The NMR Structures of the Major Intermediates of the Two-domain Tick Carboxypeptidase Inhibitor Reveal Symmetry in Its Folding and Unfolding Pathways*§

Joan L. Arolas†1,2, David Pantoja-Uceda§1,2, Salvador Ventura§, Francisco J. Blanco§,³ and Francesc X. Aviles†4

From the †Institut de Biotecnologia i Biomedicina and Departament de Bioquimica i Biologia Molecular, Universitat Autonoma de Barcelona, 08193 Bellaterra, Spain, the ‡Department of Espectroscopia y Estructura Molecular, Instituto de Quimica-Fisica Rocasolano, Consejo Superior de Investigaciones Cientificas, Serrano 119, 28006 Madrid, Spain, and the ¶Structural Biology Unit, CIC bioGUNE, Parque Tecnol´ogico de Bizkaia, Edificio 800, 48160 Derio, Spain

There is a lack of experimental structural information about folding intermediates of multidomain proteins. Tick carboxypeptidase inhibitor (TCI) is a small, disulfide-rich protein consisting of two domains that fold and unfold autonomously through the formation of two major intermediates, IIIa and IIIb. Each intermediate contains three native disulfide bonds in one domain and six free cysteines in the other domain. Here we have determined the NMR structures of these two intermediates trapped and isolated at acidic pH in which they are stable and compared their structures with that of the native protein analyzed under the same conditions. Both IIIa and IIIb were found to contain a folded region that corresponds to the N- and C-terminal domains of TCI, respectively, with structures very similar to the corresponding regions of the native protein. The remainder of the polypeptide chains of the intermediates was shown to be unfolded in a random coil conformation. Solvent exchange measurements further indicated that the two protein domains are not completely independent, but affect each other in terms of dynamics and stability, in agreement with reported inhibitory activity data. The derived results provide structural evidence for symmetric TCI folding and unfolding mechanisms that converge in IIIa and IIIb and reveal the structural basis that accounts for the strong and simultaneous accumulation of both intermediates. Altogether, this work has important implications for a better understanding of the folding mechanisms of multidomain, disulfide-rich proteins.

Understanding the sequence of folding events that lead to a biologically active protein from its amino acid sequence is one of the major challenges in structural and molecular biology. The initial theoretical efforts devoted to the study of protein folding have been reinforced in the last several years by the discovery of a wide range of pathological diseases associated with protein misfolding and the increasing pharmacological interest in protein drug design (1, 2). The structural characterization of partially folded intermediates and the analysis of the interactions that stabilize them are of fundamental importance to unveil the folding mechanisms (3, 4). But these studies are hampered by the rapid and cooperative nature of protein folding, and therefore, the short half-life of the intermediates arising during the folding reaction. It is however possible to trap and isolate discrete intermediates from the oxidative folding of disulfide-rich proteins because of the particular chemistry of disulfide bond formation that takes place as an integral part of disulfide folding (5, 6).

A large set of single domain, disulfide-rich proteins has been investigated to date in terms of oxidative folding; among them, outstanding examples such as bovine pancreatic trypsin inhibitor (BPTI), 5 ribonuclease A (RNase A), hirudin, insulin-like growth factor-1 (IGF-1), or conotoxins (7–11). These folding studies have involved the trapping (either by acidification or alkalylation), isolation by liquid chromatography, and final disulfide analysis of the intermediates that accumulate. The extent of heterogeneity of intermediates, together with their number and type (native or non-native) of disulfide bonds has been used as a general premise to classify the folding landscape of these proteins (12, 13). Structural studies of disulfide folding intermediates have been mostly focused on analogs of single domain proteins, such as BPTI, RNase A, or lysozyme, with the cys-

* This work was supported in part by the Ministerio de Educaci´on and Ciencia, Spain, Grants BIO2007-68046 (to F. X. A.) and GEN2003-20642-C09-05 (to F. J. B. and F. X. A.) and by EU-Grant 3D Repertoire, Contract LSHG-CT-2005-512028 (to F. J. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§1,2, David Pantoja-Uceda

†1,2, Joan L. Arolas

‡1,2, David Pantoja-Uceda

§1,2, Joan L. Arolas

¶3, Joan L. Arolas

The atomic coordinates and structure factors (codes 2K2X (N), 2K2Y (IIIa), and 2K2Z (IIIb)) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

The NMR chemical shifts have been deposited in the BioMagResBank under accession numbers 15729 (N), 15730 (IIIa), and 15731 (IIIb).

Both authors contributed equally to this work.

Supported by Juan de la Cierva research contracts awarded by the Ministerio de Educaci´on y Ciencia, Spain.

Supported by Ikerbasque (Basque Foundation for Science, Spain). To whom correspondence may be addressed. Tel.: 34-946572521; Fax: 34-946572502; E-mail: fblanco@cibio.gune.es.

To whom correspondence may be addressed. Tel.: 34-925811315; Fax: 34-925812011; E-mail: francescxavier.aviles@uab.es.

