Determination of a high quality NMR solution structure of the bovine pancreatic trypsin inhibitor (BPTI) and comparison with three crystal structures

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · NOVEMBER 1992

Impact Factor: 4.33 · DOI: 10.1016/0022-2836(92)90222-6 · Source: PubMed

4 AUTHORS, INCLUDING:

Kurt D. Berndt
Karolinska Institutet
64 PUBLICATIONS **2,917** CITATIONS
SEE PROFILE

Kurt Wüthrich
The Scripps Research Institute
742 PUBLICATIONS **76,162** CITATIONS
SEE PROFILE

J. Mol. Biol. (1992) **227**, 757–775

Determination of a High-quality Nuclear Magnetic Resonance Solution Structure of the Bovine Pancreatic Trypsin Inhibitor and Comparison with Three Crystal Structures

Kurt D. Berndt, Peter Güntert, Leonard P. M. Orbons† and Kurt Wüthrich

Institut für Molekularbiologie und Biophysik
Eidgenösische Technische Hochschule-Hönggerberg
CH-8093 Zürich, Switzerland

(Received 13 April 1992; accepted 2 June 1992)

A high-quality three-dimensional structure of the bovine pancreatic trypsin inhibitor (BPTI) in aqueous solution was determined by ^{1}H nuclear magnetic resonance (n.m.r.) spectroscopy and compared to the three available high-resolution X-ray crystal structures. A newly collected input of 642 distance constraints derived from nuclear Overhauser effects and 115 dihedral angle constraints was used for the structure calculations with the program DIANA, followed by restrained energy minimization with the program AMBER. The BPTI solution structure is represented by a group of 20 conformers with an average root-mean-square deviation (RMSD) relative to the mean solution structure of 0·43 Å for backbone atoms and 0·92 Å for all heavy atoms of residues 2 to 56. The pairwise RMSD values of the three crystal structures relative to the mean solution structure are 0·76 to 0·85 Å for the backbone atoms and 1·24 to 1·33 Å for all heavy atoms of residues 2 to 56. Small local differences in backbone atom positions between the solution structure and the X-ray structures near residues 9, 25 to 27, 46 to 48 and 52 to 58, and conformational differences for individual amino acid side-chains were analyzed for possible correlations with intermolecular protein–protein contacts in the crystal lattices, using the pairwise RMSD values among the three crystal structures as a reference.

Keywords: BPTI; protein structure; nuclear magnetic resonance; distance geometry; structure refinement

1. Introduction

Few proteins have undergone such detailed conformational analysis as the bovine pancreatic trypsin inhibitor (BPTI‡). Its small size and high stability have made it an ideal model for theoretical (e.g. see van Gunsteren & Karplus, 1982; Levitt, 1983) and experimental (e.g. see Creighton, 1978; Wagner, 1983) studies alike. The three-dimensional structure of BPTI has been determined at high resolution in three different crystal forms by the techniques of X-ray diffraction (Deisenhofer & Steigemann, 1975; Wlodawer *et al.*, 1987*b*) and combined neutron and X-ray diffraction (Wlodawer *et al.*, 1984). BPTI has also had a central role in the development of n.m.r. techniques for protein structure determination (Wüthrich, 1989*b*), resulting ten years ago in the determination of the secondary structure in solution (Dubs *et al.*, 1979; Wagner *et al.*, 1981; Wagner & Wüthrich, 1982*a*), and more recently in a determination of the three-dimensional polypeptide fold (Wagner *et al.*, 1987). For reasons outlined below, the earlier work on the solution structure of BPTI was based on a small number of n.m.r. constraints and therefore limited to a global

† Present address: Department of Analytical Chemistry, Spectroscopy Group, Organon Int. BV. P.O. Box 9502, NL-5340 BH Oss, The Netherlands.

‡ Abbreviations used: BPTI, bovine pancreatic trypsin inhibitor; n.m.r., nuclear magnetic resonance; 2D, two-dimensional; NOE, nuclear Overhauser effect; NOESY, 2-dimensional nuclear Overhauser enhancement spectroscopy; COSY, 2-dimensional correlated spectroscopy; E. COSY, exclusive COSY; RMSD, root-mean-square deviation; p.p.m., parts per million; c.p.u., central processing unit; $^3J_{HN\alpha}$, vicinal spin–spin coupling constant between the amide and the α-proton; $^3J_{\alpha\beta}$, vicinal spin–spin coupling constant between the α and one β-proton; REDAC, use of redundant dihedral angle constraints.

0022–2836/92/190757–19 $08.00/0

characterization of the three-dimensional arrangement of the polypeptide backbone (Wagner *et al.*, 1987). Therefore, in view of the important role of BPTI in modern protein research, the high-quality solution structure described in this paper fills an important gap.

One of the limiting factors in previous studies on the n.m.r. solution structure of BPTI was that the generally available BPTI preparations contain about 10% of closely related BPTI analogs (Siekmann *et al.*, 1987). Although this hardly ever interfered with the use of BPTI for the development of new n.m.r. techniques, it made the collection of an extensive set of conformational constraints as input for a structure determination difficult, since the assignments of many NOESY cross-peaks remained ambiguous. As a consequence, the previously described solution structure was calculated from a data set of only about 200 NOE distance constraints (Wagner *et al.*, 1987). We have now prepared a homogeneous sample of BPTI, collected an extensive set of conformational constraints and performed a new structure calculation with the program DIANA (Güntert *et al.*, 1991*a*), using the newly introduced REDAC strategy (Güntert & Wüthrich, 1991). The resulting n.m.r. solution structure is compared in detail with the aforementioned three BPTI crystal structures.

2. Methods

(a) *Protein preparation*

BPTI was obtained as a gift from Bayer AG, Leverkusen (Trasylol®). Residual protein impurities were removed by reversed-phase chromatography on a ProRPC 5/10 column (Pharmacia). The protein was dissolved in a mixture of 15% (v/v) acetonitrile in water containing 0·1% (v/v) trifluoroacetic acid and applied to the column previously equilibrated with the same solvent mixture. Elution of the protein was accomplished by a linear gradient to 33% acetonitrile at a flow rate of 1 ml/min. Analytical reverse phase chromatography of the purified inhibitor yielded a single, symmetrical peak as monitored by absorbance of the effluent at 280 nm. The absence of asymmetric shoulders on the well-resolved doublet resonance of the aromatic ring protons of Tyr23 at 6·33 p.p.m. was used to monitor the purity of BPTI. BPTI obtained in this way was found to be free of the protein contaminations which heretofore complicated the spectral interpretation. All data used for the structure calculations were obtained from spectra recorded with 2 samples of BPTI: a 10 mM solution in a mixture of 95% H_2O and 5% ^2H_2O, and a 10 mM solution in 99·99% ^2H_2O prepared after complete exchange of all labile protons (Masson & Wüthrich, 1973). The protein samples were adjusted to pH 4·6 by the addition of minute amounts of NaOH and HCl, or NaO^{2}H and ^{2}HCl, respectively.

(b) *Nuclear magnetic resonance experiments*

Two-dimensional ^{1}H n.m.r. spectra were recorded on either a Bruker AM500 or AM600 spectrometer, with the carrier placed in the center of the spectrum. Quadrature detection in the ω_1 dimension was achieved using time-proportional phase incrementation (Marion & Wüthrich, 1983). All experiments were performed in unbuffered H_2O at pH 4·6 and 36°C. For the collection of upper bounds on ^{1}H–^{1}H distances, NOESY spectra (Anil-Kumar *et al.*, 1980) were recorded at 600 MHz with $\tau_m = 40$ ms. For the NOESY spectrum in H_2O, 512 experiments were recorded in t_1 ($t_{1max} = 32$ ms) with 4096 points in t_2 ($t_{2max} = 254$ ms). For the NOESY spectrum in ^2H_2O, 610 experiments were recorded in t_1 ($t_{1max} = 51$ ms) with 2048 points in t_2 ($t_{2max} = 170$ ms). Zero-quantum coherence was suppressed in the NOESY experiments in ^2H_2O using the technique of Otting *et al.* (1990). Prior to Fourier transformation, the time domain data of the NOESY data sets were multiplied by a cosine function in both dimensions and then zero-filled to 2048 points in t_1 and 8192 points in t_2. The vicinal spin–spin coupling constants $^3J_{HN\alpha}$ were measured from the NOESY spectrum in H_2O by inverse Fourier transformation of the in-phase multiplets (Szyperski *et al.*, 1992). The vicinal spin–spin coupling constants $^3J_{\alpha\beta}$ were measured in an E. COSY spectrum (Griesinger *et al.*, 1985) recorded in ^2H_2O at 500 MHz, where 1200 experiments were recorded in t_1 ($t_{1max} =$ 120 ms) with 4096 points in t_2 ($t_{2max} = 410$ ms). The E. COSY data set was multiplied by a sine bell (DeMarco & Wüthrich, 1976) phase-shifted by $\pi/4$ in both dimensions, and then zero-filled to 2048 points in t_1 and 16,384 points in t_2. In all spectra, baseline distortions were reduced by subtraction of suitable 3rd-order polynomials from the frequency domain data, using a standard Bruker routine.

(c) *Determination of the three-dimensional structure*

The strategy followed for the solution structure determination of BPTI is similar to that for the structure determination of the *Antp*(*C39→S*) homeodomain (Güntert *et al.*, 1991*b*). The 2D n.m.r. spectra were analyzed using the program package EASY (Eccles *et al.*, 1991). ^{1}H chemical shifts for BPTI were taken from previous work (Wagner *et al.*, 1987; Tüchsen & Woodward, 1987*a*) and adjusted for the effects of inevitable slight variations in experimental conditions. Distance constraints derived from NOESY spectra were collected in several rounds of cross-peak assignments and structure calculations, using the program ASNO (K. Berndt, P. Güntert & K. Wüthrich, unpublished results). NOESY cross-peak volumes were converted to upper distance limits with the program CALIBA (Güntert *et al.*, 1991*a*,*b*). The program HABAS (Güntert *et al.*, 1989) was used to determine stereospecific assignments for β-methylene protons and to generate dihedral angle constraints based on local NOE distance constraints and experimental data on the scalar coupling constants $^3J_{HN\alpha}$ and $^3J_{\alpha\beta}$. Additional stereospecific assignments were derived using the program GLOMSA (Güntert *et al.*, 1991*a*,*b*) to compare the upper distance limits with the molecular geometry of the calculated conformers. Structure calculations were done with the program DIANA (Güntert *et al.*, 1991*a*), using the REDAC strategy (Güntert & Wüthrich, 1991) for improved convergence. The 20 conformers with smallest target function values after the final DIANA calculation were subjected to restrained energy minimization using a modified version of the program AMBER (Singh *et al.*, 1986; Weiner *et al.*, 1986), which includes pseudo-energy terms for distance constraints and dihedral angle constraints. The pseudo-energy was proportional to the 6th power of the distance constraint violations, and adjusted such that

violations of 0·1 Å (1 Å = 0·1 nm) for distance constraints or 2·5° for angle constraints corresponded to $kT/2$ at room temperature (Billeter *et al.*, 1990). The resulting 20 energy-minimized DIANA conformers are used to represent the solution structure of BPTI.

(d) *Structure comparisons*

For visual comparison of different structures, stereo views were produced either with the molecular graphics programs CONFOR (Billeter *et al.*, 1985; Xia, 1992), or Midas Plus (Ferrin *et al.*, 1988) or with the structure analysis program XAM (Xia, 1992) on a Sun 4 workstation. For pairs or groups of conformers, global superpositions and RMSD values for various subsets of atoms were computed as usual (McLachlan, 1979). The mean solution conformation was obtained by first superimposing the 20 energy-minimized DIANA conformers so as to minimize the RMSD for the backbone atoms N, C^{α} and C′ of residues 2 to 56, and then averaging the Cartesian co-ordinates of corresponding atoms in the 20 globally superimposed conformers. Displacements, D (Billeter *et al.*, 1989), were used to quantify the local precision of the solution structure and to identify local differences between crystal and solution structures. Displacements are a generalization of the usual RMSD values, since the set of atoms used for the superposition of the conformers, M_{sup}, differs from the set of atoms for which the root-mean-square deviation of the positions is actually calculated, M_{RMSD}. For example, for the the evaluation of the backbone displacement after global superposition for a given residue i in BPTI, D_{glob}, M_{sup} consists of the backbone atoms N, C^{α} and C′ of residues 2 to 56 and M_{RMSD} includes the backbone atoms N, C^{α} and C′ of residue i.

