PROTEINS: Structure, Function, and Genetics 40:482–493 (2000)

# Structures of Scrambled Disulfide Forms of the Potato Carboxypeptidase Inhibitor Predicted by Molecular Dynamics Simulations With Constraints

**Marc A. Martí-Renom,[1,2] Roland H. Stote,[3*] Enrique Querol,[1,2] Francesc X. Aviles,[1,2] and Martin Karplus[3,4*]**
[1]*Institut de Biologia Fonamental, Universitat Autunoma de Barcelona, Barcelona, Spain*
[2]*Departament de Bioquímica, Universitat Autunoma de Barcelona, Barcelona, Spain*
[3]*Laboratoire de Chimie Biophysique, Université Louis Pasteur-ISIS, Strasbourg, France*
[4]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts*

***ABSTRACT*** **The structures of two species of potato carboxypeptidase inhibitor with nonnative disulfide bonds were determined by molecular dynamics simulations in explicit solvent using disulfide bond constraints that have been shown to work for the native species. Ten structures were determined; five for scrambled A (disulfide bonds between Cys8–Cys27, Cys12–Cys18, and Cys24–Cys34) and five for the scrambled C (disulfide bonds Cys8–Cys24, Cys12–Cys18, and Cys27–Cys34). The two scrambled species were both more solvent exposed than the native structure; the scrambled C species was more solvent exposed and less compact than the scrambled A species. Analysis of the loop regions indicates that certain loops in scrambled C are more nativelike than in scrambled A. These factors, combined with the fact that scrambled C has one native disulfide bond, may contribute to the observed faster conversion to the native structure from scrambled C than from scrambled A. Results from the PROCHECK program using the standard parameter database and a database specially constructed for small, disulfide-rich proteins indicate that the 10 scrambled structures have correct stereochemistry. Further, the results show that a characteristic feature of small, disulfide-rich proteins is that they score poorly using the standard PROCHECK parameter database. Proteins 2000;40:482–493.**


**Key words: potato carboxypeptidase inhibitor; thermal unfolding; scrambled species; molecular dynamics; protein folding**

## INTRODUCTION

The *prediction* of protein structure from the sequence is currently a subject of intense interest.[1,2] Because there are many proteins where the disulfide pairing is known prior to a structure determination, a possible approach is to use this information in a molecular dynamics simulation. In a previous study,[3] it was shown that refolding PCI by molecular dynamics simulations which incorporate information pertaining to the disulfide pairing resulted in structures that were within an rmsd of 1.8–2.9 Å of the native back-bone structure.

Scrambled species of cystine-containing proteins, e.g., species that contain at least two nonnative disulfide bonds, have been observed during folding of a number of proteins, including ribonuclease,[4–6] hirudin,[7] PCI,[8] and cardiotoxin analogue III.[9] PCI is a 39 amino-acid protein that contains three disulfide bonds (see Fig. 1) and plays a role in the defense mechanism of plants.[10] In addition, by a topological analysis where disulfide bonding patterns were superimposed,[11] as well as by binding and cell-proliferation experiments,[12] it was shown that PCI is a structure-based antagonist of EGF and other related growth factors. The experimental studies of disulfide folding pathways of PCI[8] have shown that the reduced and denatured protein refolds spontaneously in vitro. Stop/go folding experiments of acid-trapped intermediates and structural analysis with MALDI mass spectrometry showed that the disulfide pathway has three stages: an initial stage of nonspecific disulfide formation (packing), a second stage with disulfide reshuffling of packed intermediates (consolidation), and a third stage which refines and consolidates the scrambled species to the native conformation. Significant populations of five trapped intermediates have been found.[8] The disulfide pairing of two of these species has been elucidated (S. Pavia, personal communication). Species A does not contain any nativelike pairing; the native pairing is Cys8–Cys24, Cys12–Cys27, and Cys18–Cys34 and species A has disulfide bonds Cys8—Cys27, Cys12–Cys18, and Cys24–Cys34. Species C contains one native disulfide bond (disulfide bonds Cys8–Cys24, Cys12–Cys18, and Cys27–Cys34). Structural characterization of the scrambled species can help in elucidating the pathway(s)

*Abbreviations*: PCI, potato carboxypeptidase inhibitor; SBMD, stochastic boundary molecular dynamics; SD, steepest descent; ps, picoseconds; fs, femptoseconds; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SASA, solvent-accessible surface area; EEF1, effective energy function.

Grant sponsor: Centre de Referëncia en Biotecnologia de la Generalitat de Catalunya; Grant sponsor: CNRS (France); Grant sponsor: Ministerio de Educaciûn y Ciencia; Grant numbers: CICYT-BIO98-0362 and BIO97-0511; Grant sponsor: C4-CESCA.

*Correspondence to: Roland H. Stote or Martin Karplus, Laboratoire de Chimie Biophysique, ISIS, CNRS-ESA7006, Université Louis Pasteur, 67000 Strasbourg, France. E-mail: rstote@chimie.u-strasbg.fr or marci@brel.u-strasbg.fr

Received 2 December 1999; Accepted 13 March 2000

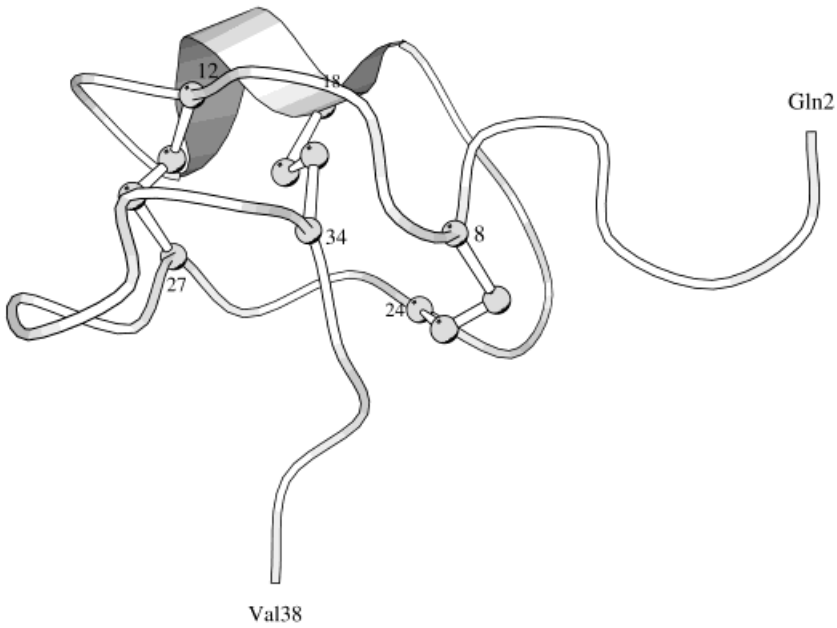


Fig. 1. A ribbon representation of the PCI crystal structure (pdb4cpa.ent)[14] highlighting the 3,10 helix and the disulfide bonds between the cysteine residues. Protein regions: loop1 from Cys 8 to Cys12, 3,10 helix from Cys12 to Cys18, loop2 from Cys18 to Cys24, and loop3 from Cys27 to Cys34. Figure generated using the MOLSCRIPT program.[41]

of PCI folding, as well as in determining the role of disulfide bonds in stabilization. Such knowledge can aid in the design and stabilization of related proteins. Since many of these proteins have important biological functions, (for example, hormones, growth factors, neuropeptides, venoms), the understanding is also of biotechnological interest. Determining the structural properties of scrambled species by experiment can be difficult. Recently, the three-dimensional structure of the native isomer and two scrambled isomers of the α-Conotoxin protein[13] have been determined. However, for the PCI species, the studies have been hindered by lack of a well-resolved NMR spectrum, although the work is still in progress. With this perspective, computer simulations have an important role in the characterization of the structural properties of scrambled species.

## MATERIALS AND METHODS

The procedure used to generate a partially unfolded PCI has been described in detail elsewhere[3]; only a brief summary is given here. The starting structure of PCI was taken from the crystal structure of PCI complexed with carboxypeptidase A.[14] A sphere of TIP3P[15,16] water with a radius of 37 Å was used to solvate the protein. High-temperature stochastic boundary molecular dynamics simulations[17–19] were used to denature PCI after the disulfide bonds had been reduced. The CHARMM program was used for all calculations.[20] After the partial unfolding of PCI, the simulation was allowed to continue at high temperature. During the high-temperature simulation, the distances between the Cα atoms and the sulfur atoms of the Cys residues were calculated; when the distances between two of the three nonnative pairs for the scrambled A species or the scrambled C species were less than 10 Å the structure was used as a starting structure for refolding. Close Cα -Cα and S-S pairing corresponding to the scrambled A species was found at 343ps, 456ps, 550ps, 620ps, and 725ps, and for the scrambled C species at 343ps, 355ps, 550ps, 620ps, and 725ps (see Table I). The protocol for the folding simulations is the same as that described in Martí-Renom et al.;[3] the only difference is that the S-S restraints were placed between the nonnative disulfide pairs rather than the native pairs as in the previous work.

