Solution Structure of Contryphan-R, a Naturally Occurring Disulfide-Bridged Octapeptide Containing D-Tryptophan: Comparison with Protein Loops[†,‡]

Paul K. Pallaghy,[§] Albina P. Melnikova,[§] Elsie C. Jimenez,[||] Baldomero M. Olivera,[||] and Raymond S. Norton*[,§]

Biomolecular Research Institute, 343 Royal Parade, Parkville 3052, Australia, and Department of Biology, University of Utah, Salt Lake City, Utah 84112

Received March 24, 1999; Revised Manuscript Received June 18, 1999

ABSTRACT: Contryphan-R is a disulfide-constrained octapeptide containing a D-tryptophan that was isolated recently from venom of the cone shell *Conus radiatus.* The polypeptide is present in two forms in solution due to cis−trans isomerization at hydroxyproline 3. The solution structure of the major form of this unusual polypeptide, determined from NMR data, consists of a well-defined fold containing a non-hydrogen-bonded chain reversal from Gly1 to Glu5, which includes a *cis*-hydroxyproline and a D-Trp, and a type I β-turn from Glu5 to Cys8. The presence of a putative salt bridge between the Glu5 carboxyl group and the N-terminal ammonium group is investigated by using various solvation models during energy minimization and is compared with the results of a pH titration. A comparison of the structure of contryphan-R with other cyclic peptide structures highlights some of the key structural determinants of these peptides and suggests that the contryphan-R fold could be exploited as a scaffold onto which unrelated protein binding surfaces could be grafted. Comparison with small disulfide-bridged loops in larger proteins shows that contryphan-R is similar to a commonly occurring loop structure found in proteins.

Short polypeptides cyclized either in the backbone or via disulfide bridges often display well-structured conformations in solution (*1*−*3*) and represent a natural solution for the generation of potent biological activity from a minimal polypeptide sequence (*4*). There is much recent interest in the biosynthesis (*5*−*7*), in vivo activity (*5*), structural determinants (*3*, *8*), and drug design potential (*8*) of this class of compounds.

A recently isolated octapeptide, contryphan-R from *Conus radiatus*, causes stiff tail syndrome in mice (*5*), although its molecular target is not yet identified. Contryphan-R contains a remarkable number of post-translational modifications, including a hydroxyproline, a C-terminal amide, and for the first time from a normally translated tryptophan codon, a D- tryptophan (Figure 1). In addition, the D-handed residue occurs at position 4 of the sequence rather than position 2, as has been observed in some bioactive vertebrate and invertebrate peptides (*5*), and contryphan-R is the first naturally occurring example of a short disulfide-bridged peptide to contain a D-handed residue (*5*). A closely related contryphan, bromocontryphan, contains a further modification, bromination of Trp7 at the 6 position (*9*). The family of known contryphans from three *Conus* species (*5*, *6*, *9*, *10*) is summarized in Figure 1.

In this paper we have determined the solution structure of contryphan-R using NMR[1] spectroscopy. The structure is well-defined, with a type I β-turn centered on Pro6−Trp7, and it is possible to identify those residues with clear or putative structural roles. We also describe the effects of different refinement protocols and solvation models on the side-chain conformations of this peptide. Structured cyclic peptides are potentially good models for the local folding of structured loops in proteins. One obvious paradigm for the folding of these short loops is that the structure is dictated by a minimal strain principle, whereby conformational space is explored to generate a set of sterically allowed backbone dihedral angles that satisfy the bridge constraint. We compare the contryphan-R structure with those of other cyclic peptides and disulfide-bonded loops embedded in proteins and discuss the implications for the folding of these segments and how structures of these embedded loops may be useful more generally in the prediction of cyclic peptide structures.

contryphan-R	GCOW**E**PWC−NH_2
bromocontryphan	GCOW**E**P**W**C−NH_2
des-[Gly 1]contryphan-R	COW**E**PWC−NH_2
contryphan-P	GCOW**D**PWC−NH_2
Leu-contryphan-P	GC**VLL**PWC−**OH**
contryphan-Sm	GCOW**Q**PWC−NH_2

FIGURE 1: Sequences of the contryphan family with variations from the GCOWXPWC-NH_2 consensus sequence highlighted. Hydroxyproline is represented by O, D-handed residues are underlined, and the double-slashed W represents 6-bromotryptophan. Contryphan-R (*5*), bromocontryphan (*9*), and des-[Gly1]contryphan-R (*5*) are from *Conus radiatus*, Leu-contryphan-P (*10*) is from *Conus purpurascens*, and contryphan-Sm (*6*) is from *Conus stercusmuscarum*. The contryphan-P sequence has been deduced from cDNA cloned from *C. purpurascens* (*10*).