3. Results

(a) *Collection of conformational constraints*

Vicinal spin–spin coupling constants, $^3J_{HN\alpha}$, were measured for 46 of the 47 non-glycine and non-proline residues in BPTI, using inverse Fourier transformation of in-phase multiplets measured in the 40 millisecond NOESY spectrum in H_2O (Szyperski *et al.*, unpublished results). The $^3J_{HN\alpha}$ value for Cys38 could not be measured because the amide proton resonance was broadened and all cross-peaks involving this resonance were overlapped with other peaks. For all six β-branched amino acids the $^3J_{\alpha\beta}$ coupling constants were measured using inverse Fourier transformation of in-phase multiplets measured in the 40 millisecond NOESY spectrum in ^2H_2O. Values of $^3J_{\alpha\beta}$ for side-chains with β-methylene protons were obtained for 26 out of 36 residues. The $^3J_{\alpha\beta}$ coupling constants of the five residues Arg1, Leu29, Gln49, Asp50 and Arg53 are indicative of conformational averaging about χ^1 (Nagayama & Wüthrich, 1981). These coupling constants were not used to generate dihedral angle constraints for the structure calculation, and in addition, all NOE distance constraints to $H^{\beta 2}$ and $H^{\beta 3}$ of these five residues were referred, with the appropriate correction, to the pseudo-atom Q^{β} (Wüthrich *et al.*, 1983).

Of the 916 upper distance limit constraints identified in the two NOESY spectra, 274 were found to be irrelevant, that is, they are either independent of the conformation or there exists no conformation that would violate the constraint. The input data set for the final structure calculation thus contained 642 relevant NOE upper distance limits (Fig. 1), and 115 dihedral angle constraints (41 for ϕ, 41 for ψ and 33 for χ^1). Stereospecific assignments were obtained for 32 diastereotopic groups of protons, and individual assignments were found for the side-chain amide protons of one Gln and three Asn residues (Table 1). In the DIANA calculations, the three disulfide bonds 5–55, 14–38 and 30–51 were

Table 1

Individual ^1H n.m.r. assignments for methylene protons, amide protons and isopropyl methyl groups in BPTI

Residue	Chemical shift δ (p.p.m.)†		
	αH	βH	Others
Pro2		0·91, 2·02	γCH_2 1·61, 1·87, δCH_2 3·60, 3·74
Phe4		3·35, 2·95	
Leu 6			δCH_3 0·95, 0·86
Pro8		1·84, 2·44	
Pro9		0·22, 0·10	γCH_2 1·27, 0·17, δCH_2 3·35, 2·92
Gly12	3·26, 3·89		
Cys14		2·80, 3·46	
Lys15		1·58, 2·08	
Arg17		1·60, 1·47	
Arg20		0·82, 1·81	
Tyr21		2·71, 2·78	
Phe22		2·82, 2·91	
Tyr23		2·73, 3·47	
Asn24		2·17, 2·87	δNH_2 7·90, 7·08
Gly28	3·61, 3·91		
Cys30		3·67, 2·67	
Gln31		1·73, 2·17	εNH_2 7·43, 6·94
Phe33		2·95, 3·09	
Val34			γCH_3 0·71, 0·80
Tyr35		2·51, 2·67	
Gly37	2·90, 4·22		
Lys41		1·66, 2·24	
Arg42		1·03, 0·37	
Asn43		3·25, 3·35	δNH_2 7·98, 7·78
Asn44			δNH_2 7·84, 3·39
Phe45		2·80, 3·41	
Ser47		4·11, 3·86	
Cys51		2·88, 3·17	
Met52		2·05, 1·96	

The chemical shifts were measured at 36 °C and pH 4·6. A complete list of the chemical shifts of all residues under these conditions has been presented (Wagner *et al.*, 1987).

† For methylene protons or isopropyl methyl groups, the chemical shift of the proton or methyl group with the lower branch number is listed first, e.g. the β_2 proton, or the γ_1 methyl in Val. For the side-chain amide protons of Asn and Gln, the first value is the chemical shift of the proton that has the shorter distance to the nearest methylene carbon (C^{β} for Asn, C^{γ} for Gln). Individual assignments of diastereotopic β-methylene protons obtained using the program HABAS are indicated by underlined chemical shift values, the other assignments resulted from analysis of the DIANA structures with the program GLOMSA (see the text for details).

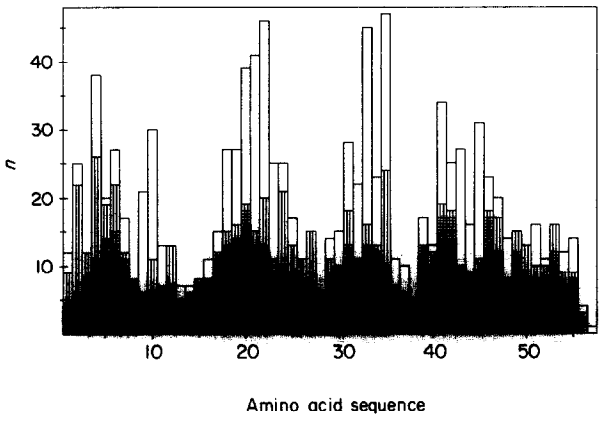


Figure 1. Plot *versus* the amino acid sequence of the number (n) and types of NOE constraints per residue used in the calculation of the solution structure of BPTI. The following code is used to define the different ranges of the NOE constraints: black, intraresidual; cross-hatched, sequential; vertically hatched, medium-range; white, long-range.

explicitly constrained by nine upper and nine lower distance limits (Williamson *et al.*, 1985). In contrast to the data set used previously to characterize the solution structure of BPTI (Wagner *et al.*, 1987), no additional constraints were used to enforce hydrogen bonds implicated by the independently identified (Wüthrich *et al.*, 1984) regular secondary structures (Williamson *et al.*, 1985).

(b) *Structure calculation*

For the final DIANA calculation, the REDAC strategy (Güntert & Wüthrich, 1991) was used. The yield of converged conformers was such that 35 out of a total of 50 calculations yielded conformers with final target function values of less than 1·0 Å^2. The complete DIANA calculation for all 50 conformers took 36·3 minutes of c.p.u. time on a Cray Y-MP computer. An analysis of the 20 conformers with smallest final target function values is afforded by Table 2, which also includes a comparison of this group of conformers before and after restrained energy minimization. The best 20 DIANA conformers before energy minimization have low

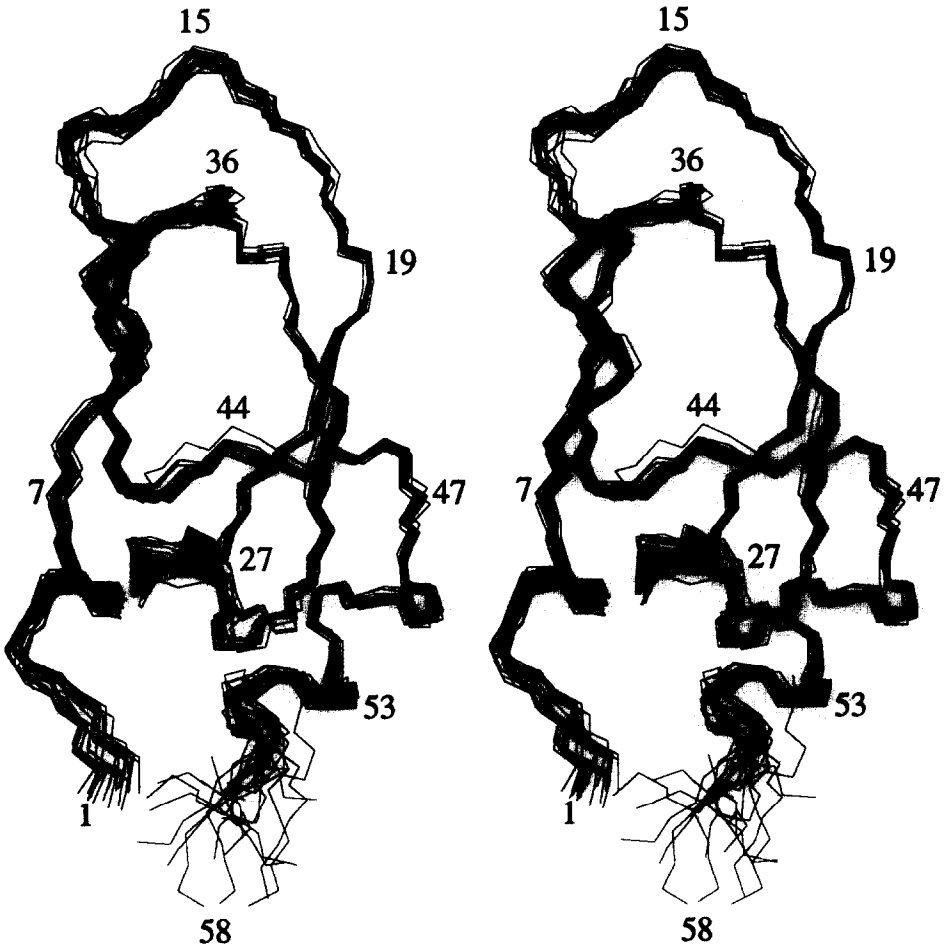


Figure 2. Stereo view of the polypeptide backbone of the 20 final, energy-refined DIANA conformers used to represent the n.m.r. structure in solution. The conformers 1 to 20 were superimposed for pairwise minimum RMSD of the backbone atoms N, C^{α} and C′ of residues 2 to 56 with the mean conformer. Some C^{α} positions are identified by the sequence numbers of the residues.

Table 2

Analysis of the 20 best BPTI structures obtained with DIANA before and after restrained energy minimization with the program AMBER

Quantity	Average value ± standard deviation (range)†			
	Before energy minimization		After energy minimization	
Target function (Å²)‡	0·40 ± 0·10	(0·21 ... 0·57)		
AMBER energy (kcal/mol)	−85 ± 50	(−171 ... 38)	−734 ± 77	(−869 ... −601)
NOE constraint violations:				
Number >0·2 Å	0·9 ± 0·8	(0 ... 2)	0	(0 ... 0)
Sum (Å)	3·0 ± 0·4	(2·2 ... 3·5)	4·6 ± 0·4	(3·6 ... 5·7)
Maximum(Å)	0·22 ± 0·06	(0·14 ... 0·42)	0·09 ± 0·00	(0·09 ... 0·10)
Dihedral angle constraint violations:				
Number >5°	0·1 ± 0·3	(0 ... 1)	0	(0 ... 0)
Sum (deg.)	5·8 ± 3·3	(2·4 ... 9·9)	9·6 ± 2·1	(6·6 ... 14·9)
Maximum (deg.)	2·2 ± 1·0	(1·0 ... 3·5)	2·3 ± 0·3	(1·7 ... 2·9)

A total of 50 conformers was calculated with DIANA using REDAC, but only the 20 conformers with the smallest final target function values were subjected to energy minimization. 1 cal = 4·184 J.

† The average of the individual values for the 20 conformers is listed.

‡ The weighting factors for the NOE distance constraints were 1, for the van der Waals' lower distance limits 2, and for dihedral angle constraints 5 Å². For the conformers after energy minimization, the target function is not defined because these conformers do not have the ECEPP standard geometry.

target function values (below 0·6 Å²) and fulfill the NOE distance constraints and dihedral angle constraints almost perfectly. The moderate AMBER energies of the conformers before energy minimization indicate that most steric overlaps were already removed by DIANA. For all 20 DIANA conformers, AMBER was able to find a low-energy conformation with nearly identical molecular geometry, with only a small increase of the sum of constraint violations. The average of the 20 pairwise RMSD values between the corresponding conformers before and after energy-minimization is only 0·25(±0·08) Å for the backbone atoms N, C^{α} and C', and 0·37(±0·12) Å for all heavy atoms in the complete polypeptide chain 1 to 58. The energy-minimized conformers exhibit smaller maximal violations of NOE distance constraints (Table 2) because the corresponding penalty term in the modified AMBER program is steeper than in DIANA (Billeter *et al.*, 1990).

