize III-A and III-B species scrambled forms accumulate along LCI oxidative folding pathway and their relative abundance differs from that observed in the absence of denaturant. These scrambled isomers display a higher difficulty to attain the native-bond pairing and this last stage of LCI folding becomes extremely slow under these conditions. In this work, we also describe the reductive unfolding pathway of LCI that undergoes an apparent all-or-none mechanism. However, low amounts of III-A and III-B intermediates can be detected, suggesting differences in protection against reduction among the disulfide bonds. The preliminary structural characterization of III-A and III-B forms made in this work suggests a correlation between the protection of native disulfide bonds inside a somehow structured conformation and the accumulation of intermediates along the folding reaction of LCI.

Gaining knowledge on the conformation of folding intermediates is the best way to understand the reasons behind their accumulation along the folding process (van den Berg et al., 1999). However, this is far from being an easy task since intermediates are usually short-lived species. In the work 5 the III-A intermediate was directly purified from the folding reaction using RP-HPLC and structurally characterized by NMR spectroscopy. The results showed that this species presents both, a highly native-like structure with a β-sheet core and an inhibitory activity that are very similar to those of native LCI. However, this intermediate is significantly more flexible than the native form, mainly in the regions around the free cysteines (Cys 22 and Cys 58). The fourth disulfide contributes to reduce flexibility, thus providing structural specificity to native LCI. III-A constitutes the first *disulfide insecure* intermediate being structurally determined. In these metastable forms defined by Scheraga and co-workers (Welker et al., 2001), the free thiol groups are as well protected as their disulfide bonds; therefore structural fluctuations that expose the thiol groups are also likely to expose the disulfide bonds and promote their reshuffling instead of oxidation of the free thiols to the native pairing. Confirming this hypothesis, direct oxidation of III-A intermediate to the fully native protein could not be attained probably due to the burial of the two free cysteines inside a native-like structure. Indeed, exchange amide experiments showed that in III-A both free cysteines and the three disulfide bonds

are located in protected regions or close to them and therefore are not solvent-accessible or have limited accessibility.

In the work 5 we also demonstrate that theoretical approaches based on topological constrains (Fold-X software; Guerois & Serrano, 2000) can predict with surprising accuracy the presence of the III-A folding intermediate, which constitutes the highest energy barrier along LCI folding process. Taking into account that these kinds of programs only use for their predictions the information related to secondary structure elements connectivity and self-organization and do not consider the presence of disulfide bonds, the successful prediction of the LCI folding hierarchy constitutes a strong indication that secondary structure plays a crucial role in the folding of LCI. We also determined the theoretical folding pathways of two 3-disulfide proteins with extremely different folding mechanisms: PCI, a protein with a very poor secondary structure content, highly constrained by its disulfide bonds, and BPTI, a protein that is highly influenced by non-covalent interactions. Our predictions were able to capture the main features lying beneath the folding of these proteins, and reinforce the hypothesis that, as it occurs with small proteins without disulfide bonds, the way a small disulfide-rich protein folds is critically influenced by its secondary structure organization.

Finally, in the last work of this thesis (work 6) we intended to characterize the III-B folding intermediate to achieve a global understanding of the LCI folding process. Due to the impossibility to purify this intermediate in a homogeneous state from the folding reaction, we decided to construct a III-B analog in which the native Cys19 and Cys43 residues were replaced with alanine residues. The folding pathway of this mutant was elucidated by kinetic analysis of the acid-trapped folding intermediates showing that the reduced protein refolds rapidly through a sequential flow of 1-, 2-, and 3-disulfide species to reach the native form in an efficient way, with a low accumulation of 2-disulfide intermediates and 3-disulfide (scrambled) isomers. The crystal three-dimensional structure of this analog, alone and in complex with carboxypeptidase A, is very similar to that of wild-type LCI, although a higher flexibility can be observed in the regions close to the mutation points. In addition, functional analysis revealed that this analog displays a native-like inhibitory activity. After obtaining these data, an immediate question is: why has LCI evolved to be a 4-disulfide protein instead of a protein with 3-disulfide bonds, with a similar structure and inhibitory efficiency and a less complicated and faster folding pathway? The data from the equilibrium unfolding experiments of this work provided an unequivocally answer to that: the disulfide bond established between Cys19-Cys43

provides stability to the protein. In this way, wild-type LCI is about 40% more stable than the III-B analog. This higher stability should be important for LCI *in vivo* function to justify the occurrence of the fourth disulfide bond.

Traditionally, the folding of small disulfide-rich proteins has been seen as a special case of protein folding, considered apart from that of small two-state proteins (Creighton, 1990). Accordingly, special techniques have been used for the characterization of the folding pathways of these proteins. This has precluded the direct comparison of the obtained data with the available data for the rest of proteins. The studies on disulfide-rich proteins have been focused on the description of the different species occurring along the process and in their kinetics of inter-conversion, but little attention has been paid to key aspects such as thermodynamics, structural characterization of folding intermediates or *in silico* prediction of folding pathways. In the works 3-6 (section II) we have addressed these issues using our protein models (i.e. the carboxypeptidase inhibitors PCI and LCI) to shed some light on the rules driving the folding of disulfide-rich proteins. On the overall, the results derived from this thesis suggest that the forces governing the folding of small disulfide-rich proteins do not differ much from those governing the folding of small proteins without disulfide bonds, i.e. native-like interactions between elements of secondary structure. This joins the discontinuity existent between the folding of disulfide-containing and non-containing proteins, and helps to advance toward a unifying theory for the folding/unfolding mechanism(s) of these proteins.

CONCLUSIONS

Section I.

1. The recombinant form of TCI is fully functional and inhibits carboxypeptidases of the A/B subfamily with inhibition constants in the nanomolar range. TCI is a protein strongly constrained by its disulfide bonds, highly stable over a wide range of pH and denaturing conditions that stimulates blood fibrinolysis *in vitro* and thus may have potential for applications in thrombotic disorders.

