Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ultrahigh-resolution structure of a BPTI mutant

Anthony Addlagatta,[a] Szymon Krzywda,[b] Honorata Czapinska,[c] Jacek Otlewski[c] and Mariusz Jaskolski[a,b]*

[a]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [b]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland, and [c]Institute of Biochemistry and Molecular Biology, University of Wroclaw, Poland

Correspondence e-mail: mariuszj@amu.edu.pl

Received 14 November 2000
Accepted 19 February 2001

PDB Reference: bovine pancreatic trypsin inhibitor, 1g6x.

The crystal structure of a mutant of bovine pancreatic trypsin inhibitor has been refined to 0.86 Å resolution using low-temperature synchrotron data. The variant contains three mutations in the binding loop (Thr11Ala, Pro13Ala, Lys15Arg) and an unrelated Met52Leu substitution. Refinement with anisotropic displacement parameters and with removal of main-chain stereochemical restraints converged with $R = 0.1035$. The use of full-matrix refinement provided an estimate of the variances in the derived parameters. Some stereochemical parameters, such as the planarity of the peptide group and the value of the $\mathrm{N}-\mathrm{C}^{\alpha}-\mathrm{C}$ angle, show a wide spread, suggesting that the standard values used as restraints in protein structure refinements may not always be entirely appropriate. Comparison with the recently determined room-temperature structure of the same mutant at 1.42 Å resolution confirms the previous observations and provides new details, such as a double conformation of the main chain at Leu29 and at Gly56–Gly57, a high proportion (over 20%) of residues in double conformations, correlation of disorder through lattice contacts and the positions of H atoms, including those in water molecules, and their involvement in $\mathrm{C}-\mathrm{H}\cdots\mathrm{O}$ and $\mathrm{N}-\mathrm{H}\cdots\pi$ hydrogen bonds.

1. Introduction

There is a growing interest in the study of macromolecular crystal structures at atomic and subatomic resolution (Dauter *et al.*, 1997; Longhi *et al.*, 1998). Atomic resolution, achieved when crystals diffract to better than 1.2 Å (Sheldrick, 1990), provides a detailed model of the protein being studied, unveiling features that are not revealed at lower resolution. Only very high resolution, especially at low temperature, can reveal the organization of solvent beyond the first hydration shell, often leading to the identification of all solvent molecules. Very high resolution is also necessary for better definition of disorder or multiple conformations and may be required to clarify all details of the mechanism of action of macromolecules (Esposito *et al.*, 2000; however, see a cautionary example regarding cryocrystallographic studies of enzymatic mechanisms; Sandalova *et al.*, 1999). Also, unusual features of protein structure masked at lower resolution by stereochemical restraints may need free or at least relaxed-restraint refinement to show up. Since very high resolution protein structures are relatively rare, the optimal procedures for their refinement are still under development and discussion (Dauter *et al.*, 1997; Longhi *et al.*, 1998; Walsh *et al.*, 1998; Ridder *et al.*, 1999; Esposito *et al.*, 2000).

The object of the present analysis is a mutated form of bovine pancreatic trypsin inhibitor (BPTI). Few proteins have undergone such a thorough analysis as BPTI, a small globular protein composed of 58 amino acids with three disulfide bridges and a well defined secondary structure and hydrophobic core. BPTI binds to its cognate serine proteases through a sequential epitope located in an exposed loop, which is supported through a disulfide bridge (Cys14–Cys38) by another, auxiliary loop. Previous studies have covered many aspects of BPTI structure, folding, stability and dynamics (for a review, see Fritz & Wunderer, 1983). They included X-ray structure determinations of several polymorphic forms of native BPTI and its mutants, some of which are at atomic resolution (Wlodawer *et al.*, 1984; Eigenbrot *et al.*, 1992; Parkin *et al.*, 1996), a neutron-diffraction study (Wlodawer *et al.*, 1984) and several careful NMR studies (Otting *et al.*, 1991, 1993; Berndt *et al.*, 1992). The engineered variant of BPTI studied here has three mutations in the binding loop, including two truncations to alanine (Thr11Ala, Pro13Ala) and a conservative substitution of the key residue, Lys15Arg, plus an unrelated Met52Leu mutation eliminating methionine from the wild-type sequence (Krowarsch *et al.*, 1999). In the sequences of the Kunitz inhibitor family (of which BPTI is a member), arginine at position 15 is more common than lysine; leucine at position 52 is also quite common, but the engineered alanines in the binding loop are either very rare (position 11) or have never been observed (position 13). Unlike other forms of BPTI, the present mutant crystallizes very easily, yielding excellent crystals that diffract to very high resolution (Czapinska *et al.*, 2000). This new crystal form ($P4_32_12$), containing one protein molecule in the asymmetric unit, has recently been characterized using room-temperature data extending to 1.42 Å resolution (Czapinska *et al.*, 2000). Although only of near-atomic resolution, this room-temperature structure (PDB code 1qlq) could be refined successfully with anisotropic displacement parameters and provided new structural information on BPTI (no change in the conformation of the binding loop despite the mutations, salt-bridged N- and C-termini and a disulfide bond in two chiralities).

Using synchrotron radiation and low-temperature conditions, ultrahigh-resolution data extending to 0.86 Å have now been obtained for the same mutant crystallized under similar conditions. The results represent dramatic progress in the study of BPTI structure, where the best models so far are characterized by resolutions of up to 1 Å but rather poor refinement statistics (*R* factors close to 0.2; Wlodawer *et al.*, 1984; Eigenbrot *et al.*, 1992; Parkin *et al.*, 1996). The current refinements, performed with anisotropic displacement parameters and with partial removal of geometrical restraints, have converged with excellent statistics ($R = 0.1035$), confirming the previous observations and revealing new details such as double-conformation main-chain segments, correlation of disorder through lattice contacts, positions of H atoms (including those in water molecules) and their involvement in C—H···O and N—H···π hydrogen bonds. Additionally, the use of full-matrix refinement provided an estimation of the variances in the derived stereochemical parameters. The standard errors, when compared with the observed spread of some of those parameters, confirm that the assumptions about those parameters (such as the planarity of the ω angle or the value of the N—C$^{\alpha}$—C angle) may not always be entirely appropriate. This has a significant bearing on the restraint dictionaries that are used in protein structure refinements at moderate resolutions.

2. Materials and methods

2.1. Crystallization and data collection

Crystals were grown by the vapour-diffusion method using ammonium sulfate as the precipitant (pH 7.5) as described previously (Czapinska *et al.*, 2000). Because of the rapidity and reproducibility of the crystallization, the crystals used for the present data collection were grown directly at the synchrotron site. Hanging-drop experiments set up in the evening produced excellent crystals for synchrotron data collection in the morning. One such crystal, measuring 0.3 × 0.2 × 0.2 mm, was cryoprotected by immersing in reservoir solution supplemented by 30% ethylene glycol, mounted in a fibre loop and flash-frozen at the EMBL X11 beamline at the DORIS ring of the DESY synchrotron operated at $\lambda = 0.9090$ Å. Using a MAR 345 mm image-plate scanner, diffraction data were collected in four runs. The maximum resolution, oscillation range and the number of images were as follows: run 1, 0.86 Å, 0.5°, 180; run 2, 1.4 Å, 1°, 90; run 3, 2.0 Å, 2°, 45; run 4, 2.9 Å, 3°, 30. Part of the (high-resolution) image area of run 1 had to be omitted from interpretation because of a shadow cast by the low-temperature nozzle. Slight mis-setting of the beam stop resulted in asymmetric low-resolution shielding, which was most pronounced in run 4. At the data-scaling stage, this run, which was collected to avoid any oversaturation, turned out to still contain overloaded reflections.

The images were integrated in *DENZO* and the intensities were merged in the *SCALEPACK* program from the *HKL* package (Otwinowski & Minor, 1997). The crystal belongs to the $P4_32_12$ space group, with unit-cell parameters $a = 51.891$, $c = 43.035$ Å. Those values, measured at 100 K, are, respectively, 1.56 and 0.87% smaller than those at 290 K (Czapinska *et al.*, 2000). Over 800 000 measurements were combined to produce 47 018 unique reflections leading to very high redundancy (17.3) even at high resolution (6.0) (Table 1). The overall $R_{\rm int}$ is very low (0.036). In the last resolution shell $R_{\rm int}$ is relatively high, but those reflections nevertheless represent significant data as indicated by the high value of $\langle I/\sigma \rangle$ (3.61) and the high percentage of significant ($I > 2\sigma$) observations (57.8%). Their inclusion in the refinement was further justified by the analysis of the convergence results, where their behaviour was normal. As pointed out by Diederichs & Karplus (1997) and Weiss & Hilgenfeld (1997), the increased value of $R_{\rm int}$ also reflects the high degree of redundancy.

2.2. Refinement protocol

The refinements, performed in *SHELX*97 (Sheldrick, 1997; Sheldrick & Schneider, 1997), started from the fractional coordinates and equivalent isotropic displacement parameters

Table 1
Statistics of data collection and processing.

Values in parentheses refer to the last resolution shell.

Crystal size (mm)	0.3 × 0.2 × 0.2
EMBL beamline	X11
λ (Å)	0.9090
Detector	MAR 345
Temperature (K)	100
Resolution (Å)	20–0.86 (0.90–0.86)
Runs	4
Reflections measured	814213
Reflections unique	47018
$R_{\rm int}$	0.036 (0.488)
Redundancy	17.3 (6.0)
Completeness (%)	94.9 (88.6)
$\langle I/\sigma \rangle$	54.1 (3.61)
$I > 2\sigma$ (%)	90.1 (57.8)

Table 2
Course of the refinement.