(c) *Assessment of the quality of the solution structure of BPTI*

The well-known three-dimensional structure of BPTI contains a 3_{10}-helix of residues 3 to 7, a β-hairpin of residues 18 to 35, a one-residue third antiparallel β-strand involving residue 44, and an α-helix of residues 47 to 56 (Deisenhofer & Steigemann, 1975; Wagner *et al.*, 1987). A visual impression of the quality of the n.m.r. structure determination is afforded by Figures 2 and 3, which show superpositions of the 20 energy-refined DIANA conformers. The RMSD values for the whole protein (residues 1 to 58) are somewhat larger than for residues 2 to 56 (Table 3) due to the scarcity of conformational constraints for the chain ends, residue 1 and residues 57 to 58 (Fig. 1). The small RMSD values for residues 2 to 56 are representative of a high-quality n.m.r. structure (Wüthrich, 1989*a*). A large number of side-chains are well-determined in the solution structure. We identified 28 "best-defined" residues, for which the local side-chain displacement values after a global backbone superposition are below a threshold of 0·8 Å. The RMSD values calculated for the backbone atoms of residues 2 to 56 and the heavy atoms of these 28 best-defined side-chains are virtually

Table 3

Quantitative characterization of the structure determination of BPTI by n.m.r. in solution and comparison with the X-ray crystal structure of form II

Atom set used for comparison	Average RMSD (Å) ⟨n.m.r.⟩†-n.m.r.†	RMSD (Å) ⟨n.m.r.⟩†-X-ray‡
Backbone atoms of residues 1–58	0·73	1·35
All heavy atoms 1–58	1·07	1·81
Backbone atoms 2–56	0·43	0·76
Same + best-defined side-chains§	0·43	0·85
All heavy atoms 2–56	0·92	1·24
Backbone atoms 18–35(β-sheet)	0·27	0·58
All heavy atoms 18–35(β-sheet)	0·58	1·03
Backbone atoms 47–56(α-helix)	0·32	0·26
All heavy atoms 47–56(α-helix)	0·85	1·15

† ⟨n.m.r.⟩ denotes the average co-ordinates of the 20 conformers that represent the solution structure of BPTI, n.m.r. the individual conformers.

‡ From Wlodawer *et al.* (1984).

§ Includes the backbone heavy atoms N, C^{α} and C' of residues 2 to 56 and the side-chain heavy atoms of the 28 residues Pro2, Phe4, Leu6, Pro8, Pro9, Thr11, Ala16, Ile18, Ile19, Tyr21, Phe22, Tyr23, Asn24, Ala25, Ala27, Cys30, Thr32, Phe33, Val34, Tyr35, Cys38, Ala40, Arg43, Phe45, Ser47, Ala48, Cys51 and Thr54. See the text for details.

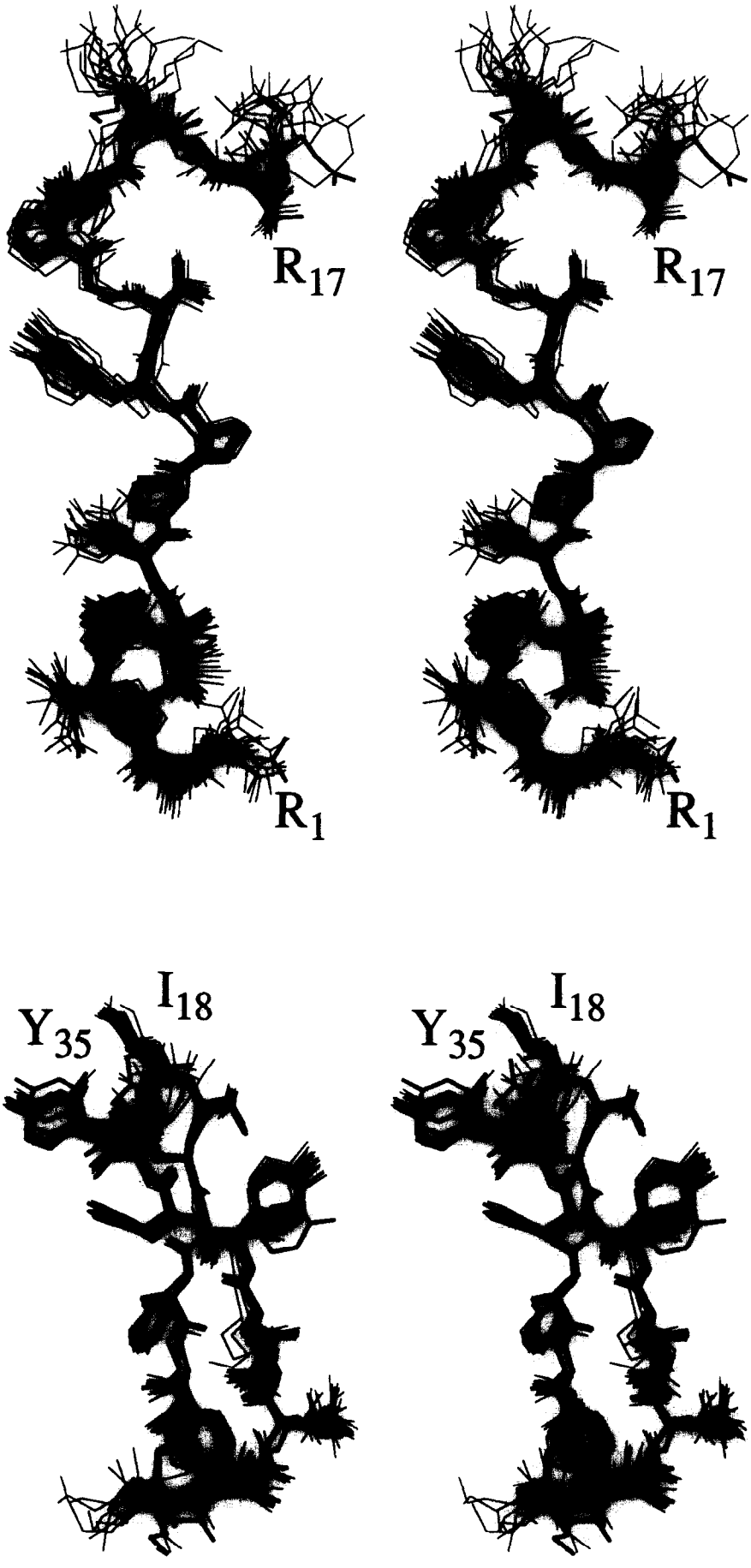


Fig. 3.

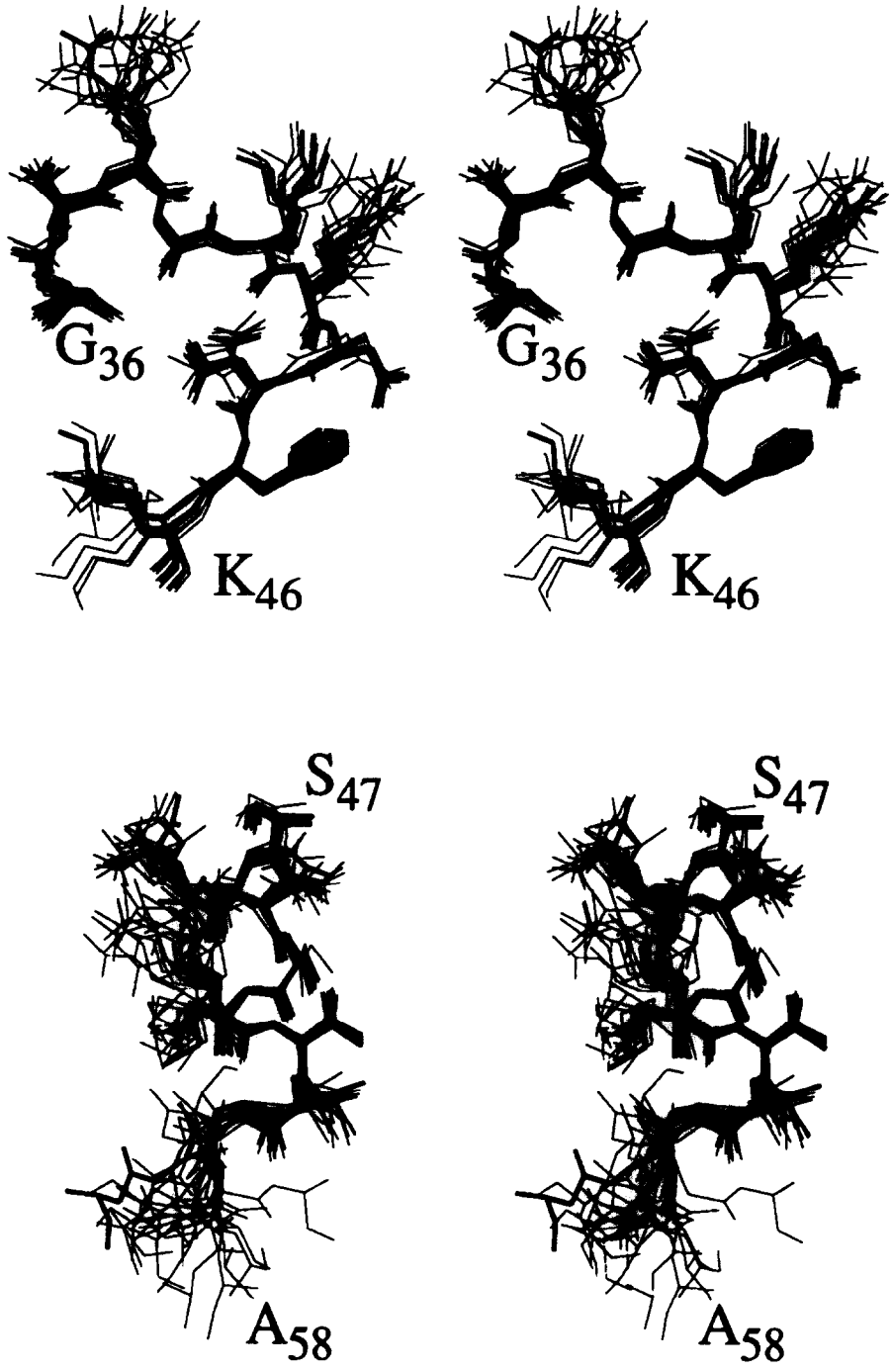


Figure 3. Stereo views of all-heavy-atom representations of the 4 segments 1 to 17, 18 to 35, 36 to 46 and 47 to 58 of the 20 final, energy-refined DIANA conformers used to represent the n.m.r. solution structure of BPTI. For each segment, the conformers 1 to 20 were individually superimposed with the mean conformer for pairwise minimum RMSD of the backbone atoms N, C^{α} and C'. Also shown for comparison, with a bold line, is the X-ray structure of form II after a similar superposition with the mean conformer. The first and last amino acid residue in each segment is identified by the sequence position and the one-letter amino acid code.

identical to those for the backbone atoms alone (Table 3). Other side-chains, in particular near the surface of the protein, are less well defined. The variable precision of the determination of individual side-chain conformations is illustrated in Figure 3, which shows an all-heavy-atom presentation of the 20 energy-refined DIANA conformers.