† This work was supported in part by NIH Grant GM 48677.

‡ The structures and NMR-derived constraints have been deposited in the Protein Data Bank under Accession Number 1qfb.

* Correspondence should be addressed to this author: E-mail ray.norton@molsci.csiro.au; FAX +61 3 9903 9655.

§ Biomolecular Research Institute.

|| University of Utah.

[1] Abbreviations: 1D, one-dimensional; 2D, two-dimensional; E-COSY, exclusive correlation spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; TOCSY, two-dimensional total correlation spectroscopy; RMS, root-mean-square.

10.1021/bi990685j CCC: $18.00 © 1999 American Chemical Society
Published on Web 08/12/1999

EXPERIMENTAL PROCEDURES

Structures were determined from 2D ^{1}H NMR spectra recorded on a Bruker DRX-600 spectrometer. The sample consisted of 4 mM synthetic contryphan-R in 90% H_2O/10% ^2H_2O at 5 °C and pH 3.0. We confirmed that the same NOE set was present at pH 5.2, where Glu5, the only titratable carboxyl group in the molecule, would be deprotonated and able to participate in ionic interactions. The methods of spectra collection and analysis and extraction of distance and angle constraints are as described previously (*11*). The TOCSY and NOESY spin−lock and mixing times were 70 and 300 ms, respectively. A series of 1D experiments was recorded over the temperature range 5−25 °C to determine the amide proton temperature coefficients, and a NOESY was run at 27 °C to resolve overlap with the water resonance. A pH titration and an amide exchange experiment were also carried out at 5 °C. Slow exchange for the amide protons of this short peptide was defined as maintenance of ≥50% peak height in a 1D spectrum measured 30 min after dissolution in ^2H_2O. A short mixing-time (50 ms) NOESY and an E-COSY (*12*) spectrum were acquired to enable the identification of stereospecific assignments and side-chain dihedral angle constraints (*13*). A $^{13}C-^1H$ heteronuclear multiple quantum coherence spectrum (*14*) was recorded to confirm the conformation of the prolyl peptide bonds via the chemical shifts of the C^β atoms (*15, 16*).

The final NMR restraint set contained 54 intraresidue, 32 sequential, 21 medium-range ($i - j < 5$) NOEs and one long-range ($i - j \geq 5$) NOE, as well as five backbone and four side-chain dihedral angle restraints. Structures were calculated as described previously (*11*), except that X-PLOR (*17*) was used throughout and the final energy minimization was performed in explicit solvent on structures without any charges neutralized. After simulated annealing in the X-PLOR distance geometry force field, the structures were placed in a cube of water (side length 35 Å) by use of Insight II (MSI, San Diego) and energy-minimized with periodic boundary conditions in the CHARMM-19 force field (*18*) in X-PLOR (see Table 2, protocol 4, for details and the other protocols for alternate strategies attempted). The 20 best structures and NMR-derived constraints have been deposited in the Protein Data Bank (*19*) (Accession Number 1qfb). PROCHECK-NMR (*20*) was used to characterize the quality of the structures.

The Protein Data Bank (*19*) (PDB) was searched for sequences matching the motif CXXXXXC by a computer program written in-house that interrogated a file containing all sequences and ID codes in the PDB. The large number of hits was reduced in redundancy on the basis of sequence and molecule name and the PDB header files were examined to determine if the two half-cystine residues did form a disulfide bond. The coordinates of the resulting chain segments were extracted from each molecule and overlaid in Insight II. Clustering was performed visually on C^α traces. Clusters of chains were then characterized via hydrogen-bond connectivities.

RESULTS AND DISCUSSION

NMR Spectroscopy. Two-dimensional ^{1}H NMR spectra were acquired on a 4 mM solution of synthetic contryphan-R in water at pH 3.0 and 278 K. The same NOE set was observed at pH 5.2. Published RP-HPLC data for this peptide show that separate peaks, eluting several minutes apart and having an intensity ratio of approximately 10:1, were due to interconverting forms of the pure peptide (*6, 10*). Two separate sets of resonances were visible in the NMR spectra with a ratio of approximately 8:1, consistent with the presence of two slowly interconverting species and corresponding well with the HPLC results. Full resonance assignments for the major conformer and most assignments for the minor conformer were made (see Supporting Information). The major and minor conformers are characterized by the conformation at the Cys2−Hyp3 peptide bond being cis and trans, respectively, as identified by the expected sequential NOEs ($d_{\alpha\alpha}$ and $d_{\alpha\delta}$ for cis and trans, respectively) and, in the case of the major conformer, the expected C^α and C^β chemical shifts (*15, 16*). The Glu5−Pro6 peptide bond is trans in both conformers. Interestingly, the Glu5 and Trp7 amide proton resonances broaden noticeably as the temperature is lowered from 20 to 5 °C (the peak heights almost halve relative to the other amide resonances), suggesting a further interconversion between conformers. It is unlikely that this is due to cis−trans isomerization at the Glu5−Pro6 peptide bond due to the slow interconversion expected for this high activation energy transition. Given the packing of Glu5 and Trp7 against the Pro6 ring, it is possible that this interconversion represents a flip between two different Pro6 ring pucker states (*21*), although conformational interconversion at the disulfide bridge (*22*) cannot be ruled out.