2. The three-dimensional crystal structures of recombinant TCI bound to either bCPA or hCPB show that TCI comprises two domains that are structurally very similar, each consisting of a short α-helix followed by a small twisted antiparallel β-sheet. These domains display high structural homology to proteins of the β-defensin-fold family.

3. TCI anchors on the surface of mammalian carboxypeptidases in a double-headed manner not previously described for other carboxypeptidase inhibitors: The C-tail of the C-terminal domain interacts with the active site of the enzyme in a way that mimics substrate binding, and the N-terminal domain binds to an exosite distinct from the active site groove.

Section II.

1. The secondary binding site of PCI in its interaction with bCPA maintains the proper positioning of the primary site (the C-tail) for an efficient binding and inhibition of the enzyme. In addition, in this secondary region the aromatic residues play a key role in the stabilization of the PCI-CPA complex, whereas polar residues contribute little to this task.

2. Non-covalent forces drive the oxidative folding of PCI at the reshuffling stage, the rate-limiting step of the process. More stable variants with improved folding efficiency can be created introducing new non-covalent intra-molecular contacts in PCI.

3. The oxidative folding of LCI is characterized by the presence of two predominant 3-disulfide species (III-A and III-B) and a heterogeneous population of scrambled isomers that strongly and consecutively accumulate along the folding reaction. III-A and III-B are

stable species that exclusively contain native disulfide bonds and slightly accumulate along the reductive unfolding process.

4. Structural characterization of the III-A intermediate shows that this species displays a highly native-like structure although it is clearly more flexible than the native form of LCI. Direct oxidation of this species to the fully native protein is restricted by the burial of its two free cysteine residues inside a native-like structure, which accounts for its strong accumulation in the LCI folding process.

5. Prediction of the theoretical folding pathways of three proteins (LCI, BPTI and PCI) with the Fold-X algorithm underlines the importance of native-like interactions between elements of secondary structure in directing the folding of disulfide-rich proteins.

6. Characterization of the III-B analog shows that this protein displays the same overall structure, a similar activity, a faster and more efficient folding process than wild-type LCI, but a lower stabilization, which indicates that the fourth disulfide bond provides LCI with both high stability and structural specificity.

PERSPECTIVES

In the section I of the present thesis we characterize TCI structurally and functionally. Taken together, the derived data introduce TCI as a candidate molecule to be used as a drug for applications in thrombotic disorders. Thus, the redesign of this molecule in order to improve its functionality and applicability would be of great interest. A first step in this direction is the minimization of this protein. Since TCI is a two-domain protein that inhibits carboxypeptidases mainly via the interaction of its C-tail with the enzyme active site, we are currently constructing a TCI form consisting of only the C-terminal domain. The analysis of this "hemi-TCI" will be useful to gain insights into the inhibition mechanism of the entire protein.

The unique architecture of the TCI fold, with two domains each containing three "independent" disulfide bonds, makes this molecule an ideal model to study the role of inter-domain contacts in the folding and unfolding of disulfide-rich proteins. In this way, we plan to characterize the pathways of oxidative folding and reductive unfolding of TCI, as well as its conformational stability by disulfide scrambling.

The section II shows how the folding of LCI is directed by interactions between elements of secondary structure rather than by the formation of its disulfide bonds. This is shown by the native-like structure displayed by the III-A intermediate and the accurate prediction of this form made by the Fold-X algorithm. To finally validate this hypothesis, we are constructing series of LCI mutants missing one, two, three or four disulfides. The analysis of these species will serve to study in depth the determinants underlying the LCI folding process and to provide new insights into the forces governing the folding of disulfide-rich proteins.

GENERAL REFERENCES

Aagaard A, Listwan P, Cowieson N, Huber T, Ravasi T, Wells CA, Flanagan JU, Kellie S, Hume DA, Kobe B & Martin JL. An inflammatory role for the mammalian carboxypeptidase inhibitor latexin: relationship to cystatins and the tumor suppressor TIG1. *Structure* 2005; 13: 309-317

Alm E & Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci USA* 1999; 96: 11305–11310

Aloy P, Catasus L, Villegas V, Reverter D, Vendrell J & Aviles FX. Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B. *Biol Chem* 1998; 379: 149-155

Anfinsen CB, Haber E, Sela M & White FH Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA* 1961; 47: 1309-1314

Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973; 181: 223-230

Antuch W, Guntert P, Billeter M, Hawthorne T, Grossenbacher H & Wuthrich K. NMR solution structure of the recombinant tick anticoagulant protein (rTAP), a factor Xa inhibitor from the tick Ornithodoros moubata. *FEBS Lett* 1994; 352: 251-257

Aviles FX, Vendrell J, Guasch A, Coll A & Huber R. Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. *Eur J Biochem* 1993; 211: 381-389. Review

Bajzar L, Manuel R & Nesheim ME. Purification and characterization of TAFI, a thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1995; 270: 14477-14484

Bajzar L, Nesheim M, Morser J & Tracy PB. Both cellular and soluble forms of thrombomodulin inhibit fibrinolysis by potentiating the activation of thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 2792-2798

Bajzar L, Jain N, Wang P & Walker JB. Thrombin activatable fibrinolysis inhibitor: not just an inhibitor of fibrinolysis. *Crit Care Med* 2004; 32: S320-324

Baker D. A surprising simplicity to protein folding. *Nature* 2000; 405: 39-42

Baldwin RL. How does protein folding get started? *Trends Biochem Sci* 1989;14: 291-294. Review

Baldwin RL. Protein folding. Matching speed and stability. *Nature* 1994; 369: 183-184

Baldwin RL. The nature of protein folding pathways: the classical versus the new view. *J Biomol NMR* 1995; 5: 103-109. Review

Baldwin RL. Protein folding from 1961 to 1982. *Nat Struct Biol* 1999; 6: 814-817. Historical Article

Barbosa-Pereira PJ, Segura-Martín S, Oliva B, Ferrer-Orta C, Aviles FX, Coll M, Gomis-Rüth FX & Vendrell J. Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol* 2002; 321: 537-547