NOE-type restraints[20] (Brooks BR, unpublished observations) were introduced between specific pairs of Cys sulfur atoms with an equilibrium distance equal to the current distance between the two cysteine sulfur atoms. Restrained molecular dynamics were run at 600K, in intervals of 10ps. Every 10ps, the restraint distances were reduced, thus slowly pulling the Cys groups together. When the Sγ-Sγ atom distances were less than or equal to 2.85 Å, disulfide bonds were introduced (Table II). After equilibrating for 25ps at 600K, the systems were cooled from 600K to 300K over 10ps, and the simulations were continued for 200ps at 300K. In all of the simulations, a time of step of 2fs was used and the temperature was controlled by a Langevin bath[19] applied to the water oxygen atoms that were in the outer 3 Å stochastic boundary. Ten folding simulations were then done using the selected structures as starting points. These simulations are referred to as scA343, scA456, scA550, scA620, and scA725 for the A species and scC343, scC355, scC550, scC620, and scC725 for the C species (Table I). The simulations are designated by the letters scA or scC corresponding to the scrambled A or C species, respectively, and by a number corresponding to the time of the starting structure. The present approach is similar in spirit to the Targeted Molecular Dynamics method[22] and the Biased Molecular Dynamics method;[23] the main difference is that the current method

**TABLE I. Simulations Performed**

| | Started form | Temperature | Disulfide | Total time |
|---|---|---|---|---|
| Control at 300K (N300)[a] | Crystal Structure | 300K | Yes | 500ps |
| No disulfide bonds at 600K (ND600)[a] | Heating from 210ps structure of N300 | 600K | No | 910ps |
| Scrambled A at 300K (scA343) | Cooling from 343ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled A at 300K (scA456) | Cooling from 456ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled A at 300K (scA550) | Cooling from 550ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled A at 300K (scA620) | Cooling from 620ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled A at 300K (scA725) | Cooling from 725ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled C at 300K (scC343) | Cooling from 343ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled C at 300K (scC355) | Cooling from 355ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled C at 300K (scC550) | Cooling from 550ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled C at 300K (scC620) | Cooling from 620ps structure of ND600 | 300K | Yes | 200ps |
| Scrambled C at 300K (scC725) | Cooling from 725ps structure of ND600 | 300K | Yes | 200ps |

[a]Simulations reported in Martí-Renom et al.[3]

**TABLE II. Distances Between Cα and Sulfur Atoms.[†] All Distances are in Å**

| | | Before NOE applied | | After NOE applied | |
|---|---|---|---|---|---|
| Simulation | Disulfide Bridges | Cα atom | Sulfur atom | Cα atom | Sulfur atom |
| scA343 | Cys8--Cys27 | 10.1 | 13.5 | 7.7 | 2.8 |
| | Cys12--Cys18 | 4.3 | 5.5 | 5.9 | 2.8 |
| | Cys24--Cys34 | 5.7 | 4.3 | 8.1 | 2.9 |
| scA456 | Cys8--Cys27 | 12.2 | 12.0 | 7.8 | 2.8 |
| | Cys12--Cys18 | 4.6 | 4.1 | 6.1 | 2.8 |
| | Cys24--Cys34 | 5.9 | 3.8 | 7.2 | 2.9 |
| scA550 | Cys8--Cys27 | 16.2 | 17.6 | 8.3 | 2.9 |
| | Cys12--Cys18 | 6.5 | 4.7 | 6.5 | 2.8 |
| | Cys24--Cys34 | 9.9 | 8.9 | 7.4 | 2.8 |
| scA620 | Cys8--Cys27 | 14.1 | 17.0 | 8.4 | 2.8 |
| | Cys12--Cys18 | 5.8 | 9.6 | 5.2 | 2.9 |
| | Cys24--Cys34 | 11.4 | 9.6 | 6.3 | 2.8 |
| scA725 | Cys8--Cys27 | 14.5 | 14.0 | 8.2 | 2.8 |
| | Cys12--Cys18 | 7.3 | 7.4 | 5.6 | 2.8 |
| | Cys24--Cys34 | 7.3 | 4.0 | 5.7 | 2.8 |
| scC343 | Cys8--Cys24 | 6.9 | 5.1 | 6.0 | 2.8 |
| | Cys12--Cys18 | 4.3 | 5.5 | 5.4 | 2.8 |
| | Cys27--Cys34 | 4.0 | 7.2 | 3.9 | 2.8 |
| scC355 | Cys8--Cys24 | 10.0 | 9.4 | 8.2 | 2.8 |
| | Cys12--Cys18 | 5.6 | 5.3 | 7.5 | 2.7 |
| | Cys27--Cys34 | 4.5 | 8.5 | 4.3 | 2.7 |
| scC550 | Cys8--Cys24 | 9.9 | 10.2 | 7.5 | 2.8 |
| | Cys12--Cys18 | 6.5 | 4.7 | 5.9 | 2.9 |
| | Cys27--Cys34 | 8.9 | 9.0 | 5.2 | 2.7 |
| scC620 | Cys8--Cys24 | 9.6 | 11.0 | 7.6 | 2.8 |
| | Cys12--Cys18 | 5.8 | 9.6 | 6.1 | 2.8 |
| | Cys27--Cys34 | 7.8 | 6.4 | 5.3 | 2.8 |
| scC725 | Cys8--Cys24 | 12.1 | 11.7 | 7.5 | 2.9 |
| | Cys12--Cys18 | 7.3 | 7.4 | 5.8 | 2.8 |
| | Cys27--Cys34 | 8.0 | 9.3 | 5.3 | 2.8 |

[†]Distances between $C_\alpha$ range from 4.6–7.4 Å in known protein structures.[42]

uses very few restraints since the target structure is not known.

After construction and equilibration of the scrambled A and C species, the stereochemical quality of the protein structures was checked using PROCHECK program.[24] The aim of PROCHECK is to assess both the overall stereochemical quality of a given protein structure, as compared with well-refined structures, and to give an indication of its local, residue-by-reside reliability. A second database consisting of only small, disulfide-rich proteins was also assembled and used to check the scrambled species. This new database contains 15 different proteins, some NMR structures, and some X-ray crystal. Individual NMR structures, rather than just the average NMR structures, were used to improve the statistics in the database. Few main-chain bond length and angles have significant deviations from original PROCHECK average values. The stereochemical parameters and additional details are presented in Appendix A.

**TABLE III. Properties of the PCI Simulations. The RMSD, RMSF and the Radius of Gyration Are Reported in Å, the SASA Is Reported in Å$^2$**

| Property | N300[f] | scA343 | scA456 | scA550 | scA620 | scA725 | scC343 | scC355 | scC550 | scC620 | scC725 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Backbone RMSD[a] | 0.00 | 2.08 | 2.34 | 3.46 | 3.38 | 3.99 | 1.91 | 1.74 | 2.41 | 2.73 | 3.58 |
| Backbone RMSF[b] | 0.99 | 0.88 | 1.28 | 1.56 | 1.49 | 1.09 | 1.06 | 1.00 | 1.30 | 1.21 | 1.16 |
| Radius of gyration[c] | 8.96 | 9.03 | 9.11 | 9.32 | 8.52 | 9.32 | 8.96 | 9.16 | 9.60 | 9.40 | 9.15 |
| SASA[d] | (2079.8)[e] | 261.4 | 390.8 | 4.0 | 193.1 | 377.5 | 338.8 | 301.0 | 479.6 | 551.0 | 335.3 |

[a]The backbone atoms RMS deviation from the N300 average structure for core residues (8Cys-34Cys) averaged over the last 50ps; the RMSD for N300 relative to the crystal structure is 1.69Å.
[b]The backbone atoms RMS fluctuation about the average structure for last 50ps.
[c]The radius of gyration for core residues (8Cys-34Cys) averaged over the last 50ps.
[d]Solvent accessible surface area (SASA) for core residues (8Cys-34Cys) averaged over the last 50ps relative to the N300 average structure value.
[e]Original N300 average structure values of SASA.
[f]Results reported in Martí-Renom et al.[3]

To further evaluate the refolded scrambled structures, two different solvation models for proteins were used; comparisons were made to results obtained for the crystal structure of native PCI, the NMR structure of PCI, as well as for the refolded structures of native PCI generated in the previous study.[3] In the first approach, the Verify 3D model of Lüthy et al.[25] was used. This empirical solvation model for proteins assumes that the total solvation free energy is a sum of contributions from the constituent groups of the protein; internal energy terms are not accounted for in this model. The second approach to the characterization of PCI and its scrambled species used an effective energy function based on the CHARMM polar hydrogen force field[16] combined with a solvent exclusion model for the solvation free energy.[26,27] In this model (EEF1), the internal energy components are considered as well as the nonbonded terms. The inclusion of internal terms was shown to be important in the previous study of native PCI, where a grossly misfolded structure obtained a good score within the model of Lüthy et al.[25] but was picked out by the EEF1 model as being a poor structure. Such comparisons were shown in our previous article[3] to be able to discriminate between the experimental PCI structure and the refolded structures.