5 The abbreviations used are: BPTI, bovine pancreatic trypsin inhibitor; TCI, tick carboxypeptidase inhibitor; ΔG°, free energy of exchange; DTT, di-thiothreitol; IGF-1, insulin-like growth factor-1; (M)CP, (metallo)carboxypeptidase; N, native protein; R, fully reduced/unfolded protein; r.m.s.d., root mean square deviation; RNase A, ribonuclease A; RP-HPLC, reversed-phase high-performance liquid chromatography; TAFI, thrombin-activatable fibrinolysis inhibitor; TCEP, tris(2-carboxyethyl)phosphine; PDB, Protein Data Bank.
Structures of the Major Folding Intermediates of TCI

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—TCI was expressed and purified as previously reported (27). Briefly, the pBAT-4-OmpA-TCI plasmid, transformed into the Escherichia coli strain BL21(DE3), was grown at 37 °C in M9 minimal medium (0.5% glycerol) supplemented with 0.2% casamino acids and induced at A600 = 1.0 with 1 mM isopropyl-β-d-thiogalactopyranoside (final concentration). The secreted TCI was refolded by overnight incubation of the supernatant with 2 mM cystine and 4 mM cysteine (final concentration) at pH 8.4 to obtain the maximum amount of native protein. The native TCI was initially purified by hydrophobic chromatography in a Sep-Pak C18 cartridge (Waters), followed by cation exchange chromatography in a Sep-Pak Accell Plus CM cartridge (Waters), and reversed-phase high performance liquid chromatography (RP-HPLC) in a 4.6-mm Protein C4 column (Vydac Grace). The protein was finally loaded onto a gel filtration column (Superdex Peptide, GE Healthcare) using 30% acetonitrile as solvent and kept lyophilized. Protein identity and purity were confirmed by mass spectrometry and automated Edman degradation. The concentration of TCI in solution was determined by measuring the absorbance at 280 nm and using a calculated absorption coefficient ε0.1%= 0.97. The native inhibitor was fully active (>95%) as determined by titration with bovine carboxypeptidase A (Sigma), assuming an equimolar interaction between inhibitor and enzyme.

Oxidative Folding and Reductive Unfolding Experiments—Native TCI (N) was reduced and unfolded in 0.1 M Tris-HCl buffer, pH 8.4, containing 6 M guanidine thiocyanate and 100 mM dithiothreitol (DTT) for 2 h at 23 °C. To initiate oxidative folding, the protein was passed through a PD-10 column (Sephadex G-25; GE Healthcare) previously equilibrated with 0.1 M Tris-HCl, pH 8.4, and immediately diluted to a final concentration of 0.5 mg/ml in the same buffer, both in the absence and presence of 0.25 mM 2-mercaptoethanol. The refolding reaction was monitored by removing aliquots of the sample at various time intervals and quenching with the same volume of 4% aqueous trifluoroacetic acid. Acid-trapped intermediates were subsequently analyzed by RP-HPLC using a linear 10–40% gradient of acetonitrile with 0.1% trifluoroacetic acid over 50 min in a 4.6-mm Jupiter C4 column (Phenomenex) at a flow rate of 0.75 ml/min. In the reductive unfolding experiments, native TCI (0.5 mg/ml) was dissolved at 23 °C in either 0.1 M Tris-HCl, pH 8.4 or 0.1 M sodium acetate buffer, pH 4.5, containing different concentrations (5, 20, and 100 mM) of DTT or tris(2-carboxyethyl)phosphine (TCEP), respectively. To monitor the unfolding reaction, time course aliquots of the samples were trapped with 4% trifluoroacetic acid and similarly analyzed by RP-HPLC.

NMR Sample Preparation—Approximately 2.5 mg of IIIa and IIIb intermediates and of fully reduced/unfolded protein (R) were obtained by reducing 15 mg of native TCI with 5 mM DTT, trapping the intermediates after 1 h of unfolding reaction with 4% trifluoroacetic acid, and isolating the species using RP-HPLC as detailed under “Oxidative Folding and Reductive Unfolding
Structures of the Major Folding Intermediates of TCI

Experiments. The NMR samples (1 mM N, IIIa, IIIb, or R) were prepared by dissolving the lyophilized protein powder in 300 µl of H2O:D2O (19:1, ratio by volume), and adjusting the pH to 3.5 with concentrated HCl and NaOH. The TCI samples were analyzed at acidic pH to maintain the free cysteines in a reduced form and prevent their oxidation. For the amide proton exchange experiments, the samples prepared at pH 3.5 were lyophilized and resuspended in 300 µl of 99.98% D2O. Measurements were always conducted using 5-mm susceptibility matched NMR tubes (Shigemi).

NMR Spectroscopy—NMR experiments were performed at 298 K on a Bruker AV 700 spectrometer. Homonuclear two-dimensional TOCSY (mixing time of 80 ms) and NOESY (mixing time of 150 ms) spectra were acquired for both, the determination of the sequence-specific polypeptide backbone chemical shift assignments and structure calculations of the native and intermediate forms of TCI. NMR data were processed with the program TOPSPIN (Bruker Biospin), and the program NMRView (33) was used for interactive spectrum analysis. Chemical shifts were measured relative to 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt used at 50 µM as internal reference for 1H. The stability of the native and IIIa and IIIb intermediates during the time necessary to perform the NMR measurements was checked by comparing one-dimensional spectra recorded before and after the two-dimensional experiments. Within the resolution provided by these spectra, no significant changes in the frequencies of the signals were detected, suggesting that the conformation of the intermediates, and the redox state of their cysteines, had not changed over time. Furthermore, in the series of two-dimensional TOCSY spectra spanning 5 days to measure the exchange rates for each of the two intermediates, no frequency changes or appearance of new signals were observed, again indicating that the intermediates were stable over this time period.

