Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Refinement and structural analysis of barnase at 1.5 Å resolution

**Carole Martin,[a] Valerie Richard,[a] Michele Salem,[a] Robert Hartley[b] and Yves Mauguen[a]***

[a]Laboratoire de Physique, CNRS, ERS 582, Centre d'Etudes Pharmaceutiques, 92296 Châtenay-Malabry CEDEX, France, and [b]Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Health, Bethesda, MD 20892, USA

Correspondence e-mail: yves.mauguen@cep.u-psud.fr

Received 26 January 1998
Accepted 14 August 1998

**PDB Reference:** barnase, 1a2p.

The structure of *Bacillus amyloliquefaciens* ribonuclease (barnase), an extracellular 110-residue enzyme initially solved at 2.0 Å resolution, has been refined at 1.5 Å using synchrotron radiation and an imaging-plate scanner. Refinement with anisotropic atomic displacement parameters resulted in increased accuracy of the structure. The final model has a crystallographic $R$ factor of 11.5% and an $R_{\rm free}$ of 17.4%. The three independent molecules in the asymmetric unit, referred to as $A$, $B$ and $C$, allowed detailed analysis of this final model and meaningful comparison with structures of barnase complexed either with nucleotide inhibitors or with its natural intracellular inhibitor, barstar. The analysis of the overall solvent structure revealed a similar number of water molecules associated with each barnase molecule; among these were 16 equivalent buried solvent molecules, the locations of which are discussed in detail and classified on the basis of their structural role. The importance of the water molecules' contribution to the barnase–barstar interaction is also highlighted. The high accuracy of the present analysis revealed the presence of a $Zn^{2+}$ ion mediating the contacts between pairs of symmetry-related $A$, $B$ or $C$ molecules; such an ion had previously only been identified for pairs of $C$ molecules.

## 1. Introduction

Proteins play major roles in the structure and especially in the function of living organisms, often by interacting with other proteins or molecules. Knowledge of accurate three-dimensional structures would permit the elucidation of the mechanisms of molecular interactions and should provide a basis for constructing enzymes with modified properties, specific inhibitors, drugs and vaccines. Barnase, a bacterial ribonuclease, and barstar, its natural intracellular inhibitor, have been used extensively as model systems in protein-folding, protein–protein and protein–nucleotide interaction studies (Fersht, 1993; Frisch *et al.*, 1997; Buckle & Fersht, 1994).

Barnase is a small extracellular ribonuclease enzyme from *B. amyloliquefaciens* which belongs to the prokaryotic subgroup of the microbial ribonuclease family. Barnase catalyzes the cleavage of single-stranded RNA *via* a two-step mechanism involving a transesterification reaction to give a 2′,3′-cyclic phosphate intermediate, followed by hydrolysis to yield a 3′-nucleotide. Barnase hydrolyses the phosphodiester bond of GpG and GpA with a higher rate than other phosphodiester bonds (Nishimura, 1960; Day *et al.*, 1992). This preference still holds for RNA hydrolysis but is nonexclusive (Rushizky *et al.*, 1963); other RNAases, such as RNAase T1

and RNAase Sa, are highly guanosine-specific (Takahashi & Moore, 1982; Zelinkova *et al.*, 1971). Barnase is active with RNA as substrate in the pH range 7.5–9.3, with a maximum activity at pH 8.5. Its isoelectric point is pH 8.4. The protein consists of a single chain of 110 amino acids and no disulfide bridges, and its molecular weight is 12.3 kDa (Lees & Hartley, 1966; Hartley & Rogerson, 1972).

The barnase structure has been solved by X-ray crystallography (Mauguen *et al.*, 1982) and refined at 2.0 Å resolution (C. Hill and Y. Mauguen, unpublished results). The solution structure has also been determined by nuclear magnetic resonance in solution (Bycroft *et al.*, 1991). Barnase is an $\alpha+\beta$ protein with three amino-terminal $\alpha$-helices and a carboxy-terminal $\beta$-sheet supporting the active site. Comparison with other microbial ribonucleases (Hartley, 1980; Hill *et al.*, 1983) and site-directed mutagenesis experiments (Meiering *et al.*, 1991) indicate that residues Lys27, Glu73, Arg83, Arg87 and His102 are involved in catalysis. Residues 56–62 form the RNA-base recognition loop with a specificity for guanine nucleotides. Different point-mutant barnase structures are available to a resolution of 2.0 Å. These mutants have been designed in a systematic site-directed mutagenesis approach undertaken by several authors (Buckle *et al.*, 1993, 1996; Chen *et al.*, 1993, 1995; Clarke *et al.*, 1995) to investigate the effect of specific interactions on the stability of the protein. The overall structures are isomorphous with the wild-type structure. The crystal structure of wild-type barnase has been solved at pH 6 and pH 9, in both cases at 2.1 Å resolution (Buckle *et al.*, 1993).