## RESULTS

### Structures A and C; Stability and Convergence

The overall properties of the various simulations are given in Table III. In all 10 simulations, the RMSF of the protein is under 1.6 Å indicating that the structures have reached an structural equilibrium during the final 50ps. The RMSD time series of the individual scrambled structures with respect to their average structure is given in Figure 2. In all cases except one, the RMSD values are stable during the 200ps simulation at 300K; only simulation scA725 has a drop in the RMSD from 2.4 Å–1.7 Å near the end of the trajectory. All ten structures are more solvent-exposed with an increase of SASA from 0.2–26.5%, and all structures but one (scA620) have a larger radius of gyration than the average simulation structure of native PCI (N300) (on average, 2.62% larger for the scrambled A species and 3.28% larger for the scrambled C species). The scA structures tend to have lower values for the SASA than the scC structures.

### Comparison of A and C With the Native Structure

Figure 3 shows the average N300 structure and the individual structures for all scrambled species A and C, as well as their average structure (scA and scC, respectively). The differences between the three average structures can be explained by the different disulfide pairing. Loop 1 of the scA structure (see legend in Fig. 1 for region definitions) is in the interior of the protein due to the disulfide bond between Cys8 and Cys27. The 3,10 helix region has moved with respect to its position in the N300 structure because of the lack of the disulfide bond between Cys12 and Cys27, which is present in the native conformation. The disulfide bond between Cys24 and Cys34 maintains the C-terminus in the native orientation. This is not the case for the average structure scC, which has a disulfide bond between Cys34 and Cys27. Loop 3 has considerable mobility and there is considerable diversity in the orientation of the C-terminal end. Another characteristic of the scC structure is the lack of a disulfide bond in the left part of the protein (as in the orientation of Fig. 3). This allows mobility between the helix 3,10 region and loop 3.

Table IV shows the back-bone RMSD matrix for the three native structures (X-ray, NMR, N300) and all the scrambled species. All the RMSD values are calculated for the core of the protein, Cys8– Cys34. Starting structures from 343ps, 550ps, 620ps, and 725ps were used to construct both scA and scC species. The RMSD between the two structures that were folded from the same structure (i.e., scA343 and scC343; scA550 and scC550; scA620 and scC620; scA725 and scC725) ranges from 1.53 Å–3.55 Å. Along the diagonal of Table IV are the RMSD values between a particular scrambled species and the starting structure from which it was folded. These results indicate that the application of a specific pair of NOE constraints is important in determining the final structures. In the family of scA structures, the average RMSD of the individual structures from the overall mean structure is 1.8 Å and the standard deviation is 0.3 Å in the family of scC structures, the corresponding values are 1.6 Å and 0.5 Å, respectively. The RMSD between the average scA struc-

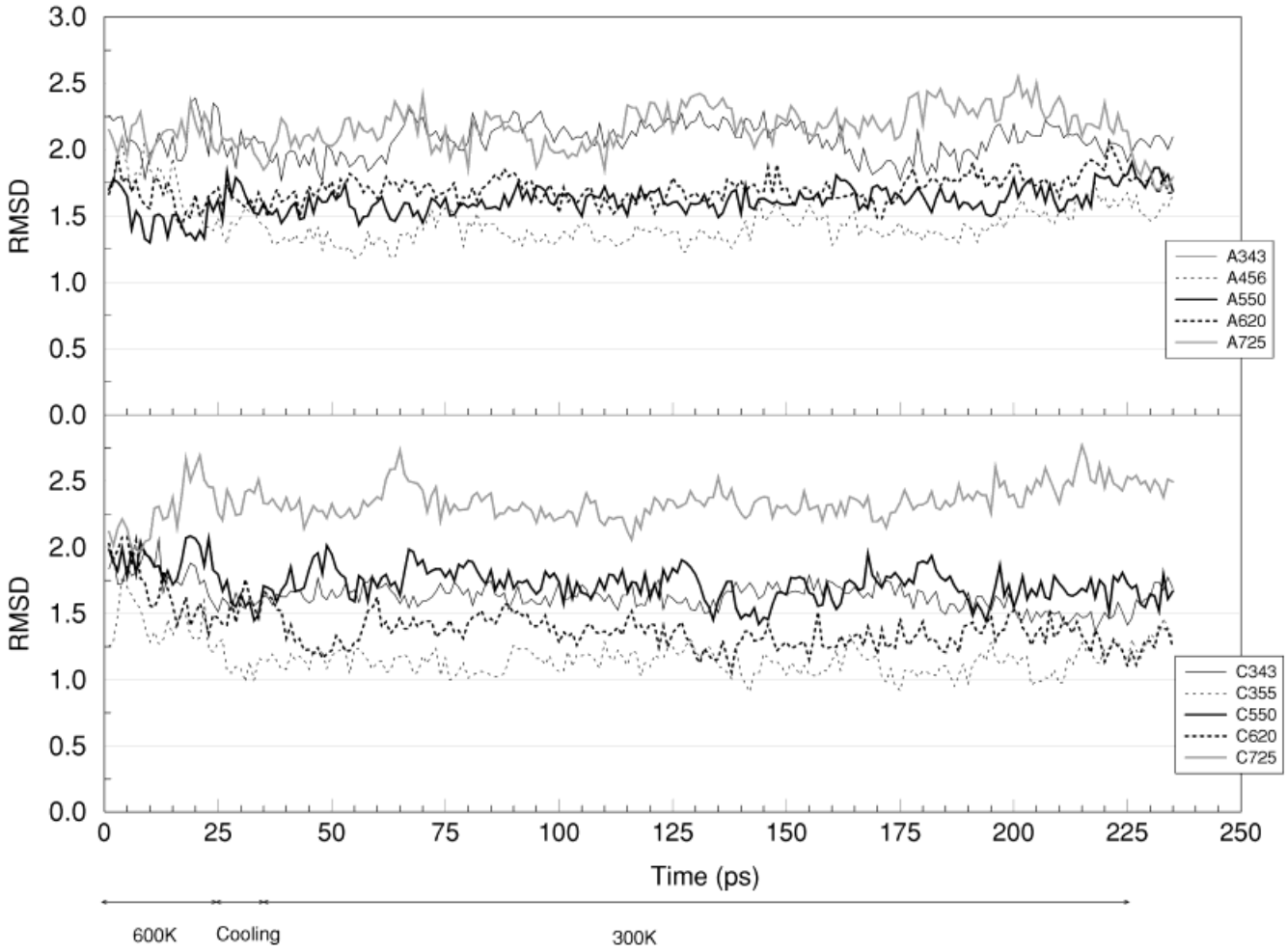


Fig. 2. RMSD as a function of time during the construction of the scrambled species. **a**: RMSD of the five different scrambled A structures from the scA structure. **b**: RMSD of the five scrambled C structures from the scC structure.

ture and the native structures are 2.8 Å, 3.0 Å, and 2.6 Å for the crystal structure, the NMR structure, and the N300 simulation structure, respectively; between the scC average structure and the native structures, the RMSD values are 2.6 Å, 2.8 Å, and 2.0 Å for the crystal, NMR, and N300 simulation structure, respectively. These results indicate that the scC structure is more nativelike than the scA structure, although there is a greater variation in the RMSD values in the scC family of structures than in the family of scA structures.