Structure Calculation—The NMR-based constraints used for the structure calculation of TCI and the two intermediates were derived from the homonuclear two-dimensional NOESY spectra. Peak picking of the spectra were carried out manually, and peak volumes were determined using the automatic integration function of NMRView (33). In the spectra of the IIIa and IIIb intermediates, only peaks corresponding to the folded domains were picked and used for the structure calculation. The three-dimensional structures were determined by combined automated NOESY cross-peak assignment (34) and structure calculation with torsion angle dynamics (52) implemented in the program CYANA (53). The standard CYANA protocol of seven iterative cycles of NOE assignment and structure calculation, followed by a final structure calculation, was applied. The consistency between the automated NOE assignment and structure calculation was based on two criteria. First, that the chemical shift assignment should cover close to 90% of the protons, and second, that the backbone root mean square deviation (r.m.s.d.) to the mean for the structure ensemble after the first cycle of structure calculation should be smaller than 3 Å (36). Stereospecific assignments for some isopropyl methyls and methylene groups were determined by the GLOMSA method (54) and before the final structure calculation by analyzing the structures obtained in the preceding seventh NOE assignment/structure calculation cycle. Pseudoatoms with appropriate distance corrections were used for distance restraints involving protons with no stereo-specific assignment (37). A set of upper and lower distance limits was introduced for each pair of cysteine residues involved in disulfide bonds (2.1/2.0 Å for S(i)-S(j) and 3.1/3.0 Å for C(i)-S(j) and S(i)-C(j)). The compatibility of the disulfide bond partner identified in each protein (32) with the structures from the initial rounds of automated calculation was evaluated before introducing these restraints in the subsequent structure calculations. Several intercysteine NOE connectivities confirm the disulfide paring of the intermediates. IIIa: Cys3(Hβ2)-Cys31(Hβ2), Cys10(HoHβ2)-Cys27(HoHβ3), and Cys40(Hβ3)-Cys52(Hβ3,Hβ2), IIIb: Cys40(Hβ2)-Cys74(Hβ2,Hβ3), and Cys84(Hβ2,Hβ3)-Cys101(Hβ2,Hβ3). Only the disulfide bond between Cys47 and Cys64 in intermediate IIIb could not be confirmed by NOEs because of signal overlap. This connectivity was however consistent with the full set of measured NOEs, which produced three-dimensional structures where the two cysteine side chains were at the distance necessary for the formation of the disulfide bond (see below). Weak constraints on (Φ,Ψ) torsion angle pairs and on side chain torsion angles between tetrahedral carbon atoms were used temporarily during the NOE assignment/structure calculation cycles to favor the allowed regions of the Ramachandran plot and staggered rotamer conformations, respectively (35). In each cycle, the structure calculation started from 100 randomized conformers and the standard CYANA-simulated annealing schedule was used with 10,000 torsion angle dynamics steps per conformer. The 20 conformers with the lowest values of the final CYANA target function were subjected to restrained energy minimization in explicit water using the AMBER 9.0 program (38). TIP3P model was used to describe a 9-Å thick cubic box of water molecules surrounding the TCI molecule under periodic boundary conditions. Initially, the protein was subjected to 2000 cycles of restrained energy minimization in vacuum. In a second step a standard protocol of preparation and equilibration of water was used. Finally, the whole system was minimized. The 20 minimized conformers were used to validate the final structure using the program PROCHECK-NMR (39). The program MOLMOL (40) was used to visualize the structures and prepare the figures.

Amide Proton Exchange—The exchange of native TCI amide protons with solvent deuterons was measured at pH 3.5 and 298 K by dissolving the lyophilized sample in 300 µl of 99.98% D2O. NMR data acquisition was started within 15 min of the initiation of the exchange reaction. A series of TOCSY spectra (80 ms mixing time, 2048 complex data points, 512 t1 increments, 8 scans per increment) were collected over the course of twelve days. The acquisition time for each experiment was 1 h and 24 min. Amide exchange experiments for the IIIa and IIIb intermediates were acquired under the same conditions over a time period of 5 days. For the reduced/unfolded form, exchange of the amide protons was complete during the acquisition of the first TOCSY spectrum. All the spectra were processed with NMRPipe (41) using the same processing scheme and parameters. The amide proton exchange rate constants were obtained from the intensity of the HN-Hα cross-peaks corresponding to each residue in the TOCSY spectra. The decay in cross-peak intensities over time was fitted to a single exponential equation.
**RESULTS**