Three other crystal structures of barnase complexed with nucleotide inhibitor have been determined. In the barnase–d(GpC) complex (Baudet & Janin, 1991), the nucleotide does not bind in a productive manner but is located outside the active site. The barnase–3′-GMP (Guillet, Lapthorn & Mauguen, 1993; Meiering *et al.*, 1993) and the barnase–$d(Cp_0Gp_1Ap_2C)$ (Buckle & Fersht, 1994) complex structures confirmed the importance of the recognition loop for the guanine specificity of the nucleotide base. The guanine base is bound to the main-chain NH group of residues Ser57 and Arg59 and to the side-chain atoms OE1 and OE2 of Glu60. Phosphate groups of the nucleotide make electrostatic interactions with barnase catalytic site basic residues Lys27, Arg83 and Arg87, and with Tyr103 for phosphate group $P_1$ and Arg83 for phosphate group $P_2$, while the active-site residues His102 and Glu73, which act as acid–base groups during catalysis, are bound to the deoxyribose.

The crystal structure of the complex of barnase with its inhibitor barstar (C40A/C82A) double mutant has been solved independently by Guillet, Lapthorn, Hartley *et al.* (1993) to 2.6 Å resolution and by Buckle *et al.* (1994) to 2.0 Å resolution. Barstar, produced by the same organism, inhibits barnase by binding to its active site, forming a very tight 1:1 complex with a $K_d$ of $10^{-14}$ *M* (Hartley, 1993; Schreiber & Fersht, 1993). The barnase-inhibition mode of barstar is a steric blockade of the enzyme active site. In the barnase–barstar complex, the two barnase phosphate-binding sites are occupied by the Asp39 side-chain carboxylate and the Gly43 main-chain carboxyl of barstar. However, barstar has no equivalent for the guanine base of an RNA substrate, and the guanine-recognition site in the barnase–barstar complex is occupied by ordered water molecules.

Studies of ribonucleases as model systems (small size, solubility, stability) have made a major contribution to understanding the structure, function and stability of enzymes. Many spectroscopic, thermodynamic and biochemical experimental data have been accumulated on barnase. In this paper we present a 1.5 Å resolution model of barnase refined using *SHELXL*93 (Sheldrick, 1993) and including anisotropic modelling of the protein atomic displacement parameters. The protein coordinates have allowed a precise comparison between free and complexed forms of this enzyme and will be useful for theoretical calculations.

## 2. Experimental methods

### 2.1. Crystallization

Barnase expression was performed by culture of *Escherichia coli* strain XL1-blue containing the recombinant plasmid pMT1002 with the wild-type barnase gene. Barnase was purified as described previously (Hartley & Rogerson, 1972). Crystallization of barnase at pH 7.5 was based on the hanging-drop method. The barnase solution, at 25 mg ml$^{-1}$ concentration and free of all salts other than ammonium sulfate, was added in a 4:3 volume ratio to a crystallization medium containing 3.0 *M* ammonium sulfate and 0.03 *M* zinc sulfate. Droplets (5–10 µl), placed on siliconized cover slips, are initially placed over a well containing a pre-equilibration solution of 1.9 *M* phosphate (1/10 dilution yields pH 6.3). These are left at room temperature between 6 h and overnight. The cover slips are then transferred to wells containing 2.9 *M* phosphate (1/15 dilution yields pH 7.5). After another 6–20 h, the drops are seeded by inserting a fine Pt wire into an appropriate seed suspension. Seeded crystals are prepared by adding saturated ammonium phosphate to a small amount of crystallization medium/barnase solution. Crystals stop growing after two or three days. Drops without any crystals after 24 h at room temperature may be seeded again. The crystals belong to space group $P3_2$ with three molecules in the asymmetric unit, which are hereafter referred to as molecules *A*, *B* and *C*. The unit-cell parameters used in the refinement, $a = b = 58.97$ (10) and $c = 81.58$ (15) Å, were determined on a computer-controlled Hilger–Watts four-circle diffractometer (Jack, 1973).

