Protein Engineering vol.6 no.8 pp.873–882, 1993

Modelling multiple disulphide loop containing polypeptides by random conformation generation. The test cases of α-conotoxin GI and endothelin I

R.Sowdhamini[1], C.Ramakrishnan and P.Balaram[2]

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560 012 India

[1]Present address: ICRF unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London, WC1E 7HX, UK

[2]To whom correspondence should be addressed

A general procedure for arriving at 3-D models of disulphide-rich polypeptide systems based on the covalent cross-link constraints has been developed. The procedure, which has been coded as a computer program, RANMOD, assigns a large number of random, permitted backbone conformations to the polypeptide and identifies stereochemically acceptable structures as plausible models based on strainless disulphide bridge modelling. *Disulphide bond modelling* is performed using the procedure MODIP developed earlier, in connection with the choice of suitable sites where disulphide bonds could be engineered in proteins (Sowdhamini,R., Srinivasan,N., Shoichet,B., Santi,D.V., Ramakrishnan,C. and Balaram,P. (1989) *Protein Engng*, 3, 95–103). The method RANMOD has been tested on small disulphide loops and the structures compared against preferred backbone conformations derived from an analysis of putative disulphide subdatabase and model calculations. RANMOD has been applied to disulphide-rich peptides and found to give rise to several stereochemically acceptable structures. The results obtained on the modelling of two test cases, α-conotoxin GI and endothelin I, are presented. Available NMR data suggest that such small systems exhibit conformational heterogeneity in solution. Hence, this approach for obtaining several distinct models is particularly attractive for the study of conformational excursions.

Key words: α-conotoxin GI/disulphide bond modelling/disulphide-rich polypeptide systems/endothelin 1

Introduction

Disulphide bonds act as an important covalent constraint in limiting the range of accessible backbone conformations in peptides and proteins. The number of disulphide bonds present in single polypeptide chains of proteins and peptides varies from 0 to as many as 12 (Klapper and Klapper, 1980). The spatial arrangement of multiple disulphide bonds in proteins has generated a considerable amount of interest and attention in the past (Meirovitch and Scheraga, 1981; Kikuchi *et al.*, 1986, 1989; Mao, 1989). Two-disulphide-bonded systems can have only three possible molecular topologies, as shown in Figure 1a. Polypeptide segments with three disulphide bonds thus can exist in 15 possible arrangements. The 15 possibilities are represented schematically in Figure 1b. Natural examples exist for all of these 15 types of arrangements (Warne and Lawskowski, 1990). In two-disulphide-bonded systems, the two covalent loops may be far apart in the primary sequence and, thus, remain independent of each other (type I). In some cases, one disulphide bond may have a smaller loop size than the other and that polypeptide stretch may be a subset of the bigger loop (type II in Figure 1a). Such an arrangement will be referred to as 'loop within loop'. On the other hand, there may be cases, where the two loops share a common polypeptide segment (type III in Figure 1a). Such loops are termed as 'interlocked loops'. Following the same convention, in the case of three-disulphide-bonded systems (Figure 1b), type I has three independent loops, types III, VII, XIII and XV have loop-within-loops and types II, IV, V, VIII and IX have both categories.

In multiple-disulphide-bonded systems, the stereochemical restraints can be considerable, especially when loop sizes are relatively small and the loops share a segment of the polypeptide chain. Many highly biologically active peptides such as endothelins (Yanagisawa *et al.*, 1988), conotoxins (Olivera *et al.*, 1985) and *Escherichia coli* stable enterotoxins (Shimonishi *et al.*, 1987) contain multiple disulphide bonds. Despite intense interest in their physiological activities, crystal structures are unavailable in most cases, although several conformational studies in solution have been reported (Gariepy *et al.*, 1986; Endo *et al.*, 1989; Kobayashi *et al.*, 1989; Pardi *et al.*, 1989; Saudek *et al.*, 1989, 1991; Aumelas *et al.*, 1991; Krystek *et al.*, 1991; Ozaki *et al.*, 1991b; Reily and Dunbar, 1991; Tamaoki *et al.*, 1991). A recent report on the high resolution crystal structure of enterotoxin ST1b provides a good example of the experimentally determined 3-D structure of a multiple-disulphide-bonded peptide (Ozaki *et al.*, 1991a). These systems provide attractive test cases for detailed 3-D modelling, which may then be extended to larger systems. This paper outlines a general approach to generating stereochemically acceptable structures for multiple-disulphide-bonded polypeptides and illustrates the application to two disulphide-rich bioactive peptides, endothelin I (Yanagisawa *et al.*, 1988) and α-conotoxin GI (Nishiuchi and Sakakibara, 1982) (Figure 2). The procedure is based on random conformation generation coupled with stereochemical modelling of S–S cross-links, using a strategy developed earlier in connection with engineering disulphide bridges into proteins (Sowdhamini *et al.*, 1989).

Methods

Generation of random conformations to polypeptide backbone

Random numbers were generated by the method of L'ecuyer (1988). These were then used to assign random values to each of the backbone torsion angles, ϕ and ψ, at each of the amino acid residues. Assignment of backbone torsions was restricted to sterically permitted regions of the (ϕ, ψ) plane. For this purpose, amino acid residues were divided into three categories: (i) non-Gly–non-Pro residues, (ii) Gly residues and (iii) Pro residues. The sterically permitted backbone conformations were chosen from the regions defined in Figure 3, for the three separate cases.

Non-Gly–non-Pro residues. Analysis of non-Gly–non-Pro residues in protein crystal structures shows that ~96.7% of the observed examples fall within the permitted regions of the

Ramachandran map (Morris *et al.*, 1992). In modelling such residues it is sufficient to choose starting (ϕ, ψ) values from the permitted regions as defined by the following three regions (see Figure 3a): (i) region 1, $-180° < \phi < -40°$ and $-70° < \psi < +180°$, (ii) region 2, $-180° < \phi < -40°$ and $-180° < \psi < -160°$ and (iii) region 3, $45° < \phi < 65°$ and $20° < \psi < 100°$. Region 1 comprises of the α_R, extended and bridge conformation, while region 2 corresponds to the partially permitted extended conformation and region 3 to the α_L conformation.