Round	Resolution (Å)	$R/R_{\rm free}$	Actions
1	20.0–1.4	0.3106/0.3505	*xyz* from 1qlq, 40 cycles of isotropic *SHELX* refinement; no H atoms, no water molecules, no multiple conformations; standard restraints
2	20.0–1.0	0.2526/0.2634	18 waters, 4 sulfates
3	20.0–1.0	0.2040/0.2292	Double conformation of the side chains of 9 residues, R1, R15, R17, K26, L29, C38, R39, K46, D50, and the main chain of 3 residues, L29, G56, G57; 2 ethylene glycols, 3 sulfates, 60 waters
4	20.0–0.86	0.2055/0.2298	Extension of resolution
5	20.0–0.86	0.1450/0.1745	Anisotropy
6	20.0–0.86	0.1329/0.1723	Double conformation of R53; 29 waters (partial), 1 sulfate
7	20.0–0.86	0.1233/0.1584	H atoms (riding)
8	20.0–0.86	0.1124/0.1430	40 waters with manual adjustment of occupancy
9	10.0–0.86	0.1058/0.1401	14 waters
10	10.0–0.86	0.1040/0.1368	9 waters
11	10.0–0.86	0.1035/0.1365	No geometry restraints on main chain; BUMP relaxed
12	10.0–0.86	0.1039	All reflections

taken from the room-temperature model (1qlq). Inspection of the first electron-density maps confirmed that no rigid-body fitting was necessary. All solvent atoms were removed and only the major conformers were included in double-conformation protein regions. The first step of the refinement was carried out to the maximum resolution corresponding to the starting model (1.4 Å) and converged with an R and $R_{\rm free}$ of 0.3106 and 0.3505, respectively (Table 2). For free R-factor calculations (Brünger, 1992), 1884 (4%) reflections were randomly selected from the full resolution range 20.0–0.86 Å. The high initial values of the R factors primarily reflect the contribution and importance of the solvent atoms, seen despite the fact that a bulk-solvent correction according to the Babinet principle (Moews & Kretsinger, 1975) was used. When the best (as confirmed in the room-temperature structure) water molecules and sulfate anions were added, the R value fell by 6% and $R_{\rm free}$ fell by 9%. In step 2, the resolution range was extended to 1.0 Å. The electron-density maps were becoming progressively clearer, allowing unambiguous manual identification of further components of the solvent area, including 61 water molecules and three sulfate anions as well as two molecules of ethylene glycol that were not present in the room-temperature structure but were introduced at the cryoprotection step. Additionally, 12 residues of the protein molecule, including the four known examples from the room-temperature structure, could be modelled in two conformations, including two segments of the main chain. This stage of the refinement decreased the R and $R_{\rm free}$ factors by about 4% and they did not change in the next step (4) when the resolution limit was extended to 0.86 Å to include all high-resolution reflections (Table 2). The isotropic refinement converged with $R = 0.2055$ and $R_{\rm free} = 0.2298$. Inspection of difference Fourier maps at this stage revealed numerous patches of positive density near the main-chain atoms and the well defined side-chain atoms. They were particularly prominent at the S atoms. This distribution of electron density near heavy atoms indicates that the structure should be refined with anisotropic displacement parameters. Introduction of anisotropic displacement parameters in the next stage was accompanied by a huge drop (of about 6%) in both R and $R_{\rm free}$ (Table 2).

The anisotropic model was used for all atoms, including solvent atoms (the best and poorest, without distinction) and for alternative conformations of the protein, even at minor occupancies. Retrospectively, it seems that it might have been more prudent, even though less convenient, to use isotropic B factors for the lower occupancy moieties in order not to overinterpret the least clear areas and to keep the observation-to-parameter ratio as high as possible. It has to be admitted, however, that according to the default settings of *SHELX*97, the anisotropic displacement parameters were restrained both to avoid excessive fluctuations from atom to atom (DELU, SIMU) and to keep the water O atoms fairly isotropic (ISOR). At this stage and in fact until the atomic model was complete, the default geometrical restraints were also kept as implemented in *SHELX*97, using the targets defined by Engh & Huber (1991) for the polypeptide geometry and averaged values derived from the Cambridge Structural Database (Allen & Kennard, 1991) for the sulfate and glycol moieties. Two sulfate ions located on twofold axes were constrained to obey the site symmetry and the occupancies of alternate conformers as well as of disordered groups showing evident correlation were also appropriately restrained.

The introduction of riding H atoms in round 7 lowered the R factor by 1% and $R_{\rm free}$ by even more (Table 2). The geometrical positions of the C—H and N—H H atoms were generated automatically in *SHELX*97. The H atoms of the seven protein hydroxyl groups were also generated automatically, but their location was verified manually for consistency with difference electron density and plausible hydrogen-bonding interactions. In all cases, a satisfactory agreement was found. For model rebuilding and verification, the program *O* (Jones *et al.*, 1991) was applied using $3F_o - 2F_c$ (Lamzin & Wilson, 1997) and $mF_o - DF_c$ (Read, 1997) maps.

Table 3
Statistics of the refinement.

Program	*SHELX*97
Resolution (Å)	10.0–0.86
Reflections	
Total	46998
$R_{\rm free}$	1883
Rejection criterion	None
Atoms	
Protein	462
Water (full/partial)	59/111
Sulfate/glycol	36/8
Parameters refined	6497
R	0.1035
$R_{\rm free}$	0.1365
R.m.s. deviations	
Bonds (Å)	0.023
Angles (Å)	0.041
Chiral volume (Å^3)	0.113
Most favoured φ/ψ (%)	95.7

In rounds 8–10, partially occupied water sites were added and adjusted manually. The criterion for adjusting occupancy was no residual density after *B*-factor refinement. An attempt to refine the water occupancies was unsuccessful, leading to divergent results for many of the refined occupancies. This is in contrast to the experience of Walsh *et al.* (1998) reported for macromolecular refinement at similar resolution. At stage 9, the low-resolution limit was adjusted to 10.0 Å. This undesirable decision was forced by the observation of high errors in the low-resolution shell, retrospectively traced to the problems with the beam stop and detector saturation during data collection.

At stage 10, the refinement using default restraints converged with R = 0.1040 and $R_{\rm free}$ = 0.1368. Since the reflection-to-parameter ratio at this stage was about 7, an attempt was made to remove or relax the restraints. This proved to be very successful with respect to geometrical restraints in the fully ordered main chain, which could be removed altogether, leading to a slight drop in both R and $R_{\rm free}$ without deterioration of the model geometry. Stage 11 included loosening of the restraints imposed on non-bonded interactions which was brought about by an increase (from 0.02 to 0.05) of the standard deviation for the BUMP parameter. This adjustment was made in order not to bias the interpretation of some non-classical interactions discussed later. In the last round (Table 2), the test reflections were included in the data set for the refinement of the final model which is characterized by R = 0.1039 and 7.23 reflections per parameter. The statistics of the refinement are summarized in Table 3. The *R*-factor analysis was performed for reflections with $F > 4\sigma(F)$, but the refinement included all reflections without exception.

All the refinements were performed using the conjugate-gradient algorithm. At the end of the refinement, one cycle of full-matrix minimization was calculated without any stereochemical restraints. The least-squares calculations were performed using the L.S. 1/BLOC 1 combination in *SHELX*97 (Sheldrick, 1997), which allowed estimation of the uncertainties in all positional parameters.

3. Results and discussion

3.1. Quality of the structure

The protein model is of very high quality and includes all atoms. Even the segments in alternate conformations are generally modelled without ambiguity. The complete main chain is visible in contiguous $3F_o - 2F_c$ density above the 1.5σ level, including fragments that assume two conformations. $3F_o - 2F_c$ density for nearly all main-chain atoms is seen even at the 4σ level. Water molecules generally have very good $3F_o - 2F_c$ density. In the best cases, water H atoms (not included in the model) are seen in difference electron-density maps above 2.5σ. Protein H atoms can be seen in H-omit maps above 2.5σ. The eight sulfate anions could be modelled in their density with confidence, although the best ones are clear at the 3σ level, while the poorest need 1σ contouring. The two glycol molecules have clear density above the 1σ level.

Only two residues (discussed below) are outside the most favoured regions of the Ramachandran plot (Ramachandran *et al.*, 1963). With 95.7% residues in the most favoured regions, this structure of BPTI is only matched by 1qlq (the same polymorph at room temperature) and by 1bpi (Parkin *et al.*, 1996). Other stereochemical parameters of the protein are discussed below.

The final refinement, performed using the full-matrix algorithm in which all the positional parameters were varied, provides the best estimate of the quality of the model. The estimated standard deviations (e.s.d.s) characterizing the C—C bond distances (a value often used as a global indicator of the quality of small-molecule organic crystal structures) are as low as 0.006 Å and have a mean value of 0.010 Å. However, this number is calculated for all full-occupancy C—C bonds, also including long side chains, where lack of specific interactions and high mobility result in some cases in poor definition and consequently in high standard errors in the atomic coordinates. An illustration of the error in bond distances in the well defined areas of the model is provided by the e.s.d.s for the carbonyl bonds of the ordered main chain. They range from 0.005 to 0.020 Å, with a mean of 0.008 Å.

When the geometry of the final model is compared with the parameters assumed as standard and used to stereochemically restrain the refinement (Engh & Huber, 1991), an r.m.s. value of 0.028 Å is calculated for those bond lengths that remained restrained in the final refinement (basically in side chains) and of 0.023 Å for all protein bonds. The latter number is somewhat higher than the deviations typical in medium-resolution refinements with fully restrained geometry, but it is more likely to reflect the inadequacies of the target values (as discussed below) than genuine errors in the model. The latter errors are more correctly estimated from the full-matrix variances to be of the order of 0.01 Å.

3.2. Refinement with removed/relaxed restraints

The number of reflections per parameter in the final refinement (7.23) is not very different from situations encountered for non-centrosymmetric small-molecule struc-

Table 4
Coordinate superpositions (calculated in *ALIGN*; Satow *et al.*, 1986) using as a target the final model with main-chain restraints removed (UNRES).

The superpositions are calculated for the fully restrained model (RES) and for the 1.4 Å room-temperature model (1qlq; Czapinska *et al.*, 2000). The results, in terms of r.m.s. (and maximal) deviations (Å), are presented for C^α and all-main-chain superpositions.

RES C^α	RES main chain	1qlq C^α
0.004 (0.012)	0.003 (0.009)	0.116 (0.254)

tures. This suggests that in principle it would be possible to carry out the refinement without any stereochemical restraints. However, such a radical approach would be inappropriate for those side chains where, owing to high mobility or disorder, there is not enough information in the diffraction data to adequately define the atomic positions. One could of course handle this problem individually, manually selecting those side chains that should be stereochemically restrained, but a more practical approach seems to be to retain the restraints for all side chains and to remove them completely from the main chain. This approach has been adopted in the present refinement, with the exception that the three residues having double-conformation main chain (Leu29, Gly56 and Gly57) were not released from restraint control. The removal of main-chain restraints was total in the sense that all stereochemical parameters (*i.e.* bond distances, bond angles, configuration of the C^α chiral centres and planarity of the sp^2 C atoms and of the amide bonds) were freed. Even with such a drastic reduction of restraints, the refinement remained very stable and quickly converged to a minimum that was in fact very close to the fully restrained model (Table 2). This is a doubly optimistic conclusion. Firstly, it indicates that relaxation or even complete removal of restraints in high-resolution protein refinements is fully justified. Secondly, it shows the tremendous power of the high-resolution diffraction data in driving the refinement of well ordered models. To an extent, this means that the repercussions of potentially inadequate target values (as discussed below) may not be as severe as they might seem when very high resolution data are used.