**Oxidative Folding and Reductive Unfolding of TCI**—The oxidative folding of fully reduced and unfolded TCI (R) was carried out in Tris-HCl buffer at pH 8.4 in the absence and presence of 0.25 mM 2-mercaptoethanol. The folding reaction was monitored by acid-trapping of the occurring intermediates and further RP-HPLC analysis as described under "Experimental Procedures." A high number of intermediates populate the initial stages of the oxidative process (up to 5 h), with similar RP-HPLC chromatograms regardless of the absence or presence of reducing agent (Fig. 1A). The addition of 2-mercaptoethanol only influences the final steps of the folding process by promoting disulfide rearrangement of scrambled isomers into the native form (N). This effect increases enormously the efficiency of the reaction with more than 80% of protein recovered as native TCI after 72 h, compared with ~20% obtained in the absence of reducing agent. As previously shown (32), TCI refolds through the sequential formation of 1-, 2-, 3-, 4-, 5-, and 6-disulfide intermediates that finally render two products: the Xa scrambled isomer and the native form. This sequential oxidation of disulfide bonds is thought to take place autonomously for each domain, leading to the predominant accumulation of two major intermediates (IIIa and IIIb), which contain three native disulfide bonds in one domain and six free cysteines in the other one (32). IIIa: [Cys3-Cys31, Cys10-Cys27, and Cys16-Cys32]; IIIb: [Cys40-Cys70, Cys47-Cys64, and Cys84-Cys171].

The reductive unfolding of TCI also seems to proceed autonomously through the formation of the IIIa and IIIb intermediates (Fig. 1B). The reactions performed in Tris-HCl buffer, pH 8.4, using DTT as reducing agent lead to an almost equivalent accumulation of both species without the presence of any other detectable intermediate. The use of the alternative reducing agent TCEP in acetate buffer, pH 4.5, significantly alters the ratio of IIIa and IIIb, diminishing the accumulation of the former intermediate. This is explained by the lower working pH, which causes reductive reactions to prevail over disulfide reshuffling, slowing down and hindering the conversion of IIIb into IIIa. Previous Stop/Go experiments demonstrated the fluctuations of IIIb into IIIa at alkaline pH (32). On the other hand, both intermediates reduce their three native disulfide bonds in a cooperative manner following an "all-or-none" mechanism, with IIIb resisting slightly higher concentrations of reducing agent than IIIa (data not shown). This result is in agreement with the higher accumulation of IIIb versus IIIa in the reductive unfolding conducted at acidic pH.

**NMR Structural Characterization of Native and Intermediate Forms of TCI**—The low field region of the one-dimensional $^1$H NMR spectrum of the native, intermediate, and reduced/
Structures of the Major Folding Intermediates of TCI

unfolded forms of TCI at pH 3.5 is shown in Fig. 2. The fully reduced and unfolded state shows most of the amide signals at 8–8.5 ppm, the typical range of random coil polypeptide chains, indicating no or little remaining structure. The IIa and IIIb intermediates show some spectral features similar to those of the native form plus an increased intensity in the random coil region, indicating the coexistence of both natively folded and unstructured regions. The assignment of the backbone $^1$H resonances of native TCI was performed following the standard sequential assignment strategy using two-dimensional homonuclear proton spectroscopy (45). The assignment of the backbone proton resonances of native TCI is complete, except for Asn whose amine protons are not observed. Further analysis of the TOCSY and NOESY spectra permitted a nearly complete $^1$H side chain assignment. Overall, the assignment covers 90.2% of the chemical shifts of non-labile protons. Two distinct sets of cross-peaks were observed in the two-dimensional spectra of the intermediates. A well resolved set of signals could be unambiguously assigned to one of the two domains of TCI: the resolved signals in the spectrum of IIa to the N-terminal domain and those in the IIIb spectrum to the C-terminal domain. The remaining signals, with chemical shifts close to random coil values and generally sharper than the others, were not resolved enough to achieve an unambiguous sequential assignment. Nevertheless, the identification of the spin system was possible for all the peaks. In intermediate IIIa we found 12(CHWDNS), 5G, 6(EQ), 6(KR), 2A, 3L, and 1V, and for intermediate IIIb we identified 16(CFYDNS), 4G, 4(EQ), 3(KR), 1A, 3L, 2V, and 2T. These results are consistent with the amino acid composition of the unassigned polypeptide chain and confirm the assignment of the folded region of the two intermediates. The overall assignment of the non-labile protons covers the 93.8 and 91.9% of the folded domains of IIa and IIIb, respectively. For all Xxx-Pro bonds present in the native and intermediate forms of TCI, a trans conformation was confirmed by intense Xxx(H$_\alpha$)-Pro(H$_\alpha$) sequential NOEs (45).