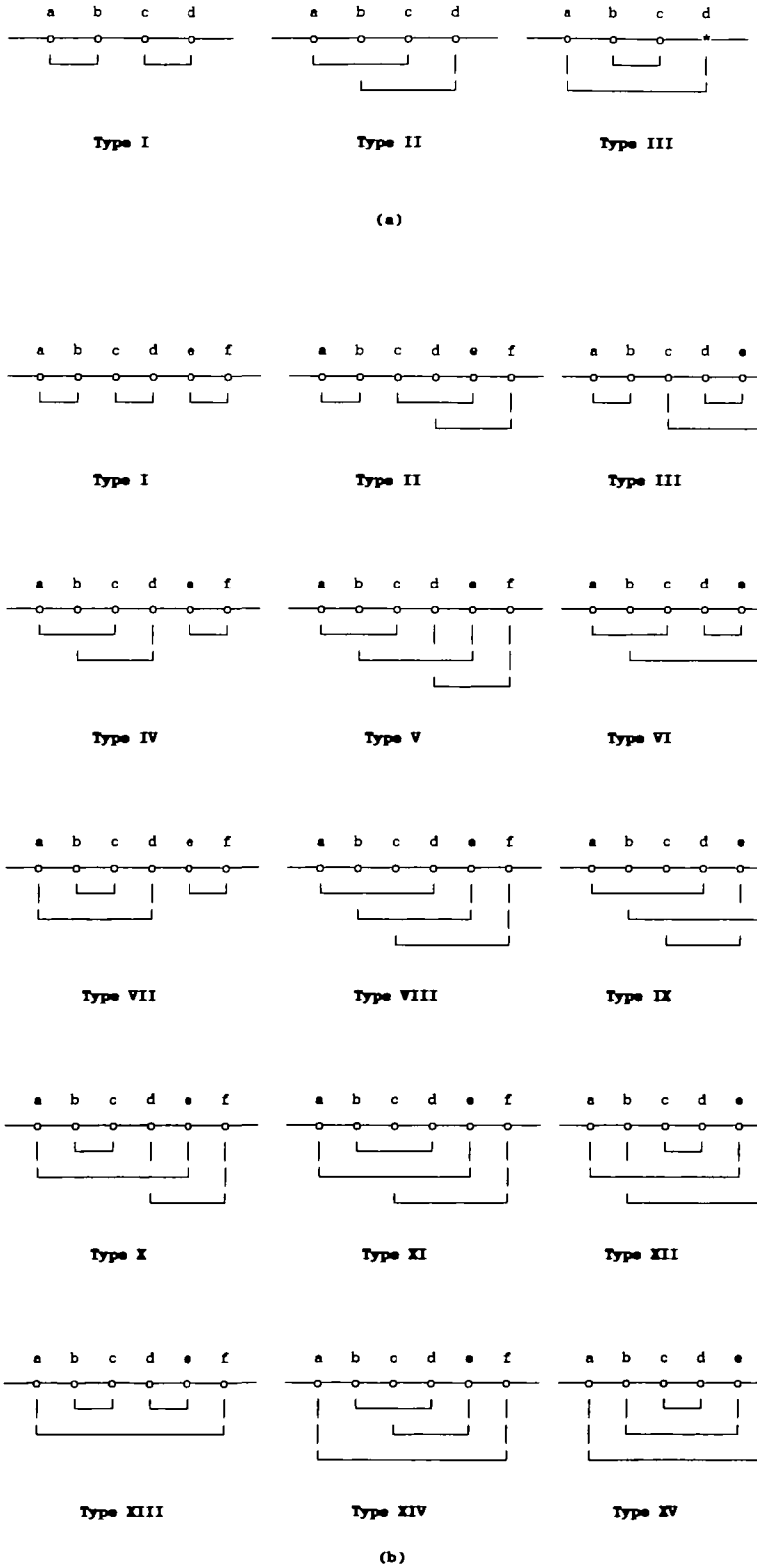

Fig. 1. Convention used for describing loop topologies of (a) two- and (b) three-disulphide-bonded systems. The polypeptide chain is represented as a horizontal line, ○ refers to Cys residues, while vertical lines indicate disulphide bond connectivity. Cys residues are marked as a–f, from the N-terminus.

Gly residues. Gly residues have the greatest conformational variability and can adopt both negative and positive (ϕ, ψ) values. The distribution of Gly residues in proteins has been analysed (Ramakrishnan and Srinivasan, 1990). Two regions were used for backbone conformation assignment for Gly residues (Figure 3b): (i) region 1, $-180° < \phi < -40°$ and $-180° < \psi < +180°$ and (ii) region 2, $+40° < \phi < +180°$ and $-180° < \psi < +180°$.

Pro residues. The amino acid proline has a high degree of restriction in ϕ, due to the presence of a pyrrolidine ring. The distribution of Pro residues in protein structures has been studied (MacArthur and Thornton, 1991). Two regions were chosen for the Pro residues (Figure 3c): (i) region 1, $-90° < \phi < -30°$ and $-70° < \psi < +180°$ and (ii) region 2, $-90° < \phi < -30°$ and $-180° < \psi < -160°$. Polypeptide chain conformations were then generated by random choice of (ϕ, ψ) values from within the limited regions of (ϕ, ψ) space (Figure 3). This procedure ensures that all conformations thus chosen have acceptable local stereochemistry.

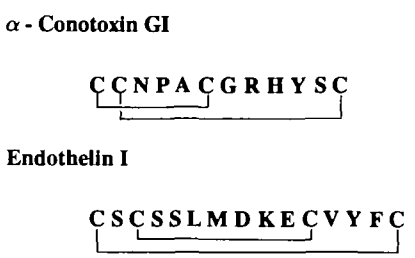

Fig. 2. Peptides chosen for modelling. The amino acid sequence in the disulphide-rich region is shown at the right. A single-letter code has been used to indicate the amino acid residues. The disulphide bond connectivity is shown by vertical lines.

One hundred (ϕ, ψ) assignments of backbone conformations by the random generation procedure for the three distinct categories are shown in Figure 3a–c. The uniform spread of (ϕ, ψ) points, clearly demonstrates that this procedure is efficient in assigning random (ϕ, ψ) values within the limits imposed.

Modelling and choice of disulphide bonds of best stereochemistry

A detailed analysis of native disulphide bonds revealed that for a disulphide bridge between residues i and j, the inter-Cα distances ($r\alpha_{ij}$) and the inter-Cβ distances ($r\beta_{ij}$) fall within the ranges of 3.8–6.8 and 3.5–4.5 Å respectively (Srinivasan *et al.*, 1990). For the present procedure, broader ranges of 3.5–7.5 and 3.3–5.0 Å for $r\alpha_{ij}$ and $r\beta_{ij}$ respectively, were chosen to identify suitable conformations. If the two distances are within limits at all the bridge positions, sulphur atoms were modelled at the two residues using the procedure MODIP (Sowdhamini *et al.*, 1989). Occasionally, even though the two Cys residues are within permissible distance limits, geometric fixing of the sulphur atoms in a strain-free manner may not be possible. If sulphur fixing is possible by MODIP at residues i and j, it normally leads to $2 \times 2 = 4$ possible conformations of the S–S bond between residues i and j.

The bridge conformations are assessed individually for their stereochemical quality. For a disulphide bridge connecting residues i and j, r_{S-S} refers to the S_i-S_j bond length, while χ_{S-S}, χ^1_i and χ^1_j refer to the three disulphide torsions, $C^\alpha_i-S_i-S_j-C^\alpha_j$, $N_i-C^\beta_i-C^\beta_i-S_i$ and $N_j-C^\alpha_j-C^\beta_j-S_j$ respectively. The limits for these parameters in stereochemical grading have been arrived at by previous analysis on native protein disulphides (Srinivasan *et al.*, 1990) and have also been described by Sowdhamini *et al.* (1989). The limits chosen are (i) r_{S-S} 2.0 ± 0.4 Å, (ii) χ_{S-S} ± 90 ± 30° and (iii) χ^1_i and χ^1_j ± 60 ± 30° or ± 180 ± 30°. Disulphides which fall

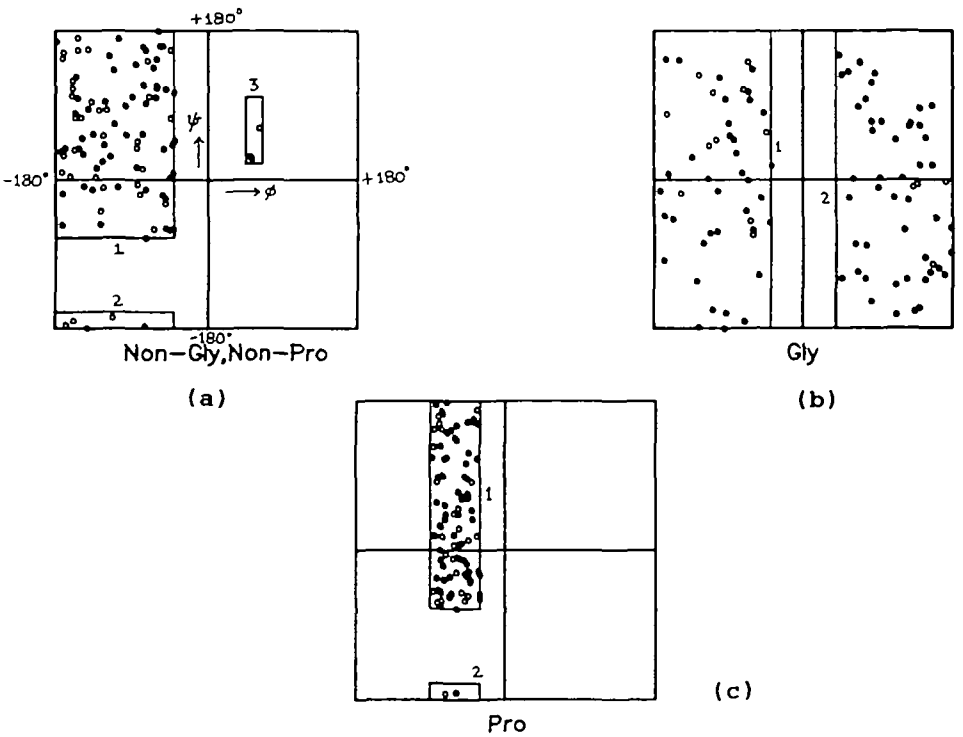

Fig. 3. The assignment of random conformations to a polypeptide. Amino acid residues have been grouped into three categories: (a) non-Gly–non-Pro residues, (b) Gly residues, (c) Pro residues. One hundred random assignments have been marked for each case. Regions within the boxes indicate the areas chosen for assignment. Good sampling is evident from a uniform spread of (ϕ, ψ) points.