The stability of the refinement without main-chain restraints is illustrated in Table 4, where the restrained (RES) and unrestrained (UNRES) models are superposed using only C^α or all backbone atoms. The r.m.s. deviations are minimal, 0.004 Å in the former and 0.003 Å in the latter case, and even the maximal deviations (0.012 and 0.009 Å, respectively) are of the order of experimental errors. In contrast, in comparison with the room-temperature 1.4 Å model the deviations are about 30 times higher, although still very small (Table 4). Even though the differences between the UNRES and RES models are not very large, it should be stressed that the removal of stereochemical restraints followed by stable refinement is methodologically important, because the unrestrained positions of the atoms reveal a picture that otherwise might be blurred by the masking effect of the presumed model.

3.3. Model stereochemistry

The high quality of the model, especially in well ordered areas, together with partially unrestrained refinement and the availability of estimates of parameter errors, allow detailed analysis of the stereochemical parameters of the protein. Some of those parameters, namely the distribution of the $N-C^\alpha-C$ angles, the deviations of the peptide bonds from strict planarity and the conformation of asparagine residues, show unusual features in comparison with the values accepted as standard and used in stereochemical dictionaries.

3.3.1. N—C^α—C angles. It has been noted previously that the $N-C^\alpha-C$ valence angles in atomic resolution protein structures show a wide distribution (Esposito *et al.*, 2000). In Fig. 1, the main-chain $N-C^\alpha-C$ angles in the present BPTI structure are analyzed. Their spread is very wide in comparison with the mean e.s.d. (0.6°). The large range of $N-C^\alpha-C$ angles is usually interpreted in correlation with the conformation of the main chain (Ashida *et al.*, 1987; Jiang *et al.*, 1995; Karplus, 1996; Esposito *et al.*, 2000). In contrast, the present plot has unimodal distribution, but this may reflect the limited size of the sample. It is seen that the $N-C^\alpha-C$ angles can deviate by as much as 5° on both sides from the target value of 111.2° and that at the high end of the distribution the deviations can be as large as 7°. The two high-end outliers correspond to a tandem of glycine residues, Gly36–Gly37, that are part of the auxiliary loop. There is no doubt that this fragment of the main chain is interpreted correctly in its well defined electron density (Fig. 2*a*). The Gly36–Gly37 tandem is preceded by an important residue in the secondary loop, Tyr35. As described below, the aromatic ring of Tyr35 is an acceptor in a short $N-H\cdots\pi$ interaction with the main-chain amide of Gly37, only two residues away. This may explain the unusual geometry and also the chemical character of the 36–37 tandem, which must allow the main chain to assume the required conformation.

3.3.2. Planarity of peptide bonds. Fig. 3 represents the distribution of the ω angles, which characterize the conformation around the peptide bond, in the present structure. This analysis is unbiased, as in the final refinement this angle was not restrained by the usual requirement of planarity. The distribution of the ω angles is very wide and is characterized by a mean value of 179.5° and a standard deviation of 9.1°, while the uncertainty of the individual ω values estimated from the inversion of the least-squares matrix is about 0.6°. Even if one disregards the extreme outliers (observed for the disordered segments of the main chain), it is still evident from Fig. 3 that deviations from strict planarity of the peptide bond amounting to ±20° are possible. This phenomenon has previously been noted in high-resolution peptide and protein structures (MacArthur & Thornton, 1996; Wilson *et al.*, 1998; Sandalova *et al.*, 1999) and calls for revision of the strictness (or even usefulness in some cases) of the planarity restraint imposed on the peptide bond. When enforced against structural evidence, this restraint could influence not only the peptide bond itself, but also (perhaps even more) the other main-chain torsion angles, which usually are not controlled by any restraints, thus leading for instance to artificially high distortions in the Ramachandran plot.

3.3.3. Conformation of asparagine residues. Apart from the estimated standard deviations determined from the least-squares matrix, the best assessment of the model quality is obtained from the analysis of the torsion angles which were not restrained at any stage of the refinement. As mentioned above, only two of the main-chain φ/ψ pairs are outside of the most favoured regions in the Ramachandran plot (Ramachandran *et al.*, 1963). They correspond to a tandem of asparagine residues (Asn43–Asn44) which show similar deformation in all reported structures of BPTI. In the literature, asparagine doublets are known to have a tendency to deviate from the canonical Ramachandran conformations. In their conformational study of asparaginyl residues in proteins, Srinivasan *et al.* (1994) found that asparagines that occur in tandem in protein structures show three main conformational preferences. According to their convention, the Asn43–Asn44 doublet in the present structure has roughly the ‘bridge’-β conformation, where the first residue is in the intermediate area between the α and β regions of the Ramachandran plot and the second residue has a general β conformation (Fig. 4). Recently, Deane *et al.* (1999) have proposed an interesting interpretation of this tendency of asparagine residues, explaining it in terms of dipole interactions between the side-chain C=O group and the main-chain carbonyl of the same or previous peptide. In the present structure such dipole interactions could also be evoked to explain the deformation of the main-chain conformation of Asn43–Asn44. One of the interactions falls in the antiparallel category (Asn43) while the other (Asn44) could be best described as ‘perpendicular’ (Fig. 4). However, it is not the ‘perpendicular’ arrangement proposed originally by Allen *et al.* (1998), but rather a variant of the sheared parallel motif where the deviation from parallelism of the C=O vectors is very severe. The usefulness of dipole–dipole interactions as a justification for conformational distortions of Asn residues is somewhat weakened when we analyze the third asparagine in the BPTI structure, Asn24. Here, the φ/ψ angles are within the most favoured Rama-

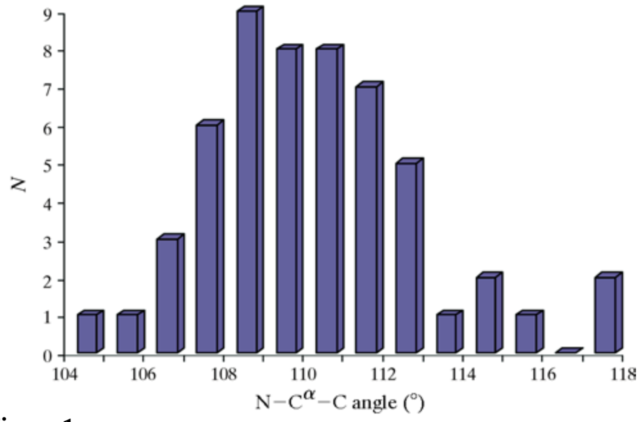


Figure 1
Distribution of the N−C$^{\alpha}$−C angles (°). The mean e.s.d. of the data points is 0.6°. The two outliers at 118° correspond to a tandem of glycine residues, Gly36–Gly37.

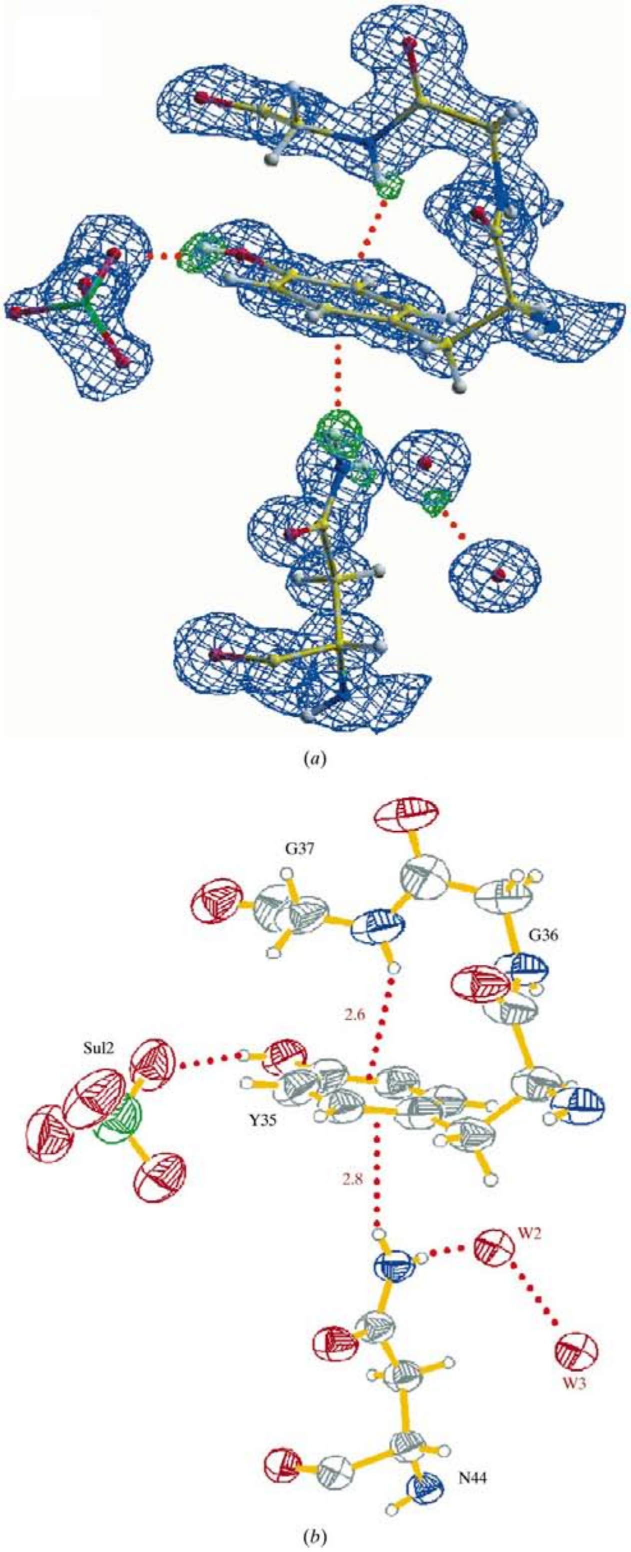


Figure 2
Molecular interactions of Tyr35. The π electrons of the aromatic ring accept two N−H hydrogen bonds from the main-chain amide of Gly37 and from the side-chain amide of Asn44. The OH group of Tyr35 is also involved in an intermolecular hydrogen bond to a sulfate anion. (*a*) Electron density of the interacting residues: $3F_o - 2F_c$ map contoured at 0.65 e Å^{-3} (1.7σ, blue) and H-omit difference map contoured at 0.10 e Å^{-3} (2.5σ, green); (*b*) the same fragment shown with anisotropic displacement ellipsoids drawn at 20% probability. Hydrogen-bond distances (Å) are from the H atoms to the centre of the aromatic ring.

chandran limits and yet the side-chain carbonyl forms an interaction with the main chain that is almost identical to that at Asn44 (Fig. 4). The conformations of those two residues differ only (by 80°) in the φ angle, but this conformational aspect has no bearing on the C═O···C═O arrangement, which is governed exclusively by the ψ and χ_1 angles.