Structure Calculation. The distance constraints were extracted from a 300 ms NOESY spectrum and input to simulated annealing and energy minimization calculations in explicit solvent. The ϕ angles of Cys2, D-Trp4, Glu5, Trp7, and Cys8 were restrained within 30° of −120°, +120°, −120°, −120°, and −60°, respectively (noting that D-Trp4 has a positive ϕ angle as expected). The χ^1 angles of Cys2, D-Trp4, Glu5, and Trp7 were restrained within 45° of −60°, −60°, −60°, and +60°, respectively. The family of structures for the major conformer in Figure 2 shows that both the backbone and side chains are well-ordered. The structural characteristics summarized in Figure 3 and Table 1 show that the structures are well-defined and have good stereochemistry. PROCHECK-NMR identifies 87.5% of the residues in the 20 models to be in the most-favored regions of the Ramachandran plot (slightly nonideal positive ϕ-angle conformations of the unconstrained Gly1 residue comprising the remainder).

Description of the Structure. Contryphan-R consists of a seven-residue loop closed by the Cys2−Cys8 disulfide bridge (Figure 2). There are two chain reversals: a non-hydrogen-bonded reversal from Gly1 to Glu5 and a type I β-turn from Glu5 to Cys8 stabilized by a hydrogen bond from the slowly exchanging Cys8 NH to Glu5 carbonyl. The cis Cys2−Hyp3 peptide bond and the D-Trp4 presumably play a role in generating the first chain reversal. The distance between the potential salt bridge partners, the Glu5 carboxyl and the N-terminal ammonium group, was 9.7 ± 1.5 Å, suggesting that a salt bridge was not formed at pH 3.0, as expected. At pH 5.2, where Glu5 would be deprotonated (see below), an identical set of NOEs was observed, indicating that the structure is unaffected by pH. There are two hydrophobic surface patches: Hyp3−D-Trp4 and Pro6−Trp7, together

FIGURE 2: Stereoview of the family of 20 NMR structures energy minimized in a box of water treated with periodic boundary conditions and superimposed over the backbone heavy atoms (N, C^{α}, and C). The side chains and amide protons are in lighter shading. This figure was produced with Insight II.

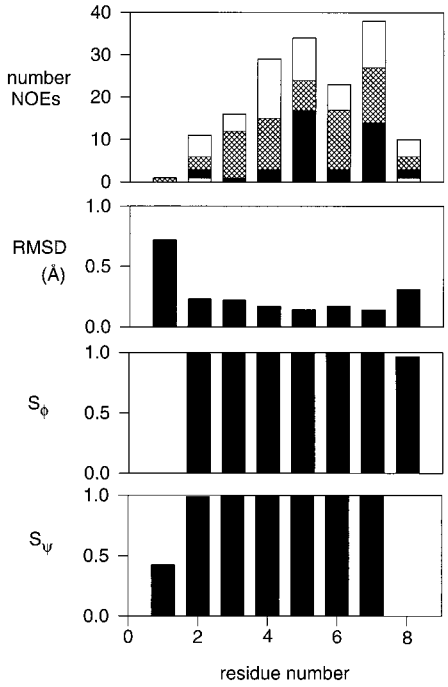

FIGURE 3: Number of NOEs, average RMS deviation from the mean structure, and ϕ and ψ angular order parameters (*23*, *24*) vs residue. The numbers of NOEs are indicated for the single long-range ($i - j \geq 5$, gray), 21 medium-range ($1 < i - j < 5$, black), 32 sequential (cross-hatched), and 54 intraresidue (open) connectivities.

with the hydrophobic component of the Glu5 side chain, both containing a (hydroxy)proline−tryptophan sequential pair. It is likely that the cis−trans isomerization at Cys2−Hyp3 is responsible for further exposing the first hydrophobic patch and thus yielding a later-eluting peak on the RP-HPLC for the minor conformer (*6*, *10*). The ring packing may be structurally important, as proposed for the linear sequence SYPFDV, which also includes neighboring proline and aromatic residues (*25*).