The chemical shifts of the assigned residues in the folded regions of the folding intermediates are very similar to the corresponding values in the spectrum of native TCI, suggesting that they have similar structures. This is confirmed by the observation that the side chain proton H$_\alpha$ of the Trp$^{25}$ has the same resonance frequency in N and IIIb, while it is closer to its random coil value in R and IIa (Fig. 2). Also, the signal of the H$_\alpha$ of Tyr$^{24}$ has a highly anomalous chemical shift both in native TCI and IIIa but not in R or IIIb, indicating a similarly folded structure around this residue in the first pair of species but not in the second one. The plot of the differences in the H$_\alpha$ chemical shifts of native TCI at pH 3.5 and pH 5.5 (supplemental Fig. S1A) demonstrates that the structure of this protein is almost identical at the two pH values, with only a few residues showing slight differences due to the proximity of side chains with acid groups that titrate in this pH range. The same figure showing the differences between the folded parts of IIa and IIIb and the equivalent regions of N at pH 3.5 demonstrates that the structures of the folded domains in the intermediates are very similar to those in the native conformation (supplemental Fig. S1B). Tertiary NOEs from different regions of the two-dimensional NOESY spectra, diagnostic of the three-dimensional structure of the TCI domains, further confirm this assessment. Cross-peaks corresponding to NOEs between residues found far away in the sequence can be identified both in the intermediate and native forms of TCI (supplemental Fig. S2).

Solution Structure of Native and Intermediate Forms of TCI—The reliable identification of the signals from the folded and unfolded regions in the intermediates, together with the large dispersion of resonances corresponding to the folded domains in the intermediates and in native TCI allowed automatic NOE assignments and structure calculations for the folded parts of the three species (see “Experimental Procedures”). These calculations were solely done on the basis of the homonuclear NOESY spectra recorded at pH 3.5. Of the total NOESY cross-peaks, 87, 83, and 81% could be assigned by the program CYANA for native TCI, IIIa and IIIb, respectively. Statistics about the quality and precision of the 20 best NMR conformers that represent the solution structures of native TCI and the folded domains of the intermediates are summarized in Table 1. The ensembles of NMR structures are well defined and are highly consistent with the experimental data, with no dis-
tance constraint violations larger than 0.35 Å. The precision of the structures is characterized by low r.m.s.d. values to the mean coordinates for the backbone and for all heavy atoms, excluding the unstructured regions at the polypeptide chain termini (residues 1 and 74–75) and at the interdomain linker (residues 37–38). The quality of the structures is also reflected by a high percentage of (H) backbone torsion angle pairs being found in the most favorable or additionally allowed regions of the Ramachandran plot.

The solution structure of native TCI at pH 3.5 (Fig. 3A) is essentially the same as that determined recently at pH 5.5 by manual NOE assignment (PDB entry 2JTO; (31)), within the precision that could be achieved with the assigned NOEs (supplemental Table S1). Also, the structures of the N-terminal domain of IIIa and the C-terminal domain of IIIb are very similar to the equivalent regions of native TCI, although the precision is not as high in these two cases due to the smaller number of assigned NOEs (Fig. 3B).

As already described, the structure of TCI consists of a one-turn α-helix (α1N or α1C) and a central triple-stranded antiparallel β-sheet (β1N, β2N and β3N or β1C, β2C and β3C) in each of the two domains. The C-terminal domain contains an additional two-turn α-helix between the last two β-strands (α2C).

**Amide Proton Exchange of Native and Intermediate Forms of TCI**—Fig. 4 shows the first TOCSY spectra recorded after the exchange reaction was initiated for native TCI and the two intermediates. 41, 23, and 13 NHs were observed in the spectrum of native TCI, IIIa, and IIIb, respectively, most of them located in the β-sheet that forms the central protein core of the two globular domains. Only residues from folded domains were observed for the intermediates, with those in the unfolded region exchanging too fast to be detected. The exchange rates, \( k_{ex} \), could be measured for 33, 16, and 9 residues of N, IIIa, and IIIb, respectively. These values could not be determined for the other residues because of different reasons: some of them exchanged so fast that could not be observed in the first TOCSY spectrum and for others the decay of the signals was too rapid, and the curve could not be fitted to yield a reliable estimate of the exchange rate. The exchange rate constants for the three TCI forms are given in supplemental Tables S2, S3, and S4. Native TCI shows essentially the same relative pattern of...
exchange along the sequence at pH 3.5 and 5.5 (31), although the exchange rate could be measured for three more residues at pH 3.5 because of the slower rate of exchange occurring at this pH.

Backbone amide proton exchange with the solvent provides a window for observing the dynamics of the main chain, from the closed incompetent state for exchange to the open competent state (46). Compared with the data at pH 5.5, there is a general decrease in the exchange rates measured at pH 3.5 for native TCI. This is, however, due to the intrinsic retardation of the exchange at this more acidic pH. When the exchange rates are normalized by the intrinsic values ($k_{\text{int}}$), there is a general increase in the exchange rates at pH 3.5, suggesting an increased backbone dynamics in the time scale at which the exchange with the solvent deuterons takes place. The same effect is qualitatively observed when comparing the folded domains of the intermediates with the corresponding regions of the native protein at the same pH, indicating that the presence of unfolded chains increases the backbone dynamics of the folded regions.

Under our working conditions (EX2 exchange regime, see “Experimental Procedures”), the exchange rates can be used to calculate the equilibrium constant between the open and closed forms for each residue of the protein, and the corresponding free energy of exchange ($\Delta G_{\text{ex}}$), which in this case is equivalent to the free energy of the opening reaction. The calculated $\Delta G_{\text{ex}}$ values for native TCI and for the folded domains of the intermediates are represented along the TCI sequence in Fig. 5. Assuming that the free energy of exchange of the slowest exchanging amide proton corresponds to the free energy of the global unfolding event, native TCI is about 2.2 kcal/mol less stable at pH 3.5 than at pH 5.5 (see also supplemental Table S2 and Ref. 31). In addition, the amide proton exchange data does not reveal any significant difference in stability between the N- and C-terminal domains of the native protein.