### 2.2. Data collection

Data were collected using synchrotron radiation at room temperature from a single crystal of native enzyme, of approximate dimensions 200 × 200 × 525 µm, on the LURE W32 beamline ($\lambda = 0.994$ Å) with a MAR Research imaging-plate scanner. Two data sets, each covering 60° rotation, were collected at different exposure times to 1.5 and 2.2 Å resolution with an oscillation angle per image of 1.2 and 2°, respectively. Data were processed using the *MOSFLM*

**Table 1**
Native data quality for the data used for refinement.

| $D_{min}$ (Å) | $R_{merge}$ | $N_{meas}$ | Mean $I/\sigma(I)$ | Completeness (%) (multiplicity) |
|---|---|---|---|---|
| 4.63 | 0.028 | 5858 | 19.3 | 100.0 (3.5) |
| 3.32 | 0.032 | 10483 | 18.7 | 99.4 (3.6) |
| 2.72 | 0.037 | 14061 | 17.3 | 99.1 (3.8) |
| 2.36 | 0.046 | 16545 | 14.5 | 98.7 (3.8) |
| 2.12 | 0.047 | 14781 | 14.2 | 98.2 (3.0) |
| 1.93 | 0.037 | 10175 | 14.7 | 96.6 (1.9) |
| 1.79 | 0.049 | 11079 | 14.4 | 97.0 (1.9) |
| 1.68 | 0.070 | 11758 | 10.2 | 95.8 (1.9) |
| 1.58 | 0.107 | 12456 | 6.9 | 95.4 (1.9) |
| 1.50 | 0.158 | 13037 | 4.7 | 94.5 (1.9) |
| Totals | 0.038 | 120253 | 11.9 | 96.9 (2.5) |

package (Leslie, 1992). A summary of data collection and processing is given in Table 1. The reduced data were 96.9% complete and contained 48313 unique reflections between 10 and 1.5 Å with an overall $R_{merge}$ of 3.8%. 89.6% of all reflections were greater than three standard deviations (71.3% in the highest resolution bin). The overall $B$ factor estimated from the Wilson plot (Wilson, 1942) is 16.15 Å$^2$.

### 2.3. Refinement

Refinement of the atomic coordinates and displacement parameters was carried out by simulated annealing and by stereochemically restrained least-squares minimization (*X-PLOR*; Brünger, 1992) using as a starting model the structure previously refined at 2.0 Å (Y. Mauguen, unpublished results). Rebuilding of the model was based on $(3F_o - 2F_c)$ and $(F_o - F_c)$ maps, and was performed on a Silicon Graphics workstation using the program packages *TURBO-FRODO* (Roussel & Cambillau, 1991) and *O* (Jones & Kjeldgaard, 1993). Solvent molecules were added to the model both manually and by the automated procedure *ARP* (Lamzin & Wilson, 1993), which is based on searching the difference Fourier map for peaks within hydrogen-bonding distance of protein atoms or other water molecules. Water molecules lying in density less than $0.5\sigma$ above the mean density in the $(3F_o - 2F_c)$ map were periodically deleted from the model. Solvent molecules were refined with occupancy 1.0.

The model coordinates from *X-PLOR* and *ARP* refinement were used as a starting model for refinement with *SHELXL*93 (Sheldrick, 1993). *SHELXL*93 refines against intensities rather than amplitudes, enabling direct use of all measured observations, including those with negative values. Protein bond and angle restraints used were those of Engh & Huber (1991). Restrained isotropic and anisotropic least-squares refinements were carried out using the conjugate-gradient algorithm (Konnert & Hendrickson, 1980) as implemented in the *CGLS* routine of *SHELXL*. In *SHELXL*, the atomic Gaussian Debye–Waller factor is given in terms of the anisotropic displacement parameters $U^{ij}$ (ADP) by the expression $\exp[-2\pi^2(h^2a^{*2}U^{11} + k^2b^{*2}U^{22} + \ldots + 2hka^*b^*U^{12})]$. ADPs can be converted to the equivalent $B$ values ($B_{eq}$), which are more commonly used in protein crystallography, by the relation $B_{eq} = 8\pi^2U_{eq}$, where $U_{eq}$ is the isotropic mean-square displacement equivalent to a set of ADPs (Trueblood *et al.*, 1996). Anti-bumping distance restraints (*SHELXL* command BUMP) were added to prevent non-bonded collisions. H atoms were included at the end of the isotropic refinement at calculated positions; isotropic displacement parameters of the H atoms were set to 1.2 times the $U_{eq}$ value of the atoms to which they were attached (1.5 for OH and methyl groups). Protein ADPs were restrained during refinement by requiring that atoms closer than 1.7 Å have the same $U^{ij}$ components with a standard deviation of 0.05 Å$^2$ (*SHELXL* command SIMU); moreover, for pairs of chemically bonded atoms the components of the anisotropic displacement in the direction of the bond (1–2 and 1–3 neighbours) were restrained to be equal with a standard deviation of 0.01 Å$^2$ (*SHELXL* command DELU). ADPs of water molecules were restrained to approximate isotropic behaviour with a standard deviation of 0.10 Å$^2$ (*SHELXL* command ISOR). 90% of the X-ray data were used for refinement cycles, 10% being omitted and used to monitor $R_{free}$ variations both with *X-PLOR* and *SHELXL*93. All data were included in the last stages of refinement.