Status of the Putative Salt Bridge. The NMR structures discussed above were subjected to restrained energy minimization in a box of explicit water molecules with periodic boundary conditions. Table 2 summarizes the effect of different solvent models and protocols on the presence of a salt bridge between the carboxylate of Glu5 and the N-terminal ammonium group and a backbone hydrogen bond from the amide of Glu5 to the carbonyl of Gly1. Calculations were carried out in the presence and absence of side-chain dihedral angle restraints. The salt bridge only formed reliably during restrained energy minimization in vacuo (a commonly used method in refinement of NMR structures) without χ angle restraints and occasionally with the χ angle restraints included (protocol 2), and the hydrogen bond formed cooperatively with the salt bridge under these conditions. The amide proton of Glu5 was found to be slowly exchanging and to have a low-temperature coefficient, which appears to support the in vacuo calculation. However, this amide proton is almost buried in the structures minimized with explicit solvent and this could account for the exchange behavior. Under the most "realistic" conditions, with explicit solvent (protocol 4, the final NMR family of structures), neither the salt bridge nor the hydrogen bond formed, although the interresidue distance did decrease slightly after restrained molecular dynamics simulations (protocols 5 and 6), during which the side-chain dihedral angle restraints were discarded to explore conformational space more fully.

To determine experimentally whether a salt bridge existed between Glu5 and the N-terminus, we carried out a pH titration on contryphan-R. The pK_a values of the Glu5 carboxyl and the N-terminal amino group were 4.5 [compared with a random coil value of 4.2−4.3 (*26*)] and 8.8 [random coil value 8.1−8.2 (*26*)], respectively. A reduced carboxyl pK_a and elevated ammonium pK_a would have been consistent with a salt bridge. The observation that the Glu5 carboxyl pK_a was higher than the value in unstructured small peptides argues against its participation in a salt bridge. Furthermore, ionization of this carboxyl had no effect on the chemical shift of the $C^{\alpha}H$ resonance of Gly1. The elevated pK_a for the N-terminal ammonium group is consistent with its interaction with another group in contryphan-R, but it is not clear which. The addition of 400 mM NaCl at pH 6.5 did not significantly perturb the chemical shifts of any resonances in the molecule.

We believe that the NMR structures generated by protocol 4 are an accurate representation of the contryphan-R structure and that a salt bridge does not occur between Glu5 and the

Table 1: Structural Characteristics for the 20 Structures of Contryphan-R from X-PLOR[a]

RMS deviations from experimental distance restraints (Å) (108)[b]	0.029 ± 0.003
RMS deviations from experimental dihedral restraints (deg) (9)[b]	0.21 ± 0.18
RMS deviations from idealized geometry	
bonds (Å)	0.015 ± 0.001
angles (deg)	3.90 ± 0.06
impropers (deg)	0.8 ± 0.1
energies (kcal mol^{-1})[c]	
E_{NOE}	5.5 ± 0.9
E_{cdih}	0.05 ± 0.07
$E_{bond} + E_{angle} + E_{improper}$	69 ± 4
E_{L-J}	-11 ± 2
E_{elec}	-191 ± 8
pairwise RMS differences (Å)	
backbone (N, C$^{\alpha}$, C)	0.47 ± 0.15
all heavy	0.70 ± 0.14

[a] The best 20 structures after solvated energy minimization in the CHARMM-19 force field. Values are mean ± standard deviation. [b] The number of restraints is shown in parentheses. [c] Force constants for calculation of the square-well potentials for the NOE and dihedral angle restraints were 50 kcal mol^{-1} Å^{-1} and 200 kcal mol^{-1} rad^{-2}, respectively.

Table 2: Gly1−Glu5 Salt Bridge and Hydrogen-Bond Distances for Different Molecular Mechanics Protocols

	distances over family (Å) ± standard deviation with (without) χ^1 dihedral restraints			
	salt bridge[a] $NH_3^+(1)-COO^-(5)$		hydrogen bond[b] NH(5)−CO(1)	
protocol				
(1) initial distance geometry[c]	9.4 ± 1.0	(8.5 ± 1.0)	4.6 ± 0.3	(4.5 ± 0.4)
(2) EM[d] in vacuo with distance-dependent dielectric[e]	7.5 ± 2.5	(4.4 ± 1.6)	4.1 ± 0.8	(2.8 ± 0.9)
(3) EM neutralized[f]	9.6 ± 1.0	(7.8 ± 1.1)	4.8 ± 0.2	(4.3 ± 0.7)
(4) EM in explicit solvent[g] (the NMR family)	9.7 ± 1.5	(8.4 ± 0.7)	4.9 ± 0.5	(5.0 ± 0.5)
(5) 20 ps MD[h] and EM in explicit solvent	−	(7.2 ± 1.1)	−	(5.0 ± 0.6)
(6) 100 ps MD and EM in explicit solvent	−	(7.0 ± 1.6)	−	(5.0 ± 0.8)