The n.m.r. structure determination includes a description of the intramolecular hydrogen bonds. A total of 30 hydrogen bonds was identified and are listed in Table 4. A hydrogen bond was identified if in ten or more of the 20 energy-refined solution conformers, the proton-acceptor distance is less than 2·4 Å and the angle between the donor–proton bond and the line connecting the acceptor and donor atoms is less than 35°. We consider a residue

Table 4
Hydrogen bonds observed in the solution structure and in the crystal structure form II of BPTI

Donor hydrogen†	Acceptor†	n.m.r.‡	Form II§
5 Cys NH	2 Pro O′	7	+
6 Leu NH	3 Asp O′	20	+
7 Glu NH	4 Phe O′	19	+
11 Thr γOH	34 Val O′	4	+
14 Cys NH	12 Gly O′	14	
16 Ala NH	36 Gly O′	0	+
18 Ile NH	35 Tyr O′	20	+
20 Arg NH	33 Phe O′	20	+
20 Arg ηNH_2	44 Asn δO	0	+
21 Tyr NH	45 Phe O′	19	+
22 Phe NH	31 Gln O′	20	+
23 Tyr NH	43 Asn δO	16	+
24 Asn NH	29 Leu O′	17	+
24 Asn δNH_2	31 Gln εO	12	+
26 Lys NH	24 Asn δO	20	+
27 Ala NH	24 Asn O′	20	
27 Ala NH	24 Asn δO	0	+
28 Gly NH	24 Asn O′	20	+
31 Gln NH	22 Phe O′	20	+
33 Phe NH	20 Arg O′	20	+
35 Tyr NH	18 Ile O′	20	+
35 Tyr ηOH	38 Cys O′	12	
36 Gly NH	11 Thr O′	15	+
43 Asn δNH_2	7 Glu O′	15	+
43 Asn δNH_2	23 Tyr O′	17	+
44 Asn NH	42 Arg O′	20	+
45 Phe NH	21 Tyr O′	20	+
49 Glu NH	49 Glu εO	0	+
50 Asp NH	47 Ser γO	16	+
51 Cys NH	47 Ser O′	19	+
52 Met NH	48 Ala O′	20	+
53 Arg NH	49 Glu O′	20	+
54 Thr NH	50 Asp O′	20	+
54 Thr γOH	50 Asp O′	12	+
55 Cys NH	51 Cys O′	18	+
56 Gly NH	52 Met O′	15	+

† All groups are listed for which a hydrogen bond was identified either in the crystal structure, in the solution structure, or both. NH and O identify the corresponding backbone atoms; side-chain atoms are indicated with Greek letters. Amide protons with exchange rates slower than 0·001 min^{-1} at pH 3·5 and 36 °C (Wagner & Wüthrich, 1982*b*), are shown in boldface type. Using this criterion, the amide protons of Leu29 and Lys41 were also found to be slowly exchanging but are not found to be hydrogen-bonded to protein atoms (see the text).

‡ Hydrogen bonds were identified whenever the proton–acceptor distance was less than 2·4 Å and the angle between the donor–proton bond and the line connecting the donor and acceptor atoms is less than 35°. The numbers indicate how many among the 20 energy-minimized DIANA conformers meet this criterion.

§ The co-ordinates of the neutron diffraction structure of crystal form II (structure identification code of the Protein Data Bank: 5PTI), were used in this analysis. A plus sign indicates that a hydrogen bond had been identified.

to be a part of a secondary structure element if its backbone is involved in the regular hydrogen bonding network characteristic of that secondary structure. Using this criterion, four components of regular secondary structure are present in the solution structure of BPTI. These include residues 3 to 7 in a 3_{10} helix, residues 18 to 35 in a β-hairpin, residue 45, which makes an antiparallel β-strand contact to residue 21, and residues 47 to 56 in an α-helix.

4. Discussion

This section is devoted to comparisons between BPTI structures in crystals and in solution. As mentioned in the Introduction, three crystal structures in three different crystal forms are known, so that structure variations associated with the different crystal lattices can be used as a reference for judging the significance of structural differences between any of the crystal structures and the solution structure.

The three high-resolution crystal structures of BPTI as determined by diffraction techniques are form I (Brookhaven Protein Data Bank entry name 4PTI; Deisenhofer & Steigemann, 1975), form II (Brookhaven Protein Data Bank entry name 5PTI; Wlodawer *et al.*, 1984) and form III (Brookhaven Protein Data Bank entry name 6PTI; Wlodawer *et al.*, 1987*b*). A superposition of the three structures (Fig. 4) shows that they are closely similar. Pairwise RMSD values of only 0·37 to 0·40 Å for the backbone atoms and 0·96 to 1·10 Å for all heavy atoms of the polypeptide chain from residues 2 to 56 quantify this visual impression. With the exception for the terminal two or three residues at both chain ends, where high values of the *B*-factors indicate significantly increased structural disorder (Fig. 5(a)), the entire polypeptide backbone is precisely defined in all three structures. The largest local deviations in backbone atom co-ordinates are found in two regions: in the vicinity of Lys15 in the center of the antiprotease loop and around Lys26 in the loop between the two strands of the β-hairpin. A more detailed comparison between the three X-ray crystal structures has been described (Wlodawer *et al.*, 1987*a*,*b*). In view of the close similarity of the three crystal structures, we decided to use the structure of form II as the representative X-ray structure for comparisons with the n.m.r. solution structure. This choice has the additional advantage that the positions of the hydrogen atoms in this crystal structure are known from neutron diffraction data (Wlodawer *et al.*, 1984).

As indicated by a variety of n.m.r. observations (surveyed by Wagner *et al.*, 1987), there is also a close similarity between the structure of BPTI in aqueous solution and in single crystals. This is quantitatively confirmed by the small RMSD values of 0·76 Å and 1·24 Å for the backbone atoms and for all heavy atoms of residues 2 to 56, respectively (Table 3). The superposition in Figure 5(a) shows that all regular secondary structure elements found in the crystal structure of BPTI are also present in the solution structure. Deviations in the polypeptide backbone folds of the two structures can be identified for residues 25 and 26 in the loop connecting the two strands of the β-hairpin, for residues 46 to 48 of the α-helix, and for the amino and carboxy-terminal two to three residues. The

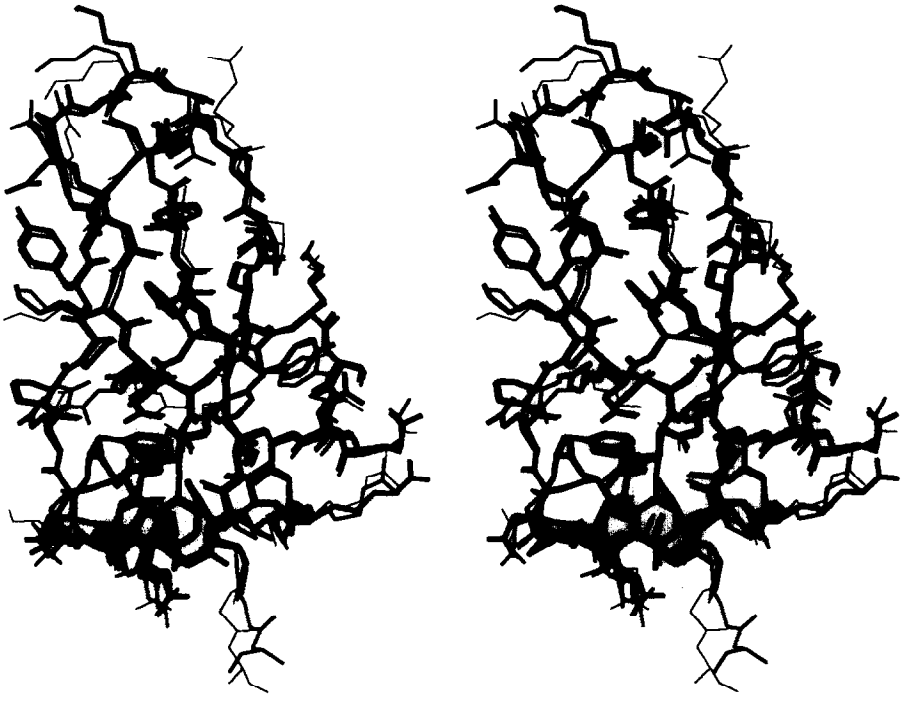


Figure 4. Stereo view of the all-heavy-atom representations of the 3 crystal structures of BPTI after superposition for minimal RMSD of the backbone atoms N, C^{α} and C′ of residues 2 to 56. Thin line: form I (Deisenhofer & Steigemann, 1975); medium line: form II (Wlodawer *et al.*, 1984); thick line: form III (Wlodawer *et al.*, 1987*b*). Same orientation of the molecule as shown in Fig. 2.

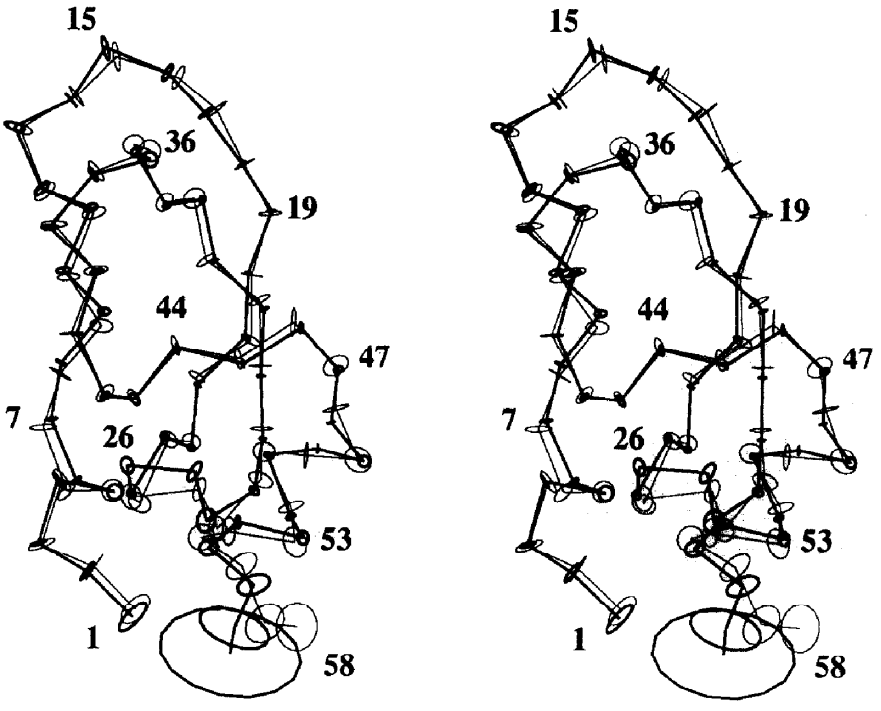
presentation of Figure 5(a), with circles around the α-carbon positions representing the mean deviation for the backbone atoms of the individual residues as calculated from the *B*-factors of the X-ray structure (Glusker & Trueblood, 1985) or the global backbone displacements of the n.m.r. solution structure, respectively, shows that these structural differences are significant in the sense that the areas of conformational space defined by the two methods do not overlap. Overall, however, in spite of these small differences, the n.m.r. solution structure and the X-ray crystal structure are found to coincide very closely. Figure 5(b) shows that this holds also for the side-chains of the residues that are best-defined in the n.m.r. structure (see also Table 3). In view of this close global coincidence of the crystal and solution structures, the remainder of this section is devoted to detailed examination of local structural differences in the polypeptide backbone and in individual amino acid side-chains.

(a) *Violations of NOE distance constraints in the X-ray crystal structure*

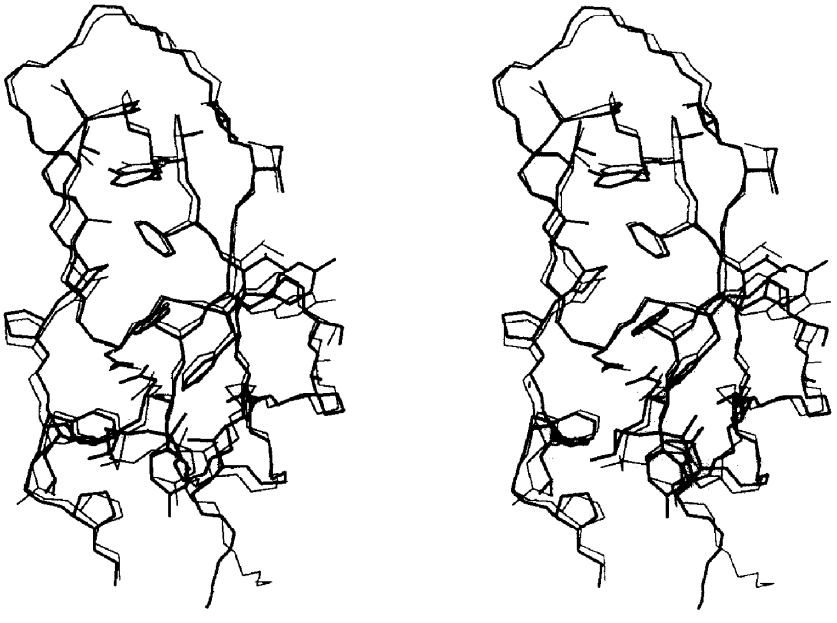
A search of the X-ray crystal structure for violations of experimental n.m.r. constraints measured in solution relates differences between the structures in solution and in crystals in a direct way with the experimental conformational constraints. Overall, 95% of the NOE distance constraints used to calculate the n.m.r. solution structure of BPTI are consistent with the X-ray crystal structure of form II, and only 28 violations greater than 0·5 Å were found, (i.e. 3 intraresidual, 7 sequential, 4 medium-range and 14 long-range NOEs). Figure 6 shows that these 28 violations are distributed over the entire protein molecule. The largest violation of 3·9 Å involves the side-chain of Arg17, which is on the protein surface and is involved in intermolecular hydrogen bonds with neighboring protein molecules in the crystal lattice. It is worth noting that this side-chain adopts different conformations in the three crystal forms (Fig. 4). There are two violated constraints in the loop between the two β-strands, namely, Asn24 Hβ2–Lys26 HN and Asn24 Hβ3–Ala27 NH. In the orientation of the BPTI molecule in Figures 5 and 6, this loop linking the two β-strands is pulled upwards in the solution structure when compared to the crystal structure.