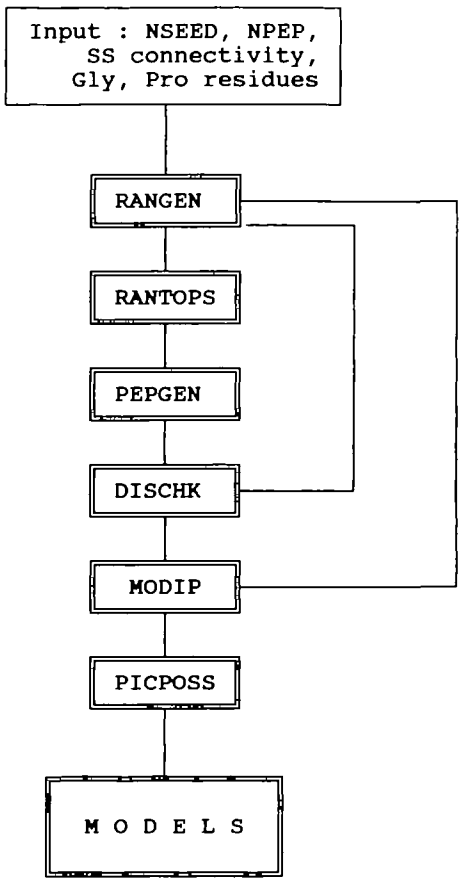

Fig. 4. Flow chart of the 'RANMOD' procedure used for modelling polypeptides rich in disulphide bonds. The required inputs include an integer for defining the seed number for random-number generation, NSEED, the number of residues in the peptide of interest, NPEP and information about the disulphide bond connectivity and the location of Gly and Pro residues (if any). The subroutine RANGEN generates random numbers while RANTOPS converts the random numbers to the (ϕ, ψ) scale. Assignment of the backbone torsion angles at each residue of the peptide is done within the 'permitted' (ϕ, ψ) space depending on the nature of the residue as shown in Figure 3. PEPGEN generates the conformation of the peptide corresponding to the random backbone conformation assignment. The two distance constraints corresponding to $C^{\alpha}-C^{\alpha}$ and $C^{\beta}-C^{\beta}$ at every bridge position are applied by the subroutine DISCHK. Those conformations which conform to the disulphide distance constraints are examined for stereochemical suitability by the subroutine MODIP. PICPOS selects best positions for the disulphide bridges modelled by MODIP.

within these ranges are assigned the highest grade of 'A'. Disulphides with a distortion observed at the two torsions χ^1_i or χ^1_j only (values outside the limits) are assigned a stereochemical grade of 'B'. On the other hand, if the r_{S-S} or χ_{S-S} value does not fall within the above-specified range, the bridge is assigned the lowest grade of 'C'. If the four disulphide possibilities are of different stereochemical grades, the one with the best grade is chosen. In cases where the grades are the same, a choice is made by examining the deviation of the values of the parameters r_{S-S}, χ_{S-S}, χ^1_i and χ^1_j from their closest ideal values.

Summary of procedure

A flow chart of the procedure is shown in Figure 4. For a peptide of known sequence and disulphide bond connectivity, the procedure assigns a very large number of backbone conformations. The program requires as input, the 'seed number' for generating random numbers (NSEED), the number of residues in the polypeptide (NPEP), details regarding the presence and exact location of the Gly and Pro residues in the primary sequence and the disulphide bond connectivity. The assignment of backbone conformations is done by generating random numbers by the subroutine RANGEN. The random numbers are then suitably magnified to the (ϕ, ψ) scale depending on the nature of the amino acid residue, by the subroutine RANTOPS. As noted earlier, the conformational assignment falls into three categories, Gly, Pro and non-Gly−non-Pro. The assignment of (ϕ, ψ) values at each amino acid residue is then followed by the generation of the polypeptide corresponding to that conformation. This is done by the subroutine PEPGEN. Each conformation is checked by DISCHK, for the r^{α}_{ij} and r^{β}_{ij} distances.

If the two distances are suitable for accommodating a disulphide bond, further model-building is performed by means of MODIP. If all disulphides could be modelled by MODIP, out of the four possible conformations of an S−S bridge, the stereochemically best one is identified by the subroutine PICPOSS. The results can be obtained as an output which lists the backbone conformational angles or the positional coordinates of the atoms or both. In the case where the $2n$ distance criterion is not satisfied ('n' refers to the number of S−S bridges present in the polypeptide), that conformation is simply discarded and not considered for any further modelling. After arriving at a backbone structure, side chain atoms are fixed using the bond lengths and angles suggested by Momany *et al.* (1975), before performing energy minimization. The preferred side chain torsions for various residue types, as suggested by Janin *et al.* (1978) have been used to construct the side chains except disulphide-bonded cysteines, glycines and alanines. All the atoms including the side chains are allowed to move position freely during energy minimization, in order to relieve any short contacts that may be present. No exhaustive conformation search procedure was employed for predicting side chain positions as the current method is primarily aimed at obtaining stereochemically acceptable backbone conformations.

This procedure has been coded as a computer program for IBM compatible PCs in FORTRAN. All the residues were maintained in the 'L' configuration. An ideal value of 180° was chosen for the dihedral angle ω (C−N bond). In the case of the highly constrained systems such as enterotoxin (which contains three disulphide bonds in 13 residues), distortion about ω was desirable in order to generate reasonable models. In such cases, random values of ω falling in the range ± 180 ± 20° were assigned (R.Sowdhamini, C.Ramakrishnan and P.Balaram, unpublished results).

Results and discussion

Testing the procedure

The ability of this procedure to generate stereochemically acceptable conformations for disulphide loops was initially tested using small systems of the type Cys−X−Y−Cys. These four-residue, 14-membered disulphide loops are known to have distinct preferences for backbone conformations, corresponding to β-turns centred around residues X and Y. For example, the 14-membered, Cys−Gly−Pro−Cys loop at the active site of *E.coli* thioredoxin (Holmgren, 1968) has been shown to adopt a type III β-turn in crystals (Holmgren *et al.*, 1975; Katti *et al.*, 1990) and by means of NMR studies (Dyson *et al.*, 1988, 1989, 1990). Theoretical studies on model 14-membered peptides have shown that β-turns are energetically preferred, while NMR and crystal structures for model peptides confirm the presence of such conformations (Venkatachalapathi *et al.*, 1982; Kishore and Balaram, 1986; Kishore *et al.*, 1988). Conformations for