Barrett AJ. Peptidases: a view of classification and nomenclature. In Turk V (ed.) Proteases: New Perspectives, Birkyhäuser Verlag Basel/Switzerland. 1999; pp. 1-12

Berti PJ & Storer AC. Alignment/phylogeny of the papain superfamily of cysteine proteiases. *J Mol Biol* 1995; 246: 273-283

Blanco-Aparicio C, Molina MA, Fernandez-Salas E, Frazier ML, Mas JM, Querol E, Aviles FX & de Llorens R. Potato carboxypeptidase inhibitor, a T-knot protein, is an epidermal growth factor antagonist that inhibits tumor cell growth. *J Biol Chem* 1998; 273: 12370-12377

Bode W & Huber R. Natural protein inhibitors and their interaction with proteinases. *Eur J Biochem* 1992; 204: 433-451. Review

Bode W, Fernandez-Catalan C, Nagase H & Maskos K. Endoproteinase-protein inhibitor interactions. *APMIS* 1999; 107: 3-10. Review

Bode W & Huber R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim Biophys Acta* 2000; 1477: 241-252. Review

Bouma BN & Meijers JC. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574. Review

Bronsoms S, Villanueva J, Canals F, Querol E & Aviles FX. Analysis of the effect of potato carboxypeptidase inhibitor pro-sequence on the folding of the mature protein. *Eur J Biochem* 2003; 270: 3641-3650

Bulychev A & Chang JY. Unfolding of hirudin characterized by the composition of denatured scrambled isomers. *J Protein Chem* 1999; 18: 771-778

Campbell W, Okada N & Okada H. Carboxypeptidase R is an inactivator of complement-derived inflammatory peptides and an inhibitor of fibrinolysis. *Immunol Rev* 2001; 180: 162-167. Review

Campbell WD, Lazoura E, Okada N & Okada H. Inactivation of C3a and C5a octapeptides by carboxypeptidase R and carboxypeptidase N. *Microbiol Immunol* 2002; 46: 131-134

Casaretto JA & Corcuera LJ. Plant proteinase inhibitors: a defensive response against insects. *Biol Res* 1995; 28: 239-249. Review

Chang JY. Controlling the speed of hirudin folding. *Biochem J* 1994a; 300: 643-650

Chang JY, Canals F, Schindler P, Querol E & Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 1994b; 269: 22087-22094

Chang JY, Schindler P, Ramseier U & Lai PH. The disulfide folding pathway of human epidermal growth factor. *J Biol Chem* 1995; 270: 9207-9216

Chang JY. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 1996; 35: 11702-11709

Chang JY. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 1997; 272: 69-75

Chang JY. Denatured states of tick anticoagulant peptide. Compositional analysis of unfolded scrambled isomers. *J Biol Chem* 1999; 274: 123-128

Chang JY & Ballatore A. Structure and heterogeneity of the one- and two-disulfide folding intermediates of tick anticoagulant peptide. *J Protein Chem* 2000a; 19: 299-310

Chang JY, Li L & Bulychev A. The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 2000b; 275: 8287-8289

Chang JY, Li L, Canals F & Aviles FX. The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. *J Biol Chem* 2000c; 275: 14205-14211

Chang JY, Li L & Lai PH. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 2001; 276: 4845-4852

Chang JY. The folding pathway of alpha-lactalbumin elucidated by the technique of disulfide scrambling. Isolation of on-pathway and off-pathway intermediates. *J Biol Chem* 2002a; 277: 120-126

Chang JY & Li L. Pathway of oxidative folding of alpha-lactalbumin: a model for illustrating the diversity of disulfide folding pathways. *Biochemistry* 2002b; 41: 8405-8413

Chang JY. Evidence for the underlying cause of diversity of the disulfide folding pathway. *Biochemistry* 2004; 43: 4522-4529

Chatrenet B & Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 1993; 268: 20988-20996

Chiti F, Taddei N, Bucciantini M, White P, Ramponi G & Dobson CM. Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo J* 2000; 19: 1441-1449

Clawson GA. Protease inhibitors and carcinogenesis: a review. *Cancer Invest* 1996; 14: 597-608. Review

Clore GM, Gronenborn AM, Nilges M & Ryan CA. Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* 1987; 26: 8012-8023

Coll M, Guasch A, Aviles FX & Huber R. Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J* 1991; 10: 1-9

Creagh EM, Conroy H & Martin SJ. Caspase-activation pathways in apoptosis and immunity. *Immunol Rev* 2003; 193: 10-21. Review

Creighton TE. Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol* 1978; 33: 231-297

Creighton TE & Goldenberg DP. Kinetic role of a meta-stable native-like two-disulphide species in the folding transition of bovine pancreatic trypsin inhibitor. *J Mol Biol* 1984; 179: 497-526

Creighton TE. Disulfide bonds as probes of protein folding pathways *Methods Enzymol* 1986; 131: 83-106

Creighton TE. Protein folding. *Biochem J* 1990; 270: 1-16. Review

Creighton TE, Zapun A & Darby NJ. Mechanisms and catalysts of disulfide bond formation in proteins. *Trends Biotechnol* 1995; 13: 18-23

Creighton TE, Darby NJ & Kemmink J. The roles of partly folded intermediates in protein folding. *Faseb J* 1996; 10: 110-118

Daggett V & Fersht A. Is there a unifying mechanism for protein folding? *Trends Biochem Sci* 2003; 28: 18-25. Review

Dill KA & Chan HS. From Levinthal to pathways to funnels. *Nat Struct Biol* 1997; 4: 10-19. Review

Dobson CM. Experimental investigation of protein folding and misfolding. *Methods* 2004; 34: 4-14

Eaton DL, Mallory BE, Tsai SP, Henzel W & Drayna D. Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma. *J Biol Chem* 1991; 266: 21833-21838