(b) *Detailed comparison of the polypeptide backbone conformation in solution and in crystals*

With the exception of the N-terminal residue and the C-terminal dipeptide segment, the polypeptide backbone of the solution structure of BPTI is well determined by the n.m.r. data (Figs 2 and 5(a); Table 3). The polypeptide backbone is also well defined in the three crystal structures (Figs 4 and 5(a)). The local precision of the backbone atom positions along the amino acid sequence can be assessed from Figure 7. For each of the three X-ray structures, the square root of the average *B*-factors for the backbone atoms is plotted *versus* the sequence (Fig. 7(a)) using dotted, short-dash and long-dash lines for the crystal forms, I, II and III,



(a)



(b)

Figure 5. Stereo views of BPTI in aqueous solution and in single crystals in the same orientation as shown in Fig. 2. (a) Superposition of the mean solution conformation (bold line) and the crystal structure of form II (thin line). The straight lines connect the C^{α} positions. The radii of the circles drawn about each C^{α} position are, respectively, equal to the mean of the backbone displacements after global superposition, D_{glob}, of the 20 energy-minimized DIANA conformers relative to the mean solution conformation, and to the backbone displacements, $\sqrt{\langle \Delta x^2 \rangle}$, calculated from the average of the crystallographic B-factors for the backbone atoms N, C^{α} and C′ using the relation $B = 8\pi^2 \langle \Delta x^2 \rangle$. (b) Superposition for minimal RMSD of the backbone atoms N, C^{α} and C′ of residues 2 to 56 of the mean solution structure (bold line) and the crystal structure form II (thin line). In addition to the backbone atoms N, C^{α} and C′ of residues 1 to 58, all heavy atoms are shown for the 28 residues with "best defined" side-chains in the solution conformation (listed as the footnotes in Table 3).

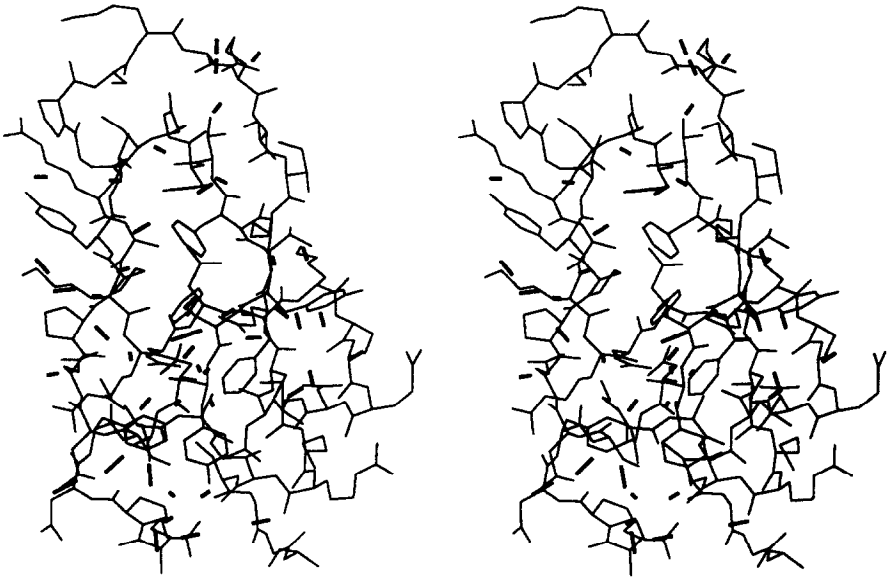


Figure 6. Location of NOE upper distance constraints measured for BPTI in solution that are violated by more than 0·5 Å in the X-ray crystal structure of BPTI form II. An all-heavy-atom representation of the X-ray crystal structure in the same orientation as shown in Fig. 2 is drawn with thin lines. NOE distance constraint violations are indicated with thick lines. The direction of these lines is that of a vector connecting the 2 hydrogen positions linked by the NOE, the length is proportional to the extent of the violation. These NOEs and the corresponding distances in the crystal structure are listed in Table 5.

respectively. The square root of the crystallographic *B*-factor is proportional to the mean displacement of its atomic co-ordinates (Glusker & Trueblood, 1985) and is useful in assessing the relative precision of the X-ray crystal structure determination. Differences between solution and crystal structures become apparent from the plots of backbone heavy-atom displacements after global superposition, D_{glob} (Fig. 7(b)), RMSD values after local superposition of tripeptide segments, $RMSD_{loc}$ (Fig. 7(c)). Continuous lines represent the average displacements between the 20 n.m.r. conformers and the mean solution conformation, and are thus a measure of the precision of the solution structure determination. The dotted, short-dash and long-dash lines represent the displacements or $RMSD_{loc}$ between the X-ray crystal structures of forms I, II and III, respectively, and the mean atom co-ordinates of the n.m.r. structure. These values are a measure of the similarity of the structures after pairwise global (Fig. 7(b)) or pairwise local (Fig. 7(c)) superpositions. It is thus seen that on the level of tripeptide segments, there is a close coincidence between each of the three crystal forms and the group of 20 n.m.r. structures (Fig. 7(c)). Differences between crystal and solution structures become apparent in the plots of displacements after global superposition (Fig. 7(b)). In addition to the terminal residues, five regions of the polypeptide chain show global backbone displacements greater than 1·0 Å in at least one of the three crystal forms (i.e. residues 9, 15 to 17, 25 to 28, 46 to 48 and 53 to 56). These can be judged significant in comparison with the displacements, D_{glob}, calculated for the group of 20 solution structures. Differences between the three crystal forms are also seen in a number of these regions. The absence of corresponding variations after local superposition (Fig. 7(c)) suggests that the observed differences are due to the displacement of an entire segment of the polypeptide chain with respect to the rest of the protein molecule, rather than to local structural variations on the level of the backbone dihedral angles.

The aforementioned observation on the global origin of apparent local structure variations is substantiated by Figure 8, where the values of the dihedral angles ϕ and ψ observed in the three crystal structures of BPTI are compared with the range of values for these angles in the 20 energy-refined DIANA conformers. As expected from the high quality of the solution structure (Fig. 2 and Table 3), the range of the backbone dihedral angles for a given residue is highly restricted for most of the polypeptide chain. A wide range of values for ϕ and ψ is found for only a few residues in the solution structure (i.e. 14, 29, 38 and 56 to 58, and 13, 28, 37 and 55 to 57, respectively). Clearly, wherever a wide range of ψ values was observed, this was accompanied by a concomitant wide range of ϕ values for the following residue. The glycyl residues 28, 37, 56 and 57 account for most of these poorly defined ψ angles. The poorly defined values of ψ and ϕ in adjacent residues may thus result from the inherent increased flexibility of glycyl residues in a polypeptide chain, or simply from a scarcity of n.m.r. constraints due to the absence of side-chain hydrogen atoms in Gly. Figure 8(a) and (b) further show that with the exception of the C-terminal dipeptide segment, all values for the backbone dihedral angles in the three X-ray crystal structures are nearly

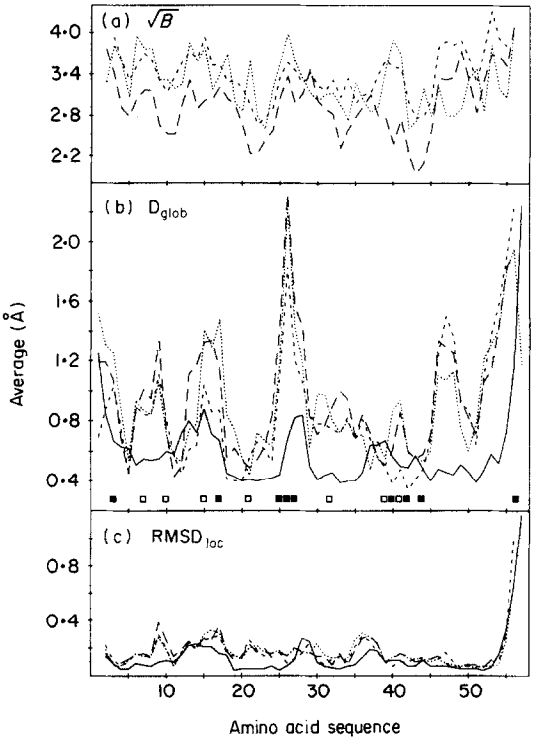


Figure 7. Comparison of local backbone conformation in the n.m.r. solution structure and the X-ray crystal structure of BPTI form II. The following quantities are plotted *versus* the amino acid sequence: (a) the *B*-factors of the 3 X-ray crystal structures of BPTI are shown as dotted, short-dash and long-dash lines for forms I, II and III, respectively. For each residue the average of the *B*-factors of the backbone atoms N, C^{α} and C′ is given and these values are connected by a continuous line. Since the mean displacements of the atomic positions are proportional to the square root of the *B*-factor (Gluster & Trueblood, 1985), the ordinate is scaled according to $\sqrt{B}$. (b) The average of the global backbone displacements, D_{glob}, of the 20 energy-minimized DIANA conformers relative to the mean n.m.r. structure and each of the 3 X-ray crystal structures (dotted, short-dash and long-dash lines for forms I, II and III, respectively). Residues involved in intermolecular hydrogen bonds in the X-ray crystal structure of form II are indicated by filled squares representing hydrogen bonds involving backbone atoms, and by open squares representing hydrogen bonds involving at least 1 amino acid side-chain. (c) The average of the local backbone RMSD for tripeptide segments, $RMSD_{loc}$, between the 20 n.m.r. conformers and the mean solution conformation, which are connected by a continuous line. The corresponding local RMSD values between each of the 3 X-ray crystal structures (forms I, II and III) and the mean n.m.r. structure are connected by dotted, short-dash and long-dash lines, respectively.

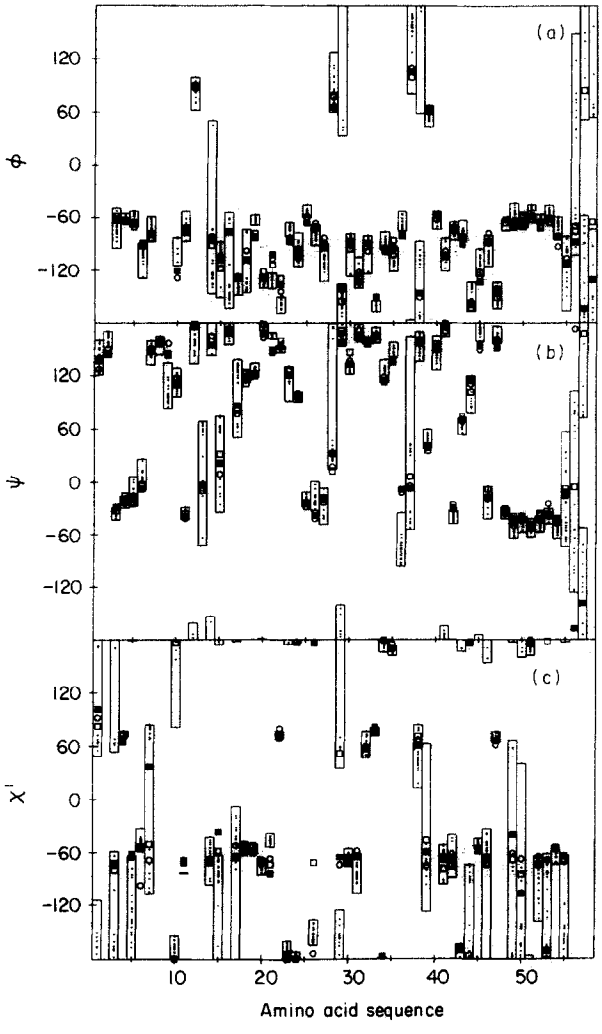


Figure 8. Plots of the dihedral angles ϕ, ψ and χ^1 *versus* the sequence of BPTI. Values for each of the 20 energy-refined DIANA conformers are represented by points and the range of these values is represented by a bar. The values for the 3 X-ray structures of the crystal forms I, II and III are given as open squares, filled squares and open circles, respectively.

identical. There is also a close correlation between the values of ϕ and ψ obtained from each of the three crystal forms of BPTI and those obtained from the group of energy-refined DIANA conformers used to represent the n.m.r. structure in solution. That is, for almost all residues, the ϕ and ψ dihedral angles of each of the three crystal forms lie within, or very near, the range of values for the 20 n.m.r. conformers.