3.4. Anisotropic displacement parameters

The pattern of atomic displacement parameters along the protein chain is very similar to that in the room-temperature 1qlq structure, but the equivalent B factors are about two times lower (Fig. 5*a*). The mean B factors of the main-chain, side-chain and solvent atoms are 12, 16 and 29 Å^2, respectively. Fig. 5(*b*) illustrates the mean anisotropy of the displacement parameters for the main-chain and side-chain atoms at both temperatures. The anisotropy of atomic displacements in the side chains is very similar at both temperatures. For the main-chain atoms, the plots are identical in the low-displacement areas, whereas higher anisotropy is observed at reduced temperature in the areas of increased mobility. This may be rationalized by the fact that lowering the temperature only reduces the dynamic, but not the static, disorder. Since any unresolved static disorder is most likely to be artificially represented by the largest components of the anisotropic displacement tensors, lowering the temperature will have a reducing effect primarily on the smallest eigenvalues in such areas. Comparison of Figs. 5(*a*) and 5(*b*) indicates that the level of anisotropy is correlated with $B_{\rm eq}$. This confirms that increased mobility and unresolved disorder have anisotropic character. An illustration of the anisotropic displacement parameters in an interesting area involving Tyr35, Gly36–Gly37, Asn43–Asn44 and some water molecules is provided in Fig. 2(*b*).

3.5. Double-conformation residues

The current structure has revealed 12 double-conformation residues (including three main-chain segments) or 21% of the total number of residues, while in the 1.4 Å room-temperature crystal structure of the same form only four residues were modelled in two conformations (Czapinska *et al.*, 2000). A similar proportion of disordered residues was reported in the atomic resolution structure of lysozyme (Walsh *et al.*, 1998). It appears that disorder in protein structures has a higher prevalence than revealed by routine medium-resolution structure determinations and that it is the high precision and high resolution of structure refinement that allow its proper detection (Esposito *et al.*, 2000). By disorder in this context we do not mean lack of organized structure, but rather clearly visible alternate conformations.

The population fractions of all the double-conformation residues were refined under the constraint of being unity when compounded over the split components. The distribution of these residues is not random. They are localized in three clusters on the surface of the molecule (Fig. 6) and additionally there is evidence that there is high degree of correlation between them. The only exception is Arg17, which does not interact with other double-conformation residues. In all three areas, the correlations involve symmetry-related molecules, sulfate anions and water molecules. One of those clusters is located on the epitope and the auxiliary loops and involves the disulfide bridge (Cys14–Cys38) that links them together. The second cluster is on the opposite end of the molecule and comprises three segments of the sequence that are close in space: the salt-bridged N-/C-termini and the adjacent β-hairpin loop formed by residues 26–29. It is the only area where the main chain is disordered in two conformations. The third cluster is a patch of basic and acidic residues centred around Asp50.

In the previously reported crystal structures of BPTI, Gln7 was frequently modelled in two conformations. As explained by Czapinska *et al.* (2000), this it not the case in the present crystal form, where Gln7 is stabilized in one conformation by altered water/counter-ion structure on the surface of the molecule.

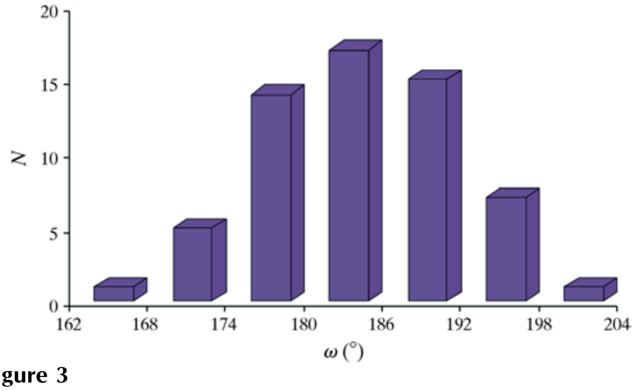


Figure 3
Distribution of the ω angles (°) for the full-occupancy peptides. The mean e.s.d. of the data points is 0.6°.

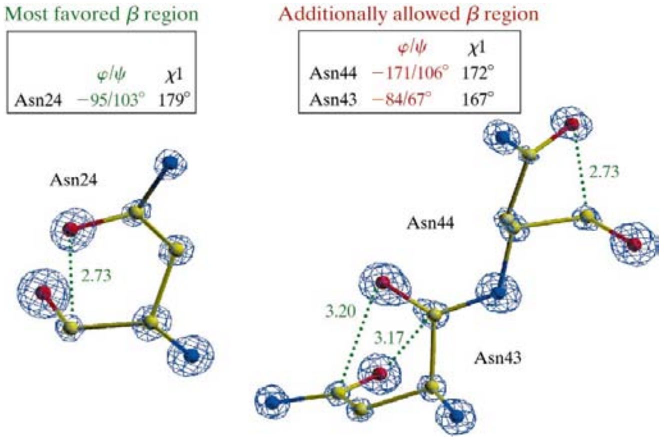
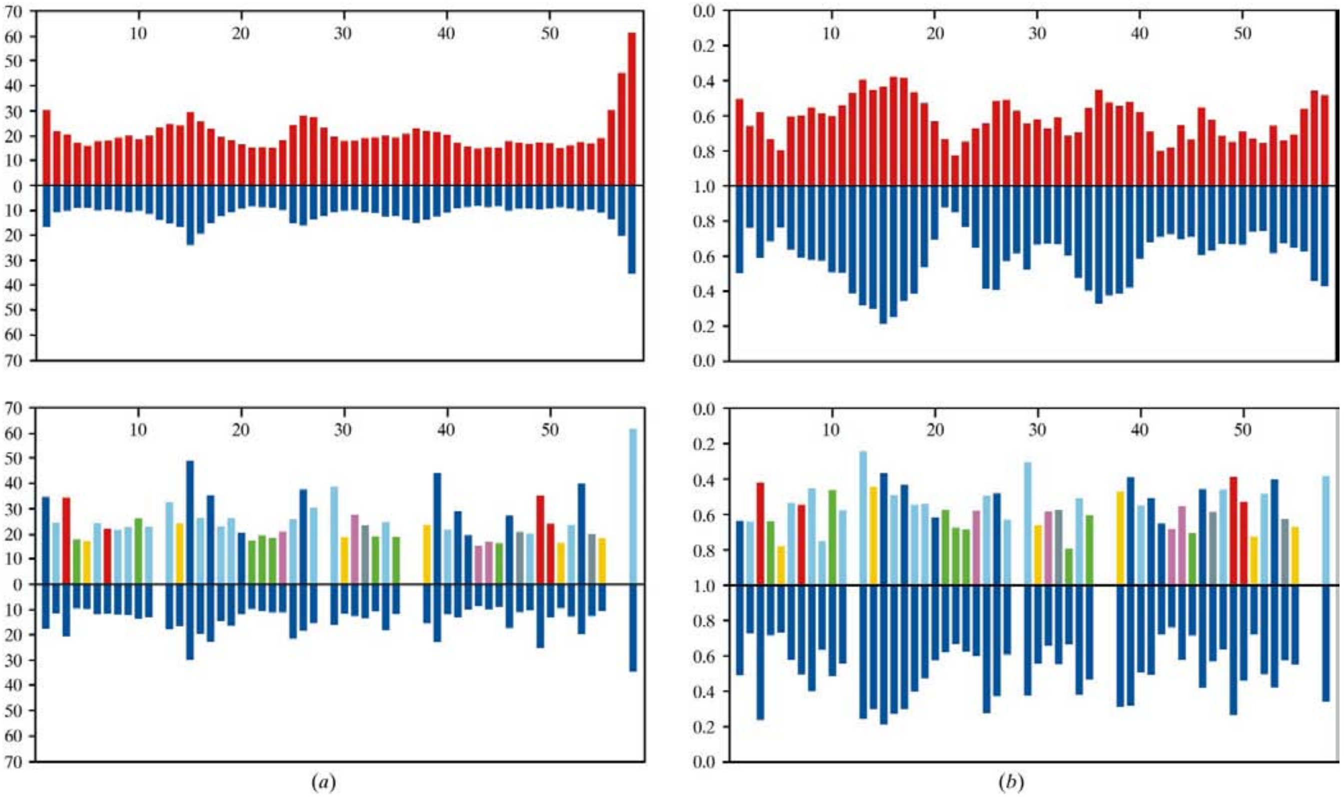


Figure 4
C═O···C═O dipole interactions between the side-chain and main-chain carbonyl groups of the three Asn residues. Those residues are among the best regions of the structure, as illustrated by the $3F_o - 2F_c$ electron density. Contour level: 2.01 e Å^{-3} (5.3σ) for Asn24 and 2.12 e Å^{-3} (5.7σ) for Asn43–Asn44. The atom–atom distances in these interactions are given in Å and are indicated by dotted lines. The conformation of the main chain is described by the φ/ψ values and that of the side chains by the χ_1 torsion angles.

3.5.1. Disordered disulfide bridge. The most intriguing feature of the room-temperature structure, *i.e.* the first crystallographic observation of a disulfide bond in BPTI coexisting in the right-handed and left-handed forms (Czapinska *et al.*, 2000), is confirmed in this low-temperature structure. The two chiralities of the Cys14–Cys38 disulfide bridge are realised by conformational twofold disorder of the side chain of Cys38. The dominating right-handed configuration (less preferred in S—S bonds) corresponds to the form of the Cys14–Cys38 bridge reported in all other crystal structures of BPTI. However, the fraction of the less-abundant left-handed conformer is relatively high (0.40) and leaves no doubt about this model. As illustrated in Fig. 7, the electron density at the three sulfur sites of the Cys14–Cys38 system clearly reflects the occupancies of these atoms. It is also interesting to note that the refined occupancy factors of the two positions of the S^{γ} atom of Cys38 in this crystal form are the same, within experimental error, at both temperatures. Their values for the right-handed/left-handed conformations at 290 and 100 K are 0.63/0.37 and 0.60/0.40, respectively. This confirms that the two forms of the Cys14–Cys38 disulfide bridge are in static disorder. The constancy of these populations is in contrast to the NMR observations reported by Otting *et al.* (1993), where the proportion of the two conformers of the Cys14–Cys38 bridge seemed to be temperature-dependent.