The loop regions of PCI are an important structural element of the protein.[3] PCI contains three loop regions delimited by Cys8 and Cys12 (loop 1), by Cys18 and Cys24 (loop 2), and by Cys27 and Cys34 (loop 3); the 3,10 helix region of the native protein is between Cy12 and Cys18. The rms differences between the loop regions in the scrambled species and the native structure were calculated. Deviations in the loop regions can arise for two reasons: one, the loop may have the correct internal structure but the orientation may differ, or, two, the internal structure of the loop may be distorted. To look at these two contributions, the rms deviations were calculated in two ways: in the first, the back-bone atoms of the crystal structure and "second" structure were superimposed and the rms deviations of the different loop regions and the helix were calculated; this gives a measure of the deviations arising from differences in global orientation. In the second calculation, the different loop and helix regions were superimposed separately and the rms deviations were calculated; this gives a measure of the internal structural distortion. The results for the three loop regions and for the 3,10 helix are given in Table V. It is evident that there are significant overall as well as internal deviations relative to the native structure. In the scrambled A species, loops 1 and 2 differ both globally and internally relative to the crystal structure. In the scrambled C species, the differences appear to be due more to the global orientation than to the internal loop structure. For example, for the scrambled C species, many of the structures have an internal rmsd for loop 1 and loop 2 of about 1 Å or less. The overall rmsd is generally 4 Å or greater. In both the scrambled A and C species, the internal rmsd for the

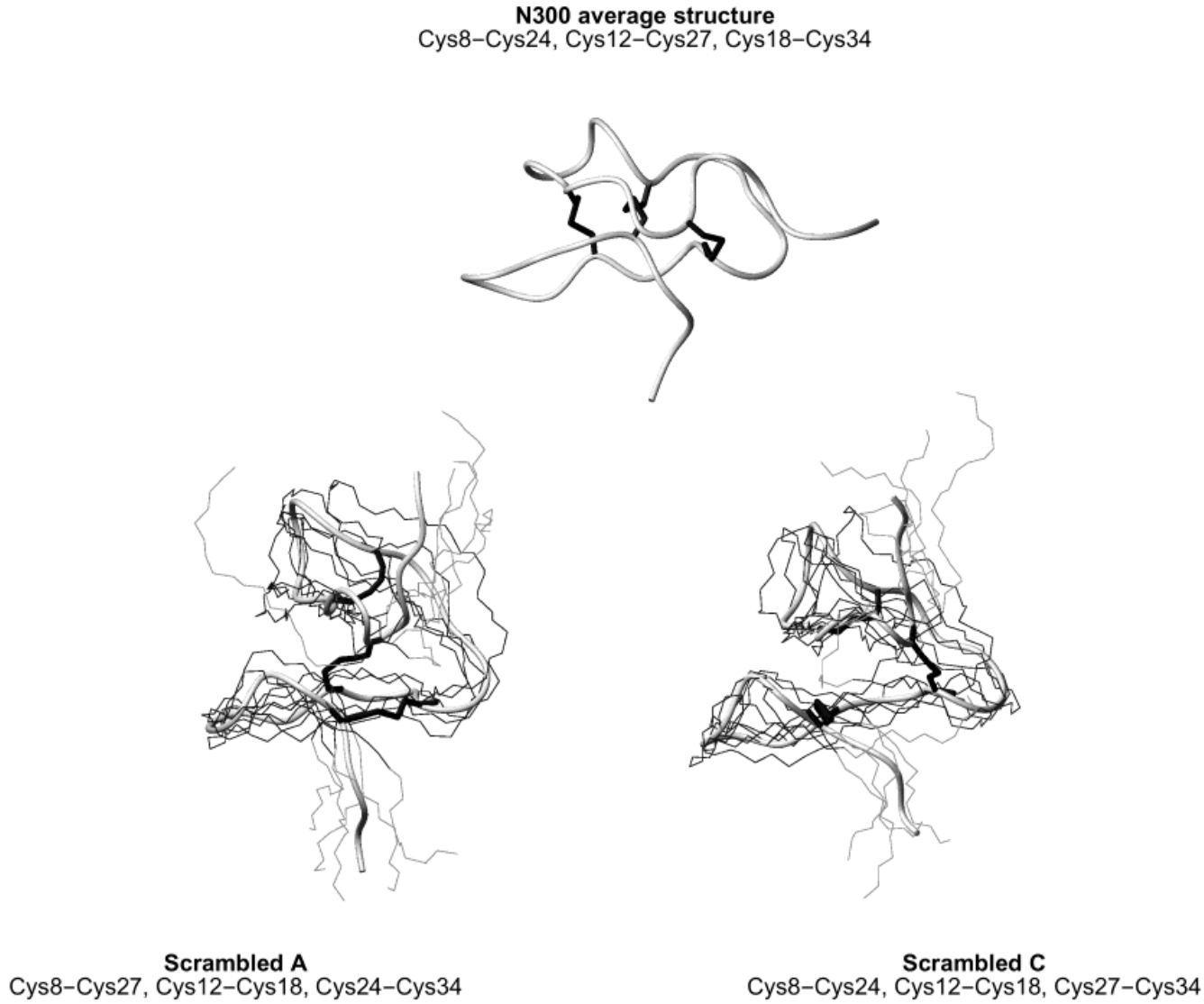


Fig. 3. A ribbon representation of the average structures for the N300 simulation and the average structures for the scrambled A and C species. The individual scrambled species are shown superimposed on the corresponding average structure.

3,10 helix region and loop 3 are about the same. Relative to the native structure, the scC has more nativelike structure in loops 1 and 2 than does the scA species.

**PROCHECK Analysis**

The stereochemical quality of the crystal structure and the structures from the simulations was checked using the PROCHECK program;[24] the results are presented in the Table VI. For part of the analysis, a database of structural parameters determined from small, disulfide-rich proteins was used.

The first part of Table VIA gives the distribution of $\phi,\psi$ angles in the different regions of the Ramachandran plot. The definitions of the different regions, i.e., *core*, *allowed*, *generous*, and *disallowed*, were the ones defined in the standard PROCHECK program (version 3.3.2), which were determined from an analysis of high-resolution protein structures in the Protein Databank.[28,29] For the native PCI structures, the percent of $\phi,\psi$ angles found in the core and allowed regions range from 50–60% and from 33.3–36.7%, respectively. Of the native structures, the crystal structure has the lowest percent in the core and the highest in the allowed and generous regions, while the simulation structure (N300) has the highest percent in the core and zero percent in the disallowed region. In the crystal structure, Trp22 is in a disallowed region and Ala4, Ala31, and Arg32 are in the generous region. During the simulation, the dihedral angles of these residues move into more acceptable regions of the Ramachandran plot. The NMR structure also has three residues in less acceptable regions of the Ramachandran plot (Cys8 and Trp28 are in the disallowed region and Asn9 is in the generous region).

**TABLE IV. Matrix Table for RMSD Between Control Structures and Scrambled Structures†**

| RMSD | Xray | NMR | N300[a] | scA343 | scA456 | scA550 | scA620 | scA725 | scC343 | scC355 | scC550 | scC620 | scC725 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xray | 0.00 | 1.16 | 1.69 | 2.31 | 2.70 | 3.43 | 3.86 | 4.16 | 2.48 | 2.32 | 3.24 | 3.36 | 3.84 |
| NMR | 1.16 | 0.00 | 1.76 | 2.39 | 2.76 | 3.77 | 3.81 | 4.40 | 2.65 | 2.58 | 3.26 | 3.51 | 4.07 |
| N300[a] | 1.69 | 1.76 | 0.00 | 2.08 | 2.34 | 3.46 | 3.38 | 3.99 | 1.91 | 1.74 | 2.41 | 2.73 | 3.58 |
| scA343 | 2.48 | 2.58 | 1.89 | 1.62 | 1.40 | 2.99 | 3.27 | 3.87 | *1.53* | 1.77 | 2.46 | 2.46 | 3.78 |
| scA456 | 3.10 | 3.39 | 2.76 | 1.97 | 2.12 | 2.61 | 2.57 | 3.36 | 1.88 | 1.50 | 2.10 | 2.00 | 3.26 |
| scA550 | 3.98 | 4.14 | 3.57 | 3.26 | 3.05 | 3.42 | 2.75 | 2.34 | 3.14 | 2.84 | *3.55* | 3.19 | 3.02 |
| scA620 | 4.75 | 4.88 | 4.43 | 4.25 | 3.81 | 2.92 | 3.48 | 2.49 | 3.08 | 2.68 | 3.12 | *2.57* | 2.55 |
| scA725 | 3.64 | 3.89 | 3.53 | 3.27 | 2.92 | 2.77 | 2.80 | 3.24 | 3.65 | 3.44 | 4.15 | 3.58 | *1.84* |
| scC343 | 2.48 | 2.58 | 1.89 | 0.00 | 1.97 | 3.26 | 4.25 | 3.27 | 1.52 | 1.88 | 2.33 | 2.01 | 3.32 |
| scC355 | 2.84 | 2.77 | 2.18 | 1.47 | 2.30 | 3.36 | 4.69 | 3.74 | 1.47 | 1.86 | 1.91 | 1.73 | 2.95 |
| scC550 | 3.98 | 4.14 | 3.57 | 3.26 | 3.05 | 0.00 | 2.92 | 2.77 | 3.26 | 3.36 | 3.25 | 1.89 | 3.63 |
| scC620 | 4.75 | 4.88 | 4.43 | 4.25 | 3.81 | 2.92 | 0.00 | 2.80 | 4.25 | 4.69 | 2.92 | 2.65 | 3.12 |
| scC725 | 3.64 | 3.89 | 3.53 | 3.27 | 2.92 | 2.77 | 2.80 | 0.00 | 3.27 | 3.74 | 2.77 | 2.80 | 3.23 |

†Values for RMSD (in Å) between average structures from last 50ps (upper-right diagonal) and starting structures for the refolding simulations (lower-left diagonal). Open box contains the RMSD values between the different A species and the shaded box contains the RMSD values between the different scrambled C structures.
[a]Results reported in Martí-Renom et al.[3]