[a] The salt bridge was characterized by the Gly1 N−Glu5 C$^{\delta}$ distance. A distance of ca. 4 Å would indicate a well-defined salt bridge. [b] The hydrogen bond was characterized by the Glu5 NH−Gly1 O distance. A distance of ca. 2.5 Å would indicate a well-defined hydrogen bond. [c] Simulated annealing in the X-PLOR distance geometry force field during structure generation in which there are no electrostatic or Lennard-Jones terms. [d] Restrained energy minimization was performed in the CHARMM-19 force field with the Powell algorithm and nonbonded cutoffs of 10 Å. [e] The in vacuo calculations in CHARMM were performed with a distance-dependent dielectric. [f] The formal charges were neutralized as described previously (*11*). [g] The TIP3 water model was employed with periodic boundary conditions and a constant dielectric constant of 1.0. [h] Restrained molecular dynamics simulations were performed in the CHARMM-19 force field with the Verlet algorithm, at 300 K and with a 1 fs time step.

N-terminus. The carboxyl pK_a value, the absence of chemical shift effects on the putative interacting partner, and the lack of effect of salt are all consistent with this. Furthermore, two of the additional contryphans isolated from related cone shells with similar physiological functions and potencies to contryphan-R have sequences with single substitutions at Glu5 by Asp (unpublished results) and Gln (*10*), respectively (Figure 1). The shorter side chain of Asp would be less likely to form a salt bridge and the Gln homologue would clearly be unable to do so, although it could form a hydrogen bond. Nevertheless, our calculations highlight the potential for oppositely charged functional groups to seek each other out during energy minimization in vacuo, even when a distance-dependent dielectric constant is used. The fact that the NMR-derived restraints were included in the minimization step did not prevent this occurring because of the lack of experimental restraints involving the relevant groups.

Structural Determinants. Six of the eight contryphan-R residues play clear or putative structural roles: the two half-cystine residues, the Hyp and Pro residues, the D-Trp4, and the hydrophobic Trp7. It is becoming apparent that the structures of small cyclic peptides are largely governed by such determinants. Wermuth et al. (*8*) have demonstrated that the structures of a series of cyclic pentapeptide homologues are dictated by the position of Gly and D-handed residues relative to the non-structurally significant components of the sequence. This is expected since the structures of small cyclic peptides typically have very few side-chain interactions and hence the backbone conformation is determined rather by bridging interactions, the number of constrained residues, and the sequence of residue types: L-handed, Gly/D-handed, or Pro. L-Handed residues have negative ϕ angle conformations, prolines have well-defined structural roles via essentially fixed negative ϕ angles, and Gly or D-handed amino acids favor positive ϕ angle conformations.

On the basis of these considerations, one might predict that the contryphan scaffold could withstand substitutions at Gly1 (outside the constrained loop), Trp4 (provided D-handedness is maintained), and Glu5, and that the remainder of the residues would be largely conserved. The family of known sequences partially supports this reasoning, with the two Cys residues, the hydrophobicity and D-handedness at position 4, Pro6, and Trp7 being conserved (Figure 1). In addition, Gly1, Hyp3, and Trp4 are largely conserved. It is likely that Trp7 and the hydophobicity at positions 3 and 4 are structurally or functionally important. It is evident from Leu-contryphan-P that Hyp3 and D-Trp4 can be substituted, although this results in a reduction in potency in mice, a modification of function in fish, and as expected, the absence of multiple conformers (*10*). As discussed above, the lack of conservation of Glu5 (even becoming neutral in contry-

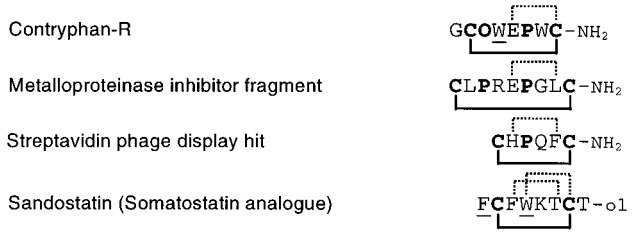

FIGURE 4: Sequences of contryphan-R and three disulfide-bridged peptides (*1*, *2*, *27*, *28*) of similar size but unrelated origin. Half-cystine and (hydroxy)proline residues are highlighted. The dotted lines represent $i/i-3$ hydrogen bonds identified by the respective structure determinations, -ol represents a C-terminus reduced to an alcohol, and the remainder of the nomenclature is as in Figure 1.

phan-Sm) suggests that the potential salt bridge between Glu5 and the N-terminal amino group does not occur or is not important structurally.