Eichinger S, Schönauer V, Weltermann A, Minar E, Bialonczyk C, Hirschi M, Schneider B, Quehenberger P & Kyrle PA. Thrombin-activatable fibrinolysis inhibitor and the risk for recurrent venous thromboembolism. *Blood* 2004; 103: 3773-3776

Ewbank JJ & Creighton TE. Pathway of disulfide-coupled unfolding and refolding of bovine alpha-lactalbumin. *Biochemistry* 1993a; 32: 3677-3693

Ewbank JJ & Creighton TE. Structural characterization of the disulfide folding intermediates of bovine alpha-lactalbumin. *Biochemistry* 1993b; 32: 3694-3707

Fersht AR. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* 1995; 92: 10869-10873

Fersht AR. Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* 1997; 7: 3-9. Review

Fricker LD & Snyder SH. Purification and characterization of enkephalin convertase, an enkephalin-synthesizing carboxypeptidase. *J Biol Chem* 1983; 258: 10950-10955

Fricker LD. Carboxypeptidase E. *Annu Rev Physiol* 1988; 50: 309-321. Review

Garcia-Saez I, Reverter D, Vendrell J, Aviles FX & Coll M. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J* 1997; 16: 6906-6913

Goldenberg DP. Native and non-native intermediates in the BPTI folding pathway. *Trends Biochem Sci* 1992; 17: 257-261

Gomis-Rüth FX, Gomez M, Bode W, Huber R & Aviles FX. The three-dimensional structure of the native ternary complex of bovine pancreatic procarboxypeptidase A with proproteinase E and chymotrypsinogen C. *EMBO J* 1995; 14: 4387-4394

Gomis-Rüth FX, Companys V, Qian Y, Fricker LD, Vendrell J, Aviles FX & Coll M. Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. *EMBO J* 1999; 18: 5817-5826

Gonzalez C, Neira JL, Ventura S, Bronsoms S, Rico M & Aviles FX. Structure and dynamics of the potato carboxypeptidase inhibitor by 1H and 15N NMR. *Proteins* 2003; 50: 410-422

Guasch A, Coll M, Aviles FX & Huber R. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J Mol Biol* 1992; 224: 141-157

Guerois R & Serrano L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J Mol Biol* 2000; 304: 967-982

Hass GM, Nau H, Biemann K, Grahn DT, Ericsson LH & Neurath H. The amino acid sequence of a carboxypeptidase inhibitor from potatoes. *Biochemistry* 1975; 14: 1334-1342

Hass GM & Derr JE. Purification and characterization of the carboxypeptidase isoinhibitors from potatoes. *Plant Physiol* 1979a; 64: 1022-1028

Hass GM & Derr JE. Distribution of carboxypeptidase isoinhibitors in the potato plant. *Plant Physiol* 1979b; 64: 1029-1031

Hass GM & Ryan CA. Carboxypeptidase inhibitor from potatoes. *Methods Enzymol* 1981a; 80: 778-791

Hass GM & Hermodson MA. Amino acid sequence of a carboxypeptidase inhibitor from tomato fruit. *Biochemistry* 1981b; 20: 2256-2260

Hendrick JP & Hartl FU. The role of molecular chaperones in protein folding. *FASEB J* 1995; 9: 1559-1569. Review

Homandberg GA, Litwiller RD & Peanasky RJ. Carboxypeptidase inhibitors from Ascaris suum: the primary structure. *Arch Biochem Biophys* 1989; 270: 153-161

Honig B. Protein folding: from Levinthal paradox to structure prediction. *J Mol Biol* 1999; 293: 283-293. Review

Hook VY, Azaryan AV, Hwang SR & Tezapsidis N. Proteases and the emerging role of protease inhibitors in prohormone processing. *FASEB J* 1994; 8: 1269-1278. Review

Hooper NM. Families of zinc metalloproteases. *FEBS Lett* 1994; 354: 1-6. Review

Hooper NM. The biological roles of zinc and families of zinc metalloproteases. In Hooper NM (ed.) Zinc Metalloproteases in Health and Disease. Taylor and Francis, London. 1996; pp. 1-21

Ikai A & Tanford C. Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* 1971; 230: 100-102

Itzhaki LS, Otzen DE & Fersht AR. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 1995; 254: 260-288

Kalafatis M, Egan JO, van't Veer C, Cawthern KM & Mann KG. The regulation of clotting factors. *Crit Rev Eukaryot Gene Expr* 1997; 7: 241-280. Review

Karplus M & Weaver DL. Protein-folding dynamics. *Nature* 1976; 260: 404-406

Kim PS & Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* 1982; 51: 459-489. Review

Kim PS & Baldwin RL. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 1990; 59: 631-660. Review

Klement P, Liao P & Bajzar L. A novel approach to arterial thrombolysis. *Blood* 1999; 94: 2735-2743

Krowarsch D, Cierpicki T, Jelen F & Otlewski J. Canonical protein inhibitors of serine proteases. *Cell Mol Life Sci* 2003; 60: 2427-2444. Review

Kuzuya M & Iguchi A. Role of matrix metalloproteinases in vascular remodeling. *J Atheroscler Thromb* 2003; 10: 275-282. Review

Lenarcic B & Turk V. Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J Biol Chem* 1999; 274: 563-566

Levin Y, Skidgel RA & Erdos EG. Isolation and characterization of the subunits of human plasma carboxypeptidase N (kininase i). Proc *Natl Acad Sci USA* 1982; 79: 4618-4622

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

Lin SL & Nussinov R. A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. *Nat Struct Biol* 1995; 2: 835-837

Liu Q, Yu L, Gao J, Fu Q, Zhang J, Zhang P, Chen J & Zhao S. Cloning, tissue expression pattern and genomic organization of latexin, a human homologue of rat carboxypeptidase A inhibitor. *Mol Biol Rep* 2000; 27: 241-246

Marino-Buslje C, Molina MA, Canals F, Aviles FX & Querol E. Overproduction of a recombinant carboxypeptidase inhibitor by optimization of fermentation conditions. *Applied Microbiology and Biotechnology* 1994; 41: 632-637