**TABLE V. Regional RMSD Between the Crystal Structure and the NMR, N300, and Refolded Structures†**

| Structure | Backbone | Loop 1 | | Helix 3-10 | | Loop 2 | | Loop 3 | |
|---|---|---|---|---|---|---|---|---|---|
| NMR | 1.7 | 0.7 | 0.4 | 1.1 | 0.4 | 1.7 | 1.3 | 1.0 | 0.5 |
| N300 | 2.0 | 1.1 | 0.3 | 1.9 | 1.1 | 1.5 | 0.7 | 2.1 | 1.2 |
| scA343 | 3.6 | 2.0 | 1.4 | 3.9 | 1.8 | 2.5 | 1.5 | 2.5 | 0.8 |
| scA456 | 4.4 | 2.3 | 1.4 | 5.1 | 1.6 | 3.5 | 1.2 | 3.0 | 0.9 |
| scA550 | 4.3 | 3.5 | 1.6 | 5.4 | 1.8 | 2.4 | 1.6 | 3.6 | 1.2 |
| scA620 | 4.9 | 4.3 | 1.7 | 6.1 | 1.7 | 4.9 | 2.5 | 2.8 | 1.4 |
| scA725 | 7.8 | 8.0 | 1.5 | 8.2 | 1.5 | 6.3 | 2.0 | 5.7 | 1.0 |
| <scA> | 4.3 | 3.8 | 1.6 | 4.2 | 1.6 | 4.2 | 1.5 | 1.8 | 0.7 |
| scC343 | 4.9 | 4.0 | 1.0 | 4.6 | 1.8 | 5.0 | 1.2 | 2.8 | 0.7 |
| scC355 | 5.0 | 4.2 | 0.7 | 5.0 | 1.4 | 5.3 | 0.8 | 2.8 | 1.0 |
| scC550 | 5.1 | 4.2 | 0.6 | 5.4 | 1.7 | 4.6 | 1.5 | 4.3 | 1.1 |
| scC620 | 5.9 | 6.4 | 1.2 | 5.5 | 1.6 | 6.6 | 1.4 | 3.8 | 1.2 |
| scC725 | 4.2 | 3.4 | 1.7 | 5.5 | 1.2 | 2.9 | 1.8 | 2.6 | 1.6 |
| <scC> | 4.2 | 4.1 | 1.0 | 4.7 | 1.4 | 4.5 | 1.0 | 2.4 | 0.9 |

†For the loop and helix entries, the first value is the RMSD after reorientation of the entire backbone to the crystal structure; the second entry is after reorientation of the specific region to the crystal structure.

In the scrambled species, the percent of dihedrals found in the core region of the Ramachandran plot is about the same as for the native structures, with the exception of structures scC343 and scC550 which have a higher percentage (<60%) and scC725 which has a lower percentage (43.3%). Except for scC725, none of the scrambled species have $\phi,\psi$ angles in the disallowed region.

The stereochemical quality index[30] classifies three parameters: the $\phi,\psi$ distribution, pooled $\chi$ 1 standard deviations, and hydrogen bond donor energies. Four classes, denoted as 1–4, are defined using statistical data for each parameter. The highest quality in each category is assigned an integer value of 1. Although this classification is most often used to determine the quality of the structure, it is employed here to determine if there are significant differences between the native PCI structure and the scrambled species. In this analysis, the database of parameters from the small disulfide-rich (SDR) proteins is also employed. The results for stereochemical quality index are presented in second part of Table VIA. In both databases, the crystal and NMR structures give quality indexes between 3 and 4 for all categories. The native simulation structure (N300) has a better quality index than the experimental structures for the $\chi$ 1 distribution in both databases. These results show that the $\chi$ 1 distribution improves upon simulation. The hydrogen bond energy index improves for the simulation structures in the SDR protein database.

The Morris scores for the proteins that make up the SDR protein database are given in Table VIB; for an NMR structure where a family of structures was available from the Protein Databank,[28,29] the scores for the individual structures were calculated and averaged to give a single score for the protein. Of the 14 structures, only four had a $\phi,\psi$ distribution score below 3; of the remaining 10 structures, five had scores in the range 3–3.5 and five structures had scores between 3.5 and 4. A score of 4 means that less than 55% of the $\phi,\psi$ angles fall in the core region of the Ramachandran plot. This is to be compared with the $\phi,\psi$ distribution score of the scA and scC structures from

**TABLE VIA. PROCHECK Results for the Native and Scrambled Structures of PCI**

| | Ramachandran Plot (%) | | | |
|---|---|---|---|---|
| | Core | Allow | Generous | Disallowed |
| X-Ray | 50.0 | 36.7 | 10.0 | 3.3 |
| NMR | 56.7 | 33.3 | 3.3 | 6.7 |
| N300 | 60.0 | 33.3 | 6.7 | 0.0 |
| A343 | 56.7 | 40.0 | 3.3 | 0.0 |
| A456 | 50.0 | 40.0 | 10.0 | 0.0 |
| A550 | 53.3 | 40.0 | 6.7 | 0.0 |
| A620 | 53.3 | 36.7 | 10.0 | 0.0 |
| A725 | 50.0 | 36.7 | 13.3 | 0.0 |
| C343 | 66.7 | 30.0 | 3.3 | 0.0 |
| C355 | 56.7 | 40.0 | 3.3 | 0.0 |
| C550 | 63.3 | 33.3 | 3.3 | 0.0 |
| C620 | 50.0 | 40.0 | 10.0 | 0.0 |
| C725 | 43.3 | 43.3 | 10.0 | 3.3 |

| | Stereochemical quality index | | | | | |
|---|---|---|---|---|---|---|
| | Original database | | | SDR* protein database | | |
| | $\phi, \Psi$ Distrib. | $\chi_1$ St. Dev. | HB Energy | $\phi, \Psi$ Distrib. | $\chi_1$ St. Dev. | HB Energy |
| X-Ray | 4 | 4 | 4 | 4 | 4 | 4 |
| NMR | 3 | 3 | 3 | 3 | 3 | 3 |
| N300 | 3 | 1 | 3 | 3 | 1 | 2 |
| A343 | 3 | 1 | 4 | 3 | 1 | 3 |
| A456 | 4 | 2 | 3 | 4 | 2 | 2 |
| A550 | 4 | 1 | 2 | 4 | 1 | 1 |
| A620 | 4 | 1 | 4 | 4 | 1 | 3 |
| A725 | 4 | 2 | 3 | 4 | 2 | 3 |
| C343 | 2 | 1 | 2 | 2 | 1 | 2 |
| C355 | 3 | 1 | 4 | 3 | 1 | 2 |
| C550 | 3 | 1 | 3 | 3 | 1 | 2 |
| C620 | 4 | 1 | 4 | 4 | 1 | 3 |
| C725 | 4 | 1 | 4 | 4 | 1 | 3 |

| | | G Factors | |
|---|---|---|---|
| | Dihedrals | Covalent | Overall |
| X-Ray | −1.1 | −3.6 | −2.0 |
| NMR | −0.7 | −2.3 | −1.3 |
| N300 | −0.8 | −2.6 | −1.6 |
| A343 | −0.8 | −0.7 | −0.7 |
| A456 | −0.9 | −1.2 | −1.1 |
| A550 | −1.0 | −3.0 | −2.0 |
| A620 | −0.9 | −2.3 | −1.6 |
| A725 | −1.1 | −1.6 | −1.4 |
| C343 | −0.7 | −0.9 | −0.8 |
| C355 | −0.6 | −0.8 | −0.7 |
| C550 | −0.7 | −3.8 | −2.1 |
| C620 | −0.7 | −2.7 | −1.6 |
| C725 | −1.0 | −1.7 | −1.3 |

*Small Disulfide-Rich.