The Cys14–Cys38 bridge belongs to a patch of disorder located on the surface of the BPTI molecule that covers part of the enzyme-binding epitope. In addition to Cys14–Cys38, it also includes Arg15, Arg17 and Arg39 (Fig. 6) and is involved in twofold symmetric packing interactions with its crystallographic copy (Fig. 7). The disorder of the side chain at the Arg15 mutation site is not extensive. The two conformers have occupancies of 0.54 and 0.46. In the room-temperature structure, this residue was modelled in one conformation but with relatively high displacement parameters (Czapinska *et al.*, 2000). One of the terminal N atoms in both conformations of Arg15 forms a hydrogen bond with the carbonyl O atom of a symmetry-related Gly37 (Fig. 7). The spatially close disordered side chain of Cys38 seems to be responsible for the flip of the guanidinium moiety. The hydrogen-bonded N atom serves as a pivot and the rest of the guanidinium group moves in and out in resonance with the Cys38 side chain. It is important to note that the minor conformation of Cys38 is closer to the major conformation of Arg15, confirming that there is no attractive interaction between these two groups.

A sulfate group (Sul6) not seen in the room-temperature structure is observed in the vicinity of the Cys14–Cys38



(*a*) (*b*)

Figure 5

Analysis of the atomic displacement parameters in the present low-temperature structure [lower histograms (blue), present results] in comparison with the 1.4 Å room-temperature structure (upper histograms, 1qlq). Upper panels are for main-chain atoms and lower panels for side-chain atoms. The character of the side chains is colour coded in the room-temperature plot according to the convention of *SHELXPRO* (Sheldrick, 1997). (*a*) analyzes the equivalent isotropic displacement parameters, $B_{\rm eq}$ (Å^2). Note that the pattern of $B_{\rm eq}$ is identical at both temperatures. (*b*) illustrates the mean anisotropy, *i.e.* the ratio of the smallest to the highest eigenvalue of the anisotropic tensor. Again, the pattern is similar at both temperatures, but the main-chain atoms in the less-ordered areas show somewhat higher anisotropy at 100 K.

disulfide bridge. It interacts with the main-chain amide N atom of Arg39 and with a terminal N atom of one of the conformers of the side chain of this residue (Fig. 7). Although it is not very clear why the Arg39 side chain is disordered in two conformations despite this strong interaction, some reasoning is possible. Sul6, though modelled in only one orientation, has an elongated density (see below), suggesting the possibility of disorder. The strong hydrogen bond with the amide group of Arg39 anchors one corner of the sulfate, while the rest of the Sul6/Arg39 system can move in response to the status of the Cys38 group which is in close proximity. Furthermore, Sul6 is separated only by a water molecule from the disordered system at the N-/C-termini of a symmetry-related molecule across a solvent channel. This suggests the possibility that these two systems of disordered residues could be linked together *via* interactions with solvent molecules. The fractional occupancies of the double-conformation residues in this cluster are within a narrow range between 0.60/0.40 and 0.54/0.46; only the value refined for Arg39 is somewhat different (0.68/0.32).

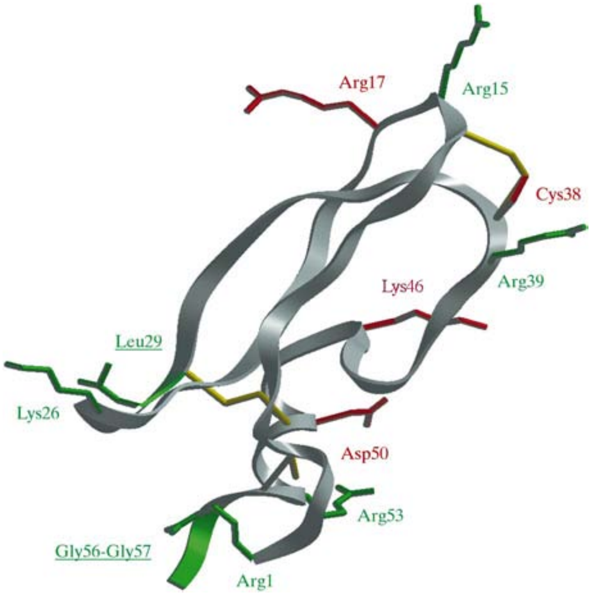


Figure 6
Residues modelled in two conformations. In addition to the four double-conformation side chains found at room temperature (red), this high-resolution low-temperature structure has an additional eight disordered residues (green). In three of them (underlined), the disorder also involves the main chain. Figure prepared in *SETOR* (Evans, 1993).

3.5.2. N- and C-termini. Apart from the room-temperature structure of the present polymorph (1qlq; Czapinska *et al.*, 2000), only one other structure (1bpi), determined at low temperature and at 1.1 Å resolution (Parkin *et al.*, 1996), showed clear C-terminal density. As observed in the room-temperature structure, there is a salt bridge between the carboxylate group of the C-terminal Ala58 and the ammonium group of the N-terminal Arg1 (Fig. 8*a*). In the present structure, the complete C-terminus is visible, but the Gly56–Gly57 tandem preceding the terminal Ala58 has been modelled and refined in two clear-cut conformations with almost equal occupancies (Fig. 8*b*). The shape of the electron density at Gly56–Gly57 leaves no doubt about this interpretation (Fig. 8*b*). This observation strongly suggests that the C-terminal disorder is confined to residues 56 and 57, while the very terminal groups are close together forming a strong and well defined interaction. Apart from the salt bridge with the C-terminal carboxylate, the N-terminal $-\mathrm{NH}_3^+$ group participates in another salt bridge with a sulfate anion and forms a hydrogen bond with a water molecule, satisfying its hydrogen-bonding potential (Fig. 8*a*). The guanidinium moiety of Arg1 is modelled in two conformations with approximately equal occupancies. One of the terminal N atoms of the side chain in one conformation forms a strong hydrogen bond with one of the conformers of the disordered Gly56 carbonyl O atom. Similarly, a terminal N atom of the second conformer of Arg1 forms an analogous interaction with the carbonyl O atom of Gly57 from the alternate conformer of the Gly56–Gly57 segment (Fig. 8*a* and 8*c*). The fact that the acceptor O atoms in these interactions are not only in different residues but also in alternate conformers of the Gly56–Gly57 segment, together with the similarity of the refined population coefficients at Arg1 and Gly56–Gly57, indicates that this system of interactions represents correlated disorder.

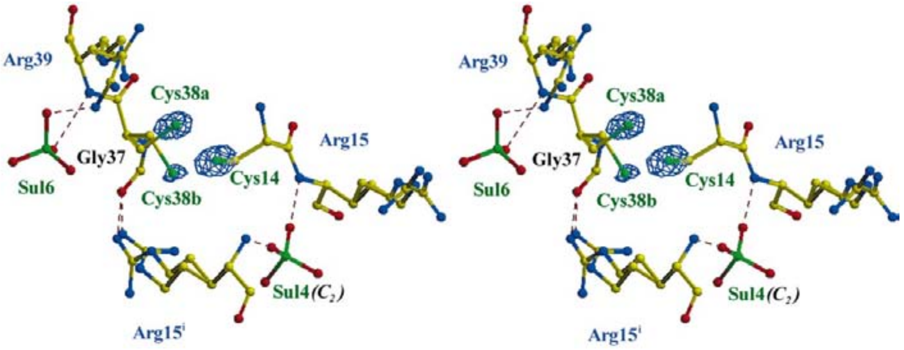


Figure 7
$3F_o - 2F_c$ electron density contoured at 1.52 e Å^{-3} (4σ) around the S atoms in the Cys14–Cys38 disulfide bridge. Note that the volumes of these densities are in proportion to the occupancies of the S atoms. This disulfide bridge is surrounded by four arginine residues, Arg15 and Arg39 from the same molecule and their symmetry-related copies (i). There are also sulfate ions in this area, one of them (Sul4) located on a twofold axis.

The coupled disorder of Gly56–Gly57 and Arg1 reaches farther and also involves intermolecular interactions. This is illustrated in Figs. 8(*c*), where the disordered side chain of a symmetry-related Lys26 is seen wedged between Gly56–Gly57 and the disordered main chain of Leu29 in the same molecule. Although the entire Leu29 residue is modelled in two conformations, it is its main chain that is in closer proximity to the disordered side chain of Lys26. This is an example of disorder being transmitted between two main-chain segments through side-chain mediation. The closest distances, representing van der Waals interactions between Lys26 and the neighbouring residues on both sides in Fig. 8(*c*) are around 3.5–4.0 Å. The disordered Lys26 side chain is in a strong salt-bridge contact with sulfate Sul8, which also forms hydrogen bonds with the disordered guanidine moiety of Arg1.

Throughout the refinement, the fractional occupancies within the Gly56–Gly57 segment and in the Arg1 side chain kept converging at practically identical values (about 0.52/0.48). In the final refinement they were therefore restrained to be represented by a single parameter. The freely refined occupancy of Leu29 converged at a value, 0.51/0.49, that is well within experimental error of the previous parameter. The occupancy ratio of the side chain of Lys26 is somewhat different (0.59/0.41), but is still close enough to be consistent with the model of correlated disorder.