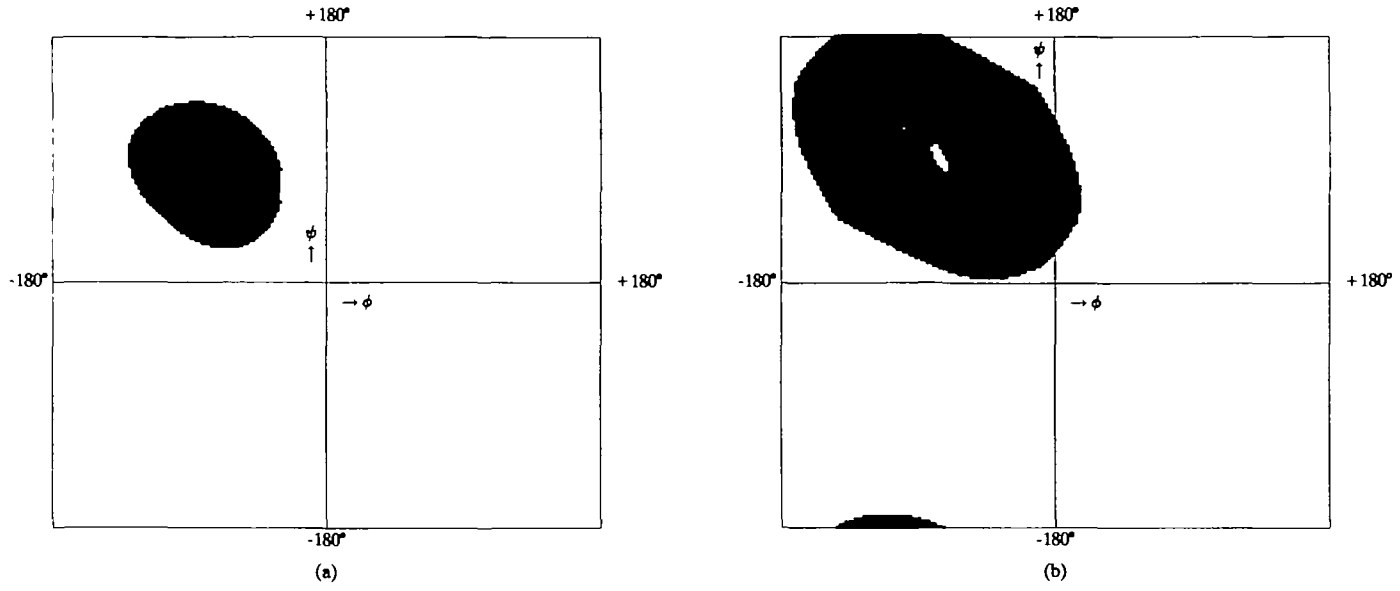

Fig. 5. A 2-D map representation of the 'permitted' (ϕ_4,ψ_1) combinations which lead to strainless disulphide modelling by MODIP. Those values which can accommodate disulphides are marked by '.'. The calculations have been performed using a grid interval of 2°. (a) The results obtained using a model system with type I β-turn at the middle. (b) The results obtained for a model system with type II β-turn at the middle.

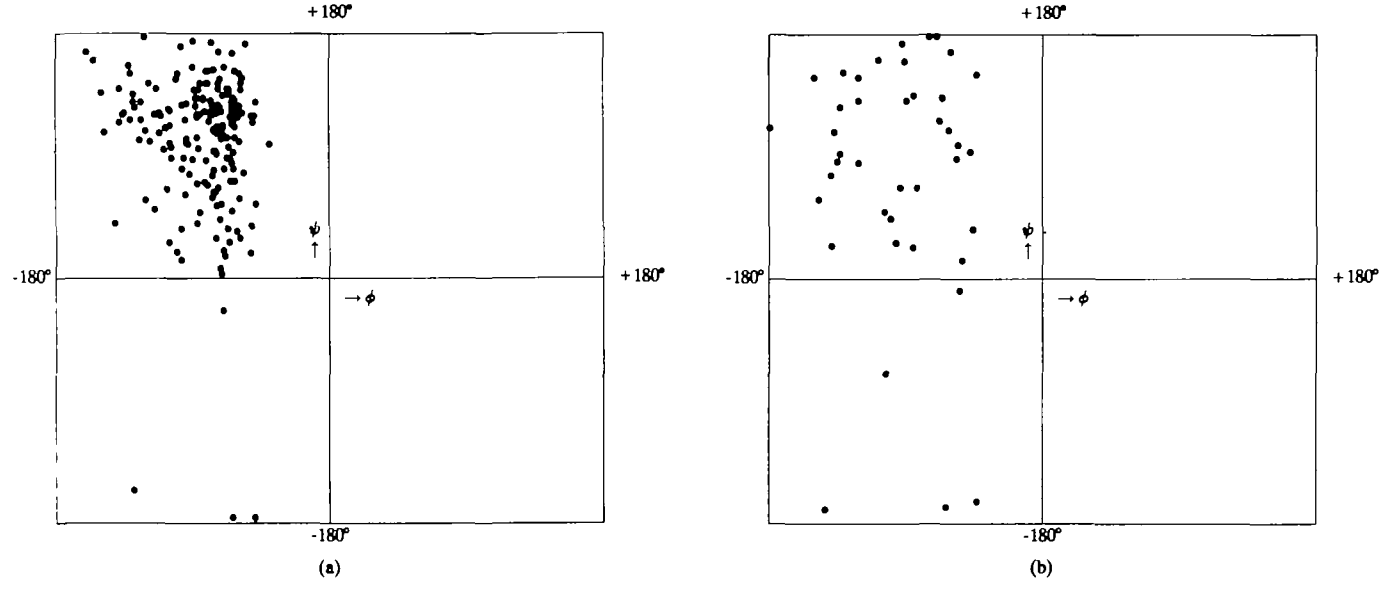

Fig. 6. 2-D map representation of (ϕ_{i+3}, ψ_i) values of 14-membered ring systems of the type Cys−X−Y−Cys. (a) Examples obtained by a search in a database of hypothetical and native disulphides derived from protein structures. (b) Examples obtained using 'RANMOD' procedure. Only selected examples which could be grouped into families have been shown. Families are characterized by classical β-turns, centred at residues X and Y. A similar pattern of clustering of points in the 2-D space can be noticed for both cases.

14-membered rings were generated using three distinct procedures.

(i) Starting with idealized type II and III β-turns centred at residues $i + 1$ and $i + 2$, the dihedral angles ψ_i and ϕ_{i+3} were systematically varied in a 2° grid and disulphide bridge closure was achieved using the MODIP procedure. The region of (ψ_i, ϕ_{i+3}) space suitable for disulphide bridge formation is shown in Figure 5 for the two cases.

(ii) The MODIP procedure was applied to a 65-protein dataset of largely non-homologous structures [for a list of proteins, see Srinivasan *et al.* (1991)]. This provides a means of identifying residue pairs which can accommodate a strainless disulphide if the residues involved are replaced by Cys. A total of 1862 such examples of various 'disulphide-loop' sizes were identified; 72 of them correspond to native disulphides. Two hundred and nine of the segments are four-residue 'disulphide loops', which were analysed for their backbone conformations.

(iii) Modelling a four-residue peptide with a disulphide connecting residues i and $i + 3$ using the RANMOD procedure.

Many of the stereochemically acceptable structures obtained by procedures (ii) and (iii) had indeed, one of the classical β-turn types and X and Y. In the case of four-residue loops obtained by procedure (ii), an overwhelming majority of them (96.4%) were characterized by a tight β-turn. Out of 209 examples, 158 of them correspond to a type I β-turn, 32 to a type II β-turn and three and five examples of types I′ and II′ β-turn respectively.

The (ψ_i, ϕ_{i+3}) values for the chosen examples which could be grouped into any one of the above families are shown in Figure 6a.

Using procedure (iii), a search of 20 000 conformations yielded 214 stereochemically acceptable conformations. Out of these, 41 structures were found to have a classical β-turn at the middle. While this corresponds to only ~20% of the total conformations obtained using RANMOD, the identification of the other 80% non-β-turn conformations, can be viewed as a virtue of this procedure. The (ψ_i, ϕ_{i+3}) values for the 41 examples obtained by the RANMOD procedure are plotted in Figure 6b. It can be seen that in both cases (Figure 6a and b) ψ_i seldom takes a positive value, when a type II or III β-turn exists at the centre. This observation is in agreement with the results obtained by procedure (i) using model systems.