**TABLE VIB. Morris Classification for Disulfide Proteins in the SDR Protein Database**

| PDB ID | # Structures | $\phi, \Psi$ Distrib | $\chi_1$ St. Dev. | HB Energy |
|---|---|---|---|---|
| 1cco | 24 | 2.7 | 2.5 | 2.4 |
| 1ixa | 1 | 4.0 | 2.0 | 3.0 |
| 1kal | 10 | 3.0 | 1.7 | 3.0 |
| 1mct | 1 | 1.0 | 1.0 | 3.0 |
| 1omb | 1 | 3.0 | 4.0 | 3.0 |
| 1omc | 21 | 2.2 | 1.3 | 2.2 |
| 1omn | 15 | 3.3 | 3.6 | 2.9 |
| 1tcg | 1 | 4.0 | 4.0 | 3.0 |
| 1tch | 1 | 4.0 | 4.0 | 3.0 |
| 1tcj | 10 | 3.8 | 4.0 | 3.4 |
| 1tck | 10 | 3.4 | 3.8 | 2.6 |
| 2tgf | 1 | 4.0 | 2.0 | 2.0 |
| 3cti | 6 | 1.5 | 1.3 | 1.8 |
| 3egf | 16 | 3.3 | 3.6 | 2.3 |

simulation; almost all of the scrambled structures fall in the range between 3 and 4. For the $\chi$ 1 standard deviation score, seven of the database proteins had average standard deviations greater than 3.5 and the rest had scores of 2.5 or less. The simulation structures all had scores between 1 and 2. For the hydrogen bond energy, all but one structure (1tcj) had scores of 3.0 or below. This compares well with the scores of the scrambled structures from the simulations. The comparisons indicate that even for the experimental structures that make up the SDR database, the $\phi$,$\psi$ distribution can be such that a large proportion fall outside the core regions of the Ramachandran plot. This is likely to be characteristic of this class of protein due to the highly constrained nature of the protein resulting from the disulfide bonds. This characteristic behavior is to be compared with the vast majority of proteins in the database, which have 65% or more of the residues in the core region.[30]

The G-factors, which characterize the covalent geometry, provide a measure of the deviation of the dihedral angles and bond lengths from the average values derived from an analysis[31] of small molecules in the Cambridge Structural Database;[32] these results are given in the last part of Table VIA. A low G factor indicates that the property corresponds to a low probability structure. For the dihedral angles, the results are consistent with the previous analysis, in that, for the most part, there is little difference between the native structures and the scrambled structures. For the covalent character, the crystal structure has a worse score than the NMR or native simulation structures. There is a noticeable improvement in the covalent score upon simulation for most of the scrambled structures, except for scC550. The general improvement is to be expected since many of the bond and angle parameters in the CHARMM force field were derived from small molecule compounds.[33] However, there is no significant difference between the scrambled species and the native species in terms of G-factors.

### Empirical Solvation Model

In recent years, numerous models have been developed for evaluating protein structures and to discriminate between native conformations and compact nonnative ones in proteins; these include empirical solvation models,[25,34–37] contact potentials,[38,39] and an effective energy function (EEF1) based on a model for the solvation free energy.[40] These models appear to discriminate between

**TABLE VII. Scores From the Model of Lüthy et al.[a] Only the Core Residues, From 11 to 27 Are Included in the Summation. Scores Obtained for the Refolded Structures of Martí-Renom et al.[b] Are Given**

| Scores | | Scores | |
|---|---|---|---|
| | | X-ray | 7.48 |
| | | NMR | 7.64 |
| | | N300 | 5.75 |
| Scrambled A and C | | Refolded Native PCI | |
| scA343 | 4.22 | A319 | 2.12 |
| scA456 | 3.43 | A358 | 2.86 |
| scA550 | 2.51 | A474 | 3.16 |
| scA620 | 4.99 | A540 | 4.79 |
| scA725 | 3.42 | A725 | 4.71 |
| scC343 | 3.23 | A820 | 4.56 |
| scC355 | 2.71 | B358 | 4.85 |
| scC550 | 3.27 | C474 | 4.81 |
| scC620 | 0.34 | D810 | 3.66 |
| scC725 | 4.76 | Misfolded | 6.12 |

[a]Lüthy et al.[25]
[b]Martí-Renom et al.[3]

grossly misfolded proteins generated by threading methods and the native conformation. In the earlier work[3] we tested the empirical solvation models of Lüthy et al.[25] and of Koehl and Delarue[36] and the EEF1 model.[40] We found that the three models appear to discriminate between the native experimental structures and the structures refolded by molecular dynamics simulations. However, in some cases the empirical solvation models gave good scores to grossly misfolded structures that contained high internal energies; the EEF1 model discriminated between the grossly misfolded structures and the near native structures. The approach taken in the previous work as well as in the present study is to apply several of these methods. In the present study, the model of Lüthy et al.[25] was used to generate 3-D profiles of the various structures. The residue-by-residue scores are determined using a 21-residue sliding window, so the scores for the first several and last several residue have little or no meaning. The Verify 3D program is accessible from the website site www.doe-mbi.ucla.edu/Services/Verify3D.html. Total scores were calculated by summing over the residue scores of the protein core (residues 11–27). The results for the crystal structure, the NMR structure, the average structure from the native protein simulation (N300 in Martí-Renom et al.,[3]) the scrambled structure determined in the present work, and the refolded structures obtained in the previous work[3] are given in Table VII. As in the previous work, the crystal structure and the mean NMR structure give the highest scores; the NMR structure scores better than the crystal structure, even though the profiles are based on X-ray structures. The average structure of the native protein (N300) has a score which is worse than the two experimental structures. The scrambled structures show similar scores, except for scC620 which has a very low score. Several of the scrambled species score less well than the native refolded structures, but otherwise there are no significant differences between the native structures and the scrambled structures. This is perhaps related to the fact that this structure has the largest SASA relative to the other scrambled species. The refolded native structures from the previous study also score less well than the experimental structures.

The EEF1 is based on the CHARMM polar hydrogen force field[16] combined with a Gaussian solvent exclusion model for the solvation free energy.[26,27] In this model, the internal energy components are considered as well as the nonbonded terms. EEF1 has been used recently to distinguish well-constructed decoys (some within 1–2 Å from the native structure) from native proteins.[40] The present structures were energy-minimized for 300 steps using the ABNR routine in the CHARMM program, and the energies were calculated for the entire protein and for the core. The results are given in Table VIII. The native structures, as in the model of Lüthy et al.,[25] all give better energies than the nonnative scrambled species, and the core energies follow a trend similar to that of the entire protein. In contrast to the results with the pure solvation models described above, the EEFI model separates structure scA620 from the rest. The energy of this scrambled A species is significantly higher than that of the others due to a high internal energy. Graphical examination of this structure shows that during the refolding of the scA620 structure, the Cys8–Cys27 disulfide bond goes through the Pro 10 ring and yields a very high internal bond energy.

Recently, the structures of three disulfide-bonded isomers of α-conotoxin GI, a 13-residue, three–disulfide-bond protein, were determined by $^1$H NMR spectroscopy; one isomer had the native disulfide pairing and the other two were scrambled species.[13] The energies for each member of the family of structures of the native isomer (XGA, 35 structures) and its scrambled isomers (XGB, 24 structures and XGC, 25 structures) were calculated using the EEF1 model after 300 steps of energy minimization. The average rmsd from the initial structures was 0.56 Å, 0.71 Å, and 0.81 Å for the XGA, XGB, and XGC structures, respectively. In this case, where only experimental structures are used, the average energy of the native structures (−359.8 kcal/mole) was lower than the average energies of the two scrambled isomers, −354.3 kcal/mole and −346.3 kcal/mole for XGB and XGC isomers, respectively. If an average structure was first calculated and then energy-minimized using the EEF1 model, the native structure still had a lower energy (−339.5 kcal/mole) than the mean XGB structure (−336.5 kcal/mole) and the mean XGC structure (−331.7 kcal/mole). The rmsd between the minimized mean structures and the initial structures was 1.59 Å, 1.52 Å, and 3.25 Å for the XGA, XGB, and XGC structures, respectively. Tables of the energies for the individual structures are given in Appendix B. Due to the short chain-length of this protein (13 residues), the Verify 3D model was not used. In comparison with the results for PCI and its scrambled isomers, a similar trend is observed: the native structure has a lower EEF1 energy than either of the two scrambled species.

**TABLE VIII. The Total Energy and Energy Components Calculated Using the EEF1† Model. Energies Are in kcal/mole**

| | Total Energy | | Internal | | van der Waals | | Electrostatic | | Solvent | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total | Core | Total | Core | Total | Core | Total | Core | Total | Core |
| X-ray | −992.6 | −783.8 | 140.6 | 109.8 | −194.0 | −118.8 | −550.9 | −393.8 | −388.3 | −381.1 |
| NMR | −1034.4 | −824.7 | 148.3 | 105.9 | −206.4 | −129.4 | −608.9 | −425.1 | −367.4 | −376 |
| N300 | −972.0 | −755.5 | 125.3 | 94.4 | −200.7 | −125.0 | −497.3 | −342.4 | −399.4 | −382.5 |
| scA343 | −948.7 | −754.8 | 126.7 | 91.1 | −171.4 | −122.0 | −486.9 | −342.8 | −417.0 | −381.1 |
| scA456 | −910.1 | −715.4 | 143.2 | 107.3 | −152.8 | −114.6 | −467.6 | −320.9 | −432.8 | −387.2 |
| scA550 | −929.5 | −724.2 | 152.5 | 117.7 | −175.1 | −105.8 | −488.6 | −346.8 | −418.2 | −389.2 |
| scA620 | 615.9 | 815.2 | 1136.4 | 1101.2 | 348.3 | 394.1 | −440.0 | −298.9 | −428.7 | −381.2 |
| scA725 | −937.0 | −745.5 | 128.1 | 88.7 | −151.4 | −101.5 | −481.3 | −338.7 | −432.5 | −393.9 |
| scC343 | −944.5 | −746.8 | 124.1 | 87.9 | −169.5 | −118.4 | −475.5 | −328.3 | −423.6 | −388.1 |
| scC355 | −932.7 | −734.6 | 125.8 | 91.8 | −161.6 | −120.4 | −461.6 | −318.1 | −435.3 | −387.9 |
| scC550 | −941.4 | −755.9 | 120.4 | 91.7 | −153.2 | −111.9 | −473.1 | −345.7 | −435.6 | −389.9 |
| scC620 | −918.2 | −748.7 | 123.6 | 86.3 | −141.5 | −111.5 | −457.1 | −329.1 | −443.2 | −394.4 |
| scC725 | −930.3 | −724.2 | 138.4 | 99.1 | −165.8 | −106.5 | −477.1 | −319.9 | −425.7 | −396.9 |