(c) *Hydrogen bonds*

Table 4 lists the intramolecular hydrogen bonds identified with identical criteria (see footnotes to Table 4) in the solution structure and the crystal structure form II of BPTI. Nearly identical hydrogen bonding patterns were observed in the three crystal structures of BPTI, which was presented as evidence for the close similarity of these structures (Wlodawer *et al.*, 1987*a*,*b*). In crystal form II, 34 hydrogen bonds were identified with our criteria. In the solution structure, 30 hydrogen bonds were identified in at least half of the 20 energy-refined DIANA conformers. Twenty-eight hydrogen bonds in the crystal structure are identical to those observed in the n.m.r. solution structure. Two crystal structure hydrogen bonds, 5 Cys NH–2 Pro O′ and 11 Thr γOH–34 Val O′, are present in

some of the solution conformers (Table 4), indicating that a hydrogen bond between these groups would be compatible with the n.m.r. data. Three hydrogen bonds, 16 Arg NH–36 Gly O′, 20 Arg ηNH_2–44 Arg δO, and 49 Glu NH–49 Glu εO identified in the crystal structure have no counterpart in the solution structure. Likewise, two hydrogen bonds, 14 Cys NH–12 Gly O′ and 35 Tyr HH–38 Cys O′, are found only in the solution structure. The amide proton of residue Ala27 is observed to hydrogen bond to the backbone carbonyl oxygen of Asn24 in the solution structure and to the side-chain carbonyl oxygen of Asn24 in the crystal structure, manifesting a local difference between the two structures.

Of particular interest are two side-chain–backbone hydrogen bonds involving the amide proton of Asp50 and the hydroxyl oxygen of Ser47, and the hydroxyl proton of Thr54 and the carbonyl oxygen of Asp50, respectively (Table 4). Such side-chain interactions with amide protons in the first turn or carbonyl oxygens in the last turn of an α-helix are thought to stabilize the helical conformation (Presta & Rose, 1988; Richardson & Richardson, 1988). The fact that Ser47 and Thr54 are among the best-defined side-chains in the n.m.r. solution structure (Table 3), in spite of their location on the surface of the molecule, is evidence that these hydrogen bonds are highly populated in the solution structure.

Amide protons involved in hydrogen bonds typically experience slowed exchange rates (Wüthrich, 1986) and provide independent evidence for regular secondary structures in proteins. Table 4 correlates hydrogen bonds observed in both the solution structure and the X-ray crystal structure of form II with amide proton exchange rates slower than 0·001 min^{-1} (Wagner & Wüthrich, 1982*b*). With the exception of a few residues at the end of helical segments (i.e. Cys5, Asp50 and Gly56), all amide protons participating in hydrogen bonds in regular secondary structures in BPTI were thus found to be slowly exchanging. Residues present in the turn between the two strands of the β-hairpin in which the amide protons are hydrogen-bonded, i.e. Lys26, Ala27 and Gly28, display more rapid exchange. Only two residues possessing slowly exchanging amide protons (Leu29 and Asn41) were not found to be hydrogen-bonded to any protein atom. These two amide protons have low solvent accessibility in the X-ray crystal structure (Wagner & Wüthrich, 1982*b*), which could explain their slow exchange behavior in the absence of a hydrogen bond. Hydrogen-bonded amide protons not involved in secondary structures, particularly those whose acceptors are side-chain atoms, often display intermediate or fast exchange rates compared to those in regular secondary structures. Overall, the near identity of the hydrogen bonds in the crystal and solution structures, in particular in the regular secondary structures (Table 4) coupled with the good agreement with the amide proton exchange data, presents clear evidence that the same molecular architectures are present in the two environments.

(d) *Comparison of amino acid side-chain conformations*

For the 28 best-defined side-chains in the solution structure (Table 3), one observes narrow ranges of χ^1 values for the 20 energy-refined DIANA conformers (Fig. 7(c)). In most instances these cover the corresponding X-ray structure values, but there are also small differences for some of these side-chains. For example, the ring of Tyr21, which is well determined in the group of solution conformers, assumes a slightly different position in the X-ray crystal structure (Fig. 3). This side-chain is also somewhat different in the three crystal structures (Fig. 4), which may be due to the fact that in the forms I and II the phenolic proton in involved in intermolecular hydrogen bonds.

The topology and chirality of the three disulfide bonds in BPTI is conserved in the three crystal forms. The disulfide bonds Cys5–Cys55 and Cys30–Cys51 form left-handed spirals (Richardson, 1981), while Cys14–Cys38 forms an unusual right-handed hook (Richardson, 1981) in which the t^2g^3 ($\chi^1 \approx 60°$) rotamer position of the side-chain of Cys38 enforces an unfavorable steric interaction between the sulfur atom and the peptide backbone. In the n.m.r. structure determination, explicit upper and lower bound constraints (Williamson *et al.*, 1985) were used to enforce the covalent structure of the disulfide bond. Additional structure calculations showed that even without these constraints, a correctly folded molecule was obtained, from which the correct disulfide pairing could be unambiguously derived. However, the disulfide bond chirality could be determined only for Cys30–Cys51, which is sandwiched between the α-helix and the β-sheet; this local structure coincides with that found in all three crystal forms (Fig. 9). The chiralities of Cys5–Cys55 and Cys14–Cys38 are largely undefined in the solution structure, with both right and left-handed forms present in the group of 20 refined DIANA conformers (Fig. 9). For these four cysteinyl residues, stereospecific assignments of the β-proton could be obtained only for Cys14 (Table 1), indicating that the lack of definition of these local structures in the solution conformation might reflect genuine structural disorder.

Discrete, multiple conformations for individual side-chains were observed in each of the three crystal forms: for Glu7 in forms I and II, Met52 in form II and Arg39 and Asp50 in form III. In the n.m.r. solution structure the observed coupling constants $^3J_{\alpha\beta 2}$ and $^3J_{\alpha\beta 3}$ for Asp50 indicate conformational averaging about χ^1 (Nagayama & Wüthrich, 1981), with the conformation range in solution covering the two forms found in the crystal structure. Arginine 39 is only poorly defined (Fig. 3), and the range of n.m.r. conformers also covers the three crystal conformations. The side-chain of Glu7 shows two preferred orientations in

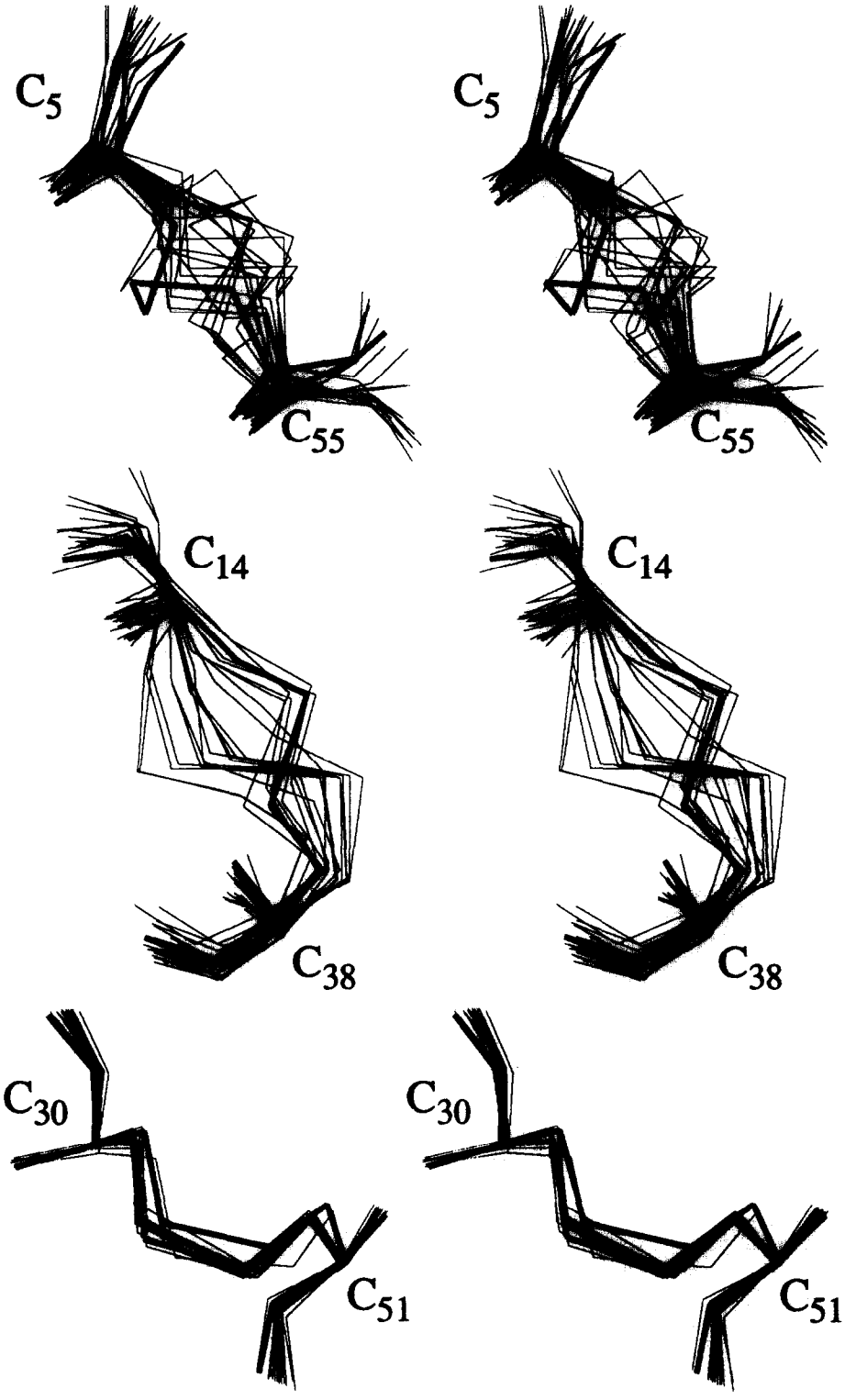


Figure 9. Stereo views of the 3 disulfide bonds observed in the X-ray crystal structure and the n.m.r. solution structure of BPTI. Shown are the all-heavy-atom representations of the crystal form II (bold lines) and the 20 refined DIANA conformers used to represent the n.m.r. solution structure (thin lines) after superposition of the backbone atoms N, C^{α} and C′ for minimal RMSD.

the group of 20 solution structures (Fig. 3), which coincide roughly with the two χ^1 rotamer positions displayed by the three crystal structures (Fig. 8(c)). Residual violations of NOE-derived distance constraints for the side-chain of Glu7 in the crystal form II (Table 5), in which both of these rotomer positions are occupied in the same crystal, show that neither of the two conformations represent a good fit of the n.m.r. data. For Met52 the more highly populated conformer in the crystal structure

Table 5

NOE distance constraints used for the n.m.r. structure determination of BPTI in solution which are violated by more than 0·5 Å in crystal structure form II