Application of the procedure

The random conformation generation procedure has been applied to the modelling of two polypeptides, α-conotoxin GI and endothelin I (see Figure 2 for details of amino acid sequence). While both are two-disulphide-bonded systems, the peptide α-conotoxin GI belongs to the type III loop topology with interlocked loops and endothelin I is of the type II topology, with a loop within loop arrangement (see Figure 1a).

α-Conotoxin GI. Conotoxins are neuropeptides, 10–30 amino acid residues long, most of which are rich in disulphide bonds (Olivera *et al.*, 1985). One of the paralytic peptides of fish-hunting cone venoms, is α-conotoxin GI which has two S–S bonds and acts on acetylcholine receptors (Gray *et al.*, 1981, 1984). Experimental studies that were performed with a view to elucidating the conformation of this peptide are summarized in Table I.

Half a million conformations of this 12-residue peptide were examined by this method to arrive at acceptable models. Although 22 structures had their Cys residues suitably separated to accommodate the two disulphide bridges by means of distance, only four of them were stereochemically ideal and suitable for strainless disulphide modelling. Table II lists the backbone torsion angles for the four models. Energy minimization was performed on one of the three models, C_1, using the AMBER package (Weiner *et al.*, 1984, 1986), kindly supplied by Professor P.A.Kollmann. The backbone torsion angles, ϕ, ψ and ω, for the model C_1 after minimization are also indicated in Table II. Figure 7b shows a stereo picture of the backbone of the energy-minimized model, C_1, with the two disulphide bridges also shown. The gross topology of the present model agrees to a great extent with the previous models, derived from experimental data (Gray *et al.*, 1985). The original model C_1 is characterized by two β-turns, a distorted type III β-turn centred at Asn4 and Pro5, a type III β-turn at Arg9 and His10 and an extended strand at residues 6–8 of the peptide. It may be noted that the stereochemistry of the two disulphides in model C_1 has improved upon energy minimization (from grades B and C to grade A). However, this is accompanied by movements in the backbone conformation, that have also distorted the β-turn centred at Asn4 and Pro5, originally present in the crude structure. Use the milder minimization techniques might permit retention of most of the original features in the backbone of the model, while improving disulphide bridge stereochemistry.

The CD spectroscopic studies of Hider (1985) showed the

Table I. Earlier studies on structure of α-conotoxin GI

Nature of study	Structural features	Reference
CD studies	Residues 5–11: α-helix	Hider (1985)
Based on structure of erabutoxin	Residues 5–8: β-turn Residues 9–12: β-turn	Gray *et al.* (1985)
NMR in DMSO-d_6[a]	No clear presence of tight turns but structure similar to that of Gray *et al.* (1985)	Kobayashi *et al.* (1989)
NMR in D_2O[b]	Tight turns centred at residues 5 and 9	Pardi *et al.* (1989)

[a]The experiment was done in 7 mM solution of the peptide containing 10% D_2O at pH 4.8 and 5°C.
[b]An 8 mM solution was used for the NMR experiment at 20°C.

Table II. Backbone torsion angles (ϕ, ψ)[c] (in degrees) and disulphide bond grades of the proposed α-conotoxin GI models

Residue no.[a]	Model C_1 (Crude)		Model C_1 (Energy refined)			Model C_2		Model C_3		Model C_4	
	ϕ	ψ	ϕ	ψ	ω	ϕ	ψ	ϕ	ψ	ϕ	ψ
2	−56	148	−	73	−174	−133	−176	−129	35	−123	5
3	−124	177	−65	99	−176	−74	−173	−63	−16	−85	169
4	−92	−45	−11	−66	180	−84	−24	−40	135	59	37
5	−78	−41	−55	−43	176	−176	−6	−40	10	−172	−65
6	−147	112	−122	60	179	−49	−50	−103	−27	−118	−10
7	−137	92	−90	101	177	−53	70	−136	157	−161	7
8	−171	166	−178	166	−178	173	175	52	−164	66	−85
9	−51	−36	−56	−34	177	−60	−18	−177	−23	−71	93
10	−48	−21	−60	−27	−171	−107	−60	−158	51	−58	−5
11	−122	−169	−105	145	−177	−44	97	−141	10	−160	−15
12	−132	6	−73	−54	−173	−75	130	−173	−62	−158	131
13	−120	29	−72	−	−	−178	32	−78	101	−172	−40
Grade of disulphide bonds[b]											
Residues 2–7	B		A			C		B		C	
Residues 3–13	C		A			C		B		C	

[a]Glu at residue 1 has not been used for a modelling study since it does not form part of the two disulphide loops.
[b]A, B and C refer to the stereochemical grades of disulphide bonds [see text and Sowdhamini *et al.* (1989) for details of grade assignment].
[c]The value of ω is maintained at 180° for all the crude models obtained from RANMOD.

presence of ~50% helical content in the peptide. In order to locate the helical region, a secondary structure prediction study was done (Geisow and Roberts, 1980) which indicated that the residues 6–10 stretch has high α-helix propensity. This led to the proposition that the residues 5–11 stretch assumes an α-helical structure, with the two ends of the chain folded back to make a disulphide bridge (Hider, 1985).

An alternative model proposed by Gray *et al.* (1985) suggests the presence of two successive β-turns (at residues 5–8 and 9–12). This is based on structural analogy with the segment of erabutoxin involved in receptor binding together with a secondary structure prediction by the Chou–Fasman method.

Kobayashi *et al.* (1989) have studied α-conotoxin GI in DMSO-d_6 by NMR. Although the sequential backbone NOEs, suggest the presence of turns at residues 2–5 and 5–8, a clear presence of the two turns is not indicated. These authors obtained relatively strong $C^{\alpha}H_i <-> NH_{i+1}$ NOEs over the segments of residues 1–4 and 9–13, with a break between residues 9 and 10. The only stretch showing a strong d_{NN} connectivity is the region between residues 6 and 8. Weaker d_{NN} NOEs are observed between residues 3 and 4. The authors conclude by suggesting that type I β-turn structures might be accommodated in the sequence around residues 3–4 and residues 5–8. Another interesting feature of this NMR study is the observation of minor conformations which exchanged extremely slowly with the major conformer such that no magnetization transfer by chemical exchange was detected in NOESY spectra, even at long mixing times (200 ms). While slow exchange in Pro-containing peptides has generally been attributed to *cis–trans* isomerization about the X–Pro bonds, it is noteworthy that no satellite signals were detectable at Asn4 which precedes the lone Pro residue. In highly constrained systems, the possibility of slow exchange between all *trans* conformers must also be considered.

However, the NMR experiments done in D_2O by Pardi *et al.* (1989) seem to indicate the presence of two tight reverse turns centred at positions Pro5 and Arg9. These authors obtained structures calculated using NMR distance constraints which have a regular β-turn centred at positions Pro5 and Ala6. A second reverse turn feature observed at residues Arg9 and His10 has been interpreted as possibly arising from two γ-turns centred at Arg9. Although such γ-turns are relatively uncommon, they have been observed in strained cyclic tetrapeptide systems (Rich *et al.*, 1983). Interestingly, one of the models listed in Table II reveals the presence of two γ-turns centred at residues 8 and 9. Pardi *et al.* (1989) suggest that the appreciably helical CD spectrum obtained by Hider (1985) may arise due to the contributions from γ-turn conformations and the two disulphide linkages. The structures proposed for conotoxin thus far differ in detail, suggesting the possibility of multiple conformations and also solvent-dependent structural variability. It is gratifying to note that several local structural features identified in NMR data are indeed observed in the conformations obtained by the RANMOD procedure. There is broad general agreement that the overall folding of the conotoxin model is reasonably well-represented by the model originally proposed by Gray *et al.* (1985). The energy-minimized conformation (C_1) obtained by the RANMOD procedure is shown in Figure 7a, while Figure 7b shows a stereo picture of the same model.