**DISCUSSION**

Protein domains are evolutionary independent units that may constitute a single domain polypeptide on their own or recombine with others to form part of a multidomain protein. Although ~75% of eukaryotic proteins contain more than one domain (47), there is limited information about the folding mechanisms of individual domains from multidomain proteins. The folding/unfolding transitions of two-domain proteins can be described by three different models: (i) the two-state model, with the two domains either folded or unfolded; (ii) the autonomous folding/unfolding of the two domains; and (iii) the folding/unfolding of interacting domains exhibiting mutual effects on their stability and/or kinetic behavior. In the first case no intermediate state exists between the fully unfolded and completely folded protein, whereas in the two latter cases one or more partially folded intermediates populate the folding reaction. In most cases, the presence of these intermediates can be inferred only by monitoring deviations from two-state behaviors at equilibrium or in time-resolved folding/unfolding experiments using indirect structural probes such as intrinsic fluorescence or ellipticity. Deciphering the conformation of each domain in the intermediates arising along the folding pathway would clearly provide a more direct assessment of their folding mechanism. Therefore, studies of multidomain proteins stabilized by disulfide bonds might provide considerable advances on this field.

The particular two-domain architecture and disulfide cross-linking of TCI represents an attractive system for structural and folding studies. We have recently reported the high-resolution structure of free TCI in solution at pH 5.5 (31). The ensemble of
NMR structures show two globular and compact domains separated by a two-residue linker. The conformation of this short segment is defined only by local constraints and displays an increased flexibility as compared with the two protein domains. In addition, we have previously characterized the oxidative folding and reductive unfolding of TCI and its dissected domains (32). The comparative folding analysis of the whole protein and its individual domains suggested that the folding pathway of this two-domain protein results from the additive contribution of their individual N- and C-terminal domains, which appear to fold autonomously. The folding and unfolding routes converge in the formation of two major intermediates (IIIa and IIIb) that contain, respectively, the three native disulfide bonds of the N- and C-terminal domains of TCI (see Fig. 6). In the present work, we have determined the solution structures of these two intermediates at pH 3.5 by NMR spectroscopy methods and made a comparison with that of the native TCI analyzed at the same acidic conditions. The results reveal a native-like structure for the N- and C-terminal domains of IIIa and IIIb, respectively, as well as a random coil conformation for the accompanying halves. The structures here solved further support the autonomous folding and unfolding ability of both TCI domains. It is common that when two domains fold/unfold autonomously in multidomain proteins, only one of them, either the most stable or the one with faster folding kinetics, but always the same, is able to keep the native structure in the intermediates (48). As far as we know, symmetric situations like those observed for TCI intermediates have not been described before, which broadens the possible folding scenarios in multidomain proteins.

Despite their high level of structural similarity (r.m.s.d. of 1.8 Å for backbone atoms) and identical disulfide bond pattern, the two domains of TCI fold at very different rates and through distinct mechanisms (32). This behavior is reminiscent of the cases of BPTI and tick anticoagulant peptide, two small, single domain proteins with similar structures and disulfide bonds but with completely different folding pathways (49). The folding of the C-terminal domain of TCI involves the formation of a smaller amount of intermediates and is much faster than that of the N-terminal one (32). Instead, the N-terminal
Structures of the Major Folding Intermediates of TCI

FIGURE 6. Schematic representation of the oxidative folding and reductive unfolding of TCI. The pathways of oxidative folding and reductive unfolding are shown by solid and dashed arrows, respectively. R and N indicate the fully reduced/unfolded and native forms of TCI. XS are ensembles of molecules with X number of disulfide bonds. IIIa and IIIb are two major intermediates with three native disulfide bonds at the N- and C-terminal domain, respectively. Xa is a predominant six-disulfide scrambled isomer accumulating at the end of the folding reaction. Three-dimensional models with the corresponding ribbon structures and disulfide bond pairings are represented for R, IIIa, IIIb, and N.

half displays a significantly higher number of solvent exchange protected residues, especially at the β1 and β2 strands and the turn in between. Also, in general, the residues at equivalent spatial positions of the N-terminal domain are more hydrophobic than those in the C-terminal region (e.g. Phe/Gly, Gly/Glu, Leu/Asn, Ala/Lys, and Tyr/Glu), suggesting that the formation of a hydrophobic core is favored in the former domain. The contact order, defined as the normalized average sequence separation between interacting residues in the folded state (50), is a measure of local versus long-range interactions in the native-state structure. This parameter is small for proteins stabilized mainly by local interactions but large when residues in a protein interact frequently with partners far away in the sequence. The contact order of the C-terminal domain is 22% while that of the N-terminal one is 19%, indicating a slightly higher extent of local contacts in the native state of the latter. Although this concept was developed for two-state folding proteins, it still could be useful to compare the formation of the first folding nucleus in multistate folding proteins. In the case of TCI, it would suggest that the N-terminal domain could adopt a compact structure earlier than the C-terminal one.