Atom 1†	Atom 2†	NOE limit (Å)‡	Violation (Å)§
Arg1 Qβ	Gly57 HN	5·5	1·4
Arg1 Hε	Cys5 Qβ	4·3	1·5
Pro2 Hγ2	Cys5 HN	4·5	0·5
Pro2 Hγ3	Cys55 Hα	4·5	1·0
Cys5 HN	Leu6 Qδ1	4·3	2·2
Glu7 Hβ3	Asn43 Hδ21	4·5	0·7[a]
Glu7 Hγ2	Asn43 Hβ3	4·5	1·9[b]
Glu7 Hγ3	Asn43 Hβ3	4·5	1·6[b]
Ala16 Hα	Arg17 QH2	5·5	3·9
Arg17 Hβ2	Val34 Qγ1	3·9	0·7
Tyr21 Hβ2	Phe22 HN	3·1	1·0
Tyr21 Hβ2	Thr32 Hα	4·5	0·8
Phe22 Hβ3	Asn43 Hδ21	4·5	0·9
Phe22 Cγ	Asn24 Hδ21	6·3	1·2
Tyr23 HN	Phe45 Hζ	4·5	0·8
Asn24 Hβ2	Lys26 HN	4·2	0·9
Asn24 Hβ3	Ala27 HN	3·2	0·5
Asn24 Hδ21	Gln31 Hβ2	4·5	1·1
Asn24 Hδ21	Gln31 Hε22	4·5	0·9
Leu29 Hα	Met52 Qε	5·2	2·2[b]
Cys30 HN	Ala48 Qβ	5·0	1·2
Tyr35 Hε2	Ala40 Hα	4·5	1·0
Tyr35 Hδ2	Gly36 HN	4·5	0·8
Arg39 HN	Arg38 Hδ3	4·5	0·7
Lys41 HN	Lys41 Qδ	4·3	0·8
Lys41 Hα	Lys41 Qδ	3·8	0·9
Lys41 Qε	Arg42 HN	5·5	1·2
Lys46 HN	Cys51 HN	4·5	1·2
Lys46 Hγ2	Ser47 HN	4·5	0·9

† These 2 columns list pairs of atoms for which the distance in the X-ray crystal structure of form II exceeds the corresponding NOE upper distance constraint by more than 0·5 Å. For hydrogen atoms, the IUB-IUPAC nomenclature is used. Where applicable, pseudo-atoms were added to the crystal structure (for the nomenclature used, see Wüthrich *et al.* (1983) or Wüthrich (1986)). Distances between pairs of atoms were evaluated and compared to the corresponding NOE distance limits measured in solution.

‡ This column lists the NOE upper distance limit for the pair of atoms that was used in the input for the calculation of the solution structure.

§ This column lists the extent to which the interatomic distance in the X-ray crystal structure of form II exceeds the corresponding NOE upper distance constraint.

[a,b] In the crystal structure of form II, 2 distinct conformations are reported for this side-chain. Violations in the conformation represented by the higher electron density are denoted with *a*, and those in the minor species with *b*.

form II is well compatible with the n.m.r. constraints, whereas the less populated conformer would violate a NOE distance constraint by 2·2 Å.

A detailed comparison of the three-dimensional structure *per se* is only warranted for those side-chains that are among the best-defined in the solution structure (Table 3), which are also well defined in the crystal structures. With few exceptions (see above), these are located in the core of the protein, and the comparative analysis of these side-chains underlines the conclusion that the architecture of the BPTI core is virtually identical in solution and in the crystals. Other side-chains near the protein surface could not be precisely characterized by the n.m.r. data (Fig. 3), although they may be well defined in the crystal structures (note that some of these side-chains were found in multiple conformations in the crystals, as discussed above). For these residues a qualitative comparison of the *conformational states* in the different environments is of considerable interest. We thus find that numerous surface side-chains with well-ordered conformations in the crystal lattice are flexibly disordered in solution (see Figs 3 and 4). In a general vein this is reminiscent of the different views of surface hydration water that one obtains in protein crystals by X-ray diffraction or in aqueous protein solutions by n.m.r. (Otting *et al.*, 1991). Much additional work will be needed to obtain a complete description of the protein surface conformation but, in addition to the aforementioned data on the short-lived surface hydration, there are experimental observations available for BPTI that implicate the involvement of individual surface side-chains in highly specified, short-lived intramolecular interactions. For example, a number of residues were found to have a pH dependence of their amide proton chemical shifts (Tüchsen & Woodward, 1985, 1987*b*; Wüthrich & Wagner, 1979) which can be attributed to their participation in a hydrogen bond with a carboxylate group (Bundi & Wüthrich, 1979). These include interactions of the amide proton of Asp3 with its own side-chain carboxylate, the amide protons of Arg42 and Asn43 with the side-chain carboxylate of Glu7, the amide proton of Lys46 with the side-chain carboxylate of Glu50, and the amide proton of Glu49 with its own side-chain carboxylate. In addition, the N-terminal amino group was found to interact with the C-terminal carboxylate (Brown *et al.*, 1978). Of these six pairwise interactions for which experimental evidence was obtained, only the hydrogen bond Asn43 NH–7 Glu εO is present in the n.m.r. solution structure determined primarily from NOE distance constraints (Table 4). As regards the other five interactions, they involve molecular fragments that appear strongly disordered in the solution structure (Figs 2 and 3) so that inclusion of these hydrogen bonds or ion pair interactions would not cause significant violations of the n.m.r. constraints used in the structure calculation.

(e) *Solvent-accessible surface and internal water molecules*

In all three crystal structures of BPTI, four water molecules were observed in interior cavities of the protein, separated from contact with the network of hydrogen-bonded external waters. These four internal water molecules, which form a total of nine highly conserved intermolecular protein–water bonds (Wlodawer *et al.*, 1987*a*), can be divided into two groups based on their location in the protein. In one group, three water molecules are located in a

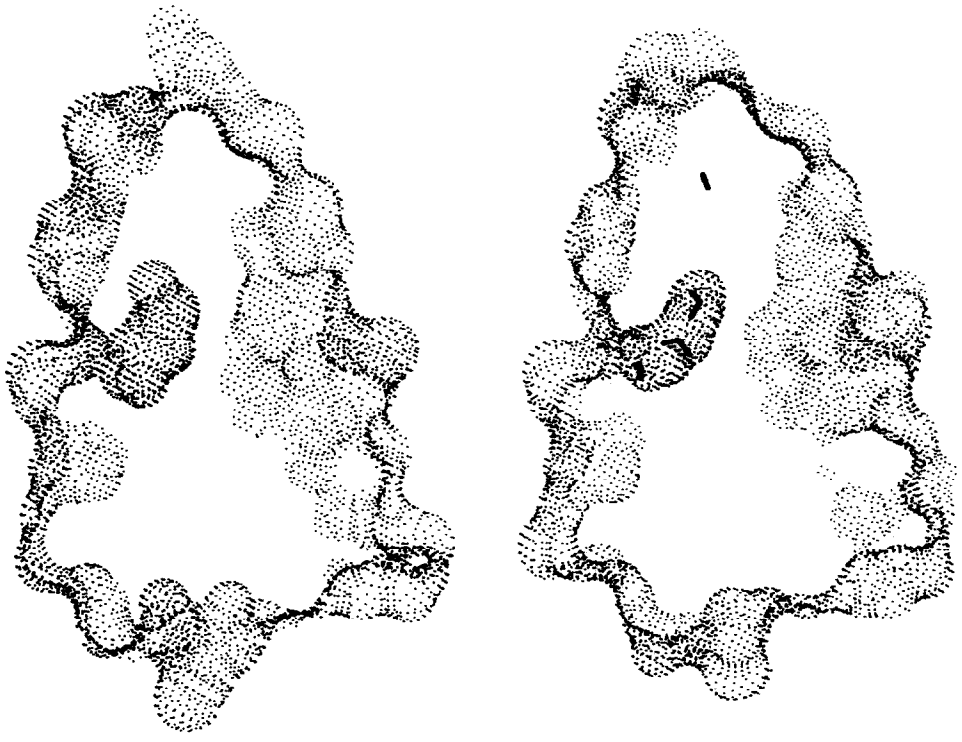


Figure 10. Visualization of the solvent-accessible surface in the mean n.m.r. solution structure of (a) BPTI and (b) the structure of the crystal form II of BPTI, showing the cavity containing 3 internal water molecules. The accessible surface, shown with stippling, was calculated with a solvent probe of radius 1·4 Å (with the use of a somewhat smaller probe, the cavity containing the single interior water was also identified (see the text)). The drawing shows a slice of thickness 6·7 Å through the BPTI molecule in the standard orientation of Fig. 2, i.e. for improved clarity the protein external surface in front of and behind the cavity has been cut away. In (b) the locations of the 4 internal water molecules as seen in crystal form II are indicated with bold lines; (a) represents the result of the n.m.r. structure calculation obtained without accounting for water molecules; note that the internal water molecules were observed directly by n.m.r. (Otting & Wüthrich, 1989).

cleft formed by residues 8 to 10 and 40 to 44, and in the other, a single water molecule is in a cage formed by two loops comprised of residues 11 to 14 and 36 to 39. These interior water molecules have also been identified in the solution structure of BPTI (Otting & Wüthrich, 1989; Otting *et al.*, 1991).

In the presently described determination of the n.m.r. solution structure of BPTI, the four internal water molecules and the water–polypeptide NOEs were not included in the input for the structure calculation. Since the internal water molecules are an integral part of the BPTI molecular architecture, an accurate solution structure should be able to accommodate these waters, in particular, in view of the fact that there are no backbone displacements greater than 0·8 Å in residues contacting these water molecules (Fig. 7). To check the n.m.r. structure for internal cavities large enough to accommodate a water molecule, the internal and external solvent-accessible surface as defined by Richards (1977), was calculated using the Midas Plus program (Ferrin *et al.*, 1988) for the mean solution structure, and the result compared to the solvent-accessible surface calculated for the stucture crystal form II. A standard solvent radius of 1·4 Å was used in these calculations. Figure 10 visualizes the results of these calculations. The mean solution structure contains an internal cavity that is very similar to the cavity formed by the three solvent water molecules W111, W112 and W113 in the crystal form II (Wlodawer *et al.*, 1984). The cavity occupied by the single internal water molecule W122 was not found in either structure using this procedure. After a slight decrease of the solvent radius used, this smaller cavity was also found in both molecules. However, a number of additional cavities were also identified in both the solution and crystal structures with this smaller probe, which do not contain water. The cavity containing three water molecules is accessible from the surface of the molecule for both the solution and the crystal structure (Fig. 10), whereas the smaller cavity, identified with a smaller probe, is not accessible from the surface in either of the two structures.

(f) *Intermolecular interactions in the crystal lattice*

The packing of a protein in the crystal lattice is an obvious source of differences between corresponding solution and crystal structures. A search

of the crystal lattice of BPTI form II for short intermolecular contacts revealed no contacts between heavy atoms shorter than 2·9 Å, and only three contacts shorter than 3·0 Å. Intermolecular protein–protein hydrogen bonds can be found in each of the three crystal forms of BPTI (Wlodawer *et al.*, 1987*a*,*b*). While similar residues are involved in the intermolecular hydrogen bonds in the three crystal forms, the partners involved are generally not conserved. Present in crystal form II are a total of eight intermolecular protein–protein hydrogen bonds, nine protein–water hydrogen bonds, and one protein–phosphate hydrogen bond (Wlodawer *et al.*, 1987*a*) identified in Figure 7(b) as open and filled squares, representing hydrogen bonds to side-chain and backbone atoms, respectively. Overall, in contrast to a recent similar study with the 434 repressor (Neri *et al.*, 1992), there is thus no evidence for short van der Waals' contacts in the crystal lattice, so that an analysis of lattice effects can be limited to possible correlations between structural differences and intermolecular hydrogen bonds. Such an analysis showed that the polypeptide segment 25 to 27, which shows the largest backbone displacements between crystal and solution structure (Fig. 5), is involved in a total of four intermolecular hydrogen bonds. These hydrogen bonds, which all involve the carbonyl oxygen atoms of residues 25 to 27, act predominately on one face of the loop. The observed difference between the X-ray and solution structures corresponds to a distortion in the direction of these interactions (Fig. 5), so that this structural difference can be rationalized by lattice effects. Otherwise, it appears that in BPTI, the crystal lattice would only affect the conformational states of individual amino acid side-chains and the state of the hydration water (Otting *et al.*, 1991).