Comparison with Other Cyclic Peptides. In Figure 4, three cyclic peptides with sequences related to that of contryphan-R but of unrelated origin are shown. The metalloproteinase inhibitor sequence (*1*) bears some resemblence to that of contryphan-R in having two Pro residues separated by two residues. When it is aligned with the Hyp and Pro of contryphan-R there is an additional residue on either side of the $P_1X_2X_3P_4X_5$ motif compared with the contryphan-R sequence. Structurally, both molecules have hydrogen-bonded β-turns from the NH of X_6 (two residues down from P_4, which is a Cys in contryphan-R) to the carbonyl oxygen of X_3 (a Glu in both contryphan-R and the metalloproteinase inhibitor) with a Pro (P_4) at position 2 of the turn. The overall structures are similar, with a non-hydrogen-bonded chain reversal (involving P_1) evident in the N-terminal half of both molecules. Interestingly, at least three conformations were found for this peptide, with the major form being trans-trans and the other two forms being cis−trans and trans−cis at the two prolyl peptide bonds (ratio 6:3:1) (*1*). The isomerization at the second Pro was found to be faster, coalescence with the trans−trans form being demonstrated at 70 °C.

The hexapeptide in Figure 4, a hit in a phage display search for streptavidin ligands (*2*), also bears some resemblance in sequence and structure to both contryphan-R and the metalloproteinase inhibitor, having a similar β-turn with a Pro at position 2 of the turn. The occurrence of this proline-stabilized β-turn, even in an artificial selection program, suggests that this turn type may be particularly beneficial for the generation of stability in short, disulfide-bridged polypeptides.

Sandostatin, a somatostatin analogue (Figure 4), has the same loop size as the above-mentioned hexapeptide. Although it lacks a proline, it does contain an engineered D-Trp. The X-ray crystal structure (*27*) has three sandostatin molecules in the asymmetric unit, the first having a β-hairpin type structure stabilized by a Phe3−Thr6 backbone hydrogen bond and the second and third having helical-like structure with a D-Trp4−Cys7 backbone hydrogen bond in addition to the Phe3−Thr6 hydrogen bond. The NMR data appear to be consistent with an equilibrium between these two structures (*28*). In both forms the D-Trp is at position 2 of the turn stabilized by the Phe3−Thr6 hydrogen bond, compared with D-Trp4 at position 3 of the non-hydrogen-bonded chain reversal, stabilized by the cis Hyp3−D-Trp4 combination in contryphan-R.

Comparison with Loops Embedded in Proteins. A search of 8306 protein sequences in the Protein Data Bank (*19*) revealed 903 protein chains containing a contryphan-R-like $C_1X_2X_3X_4X_5X_6C_7$ motif. After crystal symmetry, molecule name, sequence, and structural redundancies were eliminated and the correct half-cystine pairings were established, a total of 36 CXXXXXC chains were extracted from the coordinate files and clustered visually, based on a C^α trace overlay. Sixteen of these chains formed four obvious clusters, as shown in Figure 5, the other structures being scattered in conformations that did not cluster. The four clusters were characterized in terms of backbone hydrogen bonding, cluster 1 having a C_7 NH−X_4 O bond (with a corresponding β-turn, frequently type I) and a X_4 NH−C_1 O bond (with a corresponding β-turn). Cluster 2 has an X_6 NH−X_3 O bond (with a corresponding β-turn). Cluster 3 has a pair of hydrogen bonds from X_6 NH to X_2 O and from X_2 NH to X_6 O, forming a β-bridge (members of cluster 3 contain the