Marino-Buslje C, Venhudova G, Molina MA, Oliva B, Jorba X, Canals F, Aviles FX & Querol E. Contribution of C-tail residues of potato carboxypeptidase inhibitor to the binding to carboxypeptidase A. A mutagenesis analysis. *Eur J Biochem* 2000; 267: 1502-1509

Martinez JC & Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 1999; 6: 1010-1016

Mas JM, Aloy P, Marti-Renom MA, Oliva B, Blanco-Aparicio C, Molina MA, de Llorens R, Querol E & Aviles FX. Protein similarities beyond disulphide bridge topology. *J Mol Biol* 1998; 284: 541-548

Matthews KW, Mueller-Ortiz SL & Wetsel RA. Carboxypeptidase N: a pleiotropic regulator of inflammation. *Mol Immunol* 2004; 40: 785-793

Molina MA, Aviles FX & Querol E. Expression of a synthetic gene encoding potato carboxypeptidase inhibitor using a bacterial secretion vector. *Gene* 1992; 116: 129-138

Molina MA, Marino C, Oliva B, Aviles FX & Querol E. C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. *J Biol Chem* 1994; 269: 21467-21472

Munoz V & Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 1999; 96: 11311-11316

Mutch NJ, Moore NR, Wang E & Booth NA. Thrombus lysis by uPA, scuPA and tPA is regulated by plasma TAFI. *J Thromb Haemost* 2003; 1: 2000-2007

Myles TM, Nishimura T, Yun TH, Nagashima M, Morser J, Patterson AJ, Pearl RG & Leung L LK. Thrombin activatable fibrinolysis inhibitor, a potential regulator of vascular inflammation. *J Biol Chem* 2003; 278: 51059-51067

Nagase H. Matrix metalloproteinases. In Hooper NM (ed.) Zinc Metalloproteases in Health and Disease. Taylor and Francis, London. 1996; pp. 153-204

Nagashima M, Werner M, Wang M, Zhao L, Light DR, Pagila R, Morser J & Verhallen P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* 2000; 98: 333-342

Narayan M, Welker E, Wedemeyer WJ & Scheraga HA. Oxidative folding of proteins. *Acc Chem Res* 2000; 33: 805-812

Narayan M, Welker E & Scheraga HA. Native conformational tendencies in unfolded polypeptides: development of a novel method to assess native conformational tendencies in the reduced forms of multiple disulfide-bonded proteins. *J Am Chem Soc* 2003; 125: 2036-2037

Neurath H. Evolution of proteolytic enzymes. *Science* 1984; 224: 350-357. Review

Normant E, Martres MP, Schwartz JC & Gros C. Purification, cDNA cloning, functional expression, and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor. *Proc Natl Acad Sci USA* 1995; 92: 12225-12229

Novikova EG, Reznik SE, Varlamov O & Fricker LD. Carboxypeptidase Z is present in the regulated secretory pathway and extracellular matrix in cultured cells and in human tissues. *J Biol Chem* 2000; 275: 4865-4870

Otlewski J, Krowarsch D & Apostoluk W. Protein inhibitors of serine proteinases. *Acta Biochim Pol* 1999; 46: 531-565. Review

Pallarès I, Bonet R, García-Castellanos R, Ventura S, Aviles FX, Vendrell J & Gomis-Rüth FX. Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. *Proc Natl Acad Sci USA* 2005; 102: 3978-3983

Peters JM. Proteasomes: protein degradation machines of the cell. *Trends Biochem Sci* 1994; 19: 377-382. Review

Plaxco KW, Simons KT & Baker D. Contact order, transitionstate placement and the refolding rates of single domain proteins. *J Mol Biol* 1998; 277: 985-994

Plaxco KW, Simons KT, Ruczinski I & Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* 2000; 39: 11177-11183

Rao KR & Brew K. Calcium regulates folding and disulfide-bond formation in alpha-lactalbumin. *Biochem Biophys Res Commun* 1989; 163: 1390-1396

Rawlings ND, O'Brien E & Barrett AJ. MEROPS: the protease database. *Nucleic Acids Res* 2002; 30: 343-346

Rawlings ND, Tolle DP & Barrett AJ. MEROPS: the peptidase database. *Nucleic Acids Res* 2004a; 32: D160-164

Rawlings ND, Tolle DP & Barrett AJ. Evolutionary families of peptidase inhibitors. *Biochem J* 2004b; 378: 705-716. Review

Rees DC & Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* 1982; 160: 475-498

Rees DC, Lewis M & Lipscomb WN. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J Mol Biol* 1983; 168: 367-387

Reverter D, Ventura S, Villegas V, Vendrell J & Aviles FX. Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* 1998a; 273: 3535-3541

Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998b; 273: 32927-32933

Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000; 7: 322-328

Reverter D, Maskos K, Tan F, Skidgel RA & Bode W. Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *J Mol Biol* 2004; 338: 257-269

Reynolds DS, Gurley DS, Stevens RL, Sugarbaker DJ, Austen KF & Serafin WE. Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison of the protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases. *Proc Natl Acad Sci USA* 1989; 86: 9480-9484

Reznik SE & Fricker LD. Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci* 2001; 58: 1790-1804. Review

Rothwarf DM, Li YJ & Scheraga HA. Regeneration of bovine pancreatic ribonuclease A: identification of two nativelike three-disulfide intermediates involved in separate pathways *Biochemistry* 1998; 37: 3760-3766

Ryan CA. Chymotrypsin inhibitor I from potatoes: reactivity with mammalian, plant, bacterial, and fungal proteinases. *Biochemistry* 1966; 5: 1592-1596

Ryan CA. Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores. *Bioessays* 1989; 10: 20-24. Review

Salamanca S, Villegas V, Vendrell J, Li L, Aviles FX & Chang JY. The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* 2002; 277: 17538-17543

Salamanca S. Plegamiento y funcionalidad biológica del inhibidor de metalocarboxipeptidasas LCI (Leech Carboxypeptidase Inhibitor). *Ph D Thesis* 2003a. Universitat Autònoma de Barcelona