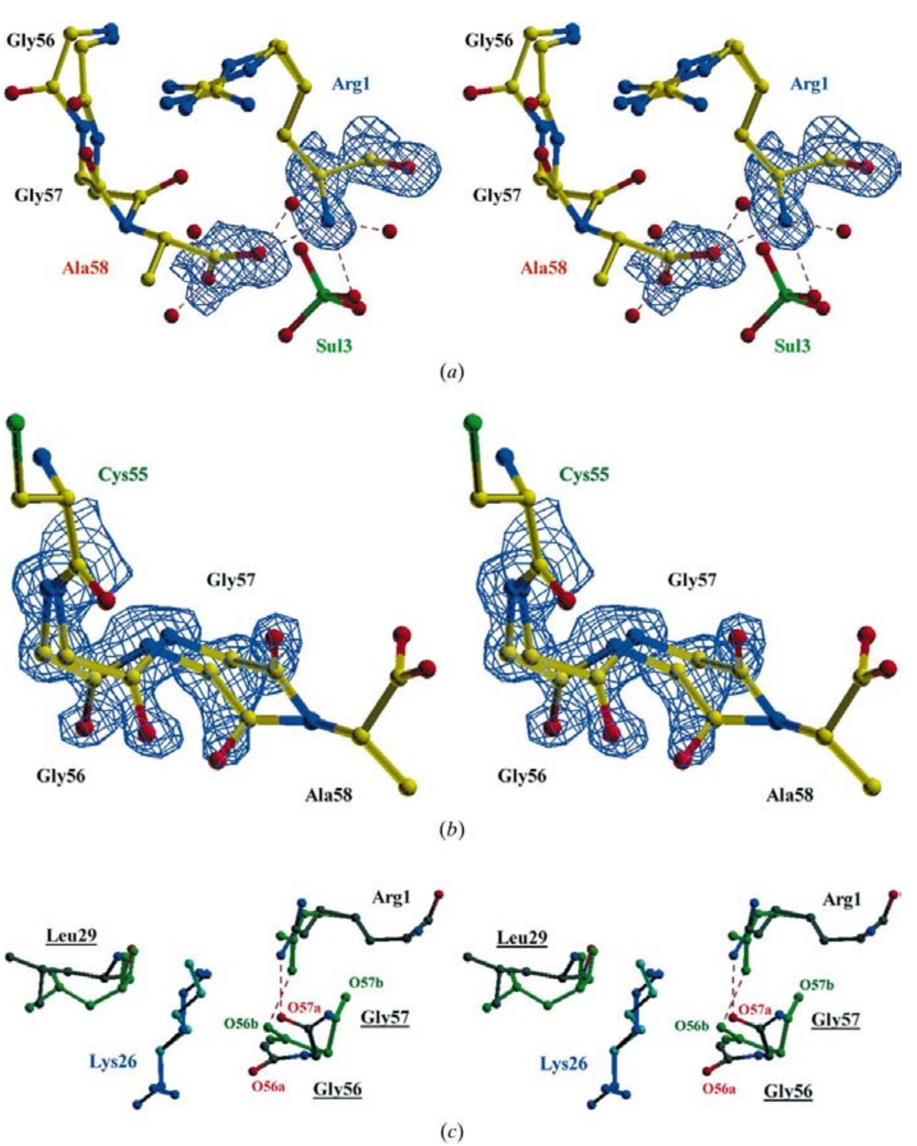


Figure 8
(*a*) $3F_o - 2F_c$ electron density contoured at 0.57 e Å^{-3} (1.5σ) at the main chain of the salt-bridged N- and C-termini. Note the disorder of the side chain of Arg1 and of the main chain Gly56–Gly57 and the ordered character of the most terminal groups. (*b*) $3F_o - 2F_c$ electron density contoured at 0.61 e Å^{-3} (1.6σ) at Gly56–Gly57, defining the disorder of the main chain. The disorder is not propagated beyond the glycine tandem. The polypeptide chain is anchored by the salt bridge at Ala58 and at Cys55 which is involved in a stable disulfide bridge. (*c*) Correlated disorder involving main chain and side chains from two molecules. Lys26 is contributed by a symmetry-related molecule, while the remaining residues are from one molecule. Disorder of the main chain (black and green) is seen at Leu29 and Gly56–Gly57. The system of correlated interactions is completed by the side chain of Arg1 (also black and green) which interacts with the Gly56–Gly57 segment.

3.5.3. Disorder around Asp50. On the molecular surface, this cluster is comprised of the side chains of Lys46, Asp50 and Arg53 (Fig. 6), but in the system of molecular interactions the side chain of Lys46 is contributed by a twofold symmetric copy of the molecule (Fig. 9). The central part of this system is Asp50, the electron density of which leaves no doubt about the conformation of the two rotamers (Fig. 9). Although there are several hydrogen-bonding distances within this system (in particular, a good two-component salt bridge between one of the Arg53 conformers and the fully occupied Sul2 and between the alternative side chain of Arg53 and the dominating rotamer of Asp50), it is difficult to propose a scheme of interactions that would be fully consistent with correlated disorder. However, the spatial proximity of these residues and their similar occupancies (within 0.60/0.40–0.51/0.49) suggest that those residues do influence each other's behaviour.

3.6. H atoms and hydrogen bonds

3.6.1. Treatment of H atoms in the refinement. The introduction of all the protein H atoms at calculated positions reduced the *R* factor by nearly 1% and $R_{\rm free}$ by over 1% (Table 2). With standard geometry, the generation of all C—H and N—H H atoms is straightforward, except for methyl and $-{\rm NH}_3^+$ groups, where the rotation about the C—(C/N)H$_3$ axis is optimized in the refinement. With hydroxyl groups the situation is somewhat different, as the rotation around the C—OH bond is relatively free and the H-atom position has to be fixed by considering the surrounding hydrogen-bond acceptors and/or its electron density. All the protein hydroxyl H atoms generated in *SHELX*97 were found to fulfill both criteria and were included in the model. The H atoms were contributing to F_c, but in the refinement they were riding on their carriers. Their displacement parameters were set at 1.2 times the $B_{\rm eq}$

values of those carriers ($1.5B_{\rm eq}$ for the rotating $-X{\rm H}_3$ groups). No H atoms were generated for the water molecules, although in many cases there was clear difference density indicating their positions. Similarly, no attempt was made to introduce the H atoms of the two glycol molecules. The difference electron-density maps calculated for the analysis of the H atoms, in particular H-omit maps, were of the type $mF_o - DF_c$, as recommended by Read (1997).

3.6.2. H atoms of water molecules. For the low-*B*-value water molecules, difference electron density reveals the positions of the H atoms. The best example is provided by the four water molecules that are buried in an internal groove (see below) where, with one exception, all the water H atoms are clearly visible in the 2.9σ difference electron-density map (Fig. 10). The patches of positive difference electron density near these water molecules did not disappear after the introduction of anisotropy, but rather became clearer. An analysis of the hydrogen-bond network around these water molecules confirms that the patches of electron density at each water molecule correspond to H atoms (Fig. 10). An interesting feature is observed in the directionality of these hydrogen bonds. Apart from well defined interactions with the protein core, the water molecules form a system of cooperative or homodromic water$\cdots$water hydrogen bonds (Saenger, 1979). W2 donates a hydrogen bond to W3, which donates a hydrogen bond to W4, which in turn donates a hydrogen bond to W5 (Fig. 10). Finally, W5 is connected to the solvent area by donating a hydrogen bond to Sul1, which is the best-defined sulfate ion in the structure.

3.6.3. C—H$\cdots$O hydrogen bonds. There is an increasing awareness of the importance of C—H$\cdots X$ hydrogen bonds in structural biology (Jeffrey & Saenger, 1991). In a growing number of recent reports on high-resolution protein structures, electron density for H atoms that are involved in such interactions could be visualized (Ridder *et al.*, 1999; Esposito *et al.*, 2000). In the present structure, electron density for H atoms is seen in many areas of the protein. In particular, many $\rm C^\alpha$ H atoms are clearly visible, especially if they are involved in C—H$\cdots$O hydrogen bonds, as in β sheets. As an example, Fig. 11(*a*) shows the antiparallel intermolecular β-sheet formed across a twofold axis. The $\rm C^\alpha$ H atoms are clearly seen even at the 3.0σ level of an H-omit difference map. From their location, it is obvious that they contribute a C—H$\cdots$O (Desiraju & Steiner, 1999) component into the system of interstrand hydrogen bonding. The fact that C—H$\cdots$O hydrogen bonds, together with the classic N—H$\cdots$O interactions, contribute to the cohesion of β-sheets has been noted by many authors (Derewenda *et al.*, 1995; Fabiola *et al.*, 1997; Ridder *et al.*, 1999; Esposito *et al.*, 2000). In the antiparallel case, it is the $\rm C^\alpha$—H donor of the preceding residue that enhances the N—H$\cdots$O bond accepted by a common carbonyl group on the other strand. A more intriguing picture emerges when the β-sheet in Fig. 11(*a*) is viewed from the side, as in Fig. 11(*b*). Here, the interacting chains, upper and lower, are seen in a projection along their general direction. When the H-omit difference density of Fig. 11(*a*) is plotted in this orientation, it is clearly misplaced by 0.2–0.3 Å from the $\rm (C^\alpha)$—H atom position generated on geometrical assumptions (Fig. 11*b*, left panel). This effect is even more pronounced when the difference map is calculated without removing the riding H atom from F_c calculation (Fig. 11*b*, right panel). This fact, together with the observation of a similar phenomenon at other (C)—H sites, suggests that the idealized angular $\rm C^\alpha$ geometry used to generate H atoms may not be exactly correct. It may be related to the chemical nature of the residues or more likely may depend on the involvement of these H atoms in relatively strong intermolecular interactions. Further analysis of accurate high-resolution protein structures will be necessary to shed more light on this problem.

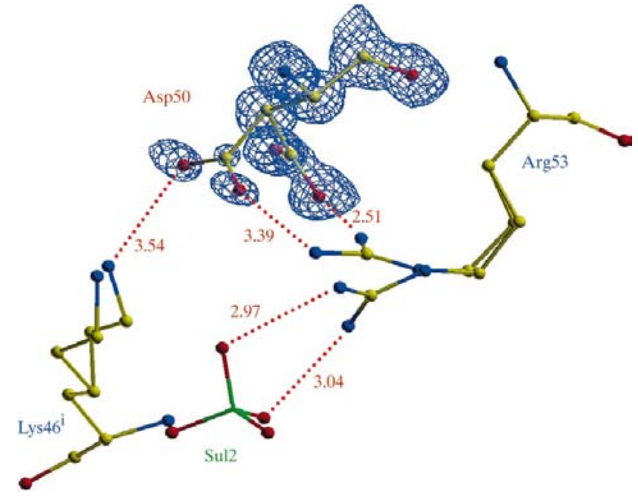


Figure 9
Disorder pattern involving Asp50 (shown in $3F_o - 2F_c$ density contoured at 0.57 e Å^{-3} or 1.5σ), Lys46 and Arg53. The system of hydrogen-bonding interactions (dotted lines, distances in Å) involves a Lys46 residue that is contributed by a twofold-related copy of the molecule (i).