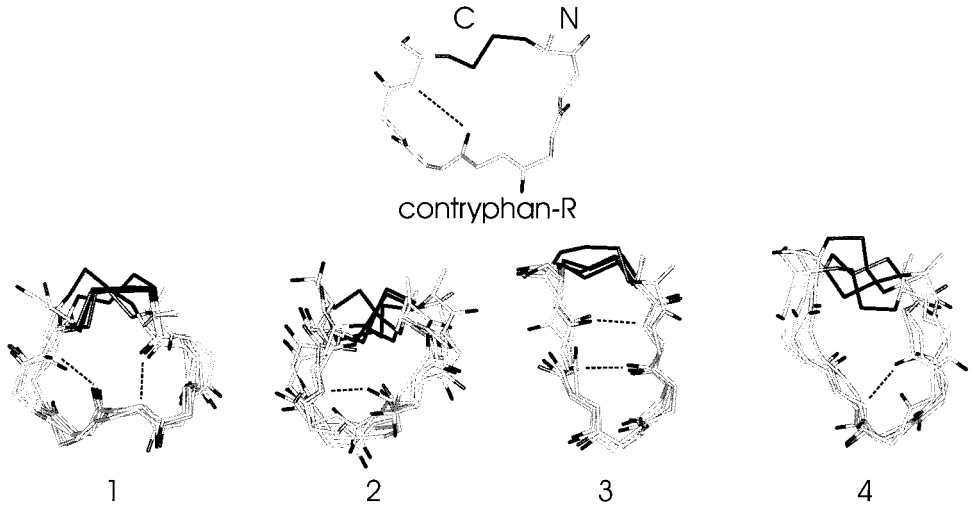

FIGURE 5: Stick representations of contryphan-R and four clusters of disulfide-bonded loops from proteins. The disulfide bridges are in dark shading, as are the carbonyl oxygen atoms. The N-termini of all the chains are at top right. The consensus hydrogen bonds discussed in the text are indicated by dotted lines. Cluster 1 consists of PDB depositions 1brv, 1akg, 1bg4, and 4cox; cluster 2 of 1byv, 4aah, 1cel, 1eit, and 1auk; cluster 3 of 1rmg, 1eg1, 1eth, and 1ovw, and cluster 4 of 1wkt, 1a9v, 1wba, and 1bcp. This figure was produced with Insight II.

chain reversal of a β-sheet hairpin structure). Cluster 4 is characterized by a X_5 NH$-X_2$ O bond (with a corresponding β-turn). Of the unclustered loops, 40% contain single $i/i-4$ hydrogen bonds, 20% have single $i/i-3$ hydrogen bonds, and 40% contain no intracycle hydrogen bonds.

The Cys2−Cys8 segment of contryphan-R is quite similar in conformation to that of the members of cluster 1 (Figure 5), sharing the C_7 NH$-X_4$ O hydrogen bond (using the CXXXXXC motif numbering). Contryphan-R does not have an X_4 NH$-C_1$ O hydrogen bond, having instead a cis Hyp/D-amino acid-induced turn at this equivalent position. The average backbone (N, C^α, C) RMS deviation of contryphan-R from the members of cluster 1 is 1.59 ± 0.18 Å. Corresponding RMS deviations from clusters 2, 3, and 4 are 2.25 ± 0.30, 2.66 ± 0.02, and 2.21 ± 0.25 Å, respectively. The members of cluster 1 include two enzymes, an α-conotoxin (otherwise unrelated to contryphan-R) and a fragment of a viral protein (Figure 5). There is no sequence similarity within cluster 1, including contryphan-R, except for a Gly in the viral fragment at X_2 (D-Trp in contryphan-R) and a Pro in the α-conotoxin at X_4 (Pro in contryphan-R). Notable sequence similarities in the other clusters are restricted to three out of four members of cluster 4 having a Pro at X_1 and two out of four members of cluster 4 having a Pro at X_5.

It is interesting to consider the determinants of local folding of these constrained segments. The sample size is not large enough to reveal subtle sequence effects, and it is also likely that tertiary interactions with other parts of the intact protein are important in dictating the local fold. Nevertheless, at a coarse level it is apparent that the conformations that occur more frequently in the PDB (i.e., the members of clusters 1−4) are all stabilized by at least one intracycle hydrogen bond. Indeed, the four clusters include all four possible β-turn-forming $i/i-3$ hydrogen bond possibilities. It is interesting that these highly populated clusters are stabilized predominantly by $i/i-3$ hydrogen bonding, as opposed to either no hydrogen bonding or $i/i-4$ hydrogen bonding, which occurs in 80% of the unclustered loops. The large subloop size generated by the $i/i-4$ hydrogen-bonded loop and the large number of conformations available to loops without intracycle hydrogen bonds are probably responsible for the lack of clustering in these cases. It is also interesting that contryphan-R has a similar conformation to one of the most highly populated clusters from these segments extracted from proteins. It seems reasonable to expect isolated cyclic peptides to adopt the highly populated conformations of segments with the same gap size embedded in proteins, especially those with intracycle hydrogen bonding, and this suggests a role for the PDB in cyclic peptide structure prediction. A useful approach for the prediction of cyclic peptide structures from sequence might be the use of a specifically designed scoring function or force field (*29*) to rank clustered options generated from loop searches of the PDB.