Endothelin I. Endothelins (ETs) are a family of regulatory peptides synthesized by selected endothelial and epithelial cells that act on nearby smooth muscles or connective tissues [for recent reviews, see Yanagisawa and Masaki (1989) and Simpson and Dunn (1990)]. Endothelin 1 (ET-1) is an acidic peptide and is one of the most potent vasoconstrictors known (Yanagisawa *et al.*, 1988). In view of the crucial biological role of endothelin as a potent vasoconstrictor and the consequent interest in antagonist design, a knowledge of the structure of this 21-residue peptide is most desirable. No crystal structure has been reported to date. The solution structure determination by NMR methods has been attempted by various groups in different solvent systems (Endo *et al.*, 1989; Saudek *et al.*, 1989, 1991; Aumelas *et al.*, 1991; Krystek *et al.*, 1991; Reily and Dunbar, 1991; Tamaoki *et al.*, 1991).

Three hundred and fifty thousand conformations were examined by the RANMOD procedure for a stereochemically acceptable structure of endothelin I. While 25 structures had the appropriate r_{ij}^{α} and r_{ij}^{β} distances, five of them could accommodate unstrained disulphide bonds by MODIP. These models were then subjected to energy minimization. Table III lists the backbone torsion angles, ϕ and ψ of the five energy-refined models. In spite of the presence of two covalent cross-links holding the molecule, there is a considerable amount of freedom in the orientation of the atoms.

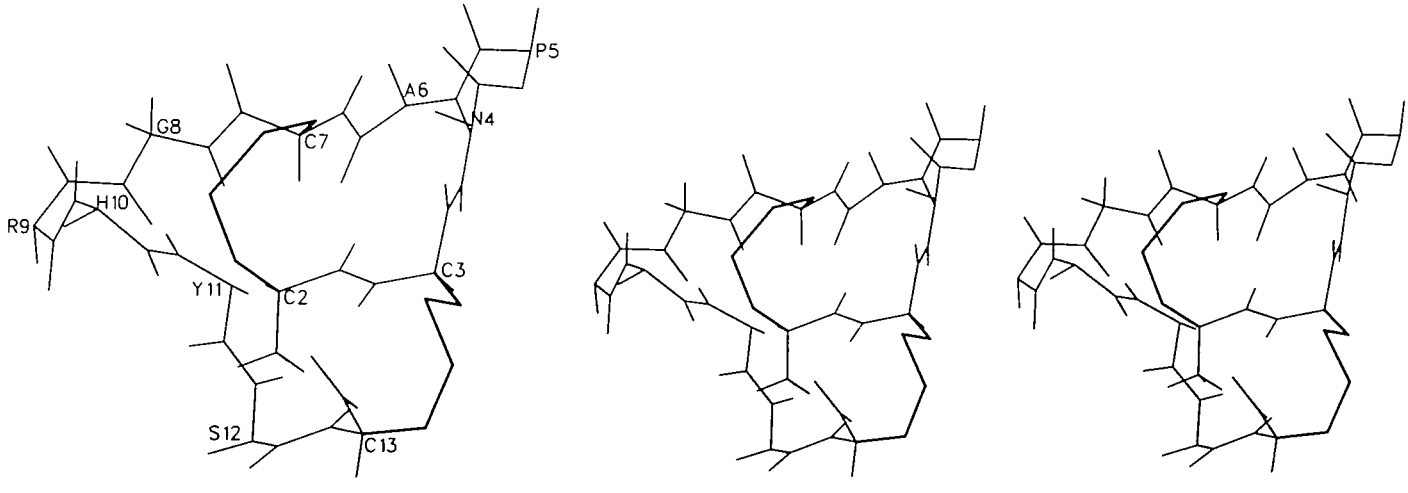

Fig. 7. Model of α-conotoxin GI obtained by the RANMOD procedure. (a) Line drawing of model C_1 obtained by the RANMOD procedure. All backbone atoms including hydrogens and C^{β} atoms have been shown. The residues have been labelled at C^{α} positions. Glu1 has not been shown since this residue is not crucial for activity and, hence, not considered for modelling. Disulphide bonds have been shown in thick lines. (b) Stereo picture of the model of C_1.

Table III. Results of the energy minimization of endothelin models compared with NMR structures

Feature	NMR structure model N_b	NMR structure model N_a	Model I	Model II	Model III	Model IV	Model V
ϕ_1,ψ_1	−157,127	−,−[a]	−,−117	−,−118	−,−111	−,−175	−,13
ϕ_2,ψ_2	−37,104	−87,56	−78,63	−143,−45	−126,−46	−156,92	−153,73
ϕ_3,ψ_3	162,112	−101,119	−146,−176	−129,108	−82,66	−83,131	−77,−125
ϕ_4,ψ_4	−137,98	−73,35	−71,92	−106,160	−163,87	−83,49	−139,101
ϕ_5,ψ_5	−117,119	−9,44	−75,57	−119,−94	−157,−87	−151,78	−156,−73
ϕ_6,ψ_6	−93,29	−36,9	−160,−49	−148,74	−146,−76	−82,63	−123,−141
ϕ_7,ψ_7	118,61	−119,−39	−129,72	−150,−170	−92,63	−102,−170	−75,−58
ϕ_8,ψ_8	163,141	−83,78	−64,160	−142,63	−82,167	−76,65	−98,134
ϕ_9,ψ_9	−94,32	−11,4	−133,52	−131,−108	−150,−61	69,9	−125,107
ϕ_{10},ψ_{10}	−111,−12	−87,−57	53,62	−81,−86	−82,54	−73,80	−87,65
ϕ_{11},ψ_{11}	−151,32	−68,−24	−122,50	−79,116	−78,75	−146,51	−164,62
ϕ_{12},ψ_{12}	−153,23	−68,−43	−142,77	−71,95	−63,−47	−76,147	−143,−85
ϕ_{13},ψ_{13}	−143,54	−53,−27	−142,−72	−23,−42	−111,−139	−139,160	−91,−29
ϕ_{14},ψ_{14}	178,64	−99,−6	−98,55	−51,−34	−64,−36	−86,−44	−139,25
ϕ_{15},ψ_{15}	−155,68	−127,−70	−155,−	−74,−	−141,−	−60,−	−142,−
Disulphide geometry[b]							
(Residues 1−15)	−	−	A,+	B,±	A,+	A,−	A,+
(Residues 3−11)	−	−	−	B,−	B,±	A,−	A,+
Total energy (in kJ/mol)			69.03	74.16	94.97	87.43	69.46
r.m.s. deviation (in Å)[c]							
(N_{vii})	0.0	4.0	4.29	4.88	4.53	3.51	4.79
(N_v)	4.0	0.0	3.75	4.18	4.00	4.23	4.95

[a]The (ϕ,ψ) value was not reported for the NMR model N_a.
[b]'−' represents cases where S fixing was not possible by MODIP. 'A' or 'B' refers to the grade assigned to the disulphide bond by MODIP, while the symbols, '+', '−' or '±' refer to the chirality of the best possible S−S bond conformation (magnitude of the torsion angle, X_{S-S}).
[c]Only the atoms, N, C^α, C and O were considered for best superimposition.