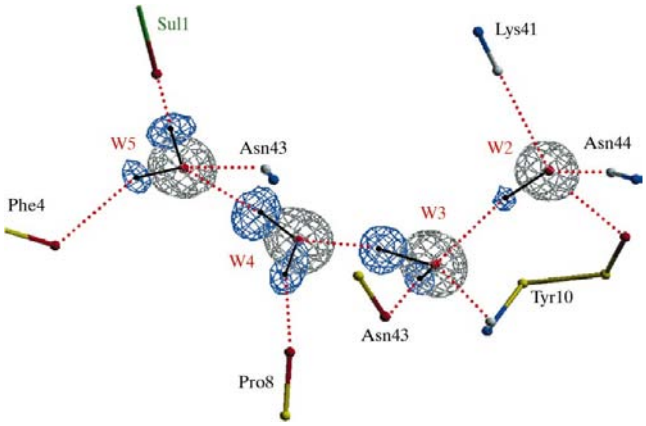
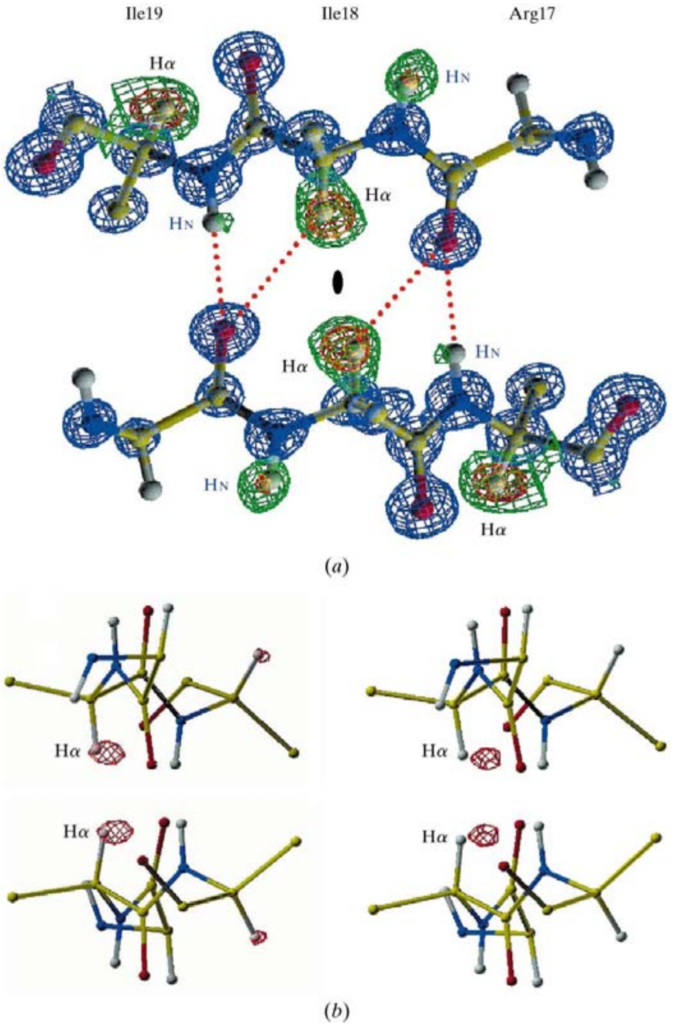


Figure 10
Electron density and hydrogen bonding around four internal water molecules. Note the excellent definition of the H atoms visible in a difference Fourier map contoured at 0.12 e Å^{-3} (2.9σ). Also note the involvement of residues Asn43 (main chain, donor and acceptor) and Asn44 (side chain, donor) in binding to this homodromic chain of water molecules. Waters W2, W3 and W5 have tetrahedral coordination; W4 donates two hydrogen bonds and accepts one. The positions of the H atoms are marked to indicate the centroids of the electron density. They were not included in the model.

Another example of the contribution of $C^{\alpha}-H\cdots O$ interactions to β-sheet integrity is provided in Fig. 12, which illustrates a long intramolecular β-sheet formed by the Ile18–Gly36 β-hairpin. In the centre of this β-sheet, the $C^{\alpha}-H\cdots O$ bonds are particularly strong, with $H\cdots O$ distances of 2.3–2.4 Å. At the ends of the β-sheet where the chains separate, the $C^{\alpha}-H\cdots O$ bonds are broken first, while the stronger $N-H\cdots O$ interactions are still preserved.

3.6.4. $N-H\cdots\pi$ hydrogen bonds. The importance of $N-H\cdots\pi$ hydrogen bonds in protein structures has been indicated in earlier reports (Levitt & Perutz, 1988; Perutz, 1993), also with reference to BPTI (Wlodawer *et al.*, 1984; Parkin *et al.*, 1996), where Tyr35 has been known to form such interactions. It precedes the Gly36-Gly37 tandem (discussed above) in the secondary loop. Its mutation to glycine in an earlier study resulted in the collapse of this loop and in drastic changes in the overall structure of BPTI (Housset *et al.*, 1991). This indicates that tyrosine at position 35 plays a major role in maintaining the overall topology of the structure. The significance of the aromatic character of Tyr35 is explained in Fig. 2. The phenyl ring is wedged between two amide N—H groups: one from the main chain of Gly37 and the other from the side chain of Asn44. There is no doubt about the orientation of the Asn44 amide group, as its H atoms are clearly visible in the H-omit map (Fig. 2*a*). The hydrogen-bonding potential of the two amide N—H donors facing Tyr35 can be satisfied only if one considers the π electrons of the Tyr35 aromatic ring to play the acceptor role (Desiraju & Steiner, 1999). Thus, the aromatic moiety supports the core of the structure by accepting two symmetrically disposed $N-H\cdots\pi$ hydrogen bonds. The amide H atom of Gly37 points to the centre of the aromatic ring at a distance of 2.63 Å and the angle around the H atom is 150° (Fig. 2). From the opposite face of the aromatic ring, the amide N—H of the side chain of Asn44 approaches the centroid of the aromatic ring at a distance of 2.85 Å and the angle at the H atom is 148°. The H atoms involved in these interactions are visible in the H-omit map (Fig. 2*a*). The angles formed by the planes of the amides (including the H atoms) of Gly37 (main chain) and of Asn44 (side chain) with the aromatic ring of Tyr35 are 89 and 78°, respectively. They ensure that these interactions are hydrogen bonds in nature and not stacking interactions (Mitchell *et al.*, 1993, 1994).



(*a*)

(*b*)

Figure 11
An antiparallel intermolecular β-sheet formed by symmetry-related β strands that are a direct continuation of the binding loop. (*a*) $3F_o - 2F_c$ density (blue) and H-omit difference map contoured at two levels. The C^{α} H atoms, seen even at very high contour level (0.12 e Å^{-3} or 3.0σ, red) clearly contribute to the interstrand hydrogen bonding. In (*a*) the β-sheet is viewed along its twofold axis. In (*b*) the figure has been rotated 90° and the twofold axis runs across the page between the two interacting strands. They are shown in two situations: left, together with a difference map (contour at 0.15 e Å^{-3} or 3.8σ) calculated with the H atoms omitted; right, with a difference map (contour at 0.10 e Å^{-3} or 2.5σ) calculated with the C^{α} H atoms contributing to F_c at their geometrically generated positions.

3.7. Solvent region

3.7.1. Water molecules. The final model comprises 59 fully and 111 partially occupied water molecules. The Matthews coefficient (Matthews, 1968) of the present crystal form, after taking into account the ethylene glycol molecules and sulfate anions, indicates about 190 water molecules in the asymmetric unit. This means that the present model accounts for nearly 90% of the water sites. It has to be remembered, however, that this crystal form is characterized by a rather high (considering the small size of BPTI) solvent content of 46%(*v*/*v*). The occupancy of the water molecules was adjusted manually depending on the refined *B* factors and the quality of their

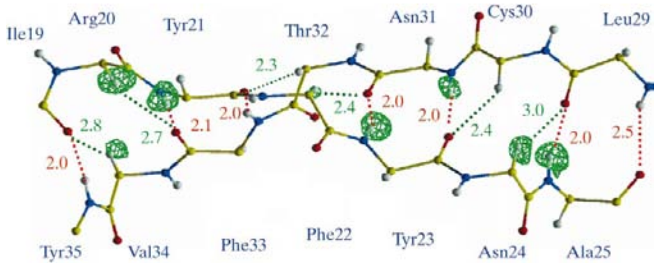


Figure 12
An antiparallel intramolecular β-sheet. Hydrogen bonds (dotted lines) are shown with their $H\cdots O$ distances (Å) for $N-H\cdots O$ (red) and $C^{\alpha}-H\cdots O$ (green) interactions. Note the excellent definition of the $(C^{\alpha})-H$ atoms in the H-omit difference map contoured at 0.10 e Å^{-3} (2.6σ).

electron density. The lowest accepted occupancy factor was 0.4. A test refinement in which the occupancies of the water molecules were allowed to vary was unsuccessful as the refinement was not stable.

Among the best-ordered waters are the five molecules that are internalized in this crystal form of BPTI (Czapinska *et al.*, 2000). W1, which is locked in a cavity formed by the primary and secondary loops and the bridging Cys14–Cys38 disulfide, is completely isolated from the rest of the solvent system. It has perfect tetrahedral coordination provided by the N—H groups of Cys14 and Cys38 and the C=O groups of Ala11 and Cys38 and is surrounded by several residues whose side chains are disordered. The other set comprises four water molecules that are perfectly defined and completely buried in the core (see above). The groove where they are located is formed exclusively by well ordered residues, including the non-canonical asparagine residues (Asn43-Asn44) discussed above. It is possible that the less-favoured conformation of these asparagines is caused in part by their involvement in this structure. The displacement factors of about 10 Å^2 for each of these water molecules, comparable with the average *B* factor of the main-chain atoms, illustrate their quality.

3.7.2. Sulfate ions. A total of eight sulfate anions have been modelled, four with full occupancy, two on a twofold axis, one in a general position with occupancy 0.5 and one sharing its site with three water molecules. All the sulfate ions found at room temperature, three with full occupancy (Sul1–Sul3) and one on a twofold axis (Sul4), as well as their interactions (Czapinska *et al.*, 2000) are conserved in this model. This low-temperature model includes another twofold-symmetric ion, Sul5. It is hydrogen bonded to the main-chain amide of Asp3 from two molecules. With two Asp3 carboxylate groups surrounding it, the environment of Sul5 is acidic or negatively charged, but the electron density proves beyond doubt that this is a sulfate moiety (Fig. 13*a*). As in the room-temperature structure, Sul4 is embedded in the arginine cage involving Arg15 and Arg39 and their symmetry equivalents. Of the other three sulfates, Sul7 is the most interesting. At the modelling stage it was built into a clear tetrahedrally shaped density with three slightly elongated S—O bonds. Around it, there are two strong hydrogen-bond donors from the main-chain amides of Ala11 and Gly12, interacting with two of the elongated corners. Refinement with full occupancy resulted in strongly negative difference density around the S atom. When the SO_4^{2-} group was replaced with four H_2O molecules at the corners of the tetrahedron, sufficiently far from the centre in order not to violate the BUMP restraint in *SHELX*, a large positive peak was found in the difference map at the centre of the tetrahedron and a negative peak at one of the corners (corresponding to the shortest 'S—O' distance). These results were interpreted as indicating that three corners of the tetrahedron represent fully occupied O atoms while the centre and the fourth corner represent partially occupied S and O atoms, respectively. Consequently, this site was interpreted as being shared by a partial-occupancy sulfate anion and a set of three water molecules of complementing occupancy. A similar situation where a sulfate site is shared by an overlapping network of water molecules has been reported previously in the ultrahigh-resolution structure of ribonuclease A (Esposito *et al.*, 2000). Refinement of the occupancies of these three water molecules and the sulfate group converged at 0.64 and 0.36, respectively.