How can we correlate this information with the folding of TCI? For the C-terminal domain, the intermediates that accumulate during the initial stages of the folding performed in the absence of redox agents correspond to a reduced number of three-disulfide intermediates, whose composition does not evolve significantly over time (32). The addition of a redox compound promoting disulfide reshuffling causes a strong acceleration of the folding process. Loosely packed conformations with high backbone dynamics would favor this kind of folding reaction. By contrast, the individual N-terminal domain folds presence of unfolded regions in IIIa and IIIb increases the backbone dynamics of their folded regions. This indicates that although the respective folding pathways are controlled by the primary sequence of each domain, the two moieties are not completely independent of each other in the final folded structure, providing TCI with additional conformational freedom relative to that displayed by the individual domains. This finding is consistent with the stability and inhibitory properties of the intermediates, for which the unfolded domain has a destabilizing effect on the folded part and significantly contributes to the inhibition of metalloproteases (32). The higher dynamics of the C-terminal domain with respect to the N-terminal one suggests that the motional mode could be important for the functionality of the inhibitor because a rapid sampling in the conformational space to select the most suitable conformation would enhance the affinity for the target protease, especially in the first binding steps. Accordingly, the structure of the C-terminal-half suffers the strongest conformational changes upon inserting the C terminus into the active site groove of human CPB (31).

On the other hand, it seems clear that both domains can attain their native structure autonomously because no stable intermediates with disulfide bonds cross-linking the two TCI halves are observed. A recent study analyzing the relationship between packing density as well as interface structure and interdependence in the folding of domains from multidomain proteins suggests that the domains displaying a flexible linker and a small interface tend to fold autonomously, whereas the folding of domains sharing a large interface is coupled to some extent (48). In agreement with this report, the autonomous assembly of TCI would be facilitated by the presence of a short but flexible linker connecting the two globular domains and the

more slowly through the accumulation of 1-, 2-, and 3-disulfide intermediates that finally render two products: the native domain and a predominant Xa-like scrambled isomer (32). The burial of particular disulfides into a more compact initial structure with reduced conformational dynamics and disfavored disulfide exchange might promote this effect. As discussed above, the faster compaction of the N-terminal domain could promote a faster and preferential formation of native disulfides in a fraction of the polypeptide population at the beginning of the folding reaction, thus explaining why IIIa and IIIb coexist even if the folding of the C-terminal domain is much faster and more efficient.

Based on the exchange data, the backbone dynamics of the individual TCI domains is higher in the intermediates than in the native protein at the time scale at which exchange takes place. Therefore, the
structures of the major folding intermediates of TCI

4. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M., and Fersht, A. R. (2005) Nature 437, 1053–1056
5. Creighton, T. E. (1997) Biol. Chem. 378, 731–744
6. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000) Biochemistry 39, 4207–4216
7. Bulaj, G., and Olivera, B. M. (2008) Antioxid. Redox. Signal 10, 141–156
8. Chang, J. Y., Schindler, P., and Chatrenet, B. (1995) J. Biol. Chem. 270, 11992–11997
9. Darby, N. J., Morin, P. E., Talbo, G., and Creighton, T. E. (1995) J. Mol. Biol. 249, 463–477
10. Guo, Z. Y., Qiao, Z. S., and Feng, Y. M. (2008) Antioxid. Redox Signal 10, 127–140
11. Welker, E., Narayan, M., Wedemeyer, W. J., and Scheraga, H. A. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 2312–2316
12. Arolas, J. L., Aviles, F. X., Chang, J. Y., and Ventura, S. (2006) Trends Biochem. Sci. 31, 292–301
13. Chang, J. Y. (2004) Biochemistry 43, 4522–4529
14. Kortemme, T., Hollecker, M., Kemmink, J., and Creighton, T. E. (1996) J. Mol. Biol. 257, 188–198
15. Laity, J. H., Lester, C. C., Shimotakahara, S., Zimmerman, D. E., Montelione, G. T., and Scheraga, H. A. (1997) Biochemistry 36, 12683–12699
16. Noda, Y., Yokota, A., Horii, D., Tominatinga, T., Tanisaka, Y., Tachibana, H., and Segawa, S. (2002) Biochemistry 41, 2130–2139
17. Shimotakahara, S., Ross, C. B., Laity, J. H., Zimmerman, D. E., Scheraga, H. A., and Montelione, G. T. (1997) Biochemistry 36, 6915–6929
18. van Mierlo, C. P., Kemmink, J., Neuhaus, D., Darby, N. J., and Creighton, T. E. (1994) J. Mol. Biol. 235, 1044–1061
19. Yokota, A., Hirai, K., Miyachi, H., Imura, S., Noda, Y., Inoue, K., Akasaka, K., Tachibana, H., and Segawa, S. (2004) Biochemistry 43, 6663–6669
20. Arolas, J. L., D’Silva, L., Popowicz, G. M., Aviles, F. X., Holak, T. A., and Ventura, S. (2005) Structure 13, 1193–1202
21. Cemazar, M., Joshi, A., Daly, N. L., Mark, A. E., and Craik, D. J. (2008) Structure 16, 842–851
22. Cemazar, M. A., Daly, N. L., Haggblad, S., Lo, K. P., Yulaningsih, E., and Craik, D. J. (2006) J. Biol. Chem. 281, 8224–8232
23. Daly, N. L., Clark, R. J., and Craik, D. J. (2003) J. Biol. Chem. 278, 6314–6322
24. Sato, A., Koyama, S., Yamada, H., Suzuki, S., Tamura, K., Kobayashi, M., Niwa, M., Yasuda, T., Kyogoku, Y., and Kobayashi, Y. (2000) J. Pept. Res. 56, 218–230
25. van den Berg, B., Chung, E. W., Robinson, C. V., and Dobson, C. M. (1999) J. Mol. Biol. 290, 781–796
26. Weissman, J. S., and Kim, P. S. (1991) Science 253, 1386–1393
27. Arolas, J. L., Lorenzo, J., Royva, A., Castella, J., Aviles, F. X., and Sommerhoff, C. P. (2005) J. Biol. Chem. 280, 3441–3448
28. Arolas, J. L., Vendrell, J., Aviles, F. X., and Fricker, L. D. (2007) Curr. Pharm. Des. 13, 349–366
29. Arolas, J. L., Popowicz, G. M., Lorenzo, J., Sommerhoff, C. P., Huber, R., Aviles, F. X., and Holak, T. A. (2005) J. Mol. Biol. 350, 489–498
30. Sanglas, S., Valnickova, Z., Arolas, J. L., Pallarés, I., Guevara, T., Solà, M., Kristensen, T., Enghlund, J. J., Aviles, F. X., and Gomis-Rüth, F. X. (2008) Mol. Cell. 31, 598–606
31. Pantoja-Uceda, D., Arolas, J. L., García, P., López-Hernández, E., Padró, D., Aviles, F. X., and Blanco, F. J. (2008) Biochemistry 47, 7066–7078
32. Arolas, J. L., Bronsmans, S., Ventura, S., Aviles, F. X., and Calvete, J. J. (2006) J. Biol. Chem. 281, 22906–22916
33. Johnson, B. A., and Blevins, R. A. (1994) J. Biomol. NMR 4, 603–614
34. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) J. Mol. Biol. 319, 209–227
35. Pantoja-Uceda, D., Lopez-Mendez, B., Koshiba, S., Inoue, M., Kigawa, T., Terada, T., Shirouzu, M., Tanaka, A., Seki, M., Shinozaki, K., Yokoyama, S., and Guntert, P. (2005) Protein Sci. 14, 224–230
36. Jee, J., and Guntert, P. (2003) J. Struct. Funct. Genomics 4, 179–189
37. Wüthrich, K., Billette, M., and Braun, W. (1983) J. Mol. Biol. 169, 949–961
38. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) J. Comput. Chem. 26, 1668–1688

Acknowledgments—We thank Prof. Robert Huber and Dr. Josep Vendrell for helpful discussions and critical reading of the manuscript. We also thank CNIO (Madrid, Spain) for NMR time in the 700-MHz spectrometer and computer time in the SGI® Altix® 4700 blade-based supercomputer.

REFERENCES
1. Cheng, Y., LeGall, T., Oldfield, C. J., Mueller, J. P., Van, Y. Y., Romero, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2006) Trends Biotechnol. 24, 435–442
2. Ross, C. A., and Poirier, M. A. (2004) Nat. Med. 10, (suppl.), S10–S17
3. Creighton, T. E., Darby, N. J., and Kemmink, J. (1996) Faseb. J. 10, 110–118
Structures of the Major Folding Intermediates of TCI

39. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) J. Biomol. NMR. 8, 477–486
40. Koradi, R., Billeter, M., and Wuthrich, K. (1996) J. Mol. Graph 14, 51–55, 29–32
41. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) J. Biomol. NMR. 6, 277–293
42. Johnson, B. A. (2004) Methods Mol. Biol. 278, 313–352
43. Linderstrøm-Lang, K. U. (1955) Chem. Soc. Spec. Publ. 2, 1–20
44. Hvidt, A., and Nielsen, S. O. (1966) Adv. Protein Chem. 21, 287–386
45. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, pp. 117–161, Wiley-Interscience, New York
46. Jardetzky, O., and Schmitt, T. H. (1996) Chemtracts Biochem. Mol. Biol. 6, 1–17
47. Batey, S., and Clarke, J. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 18113–18118
48. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007) Nat. Rev. Mol. Cell Biol. 8, 319–330
49. Chang, J. Y. (1996) Biochemistry 35, 11702–11709
50. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) J. Mol. Biol. 277, 985–994
51. Mukherjee, S., Mohan, P. M., Kuchroo, K., and Chary, K. V. (2007) Biochemistry 46, 9911–9919
52. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) J. Mol. Biol. 273, 283–298
53. Guntert, P. (2004) Methods Mol. Biol. 278, 353–378
54. Guntert, P., Braun, W., and Wuthrich, (1991) J. Mol. Biol. 217, 517–530