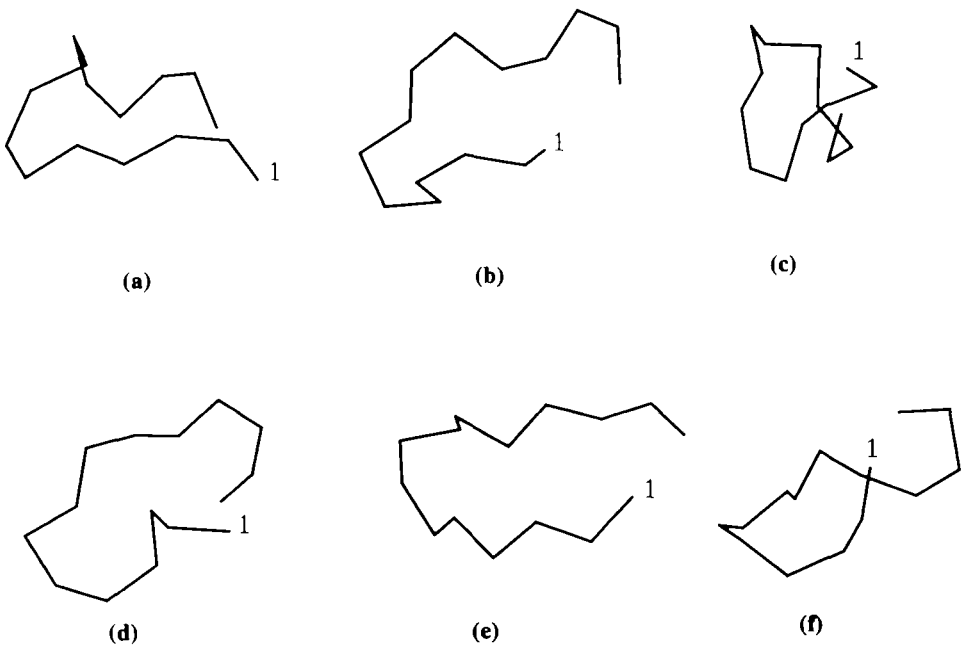

Fig. 8. Cα trace of the five models of endothelin I, obtained by the random conformation generation procedure, compared with that of the NMR model N_b (Endo *et al.*, 1989). (a) NMR model N_b. (b)−(f) Models I−V of endothelin I. The five RANMOD models have been plotted after best superimposition with the NMR model N_b.

Details in the form of backbone torsion angles for the NMR structures have been reported only for two structures, one by Krystek *et al.* (1991) (which will be referred to as Model N_a) and the other by Endo *et al.* (1989) (which will be referred to as Model N_b). Disulphide modelling was not possible for the residues 1−15 S−S bridge in the case of NMR model N_a, since the (ϕ, ψ) value for residue 1 has not been reported (Krystek *et al.*, 1991). The NMR model N_b (Endo *et al.*, 1989), built

Table IV. Disulphide bridge geometries of the bioactive peptide models obtained by RANMOD (after energy minimization)

Model No.	S–S bridge	Lengths (in Å)			Torsion angles (in degrees)			Grade
		r^{α}_{ij}	r^{β}_{ij}	r_{S-S}	χ_{S-S}	χ^1_i	χ^1_j	
α-Conotoxin GI (Residues 2–13)								
I	Residues 2–7	4.21	4.05	1.91	+92.8	−153.0	167.9	A
	Residues 3–13	5.07	3.60	1.97	+64.5	44.6	55.2	A
Endothelin I (Residues 1–15)								
I	Residues 1–15	6.40	3.58	2.21	+77.7	49.3	−80.5	A
	Residues 3–11	6.83[a]	3.57	2.17	−84.9	−67.3	−47.3	A
II	Residues 1–15	4.90	3.90	1.96	−84.5	−73.1	20.5	B
				2.03	+88.8	−73.1	104.1	B
	Residues 3–11	5.23	4.42	1.81	−117.2	86.8	112.7	B
III	Residues 1–15	4.48	3.70	1.94	+66.2	171.5	−174.7	A
	Residues 3–11	5.10	3.84	1.91	−77.7	164.4	−93.3	B
				2.19	+93.6	76.5	−93.3	B
IV	Residues 1–15	5.32	3.52	2.30	−78.9	−32.4	−53.4	A
	Residues 3–11	5.46	3.94	1.96	−87.6	51.7	178.2	A
V	Residues 1–15	6.16	3.62	2.30	+84.6	−167.1	−50.7	A
				1.98	+66.9	58.3	86.5	A
	Residues 3–11	5.39	4.23	1.81	+100.0	44.3	47.9	A

[a]The $C^{\alpha}-C^{\alpha}$ distance of this disulphide has increased upon energy minimization (above the limits used in MODIP). Hence, this disulphide has been modelled by MODIP after $C^{\alpha}-C^{\alpha}$ distance relaxation.

using reported backbone torsion angles and standard parameters, has unfavourable distances for sulphur fixing (r^{α}_{1-15} = 4.8 Å, r^{β}_{1-15} = 5.5 Å; r^{α}_{3-11} = 6.6 Å, r^{β}_{3-11} = 5.5 Å), while the residues 3–11 positions in model N_a (Krystek *et al.*, 1991), also obtained by a similar method, were not stereochemically suited for S–S bridge formation. Figure 8 shows line drawings of the five endothelin energy-minimized models along with the NMR model N_b for the sake of comparison.

Conformational flexibility in endothelin I has been well documented in the experimental data. Although the region of residues 1–15 is considered more structurally well-defined, this is probably with respect to the floppy C-terminal (residues 16–21) region. It may be noted that a superposition of the reported structures, N_a and N_b, results in a root mean square (r.m.s.) deviation value between the two NMR structures of 4 Å, when taking only the N, C^{α}, C and O atoms into consideration. This suggests that the models obtained theoretically may provide alternative possibilities for evaluating experimental data.

Conclusions

The utility of combining a rigorous disulphide modelling approach with a simple random search procedure to generate stereochemically acceptable conformations has been demonstrated. The procedure ensures strainless disulphide bridges, as seen in Table IV, even for fairly constrained systems. Comparison of the models of enterotoxin obtained by RANMOD with the crystal structure shows that the overall fold has been predicted correctly in several models. Indeed, one of the models bears reasonable resemblance to the crystal structure (R.Sowdhamini, C.Ramakrishnan and P.Balaram, unpublished results).

For a 15-residue polypeptide constrained by two S–S bridges (disulphide-rich region of endothelin I), good starting structures are obtained by typically searching 350 000 conformations. This takes ~50 min in a PC-AT 386. Hence, a fairly large number of conformations can be searched in a reasonable amount of time. In contrast, if a systematic search is performed within the permitted (ϕ, ψ) region using a 10° grid, it would require examining 9.32×10^{39} conformations resulting in the use of an enormous amount of computer time.

An attractive feature of the procedure is that several distinct and dissimilar conformations are obtained as plausible models, although there is no way of ensuring if the sampling in the conformational space is complete. The systems examined undoubtedly exhibit conformational heterogeneity in solution and direct agreement between a single modelled conformation and experimental results is unlikely. Generation of a large number of stereochemically acceptable structures and a subsequent experimental databased filter (like NMR distance constraints or CD, Raman spectroscopic data) followed by molecular dynamics simulations can expand the scope of the present procedure. Availability of different models of comparable energy might also explain distinct multiple conformations in solution.

For systems larger than those examined in this study, obtaining a single correct model is desirable. This approach can be extended to the tertiary fold modelling of larger S–S-rich proteins. RANMOD is being used to model the fold of BPTI and nerve growth factor, starting from their secondary structures. Results obtained so far indicate that successful modelling can be achieved by introducing proper filters such as intersegment short contact checks (R.Sowdhamini, unpublished results). Indeed, in the case of BPTI, several models obtained by RANMOD have native-like folds in the residues 14–51 region containing two disulphides. The application of pseudopotential functions for the identification of native folds among misfolded models (Hendlich *et al.*, 1990; Sippl, 1990) on RANMOD-derived models is worthy of investigation.

Acknowledgements

This work was supported by an extramural research grant from the Astra Research Centre, India. R.S. acknowledges the receipt of a senior research fellowship from the Council of Scientific and Industrial Research, India. We would like to thank Dr N.Srinivasan for useful discussions and Dr N.V.Joshi for providing a routine for the generation of random numbers.