5. Conclusion

The preceding Discussion is dominated by arguments leading to the conclusion that BPTI has nearly identical molecular architectures in solution and in three crystal forms. Close coincidence of the crystal and solution structures of BPTI had been implicated by less complete n.m.r. data (e.g., see Wagner *et al.*, 1987). The general result obtained here for BPTI is also in line with earlier conclusions drawn from comparisons of crystal and solution structures of, for example, secretory trypsin inhibitors (Williamson *et al.*, 1985), tendamistat (Kline *et al.*, 1986; Billeter *et al.*, 1989), interleukin-1β (Clore & Gronenborn, 1991), plastocyanin (Moore *et al.*, 1991), and 434 repressor (Neri *et al.*, 1992). To fully appreciate the significance of the results of these comparative studies, one must remember that they extend only over the polypeptide backbone, a limited set of "best-defined" amino acid side-chains in the n.m.r. structures, which are almost exclusively located in the core of the protein and, in the case of BPTI (Otting & Wüthrich, 1989) and interleukin-1β (Clore *et al.*, 1990), interior water molecules. For amino acid side-chains located on the protein surface, the n.m.r. data usually indicate pronounced dynamic structural disorder (Fig. 3). For BPTI, one has additional information demonstrating that a high degree of structural order for certain surface hydration water molecules observed by X-ray diffraction in protein crystals cannot readily be correlated with n.m.r. data on the life times of surface hydration water molecules in aqueous solution (Otting *et al.*, 1991). As discussed (e.g., see Wüthrich, 1991), n.m.r. solution structures of globular proteins typically contain both a well-defined molecular core for which a precise description of the atom positions can be obtained, as well as a motionally disordered array of surface side-chains and surface hydration water molecules, which are, as far as a direct comparison is technically feasible, fundamentally different from the view of the corresponding features of the protein obtained by diffraction experiments in crystals. The presently described high-quality n.m.r. solution structure of BPTI provides a new illustration of this duality of finding, in the same protein molecule, a well-defined core that is virtually identical to the corresponding molecular region in protein crystals, and a surface for which n.m.r. and crystallographic studies provide complementary information. The present paper also presents yet another illustration that both techniques can handle the determination of the core structure rather well. Clearly, future interests will have to be focused on improved understanding of protein surfaces and their impact on the functional properties.

We thank Bayer AG, Leverkusen, Germany, for a generous gift of Trasylol®. We also acknowledge M. Billeter, E. Liepinsh and G. Otting for helpful discussions and R. Marani for the careful processing of the manuscript. Financial support by the Schweizerischer Nationalfolds (project 31.25174.88) and the use of the Cray Y-MP computer of the ETH Zürich are gratefully acknowledged.

References

Anil-Kumar, Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1–6.

Billeter, M., Engeli, M. & Wüthrich, K. (1985). Interactive program for investigation of protein structures based on ^{1}H NMR experiments. *J. Mol. Graph.* **3**, 79–83, 97–98.

Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). Comparison of the high-resolution structures of the α-amylase inhibitor Tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. Mol. Biol.* **206**, 677–687.

Billeter, M., Schaumann, Th., Braun, W. & Wüthrich, K. (1990). Restrained energy refinement with two different algorithms and force fields of the structure of the α-amylase inhibitor Tendamistat determined by NMR in solution. *Biopolymers*, **29**, 695–706.

Brown, L. R., DeMarco, A., Richarz, R., Wagner, G. & Wüthrich, K. (1978). The influence of a single salt bridge on static and dynamic features of the globular solution conformation of the basic pancreatic trypsin inhibitor: ^{1}H and ^{13}C nuclear-magnetic resonance studies of the native and the transaminated inhibitor. *Eur. J. Biochem.* **88**, 87–95.

Bundi, A. & Wüthrich, K. (1979). Use of amide ^{1}H-NMR titration shifts for studies of polypeptide conformation. *Biopolymers*, **18**, 299–311.

Clore, G. M. & Gronenborn, A. M. (1991). Comparison of the solution nuclear magnetic resonance and X-ray crystal structures of human recombinant Interleukin 1β. *J. Mol. Biol.* **221**, 47–53.

Clore, G. M., Bax, A., Wingfield, P. T. & Gronenborn, A. M. (1990). Identification and localization of bound internal water in the solution structure of interleukin 1β by heteronuclear three-dimensional ^{1}H rotating frame Overhauser ^{15}H–^{1}H multiple quantum coherence NMR spectroscopy. *Biochemistry*, **30**, 2315–2323.

Creighton, T. E. (1978). Experimental studies of protein folding and unfolding. *Prog. Biophys. Molec. Biol.* **33**, 231–297.

Deisenhofer, J. & Steigmann, W. (1975). Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1·5 Å resolution. *Acta Crystallogr. sect. B*, **31**, 238–250.

DeMarco, A. & Wüthrich, K. (1976). Digital filtering with a sinusoidal window function: an alternative technique for resolution enhancement in FT NMR. *J. Magn. Reson.* **24**, 201–204.

Dubs, A., Wagner, G. & Wüthrich, K. (1979). Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. *Biochim. Biophys. Acta*, **577**, 177–194.

Eccles, C., Güntert, P., Billeter, M. & Wüthrich, K. (1991). Efficient analysis of protein 2D NMR spectra using the software package EASY. *J. Biomol. NMR*, **1**, 111–130.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS display system. *J. Mol. Graph.* **6**, 13–27.

Glusker, J. P. & Trueblood, K. N. (1985). *Crystal Structure Analysis*, Oxford University Press, New York.

Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1985). Two-dimensional correlation of connected NMR transitions. *J. Amer. Chem. Soc.* **107**, 6394–6396.

Güntert, P. & Wüthrich, K. (1991). Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J. Biomol. NMR*, **1**, 447–456.

Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). Automated stereospecific ^{1}H NMR assignments and their impact on the precision of protein structure determinations in solution. *J. Amer. Chem. Soc.* **111**, 3997–4004.

Güntert, P., Braun, W. & Wüthrich, K. (1991*a*). Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting progams CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 517–530.

Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1991*b*). Structure determination of the *Antp(C39→S)* homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol.* **217**, 531–540.

Kline, A. D., Braun, W. & Wüthrich, K. (1986). Studies by ^{1}H nuclear magnetic resonance and distance geometry of the solution conformation of the α-amylase inhibitor Tendamistat. *J. Mol. Biol.* **189**, 377–382.

Levitt, M. (1983). Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* **168**, 595–620.

Marion, D. & Wüthrich, K. (1983). Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin–spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**, 967–974.

Masson, A. & Wüthrich, K. (1973). Proton magnetic resonance investigation of the conformational properties of the basic pancreatic trypsin inhibitor. *FEBS Letters*, **31**, 114–118.

McLachlan, A. D. (1979). Gene duplication in the structural evolution of chymotrypsin. *J. Mol. Biol.* **128**, 49–79.

Moore, J. M., Lepre, C. A., Gippert, G. P., Chazin, W. J., Case, D. A. & Wright, P. E. (1991). High-resolution solution structure of the reduced french bean plastocyanin and comparison with the crystal structure of poplar plastocyanin. *J. Mol. Biol.* **221**, 533–555.

Nagayama, K. & Wüthrich, K. (1981). Structural interpretation of vicinal proton–proton coupling constants $^{3}J_{\alpha\beta}$ in the basic pancreatic trypsin inhibitor measured by two-dimensional *J*-resolved NMR spectroscopy. *Eur. J. Biochem.* **115**, 653–657.

Neri, D., Billeter, M. & Wüthrich, K. (1992). Determination of the NMR solution structure of the DNA-binding domain(1–69) of the 434 repressor and comparison with the X-ray crystal structure. *J. Mol. Biol.* **223**, 743–767.

Otting, G. & Wüthrich, K. (1989). Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Amer. Chem. Soc.* **111**, 1871–1875.

Otting, G., Orbons, L. P. M. & Wüthrich, K. (1990). Suppression of zero-quantum coherence in NOESY and soft NOESY. *J. Magn. Reson.* **89**, 423–430.

Otting, G., Liepinsh, E. & Wüthrich, K. (1991). Protein hydration in aqueous solution. *Science*, **254**, 974–980.

Presta, L. G. & Rose, G. D. (1988). Helix signals in proteins. *Science*, **240**, 1632–1641.

Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.

Richardson, J. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of a helices. *Science*, **240**, 1648–1652.

Siekmann, J., Wenzel, H. R., Schröder, W., Schutt, H., Truscheit, E., Arens, A., Rauenbusch, E., Chazin, W. J., Wüthrich, K. & Tschesche, H. (1987). Pyroglutamyl-aprotinin, a new aprotinin homologue from bovine lungs—isolation, properties, sequence analysis and characterization using ^{1}H nuclear magnetic resonance in solution. *Biol. Chem. Hoppe-Seyler*, **368**, 1589–1596.

Singh, U. C., Weiner, P. K., Caldwell, J. W. & Kollman, P. A. (1986). *Amber 3.0*, University of California at San Francisco.

Szyperski, T., Güntert, P., Otting, G. & Wüthrich, K. (1992). Determination of scalar coupling constants by

inverse Fourier transformation of in-phase multiplets. *J. Magn. Reson.* in the press.

Tüchsen, E. & Woodward, C. (1985). Hydrogen exchange kinetics of amide protons at the bovine pancreatic trypsin inhibitor protein–solvent interface. *J. Mol. Biol.* **185**, 405–419.

Tüchsen, E. & Woodward, C. (1987*a*). Assignment of asparagine-44 side chain primary ^{1}H NMR resonances and the peptide amide ^{1}H resonance of glycine-37 in basic pancreatic trypsin inhibitor. *Biochemistry,* **26**, 1918–1925.

Tüchsen, E. & Woodward, C. (1987*b*). Hydrogen exchange kinetics of surface peptide amides in bovine pancreatic typsin inhibitor. *J. Mol. Biol.* **193**, 793–802.

van Gunsteren, W. F. & Karplus, M. (1982). Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. *Biochemistry,* **21**, 2259–2274.

Wagner, G. (1983). Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. *Quart. Rev. Biophys.* **16**, 1–57.

Wagner, G. & Wüthrich, K. (1982*a*). Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **155**, 347–366.

Wagner, G. & Wüthrich, K. (1982*b*). Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. *J. Mol. Biol.* **160**, 343–361.

Wagner, G., Anil-Kumar & Wüthrich, K. (1981). Systematic application of two-dimensional ^{1}H nuclear-magnetic-resonance techniques for studies of proteins. 2. Combined use of correlated spectroscopy and nuclear Overhauser spectroscopy for sequential assignments of backbone resonances and elucidation of polypeptide secondary structures. *Eur. J. Biochem.* **114**, 375–384.

Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gō, N. & Wüthrich, K. (1987). Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using the algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611–639.

Weiner, P. K., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An all-atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230–252.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). Solution conformation of proteinase inhibitor IIa from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315.

Wlodawer, A., Walter, J., Huber, H. & Sjölin, L. (1984). Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J. Mol. Biol.* **180**, 301–329.

Wlodawer, A., Deisenhofer, J. & Huber, H. (1987*a*). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **193**, 145–156.

Wlodawer, A., Nachman, J., Gilliland, G. L., Gallagher, W. & Woodward, C. (1987*b*). Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **198**, 469–480.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids.* Wiley, New York.

Wüthrich, K. (1989*a*). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science,* **243**, 45–50.

Wüthrich, K. (1989*b*). The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination. *Acc. Chem. Res.* **22**, 36–44.

Wüthrich, K. (1991). Six years of protein structure determination by NMR spectroscopy: what have we learned? *Protein Conformation* (Ciba Foundation Symposium 161), pp. 136–149, Wiley, Chichester.

Wüthrich, K. & Wagner, G. (1979). Nuclear magnetic resonance of labile protons in the basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **130**, 1–18.

Wüthrich, K., Billeter, M. & Braun, W. (1983). Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.* **169**, 949–961.

Wüthrich, K., Billeter, M. & Braun, W. (1984). Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton–proton distances. *J. Mol. Biol.* **180**, 715–740.

Xia, T. H. (1992). Software for the determination and visual display of NMR structures of proteins: the distance geometry program DGPLAY and the computer graphics programs CONFOR and XAM. Diss. ETH-Zürich.

Edited by P. E. Wright