†Lazaridis and Karplus.[26,27]

## CONCLUSIONS

A molecular dynamics simulation approach[3] for refolding to the native state by the introduction of disulfide constraints is used here to predict the structures of two scrambled disulfide species of the PCI. Five scrambled species have been isolated experimentally; for two (A and C) of these the disulfide pairing is known. Using this information, 10 molecular dynamics simulations were performed to predict the structures of the scrambled A and C species. The overall RMSD from the native structure of the nonscrambled species (X-ray crystal, NMR, and N300 simulation) ranged between 2.7 Å and 3.2 Å for the scrambled A species and between 2.1 Å and 2.9 Å for the scrambled C species. The actual folds are significantly different due the different disulfide pairing. The scrambled species are all more solvent-exposed than the native structure and have a larger radius of gyration than the native structure.

The molecular dynamics simulation results are in general agreement with the limited experimental results on the unfolding and refolding of PCI.[8] They indicate that the scrambled species are less compact and less structured than the native species. They also suggest that the scrambled A species is more structured than scrambled C species. From the simulations, the five scrambled A structures have smaller SASA values and radii of gyration than the five C scrambled species and also show a somewhat smaller variation in rmsd than the scrambled C structures. However, analysis of the loop regions and 3,10 helix indicates that the scrambled C species has more nativelike character than the scrambled A species. These factors, as well as the fact that the scrambled C species contains one native disulfide bond, may contribute to the faster conversion of the scrambled C species to the native structure. Testing of these predictions will have to wait for the determination of the structures of the scrambled A and C species of PCI. This is currently underway by NMR spectroscopy.

## ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with Silvana Pavia and Dr. Francesc Canals. The Institut de Dèveloppment et des Ressources en Informatique Scientifique (IDRIS) provided an allocation of CRAY computer time that was partially used for this work. Computer resources provided by C4-CESCA (Barcelona) and the Parallel Computing Center at the Université Louis Pasteur are gratefully acknowledged. M.A. Martí-Renom was a fellowship recipient from Universitat Autúnoma de Barcelona (FI-DGR/UAB).

## REFERENCES

1. Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen J. Critical assessment of methods of protein structure prediction (CASP): round II. Proteins 1997;Suppl 1:2–6.
2. Sternberg MJ, Bates PA, Kelley LA, MacCallum RM. Progress in protein structure prediction: assessment of CASP3. Curr Opin Struct Biol 1999;9:368–73.
3. Martí-Renom MA, Stote RH, Querol E, Avilès FX, Karplus M. Refolding of potato carboxypeptidase inhibitor by molecular dynamics simulations with disulfide bond constraints. J Mol Biol 1998;284: 145–172.
4. 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;47:1309–1515.
5. Haber E, Afinsen CB. Side-chain interactions governing the pairing of half cystine residues in ribonuclease. J Biol Chem 1962;237:1839–1843.
6. Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–230.
7. Chang J-Y. The properties of scrambled hirudins. J Biol Chem 1995;270:25661–25666.
8. Chang J-Y, Canals F, Schindler P, Querol E, Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. J Biol Chem 1994;269:22089–22094.
9. Chang J-Y, Kumar TKS, Yu C. Unfolding and refolding of cardiotoxin III elucidated by reversible conversion of the native and scrambled species. Biochemistry 1998;37:6745–6751.
10. Hass GM, Ryan CA. Carboxypeptidase inhibitor from potatoes. Methods Enzymol 1981;80:778–791.
11. Mas JM, Aloy P, Martí-Renom MA, Oliva B, Blanco-Aparicio C, Molina MA, de Llorens R, Querol E, Avilés FX. Protein similarities beyond disulphide bridge topology. J Mol Biol 1998;284:541–548.
12. Blanco-Aparicio C, Molina MA, Fernandez-Salas E, Frazier ML, Mas JM, Querol E, Avilés 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;20: 12370–12377.
13. Gehrmann J, Alewood PF, Craik DJ. Structure determination of the three disulfide bond isomers of a-conotoxin GI: a model for the role of disulfide bonds in structural stability. J Mol Biol 1998;278: 401–415.
14. Ress 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.
15. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein Jr. ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
16. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996;105:1902–1921.
17. Brünger A, Brooks III CL, Karplus M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. Chem Phys Lett 1984;30:495–500.
18. Brooks III CL, Karplus M, Pettitt BM. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Adv Chem Phys 1988;71:259.
19. Brooks III CL, Karplus M. Solvent effects on protein motion and protein effects on solvent motion. J Mol Biol 1989;208:159 –181.
20. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.
21. Reference deleted in proofs.
22. Schlitter J, Engles M, Kruger P, Jacoby E, Wollmer A. Targeted molecular dynamics simulation of conformational change. Application to the T = R transition in insulin. Mol Simulat 1993;10:291–308.
23. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 1999;288:441–459.
24. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
25. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensioanl profiles. Nature 1992;356:83–85.
26. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
27. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
28. Abola EE, Bernstein FC, Bryant SH, Koetzle TF, Weng J. Protein Data Bank. Crystallographic Databases-Information Content, Software Systems, Scientific Applications. 1987;107–132.
29. Abola EE, Sussman JL, Prilusky J, Manning NO. Protein data bank archives of 3-D macromolecular structures. Methods Enzymol 1997;277:556–571.
30. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins 1992; 12:345–364.
31. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
32. Allen FH, Kennard O. 3D search and research using the Cambridge structural database. Chemical Design Automation News, 1993;8:1, 31–37.
33. MacKerell ADJ, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem 1998;102:3586–3616.
34. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known 3-D structure. Science 1991;253: 164–170.
35. Casari G, Sippl MJ. Structure-derived hydrophobic potential. J Mol Biol 1992;244:725–732.
36. Koehl P, Delarue M. Polar and nonpolar environment in the protein core: implications for folding and binding. Proteins 1994; 20:264–278.
37. Samudrala R, Moult J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 1998;275:895–916.
38. Park B., Levitt M. Energy functions that discriminate X-ray and near native folds from well-constructed decoys: J Mol Biol 1996;258: 367–92.
39. Park BH, Huang ES., Levitt M. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. J Mol Biol 1997;266:831–46.
40. Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1999;288:477–487.
41. Kraulis PJ. Molscript: a program to produce both detailed and schematic plot structures. J Appl Crystallogr 1991;24:946–950.
42. Thornton JM. Disulfide bridges in globular proteins. J Mol Biol 1981;151:261–287.

## APPENDIX A

All the parameters were calculated using the subprograms anglen and secstr in the PROCHECK package. Once the parameters were calculated, the new values were tabulated and introduced in the files mplot.inc and pplot.inc of the PROCHECK program.

**TABLE A-I. Stereochemical Parameters for the Small-Disulfide-Rich Protein Database for the PROCHECK Program**[†]

| Stereochemical parameter | Mean value | Standard deviation |
|---|---|---|
| $\phi$-$\Psi$ in most favoured regions of Ramachandram plot | >90% | |
| $\chi^1$ dihedral angle: (in°) | | |
| gauche minus | 58.65 | 21.19 |
| trans | 194.18 | 20.42 |
| gauche plus | −65.85 | 20.64 |
| $\chi^2$ dihedral angle (in°) | 177.50 | 24.61 |
| Proline $\phi$ torsion angle (in°) | −66.21 | 10.89 |
| $\chi^3$; S-S bridge (in°) | | |
| right-handed | 94.35 | 30.07 |
| left handed | −82.08 | 27.89 |
| Disulphide bond separation (in Å) | 2.02 | 0.04 |
| $\omega$ dihedral angle (in°) | 176.03 | 5.80 |
| Main-chain hydrogen bond energy (Kcal/mol) | −1.60 | 0.70 |
| C$\alpha$ chirality | | |
| C$\alpha$-N-C-C$\beta$ Torsion angle (in°) | 34.80 | 3.08 |

[†]These values are extracted from PDB files: 1cco.pdb, 1ixa.pdb, 1xal. pdb, 1mct.pdb (chain I), 1omb.pdb, 1omc.pdb, 1omn.pdb, 1tcg.pdb, 1tch.pdb, 2tgf.pdb, 3cti.pdb, 3egf.pdb, pmpcfu.pdb, pmpcnf.pdb, and pmp2sakder.pdb.