Salamanca S, Li L, Vendrell J, Aviles FX & Chang JY. Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 2003b; 42: 6754-6761

Scheraga HA, Konishi Y & Ooi T. Multiple pathways for regenerating ribonuclease A. *Adv Biophys* 1984; 18: 21-41. Review

Scheraga HA, Wedemeyer WJ & Welker E. Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. *Methods Enzymol* 2001; 341: 189-221

Schneider M & Nesheim M. Reversible inhibitors of TAFIa can both promote and inhibit fibrinolysis. *J Thromb Haemost* 2003; 1: 147-154

Shin HC & Scheraga HA. Catalysis of the oxydative folding of bovine pancreatic ribonuclease A by protien disulfide isomerase. *J Mol Biol* 2000; 300: 995-1003

Sidyelyeva G & Fricker LD. Characterization of Drosophila carboxypeptidase D. *J Biol Chem* 2002; 277: 49613-49620

Silveira A, Schatteman K, Goossens F, Moor E, Scharpe S, Stromqvist M, Hendriks D & Hamsten A. Plasma procarboxypeptidase U in men with symptomatic coronary artery disease. *Thromb Haemost* 2000; 84: 364-368

Skidgel RA. Structure and function of mammalian zinc carboxypeptidases. In Hooper NM (ed.) Zinc Metalloproteases in Health and Disease. Taylor and Francis, London. 1996; pp. 241-284

So AK, Varisco PA, Kemkes-Matthes B, Herkenne-Morard C, Chobaz-Peclat V, Gerster JC & Busso N. Arthritis is linked to local and systemic activation of coagulation and fibrinolysis pathways. *J Thromb Haemost* 2003; 1: 2510-2515

Song L & Fricker LD. Purification and characterization of carboxypeptidase D, a novel carboxypeptidase E-like enzyme, from bovine pituitary. *J Biol Chem* 1995; 270: 25007-25013

Song L & Fricker LD. Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. *J Biol Chem* 1997; 272: 10543-10550

Springman EB, Dikov MM & Serafin WE. Mast cell procarboxypeptidase A. Molecular modeling and biochemical characterization of its processing within secretory granules. *J Biol Chem* 1995; 270: 1300-1307

Tan F, Chan SJ, Steiner DF, Schilling JW & Skidgel RA. Molecular cloning and sequencing of the cDNA for human membrane-bound carboxypeptidase M. Comparison with carboxypeptidases A, B, H, and N. *J Biol Chem* 1989; 264: 13165-13170

Tanford C. Contribution of hydrophobic interactions to the stability of globular conformation of proteins. *J Am Chem Soc* 1962; 84: 4240

Thornberry NA & Lazebnik Y. Caspases: enemies within. *Science* 1998; 281: 1312-1316. Review

Torres AM & Kuchel PW. The β-defensin-fold family of polypeptides. *Toxicon* 2004; 44: 581-588. Review

Tsong TY, Baldwin RL & Elson EL. The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc Natl Acad Sci USA* 1971; 68: 2712-2715

Turk V & Bode W. The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* 1991; 285: 213-219. Review

Vallee BL & Neurath H. Carboxypeptidase, a zinc metalloenzyme. *J Biol Chem* 1955; 217: 253-261

van den Berg B, Chung EW, Robinson CV & Dobson CM. Characterisation of the dominant oxidative folding intermediate of hen lysozyme. J Mol Biol 1999; 290: 781-796

van Tilburg NH, Rosendaal FR & Bertina RM. Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. *Blood* 2000; 95: 2855-2859

Vendrell J, Cuchillo CM & Aviles FX. The tryptic activation pathway of monomeric procarboxypeptidase A. *J Biol Chem* 1991; 265: 6949-6953

Vendrell J & Aviles FX. Carboxypeptidases. In Turk V (ed.) Proteases: New Perspectives, Birkyhäuser Verlag Basel/Switzerland. 1999; pp. 1-12

Vendrell J, Querol E & Aviles FX. Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* 2000; 1477: 284-298. Review

Vendrell J, Aviles FX & Fricker LD. Metallocarboxypeptidases. In Handbook of Metalloproteins. Messerschmidt A, Bode W, Cygler M. Wiley & Sons, Ldt, Chichester 2004; pp. 176-189

Venhudova G, Canals F, Querol E & Aviles FX. Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 2001; 276: 11683-11690

Ventura S, Villegas V, Sterner J, Larson J, Vendrell J, Hershberger CL & Aviles FX. Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J Biol Chem* 1999; 28: 19925-19933

Villanueva J, Canals F, Prat S, Ludevid D, Querol E & Aviles FX. Characterization of the wound-induced metallocarboxypeptidase inhibitor from potato. cDNA sequence, induction of gene expression, subcellular immunolocalization and potential roles of the C-terminal propeptide. *FEBS Lett* 1998; 440: 175-182

Villegas V, Martinez JC, Aviles FX & Serrano L. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J Mol Biol* 1998; 283: 1027-1036

Walker JB, Hughes B, James I, Haddock P, Kluft C & Bajzar L. Stabilization versus inhibition of TAFIa by competitive inhibitors in vitro. *J Biol Chem* 2003; 278: 8913-8921

Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-2718

Wedemeyer WJ, Welker E, Narayan M & Scheraga HA. Disulfide bonds and protein folding. *Biochemistry* 2000; 39: 4207-4216. Review

Weissman JS & Kim PS. Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991; 253: 1386-1393

Welker E, Narayan M, Volles MJ & Scheraga HA. Two new structured intermediates in the oxidative folding of RNase A. *FEBS Lett* 1999; 460: 477-479

Welker E, Narayan M, Wedemeyer WJ & Scheraga HA. Structural determinants of oxidative folding in proteins. *Proc Natl Acad Sci USA* 2001; 98: 2312-2316

Wetlaufer DB. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 1973; 70: 697-701