CONCLUSIONS

Contryphan-R is a structured peptide with a type I β-turn, although it does exist in at least two forms due to cis−trans isomerization at the Cys2−Pro3 peptide bond. The structure should be sufficiently robust to withstand substitution at several sites and should prove interesting as a new scaffold for molecular design. Contryphan-R appears to be a useful test case for probing solvent models and computational protocols. In particular, the distance-dependent dielectric solvent model does not, in this case, reproduce the results of the explicit solvent model. This may have general consequences for the conformations of surface residues refined in this way since the NMR data are often sparse at the molecular surface, partly due to genuine surface mobility. It is interesting that contryphan-R is most similar to one of the most populated clusters of protein-embedded loops and this suggests a role for the PDB in cyclic peptide structure prediction.

ACKNOWLEDGMENT

We thank Richard Jacobsen for assistance with peptide synthesis.

SUPPORTING INFORMATION AVAILABLE

Two tables of ^{1}H chemical shifts for the major and minor conformers of contryphan-R under the solution conditions used for the structure determination. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Francart, C., Wieruszeski, J.-M., Tartar, A., and Lippens, G. (1996) *J. Am. Chem. Soc. 118*, 7019−7027.
2. Katz, B. A. (1995) *Biochemistry 34*, 15421−15429.
3. Gibbs, A. C., Kondejewski, L. H., Gronwald, W., Nip, A. M., Hodges, R. S., Sykes, B. D., and Wishart, D. S. (1998) *Nat. Struct. Biol. 5*, 284−288.
4. Kieber-Emmons, T., Murali, R., and Greene, M. I. (1997) *Curr. Opin. Biotechnol. 8*, 435−441.
5. Jimenez, E. C., Olivera, B. M., Gray, W. R., and Cruz, L. J. (1996) *J. Biol. Chem. 271*, 28002−28005.
6. Jacobsen, R., Jimenez, E. C., Grilley, M., Watkins, M., Hillyard, D., Cruz, L. J., and Olivera, B. M. (1998) *J. Pept. Res. 51*, 173−179.
7. Shikata, Y., Watanabe, T., Teramoto, T., Inoue, A., Kawakami, Y., Nishizawa, Y., Katayama, K., and Kuwada, M. (1995) *J. Biol. Chem. 270*, 16719−16723.
8. Wermuth, J., Goodman, S. L., Jonczyk, A., and Kessler, H. (1997) *J. Am. Chem. Soc. 119*, 1328−1335.
9. Jimenez, E. C., Craig, A. G., Watkins, M., Hillyard, D. R., Gray, W. R., Gulyas, J., Rivier, J. E., Cruz, L. J., and Olivera, B. M. (1997) *Biochemistry 36*, 989−994.
10. Jacobsen, R. B., Jimenez, E. C., De la Cruz, R. G. C., Gray, W. R., Cruz, L. J., and Olivera, B. M (1999) *J. Pept. Res.* (in press).
11. Pallaghy, P. K., Scanlon, M. J., Monks, S. A., and Norton, R. S. (1995) *Biochemistry 34*, 3782−3794.
12. Griesinger, C., Sørenson, O., and Ernst, R. R. (1987) *J. Magn. Reson. 75*, 474−492.
13. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gõ, N., and Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611−639.
14. Bax, A., Griffey, R. H., and Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301−315.
15. Dorman, D. E., and Bovey, F. A. (1973) *J. Org. Chem. 38*, 2379−2383.
16. Wüthrich, K. (1976) *NMR in biological research: peptides and proteins*, North-Holland, pp 275−278, Amsterdam.
17. Brünger, A. T. (1992) X-PLOR Version 3.1: A system for X-ray crystallography and NMR, Yale University, New Haven, CT.
18. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. J. (1983) *Comput. Chem. 4*, 187−217.

19. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
20. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
21. Milner-White, E. J., Bell, L. H., and MacCallum, P. H. (1992) *J. Mol. Biol. 228*, 725−734.
22. Otting, G., Liepinsh, E., and Wüthrich, K. (1993) *Biochemistry 32*, 3571−3582.
23. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.
24. Pallaghy, P. K., Duggan, B. M., Pennington, M. W., and Norton, R. S. (1993) *J. Mol. Biol. 234*, 405−420.
25. Yao, J., Dyson, H. J., and Wright, P. E. (1994) *J. Mol. Biol. 243*, 754−766.
26. Gooley, P. R., Blunt, J. W., Beress, L., and Norton, R. S. (1988) *Biopolymers 27*, 1143−1157.
27. Pohl, E., Heine, A., Sheldrick, G. M., Dauter, Z., Wilson, K. S., Kallen, J., Huber, W., and Pfaffli, P. J. (1995) *Acta Crystallogr. D51*, 48−59.
28. Melacini, G., Zhu, Q., and Goodman, M. (1997) *Biochemistry 36*, 1233−1241.
29. Keasar, C., and Rosenfeld, R. (1998) *Folding Des. 3*, 379−388.

BI990685J