## APPENDIX B

EEF1 energies for the three disulfide-bonded isomers of α-conotoxin GI. The native structure is denoted as XGA, the two disulfide bond isomers as XGB and XGC.

**TABLE B-I. EEF1 Energies for the α-Conotoxin GI XGA Native Isomer (2–7;3–13). Energies Are for the Individual Structures in the Family After 300 Steps of Minimization (ABNR), in kcal/mole**

| XGA-Total | Total | Internal | vdw | electro. | solvent |
|---|---|---|---|---|---|
| 1 | −374.0 | 48.6 | −49.5 | −198.0 | −175.1 |
| 2 | −359.5 | 44.5 | −46.7 | −180.7 | −176.6 |
| 3 | −357.7 | 46.0 | −46.7 | −181.0 | −176.0 |
| 4 | −359.0 | 47.2 | −52.0 | −182.0 | −172.2 |
| 5 | −359.8 | 44.3 | −46.6 | −180.5 | −177.0 |
| 6 | −359.3 | 44.6 | −50.9 | −178.5 | −174.5 |
| 7 | −360.0 | 42.0 | −50.3 | −176.7 | −174.9 |
| 8 | −357.7 | 41.8 | −46.0 | −174.5 | −178.9 |
| 9 | −357.0 | 41.3 | −45.9 | −173.9 | −178.3 |
| 10 | −357.8 | 50.4 | −53.9 | −183.9 | −170.4 |
| 11 | −358.1 | 45.0 | −47.9 | −176.5 | −178.8 |
| 12 | −360.6 | 41.8 | −49.6 | −177.1 | −175.7 |
| 13 | −359.0 | 46.2 | −48.0 | −179.3 | −177.9 |
| 14 | −363.4 | 49.8 | −44.1 | −190.0 | −179.0 |
| 15 | −360.8 | 47.2 | −51.0 | −180.1 | −176.8 |
| 16 | −360.5 | 44.5 | −45.7 | −179.1 | −180.2 |
| 17 | −356.7 | 41.6 | −45.5 | −174.7 | −178.2 |
| 18 | −361.9 | 46.5 | −47.2 | −183.7 | −177.5 |
| 19 | −358.9 | 49.3 | −42.4 | −185.1 | −180.7 |
| 20 | −358.5 | 42.3 | −48.4 | −176.5 | −175.9 |
| 21 | −360.6 | 43.2 | −45.8 | −181.0 | −176.9 |
| 22 | −355.2 | 42.7 | −45.3 | −173.2 | −179.4 |
| 23 | −361.8 | 45.0 | −51.6 | −179.6 | −175.6 |
| 24 | −366.6 | 46.4 | −47.9 | −188.4 | −176.7 |
| 25 | −358.8 | 50.2 | −46.5 | −186.3 | −176.2 |
| 26 | −352.3 | 45.5 | −48.3 | −173.2 | −176.3 |
| 27 | −357.7 | 42.9 | −44.1 | −176.6 | −179.9 |
| 28 | −362.0 | 46.1 | −46.9 | −182.8 | −178.5 |
| 29 | −355.6 | 45.3 | −47.8 | −178.2 | −174.9 |
| 30 | −360.0 | 44.7 | −46.4 | −181.4 | −176.8 |
| 31 | −360.0 | 48.4 | −47.9 | −183.2 | −177.3 |
| 32 | −359.1 | 49.8 | −53.0 | −181.2 | −174.6 |
| 33 | −364.1 | 47.7 | −47.2 | −187.7 | −176.9 |
| 34 | −359.1 | 44.0 | −48.0 | −178.8 | −176.4 |
| 35 | −360.0 | 48.4 | −47.9 | −183.2 | −177.3 |

**TABLE B-II. EEF1 Energies for the α-Conotoxin GI XGB Isomer (2–13;3–7). Energies for the Individual Structures in the Family. Energies After 300 Steps of Minimization (ABNR) in kcal/mole**

| XGB-Total | Total | Internal | vdw | Electro. | Solvent |
|---|---|---|---|---|---|
| 1 | −360.8 | 46.2 | −45.2 | −183.4 | −178.3 |
| 2 | −352.3 | 50.0 | −49.4 | −178.6 | −174.3 |
| 3 | −363.5 | 48.4 | −54.8 | −183.5 | −173.7 |
| 4 | −344.1 | 54.0 | −47.0 | −175.7 | −175.4 |
| 5 | −363.5 | 48.9 | −55.1 | −184.0 | −173.3 |
| 6 | −354.0 | 55.9 | −53.3 | −185.3 | −171.4 |
| 7 | −352.1 | 48.7 | −45.3 | −179.1 | −176.4 |
| 8 | −352.1 | 48.7 | −45.3 | −179.1 | −176.4 |
| 9 | −349.1 | 50.7 | −46.4 | −179.1 | −174.2 |
| 10 | −360.3 | 52.5 | −56.0 | −189.6 | −167.3 |
| 11 | −351.8 | 50.1 | −48.2 | −182.3 | −171.4 |
| 12 | −359.0 | 49.0 | −49.9 | −187.0 | −171.2 |
| 13 | −362.9 | 42.7 | −42.3 | −185.8 | −177.5 |
| 14 | −348.1 | 50.0 | −44.6 | −176.1 | −177.3 |
| 15 | −347.1 | 58.7 | −43.7 | −187.6 | −174.5 |
| 16 | −366.1 | 46.8 | −47.4 | −191.7 | −173.8 |
| 17 | −364.4 | 55.4 | −53.9 | −197.3 | −168.6 |
| 18 | −356.5 | 54.4 | −46.2 | −192.2 | −172.5 |
| 19 | −362.6 | 44.9 | −50.9 | −183.0 | −173.5 |
| 20 | −353.1 | 41.6 | −51.2 | −170.0 | −173.5 |
| 21 | −345.9 | 52.0 | −45.3 | −174.2 | −178.3 |
| 22 | −344.0 | 50.9 | −51.7 | −170.6 | −172.6 |
| 23 | −338.6 | 55.7 | −46.8 | −173.3 | −174.2 |
| 24 | −351.6 | 54.0 | −40.9 | −190.0 | −174.8 |

**TABLE B-III. EEF1 Energies for the α-Conotoxin GI XGC Isomer (2–3;7–13). Energies for the Individual Structures in the Family. Energies After 300 Steps of Minimization (ABNR) in kcal/mole**

| XGC-Total | Total | Internal | vdw | Electro. | Solvent |
|---|---|---|---|---|---|
| 1 | −333.2 | 61.1 | −50.3 | −176.4 | −167.7 |
| 2 | −353.8 | 56.5 | −55.7 | −182.7 | −171.9 |
| 3 | −346.4 | 53.2 | −56.3 | −167.7 | −175.6 |
| 4 | −334.3 | 51.1 | −48.1 | −160.6 | −176.8 |
| 5 | −336.4 | 54.0 | −55.9 | −162.5 | −172.0 |
| 6 | −345.2 | 53.7 | −51.6 | −168.6 | −178.7 |
| 7 | −330.2 | 51.4 | −39.7 | −157.0 | −185.0 |
| 8 | −342.8 | 47.0 | −53.6 | −162.9 | −173.3 |
| 9 | −333.3 | 53.7 | −48.6 | −163.2 | −175.3 |
| 10 | −347.2 | 50.8 | −50.4 | −171.8 | −175.8 |
| 11 | −337.2 | 62.6 | −48.7 | −177.5 | −173.6 |
| 12 | −357.2 | 52.6 | −67.2 | −185.2 | −157.4 |
| 13 | −347.1 | 52.4 | −59.1 | −171.2 | −169.2 |
| 14 | −356.5 | 50.5 | −53.1 | −177.1 | −176.8 |
| 15 | −353.4 | 57.5 | −54.6 | −193.9 | −162.4 |
| 16 | −365.0 | 51.7 | −51.2 | −200.3 | −165.2 |
| 17 | −347.8 | 56.0 | −49.0 | −182.1 | −172.7 |
| 18 | −350.7 | 56.1 | −51.1 | −185.0 | −170.7 |
| 19 | −343.6 | 56.7 | −54.5 | −177.8 | −168.0 |
| 20 | −339.5 | 53.6 | −50.1 | −168.0 | −175.0 |
| 21 | −352.8 | 54.2 | −54.7 | −179.7 | −172.6 |
| 22 | −340.3 | 56.5 | −51.0 | −172.6 | −173.2 |
| 23 | −357.1 | 51.7 | −55.8 | −184.0 | −169.0 |
| 24 | −348.3 | 49.5 | −48.8 | −175.6 | −173.3 |
| 25 | −357.4 | 53.5 | −51.1 | −186.3 | −173.4 |