In the room-temperature 1qlq structure the site of Sul5 is occupied by a relatively low-*B* water molecule. A single water molecule was also placed at the site of Sul6. In the area of the Sul7/water system two waters were modelled, corresponding to two of the three fractional water molecules in the present model. Nothing was built in the site of Sul8. All the above room-temperature water molecules had full occupancy.

3.7.3. Ethylene glycol molecules. Two ethylene glycol molecules were built in the electron density, one in *gauche* (Edo1) and one in *trans* (Edo2) conformation (Fig. 13*b*). Ethylene glycol was not present in the crystallization buffer and these molecules were incorporated into the crystal lattice at the cryoprotection stage. Although both molecules are located in solvent channels, one of the hydroxyl groups (1) of Edo1 forms good hydrogen-bond interactions with protein atoms. It donates a hydrogen bond to the main-chain carbonyl

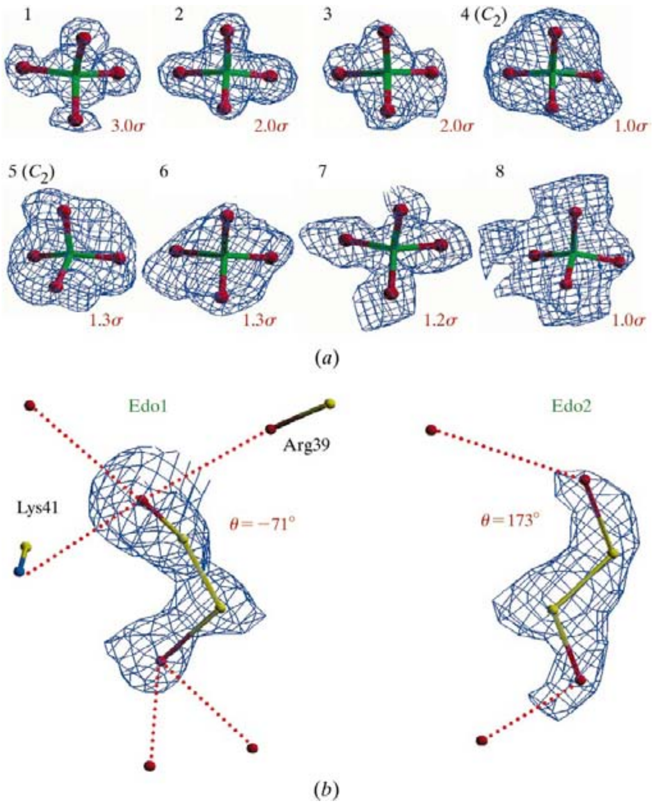


Figure 13
(*a*) The eight sulfate anions and their electron density. The first four anions correspond to those found in the room-temperature structure (1qlq). Residues Sul5–Sul8 are only present in the low-temperature model. The different contouring levels reflect the non-uniform quality of these ions. The best one is Sul1, the poorest Sul8. (*b*) The two ethylene glycol molecules in their $3F_o - 2F_c$ electron density (contour level: 1.0σ for Edo1; 1.3σ for Edo2). Also shown are the hydrogen bonds (dotted lines) formed by their hydroxyl groups and the values of the O—C—C—O torsion angles. In all maps, the contour level of 1σ corresponds to 0.38 e Å^{-3}.

of Arg39 while accepting another one from the well ordered NH_3^+ group of Lys41. There are also water molecules at hydrogen-bonding distances from all four OH groups of the glycol molecules (Fig. 13*b*). The conformational torsion angles of the glycol molecules correspond to nearly ideal values even though they were not restrained during the refinement. In the room-temperature structure 1qlq, the position of hydroxyl 1 of Edo1 was filled by a full-occupancy water molecule. No other solvent molecules were located within the areas occupied by the present glycol molecules.

4. Conclusions

The experiences gained during the refinement of this ultra-high-resolution protein structure allow formulation of some general comments and lessons.

(i) With resolution of 0.9 Å or better and typical protein crystal density (Matthews volume about 2 $Å^3\,Da^{-1}$), an observation-to-parameter ratio of 7 is easily achievable even with anisotropic displacement parameters. This level of over-determinacy is not very different from situations encountered in the refinement of non-centrosymmetric small-molecule crystal structures.

(ii) With so many reflections per parameter, it is entirely legitimate to remove all geometrical restraints from the well ordered main-chain regions. Additionally, the repulsive-interaction restraints ('bumps') should be relaxed or even removed in order not to bias the interpretation of non-classical intermolecular interactions. However, the geometrical restraints should be retained for the side chains, where disorder and higher displacement parameters may result in weaker definition by the diffraction data. Also, typical displacement-parameter restraints should be retained.

(iii) The standard target values and their weights used to stereochemically restrain the refinement may need revision. It is already obvious that the strictness of the peptide-bond planarity is usually exaggerated, but other parameters, such as the $N-C^{\alpha}-C$ angle, may also need adjustment.

(iv) Protein refinements at this resolution may produce a high-quality model. Typically, the estimated standard deviations in bond lengths should be in the range 0.005–0.010 Å. On the other hand, an r.m.s. deviation from target bond lengths, at least with the current target values, of the order of 0.02 Å should be considered normal and to correctly reflect the approximate nature of those targets.

(v) As a result of better definition and better quality of electron-density maps, the proportion of alternate-conformation residues in atomic resolution protein refinement (about 20%) may be higher than in typical medium-resolution refinements.

(vi) Precise high-resolution refinements allow better definition of molecular interactions involving H atoms and very often allow direct visualization of these atoms in electron-density maps. With the improvement of quality, it will become possible to verify and analyze in greater detail non-classical interactions, such as $C-H\cdots O$ hydrogen bonds (particularly within β-sheets) and $X-H\cdots\pi$ bonds. However, the placement of geometrically generated H atoms may not be adequate when these atoms are involved in significant interactions, such as C—H hydrogen bonds.

(vii) Owing to the rapidity and reproducibility of the crystallization, the crystals used for the present data collection were grown directly at the synchrotron site. Thus, this particular mutant of BPTI can be used as a test protein in various crystallographic studies requiring synchrotron radiation and as a standard in high-resolution studies.

The research of Mariusz Jaskolski was supported in part by an International Research Scholar's award from the Howard Hughes Medical Institute and that of Anthony Addlagatta by an award from the Mianowski Fund. Travel support provided under an exchange agreement between DESY and the Polish Academy of Sciences is gratefully acknowledged.

References

Allen, F. H., Baalham, C. A., Lommers, J. P. M. & Raithby, P. R. (1998). *Acta Cryst.* B**54**, 320–329.

Allen, F. H. & Kennard, O. (1991). *Chem. Des. Autom. News.* **8**, 31–37.

Ashida, T., Tsunogae, Y., Tanaka, I. & Yamane, T. (1987). *Acta Cryst.* B**43**, 212–218.

Berndt, K. D., Güntert, P., Orbons, L. P. & Wüthrich, K. (1992). *J. Mol. Biol.* **227**, 757–775.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Czapinska, H., Otlewski, J., Krzywda, Sz., Sheldrick, G. M. & Jaskolski, M. (2000). *J. Mol. Biol.* **295**, 1237–1249.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.

Deane, C. M., Allen, F. H., Taylor, R. & Blundell, T. L. (1999). *Protein Eng.* **12**, 1025–1028.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Desiraju, G. R. & Steiner, T. (1999). *The Weak Hydrogen Bond.* Oxford University Press.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–274.

Eigenbrot, C., Randal, M. & Kossiakoff, A. A. (1992). *Proteins Struct. Funct. Genet.* **14**, 75–87.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.

Evans, S. V. (1993). *J. Mol. Graph.* **11**, 134–138.

Fabiola, G. F., Krishnaswamy, S., Nagarjan, V. & Pattabhi, V. (1997). *Acta Cryst.* D**53**, 316–320.

Fritz, H. & Wunderer, G. (1983). *Arzneim. Forsch.* **33**, 479–494.

Housset, D., Kim, K.-S., Fuchs, J., Woodward, C. & Wlodawer, A. (1991). *J. Mol. Biol.* **220**, 757–770.

Jeffrey, G. A. & Saenger, W. (1991). *Hydrogen Bonding in Biological Structures.* New York: Springer–Verlag.

Jiang, X., Cao, M., Teppen, B., Newton, S. & Schafer, L. (1995). *J. Phys. Chem.* **99**, 10521–10525.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Karplus, P. A. (1996). *Protein Sci.* **5**, 1406–1420.

Krowarsch, D., Dadlez, M., Buczek, O., Krokoszynska, I., Smalas, A. O. & Otlewski, J. (1999). *J. Mol. Biol.* **289**, 175–186.

Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.

Levitt, M. & Perutz, M. F. (1988). *J. Mol. Biol.* **201**, 751–754.

Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.

MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **246**, 1180–1195.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mitchell, J. B. O., Nandi, C. L., Ali, S., McDonald, I. K., Thornton, J. M., Price, S. L. & Singh, J. (1993). *Nature (London)*, **366**, 413.
Mitchell, J. B. O., Nandi, C. L., McDonald, I. K., Thornton, J. M. & Price, S. L. (1994). *J. Mol. Biol.* **239**, 315–331.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–225.
Otting, G., Liepinsh, E. & Wüthrich, K. (1991). *Science*, **254**, 974–980.
Otting, G., Liepinsh, E. & Wüthrich, K. (1993). *Biochemistry*, **32**, 3571–3582.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 306–327.
Parkin, S., Rupp, B. & Hope, H. (1996). *Acta Cryst.* D**52**, 19–29.
Perutz, M. F. (1993). *Philos. Trans. R. Soc. Ser. A*, **345**, 105–112.
Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.
Read, R. J. (1997). *Methods Enzymol.* **277**, 110–128.
Ridder, I. S., Rozeboom, H. J. & Dijkstra, B. W. (1999). *Acta Cryst.* D**55**, 1273–1290.
Saenger, W. (1979). *Nature (London)*, **279**, 343–344.
Sandalova, T., Schneider, G., Kack, H. & Lindqvist, Y. (1999). *Acta Cryst.* D**55**, 610–624.
Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. (1986). *J. Mol. Biol.* **190**, 593–604.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1997). *The SHELX97 Manual.* University of Göttingen, Göttingen, Germany.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **227**, 319–343.
Srinivasan, N., Anuradha, V. S., Ramakrishnan, R., Sowdhamini, R. & Balram, P. (1994). *Int. J. Pept. Protein Res.* **44**, 112–122.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Pontius, J., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W., Vriend, G., Thornton, J. M., Laskowski, R. A. & MacArthur, M. W. (1998). *J. Mol. Biol.* **276**, 417–436.
Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). *J. Mol. Biol.* **180**, 301–329.