Wu J, Yang Y & Watson JT. Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci* 1998; 7: 1017-1028

Zwanzig R, Szabo A & Bagchi B. Levinthal's paradox. *Proc Natl Acad Sci USA* 1992; 89: 20-22

SUMMARY OF THE PRESENT INVESTIGATION IN CATALAN

Treball 1. Aïllament, clonatge de l'ADNc, expressió recombinant i caracterització d'un inhibidor de carboxipeptidases de la paparra *Rhipicephalus bursa*

S'ha aïllat de la paparra *Rhipicephalus bursa* i seqüenciat N-terminalment un nou inhibidor proteic de metal·locarboxipeptidases anomenat TCI (per "inhibidor de carboxipeptidases de paparra"). Des de l'ARNm d'aquesta paparra s'ha clonat completament l'ADNc que codifica per aquesta proteïna mitjançant RT-PCR i tècniques RACE. L'ADNc del TCI conté un marc de lectura obert que codifica per una proteïna precursora de 97 aminoàcids. Aquesta inclou un pèptid senyal de 22 residus i el TCI madur, una proteïna de 75 aminoàcids amb 12 cisteïnes. La seqüència aminoacídica del TCI, deduïda des de l'ADNc, no presenta homologia amb cap altra proteïna coneguda. Tanmateix, l'extrem C-terminal del TCI s'assembla als extrems C-terminals d'altres inhibidors proteics de metal·locarboxipeptidases, fet que suggereix un mecanisme d'inhibició comú. El TCI que s'ha expressat de forma recombinant en *Escherichia coli* és plenament funcional i inhibeix carboxipeptidases de la família A/B amb constants d'inhibició en el rang nanomolar. Els estudis estructurals mitjançant dicroisme circular i ressonància magnètica nuclear indiquen que el TCI és una proteïna fortament limitada pels seus ponts disulfur, molt estable enfront un ampli rang de pH i altament resistent a condicions desnaturalitzants. El TCI recombinant és un inhibidor que s'uneix fortament a la carboxipeptidasa B plasmàtica, també coneguda com TAFIa, estimulant així la fibrinòlisi *in vitro*, fet que li confereix potencial en aplicacions de prevenció o tractament de desordres trombòtics.

Treball 2. Les estructures tridimensionals de l'inhibidor de carboxipeptidases de paparra acomplexat amb carboxipeptidases A/B revelen un nou mode d'unió doble

L'inhibidor de carboxipeptidases de paparra (TCI) és un inhibidor proteic de metal·locarboxipeptidases present a la paparra *Rhipicephalus bursa*. Les estructures tridimensionals cristal·lines del TCI recombinant unit a la carboxipeptidasa A bovina i a la carboxipeptidasa B humana s'han determinat i refinat a 1.7 i 2.0 Å de resolució, respectivament. El TCI està format per dos dominis que són estructuralment molt semblants a pesar de tenir un baix grau d'homologia seqüencial. Cada domini inclou una hèlix α curta seguida per una petita fulla β antiparal·lela girada i presenta una alta homologia estructural enfront proteïnes de la família de les β-defensines. El TCI s'ancora a la superfície de les carboxipeptidases de mamífer en un mode d'unió doble que no s'havia observat abans per cap altre inhibidor de carboxipeptidases. Així, els últims tres aminoàcids C-terminals interaccionen amb el centre actiu de l'enzim d'una forma que mimetitza la unió del substrat, mentre que el domini N-terminal s'uneix a un lloc d'unió extern diferent al centre actiu. Les estructures d'aquests complexes poden ser importants de cara a les aplicacions del TCI com a droga trombolítica i com a base per al disseny d'inhibidors de carboxipeptidases bivalents.

Treball 3. El lloc d'unió secundari de l'inhibidor de carboxipeptidases de patata. Contribució a la seva estructura, plegament i propietats biològiques

La contribució de cadascun dels aminoàcids del lloc d'unió secundari de l'inhibidor de carboxipeptidases de patata (PCI) de cara a les propietats generals d'aquesta proteïna s'ha examinat utilitzant mutagènesi "alanine-scanning". Els estudis estructurals i enzimàtics fets per una sèrie de mutants del PCI demostren que el posicionament correcte del lloc d'unió primari per una unió i inhibició eficient de la carboxipeptidasa A depèn significativament del lloc d'unió secundari. Els aminoàcids aromàtics d'aquesta regió secundària tenen un paper clau en l'estabilització del complex format entre el PCI i l'enzim, mentre que els aminoàcids polars hi contribueixen poc. S'ha realitzat un estudi comparatiu del plegament oxidatiu d'aquests mutants del PCI utilitzant una aproximació de captura dels ponts disulfur. Els resultats d'aquest estudi, juntament amb la caracterització estructural d'algun d'aquests mutants, clarament indiquen que les forces no covalents dirigeixen el replegament d'aquesta petita proteïna rica en ponts disulfur a l'etapa de "reshuffling" que és el pas limitant del procés de plegament del PCI. A més, revelen que es poden obtenir variants del PCI més estables i alhora amb millor eficiència de plegament mitjançant la introducció de nous contactes intramoleculars no covalents. En conjunt, els resultats derivats d'aquest treball ajuden a entendre els determinants de plegament del lloc d'unió primari i secundari del PCI i la seva contribució en la inhibició d'una carboxipeptidasa, aportant pistes sobre l'evolució del PCI i un millor coneixement d'aquesta proteïna de cara al seu disseny biotecnològic.