References

Aumelas,A., Chiche,L., Mahe,E., Le-nguyen,D., Sizun,P., Berthault,P. and Perly,B. (1991) *Int. J. Peptide Protein Res.*, **37**, 315–324.
Dyson,H.J., Holmgren,A. and Wright,P.E. (1988) *FEBS Lett.*, **228**, 254–258.
Dyson,H.J., Holmgren,A. and Wright,P.E. (1989) *Biochemistry*, **28**, 7074–7087.
Dyson,H.J., Gippert,G.P., Case,D.A., Holmgren,A. and Wright,P.E. (1990) *Biochemistry*, **29**, 4129–4136.
Endo,S., Inooka,H., Ishibashi,Y., Kitada,C., Mizuta,E. and Fujino,M. (1989) *FEBS Lett.*, **257**, 149–154.
Gariepy,J., Lane,A., Frayman,F., Wilbur,D., Robien,W., Schoolnik,G.K. and Jardetsky,O. (1986) *Biochemistry*, **25**, 7854–7866.
Geisow,M.J. and Roberts,R.D.B. (1980) *Int. J. Biol. Macromol.*, **2**, 387–389.
Gray,W.R., Luque,A., Olivera,B.M., Barret,J. and Cruz,L.J. (1981) *J. Biol. Chem.*, **256**, 4734–4740.
Gray,W.R., Luque,F.A., Galyean,R., Atherton,E., Sheppard,R.C., Stone,B.L., Reyes,A., Alford,J., McIntosh,M., Olivera,B.M., Cruz,L.J. and Rivier,J. (1984) *Biochemistry*, **23**, 2796–2802.
Gray,W.R., Middlemas,D.M., Zeikus,R., Olivera,B.M. and Cruz,L.J. (1985) In Deber,C.M., Hruby,V.J. and Kopple,K.D. (eds), *Peptides*. Pierce Chem. Co., Rockford, pp. 823–832.
Hendlich,M., Lackner,P., Weitckus,S., Floeckner,H., Froschauer,R., Gottsbacher,K., Caseri,G. and Sippl,M.J. (1990) *J. Mol. Biol.*, **216**, 167–180.
Hider,R.C. (1985) *FEBS Lett.*, **184**, 181–184.
Holmgren,A. (1968) *Eur. J. Biochem.*, **6**, 475–484.
Holmgren,A., Sjoberg,B.-O., Eklund,H. and Branden,C.-I. (1975) *Proc. Natl Acad. Sci. USA*, **172**, 2305–2309.
Janin,J., Wodak,S., Levitt,M. and Maigret,B. (1978) *J. Mol. Biol.*, **125**, 357–386.
Katti,S.K., LeMaster,D.M. and Eklund,H. (1990) *J. Mol. Biol.*, **212**, 167–184.
Kikuchi,T., Nemethy,G. and Scheraga,H.A. (1986) *J. Comput. Chem.*, **7**, 67–88.
Kikuchi,T., Nemethy,G. and Scheraga,H.A. (1989) *J. Comput. Chem.*, **10**, 287–294.
Kishore,R. and Balaram,P. (1986) In Holmgren,A., Branden,C.-I., Jornvall,H. and Sjoberg,B.-M. (eds), *Thioredoxin and Glutaredoxin Systems—Structure and Function*. Raven Press, New York, pp. 57–64.
Kishore,R., Raghothama,S. and Balaram,P. (1988) *Biochemistry*, **27**, 2462–2471.
Klapper,M.H. and Klapper,I.Z. (1980) *Biochim. Biophys. Acta*, **626**, 97–105.
Kobayashi,Y., Ohkubo,T., Kyogoku,Y., Nishiuchi,Y., Sakakibara,S., Braun,W. and Go,N. (1989) *Biochemistry*, **28**, 4853–4860.
Krystek,S.R.,Jr, Bassolino,D.A., Novotny,J., Chen,C., Marschner,T.M. and Andersen,N.H. (1991) *FEBS Lett.*, **281**, 212–218.
L'ecuyer,P. (1988) *Commun. ACM*, **31**, 742–749.
MacArthur,M.W. and Thornton,J.M. (1991) *J. Mol. Biol.*, **218**, 397–412.
Mao,B. (1989) *J. Am. Chem. Soc.*, **111**, 6132–6136.
Meirovitch,H. and Scheraga,H.A. (1981) *Macromolecules*, **14**, 1250–1259.
Momany,F.A., McGuire,R.F., Burgess,A.W. and Scheraga,H.A. (1975) *J. Phys. Chem.*, **79**, 2361–2366.
Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) *Proteins: Struct. Funct. Genet.*, **12**, 345–364.
Nishiuchi,Y. and Sakakibara,S. (1982) *FEBS Lett.*, **148**, 260–262.
Olivera,B.M., Gray,W.R., Zeikus,R., McIntosh,J.M., Varga,J., Rivier,J., de Santos,V. and Cruz,L.J. (1985) *Science*, **230**, 1338–1343.
Ozaki,H., Sato,T., Kubota,H., Hata,Y., Katsube,Y. and Shimonishi,Y. (1991a) *J. Biol. Chem.*, **266**, 5934–5941.
Ozaki,H., Kubota,H., Sato,T., Hidaka,Y., Tamaoki,H., Kobayashi,Y., Kyogoku,Y., Sugimura,T., Tai,A. and Shimonishi,Y. (1991b) *Bull. Chem. Soc. Jpn*, **64**, 1136–1144.
Pardi,A., Galdes,A., Florance,J. and Maniconte,D. (1989) *Biochemistry*, **28**, 5494–5501.
Ramakrishnan,C. and Srinivasan,N. (1990) *Curr. Sci.*, **59**, 851–866.
Reily,M.D. and Dunbar,J.B.,Jr (1991) *Biochem. Biophys. Res. Commun.*, **178**, 570–577.
Rich,D.H., Kawai,M. and Jasensky,R.D. (1983) *Int. J. Peptide Protein Res.*, **21**, 35–42.
Saudek,V , Hoflack,J. and Pelton,J.T. (1989) *FEBS Lett* , **257**, 145–148
Saudek,V., Hoflack,J. and Pelton,J.T. (1991) *Int. J. Peptide Protein Res.*, **37**, 174–179.
Shimonishi,Y., Hidaka,Y., Koizumi,M., Hane,M., Aimoto,S., Takeda,T., Miwatani,T. and Takeda,Y. (1987) *FEBS Lett.*, **215**, 165–170.
Simpson,M. and Dunn,M. (1990) *FASEB J.*, **4**, 2989–3000.
Sippl,M.J. (1990) *J. Mol. Biol.*, **213**, 859–883.
Sowdhamini,R., Srinivasan,N., Shoichet,B., Santi,D.V., Ramakrishnan,C. and Balaram,P. (1989) *Protein Engng*, **3**, 95–103.
Srinivasan,N., Sowdhamini,R., Ramakrishnan,C. and Balaram,P. (1990) *Int. J. Peptide Protein Res.*, **36**, 147–155.
Srinivasan,N., Sowdhamini,R., Ramakrishnan,C. and Balaram,P. (1991) In Balaram,P. and Ramaseshan,S. (eds), *Molecular Conformation and Biological Interactions*. Indian Academy of Sciences, Bangalore, pp. 59–73.
Tamaoki,M., Kobayashi,Y., Nishimura,S., Ohkubo,T., Kyogoku,Y., Nakajima,K., Kumagaye,S., Kimura,T. and Sakakibara,S. (1991) *Protein Engng*, **4**, 509–518.
Venkatachalapathi,Y.V., Prasad,B.V.V. and Balaram,P. (1982) *Biochemistry*, **21**, 5502–5509.
Warne,N.W. and Lawskowski,M.,Jr (1990) *Biochim. Biophys. Res. Commun.*, **172**, 1364–1370.
Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S.,Jr and Weiner,P. (1984) *J. Am. Chem. Soc.*, **106**, 765–784.
Weiner,S.J., Kollman,P.A., Nguyen,D.T. and Case,D.A. (1986) *J. Comput. Chem.*, **7**, 230–252.
Yanagasiwa,M. and Masaki,T. (1989) *Biochem. Pharmacol.*, **38**, 1877–1883.
Yanagisawa,M., Kurihara,H., Kimura,S., Tomobe,M., Kobayashi,M., Mitsui,Y., Yazaki,Y., Goto,K. and Masaki,T. (1988) *Nature*, **332**, 411–415.

Received on January 12, 1993; revised on April 28, 1993; accepted on July 22, 1993