Treball 4. Paper dels intermediaris cinètics en el plegament de l'inhibidor de carboxipeptidases de sangonera

En aquest treball s'han caracteritzat els camins de plegament oxidatiu i desplegament reductor de l'inhibidor de carboxipeptidases de sangonera (LCI; quatre ponts disulfur) a través de l'anàlisi estructural i cinètic dels intermediaris de plegament atrapats en medi àcid. El plegament oxidatiu de l'LCI reduït i desnaturalitzat és desenvolupa ràpidament a través d'un flux seqüencial d'espècies amb un, dos, tres i quatre ponts disulfur fins assolir la forma nativa. Dins dels intermediaris de plegament de l'LCI hi trobem dues espècies predominants de tres ponts disulfur (anomenades III-A i III-B) i una població heterogènia d'isòmers "scrambled" (quatre ponts disulfur no natius) que s'acumulen consecutivament al llarg de la reacció de plegament. Aquest estudi revela que les formes III-A i III-B contenen exclusivament ponts disulfur natius i corresponen a espècies estables i parcialment estructurades que s'interconverteixen entre elles assolint l'equilibri abans que la formació dels isòmers "scrambled" tingui lloc. Aquests dos intermediaris actuen com a trampes cinètiques durant el plegament oxidatiu i per tant s'acumulen menys quan se'ls desestabilitza, fet que accelera significativament les cinètiques de plegament de l'LCI. Les formes III-A i III-B semblen tenir els ponts disulfur i els tiols lliures protegits del solvent d'una forma

semblant; es necessiten importants reorganitzacions estructurals a través de la formació d'isòmers "scrambled" per formar LCI natiu. El camí de plegament reductor de l'LCI segueix un mecanisme de "tot-o-res" aparent, tot i que es poden detectar petites quantitats d'intermediaris III-A i III-B, suggerint diferències en la protecció dels diferents ponts disulfur enfront reducció. La caracterització de les formes III-A i III-B mostra que el primer intermediari s'assembla estructuralment i funcionalment a l'LCI natiu, mentre que la forma III-B s'assembla més als isòmers "scrambled".

Treball 5. Caracterització estructural per RMN i prediccions computacionals de l'intermediari més important en el plegament oxidatiu de l'inhibidor de carboxipeptidases de sangonera

L'intermediari III-A conté tres ponts disulfur natius i és el pas limitant més important del plegament oxidatiu de l'inhibidor de carboxipeptidases de sangonera (LCI; quatre ponts disulfur). En aquest treball s'ha purificat l'intermediari III-A directament de la reacció de plegament i se l'ha caracteritzat estructuralment per RMN. Els resultats mostren que aquesta espècie presenta una estructura força nativa, tot i que li manquen alguns elements d'estructura secundària per la qual cosa és més flexible que l'LCI natiu. III-A representa el primer exemple d'un intermediari "disulfide insecure" ("ponts disulfur insegurs") determinat estructuralment. L'oxidació directa d'aquesta espècie cap a la proteïna totalment nativa sembla estar restringida per l'enterrament de les seves dues cisteïnes dins d'una estructura semblant a la nativa. També s'han utilitzat aproximacions teòriques basades en constriccions topològiques que prediuen força bé la presència d'aquest intermediari de plegament. En conjunt, els resultats derivats d'aquest treball suggereixen que les interaccions natives entre segments d'estructura secundària, i no pas la formació dels ponts disulfur, dirigeixen el plegament de l'LCI.

Treball 6. Caracterització d'un anàleg d'un dels intermediaris més importants en el plegament oxidatiu de l'inhibidor de carboxipeptidases de sangonera

El camí de plegament oxidatiu de l'inhibidor de carboxipeptidases de sangonera (LCI; quatre ponts disulfur) es desenvolupa a través de la formació de dos intermediaris majoritaris (III-A i III-B) que contenen tres ponts disulfur natius i actuen com a fortes trampes cinètiques al llarg de la reacció de plegament. A l'intermediari III-B li falta el pont disulfur que està format entre la Cys19 i la Cys43 i que uneix el cor de làmines β amb la hèlix α al LCI. En aquest treball s'ha construït un anàleg d'aquest intermediari substituint la Cys19 i la Cys43 per alanines. El plegament oxidatiu d'aquest anàleg es desenvolupa ràpidament a través d'un flux seqüencial d'espècies amb un, dos i tres ponts disulfur fins arribar a la forma nativa. Aquest procés té lloc d'una forma eficient degut a la baixa acumulació d'intermediaris amb dos ponts disulfur i d'isòmers "scrambled" amb tres ponts. L'estructura tridimensional d'aquest anàleg, sol i acomplexat amb la carboxipeptidasa A,

s'ha determinat mitjançant cristal·lografia de raig X a 2.2 Å de resolució. En conjunt, la seva estructura és molt semblant a la de la forma salvatge de l'LCI, tot i que els aminoàcids de la regió adjacent als llocs de mutació mostren una major flexibilitat. Aquesta manca d'ordre estàtic s'observa sobretot en la forma no acomplexada i es redueix notablement quan l'anàleg s'uneix a la carboxipeptidasa A. L'estructura del complex també demostra que l'anàleg i la forma salvatge de l'LCI s'uneixen a l'enzim de la mateixa manera. Això era d'esperar ja que les dues espècies inhibeixen d'una forma semblant les diferents carboxipeptidases. Experiments de desplegament en equilibri mostren que aquest mutant està desestabilitzat aproximadament 1.5 kcal·mol^{-1} (40%) respecte a la forma salvatge. Aquests resultats indiquen que el quart pont disulfur li dóna estabilitat i especificitat estructural a l'LCI.

Acknowledgements

This thesis has been possible thanks to much help and encouragement from friends and colleagues. I would like to express my sincere gratitude to:

My supervisor, Prof. Francesc X. Avilés for his help and support during the thesis, and for introducing me into the intricate world of scientific research.

My other supervisors, Drs. Salvador Ventura, Julia Lorenzo, Sílvia Bronsoms, Ana Rovira and Gabriela Venhudová who, with enthusiasm, have guided me in the mysteries of laboratory work, for their support and always stimulating discussions.

My friends and colleagues from the IBB and "Pros" for their invaluable help throughout these years and all nice times together that leave me with unforgettable memories.

Profs. Christian Sommerhoff, Tad Holak, Robert Huber and Wolfram Bode, and the people from their groups, for their warm welcome in München and the pleasant work collaboration.

Finally my parents and friends who have been always close to me.