## 3. Results

### 3.1. Accuracy of the model

The progress of the refinement is summarized in Table 2. The initial $R$ factor was 33.6% between 10 and 1.5 Å. Conventional refinement (*X-PLOR* and *ARP*) of the three molecules (2624 non-H protein atoms, 375 water molecules and three $Zn^{2+}$ cations) converged to an $R$ factor of 17.8% for 45423 unique reflections with $F_o > 3\sigma(F_o)$ in the range 7–1.5 Å. Difference electron-density peaks observed near the zinc ion bound to molecule C (Fig. 1) before introduction of ADPs represent its thermal vibration. Subsequent rounds of refinement were carried out with *SHELXL*93. The gradual addition of 40 water molecules, insertion of calculated H-atom positions and anisotropic refinement for the zinc ion bound to molecule C reduced both the $R$ and $R_{free}$ values and reduced the maximum of the $F_o - F_c$ electron density from 0.56 to 0.38 e Å$^{-3}$. The criteria for retaining solvent sites were that after refinement they had acceptable $B_{eq}$ values (<70 Å$^2$), they displayed well defined spherical $2F_o - F_c$ density at least a standard deviation in height and that they were within 2.3 and 3.5 Å distance of a plausible hydrogen-bonding partner. At this stage, the reduction in both conventional $R$ and $R_{free}$ resulting from restrained anisotropic refinement for non-H protein atoms were 4.5 and 2.9%, respectively. Another 1% drop in both $R$ and $R_{free}$ was obtained by the anisotropic refinement of water molecules restrained to approximate isotropic behaviour. It is usually considered that restrained ADP refinement is unlikely to be justified for crystals that do not diffract to better than 1.5 Å. Its validity depends on the quality of the data and can be monitored with the variation of $R_{free}$ (Sheldrick & Schneider, 1997). Despite the low value of the data-to-parameter ratio at the end of the refinement (1.8), the 4.5% drop in $R_{free}$ between the isotropic and the aniso-

**Table 2**
Summary of the different stages of refinement.

| Action taken | $N_P$† | $N_H$‡ | $N_W/N_X$§ | $N_{par}$¶ | $N_{data}$†† | $R$‡‡ (%) | $R_{free}$§§ (%) |
|---|---|---|---|---|---|---|---|
| Initial model [10–1.5 Å; $F_o > 3\sigma(F_o)$] | 2592 | 0 | 0/1 | 10373 | 41190 | 33.6 | — |
| Final *X-PLOR*/*ARP* model [7–1.5 Å; $F_o > 3\sigma(F_o)$] | 2624 | 0 | 375/3 | 12009 | 40880 | 17.4 | 21.9 |
| Riding H atoms added, solvent added, zinc ion anisotropic (*SHELXL*93) (8–1.5 Å) | 2624 | 2499 | 410/3 | 12155 (10438) | 43306 | 16.6 | 21.1 |
| Minor adjustments, all non-H protein atoms anisotropic | 2624 | 2499 | 416/3 | 25309 (30506) | 43306 | 12.1 | 18.2 |
| Waters anisotropic | 2624 | 2499 | 415/3 | 27380 (32995) | 43306 | 11.1 | 17.4 |
| All reflections (8–1.5 Å) | 2624 | 2499 | 415/3 | 27380 (33063) | 48118 | 11.5 | |

† $N_P$ = Number of protein atoms (including partially occupied atoms). ‡ $N_H$ = Number of H atoms. § $N_W$ = Number of waters, $N_X$ = number of zinc ions. ¶ $N_{par}$ = Number of least-squares parameters and, in parentheses, final number of restraints for *SHELXL*93. †† $N_{data}$ = Number of reflections used in refinement. ‡‡ $R = 100\,(\sum ||F_o| - |F_c||/\sum |F_o|)$. §§ $R_{free}$ = $R$ calculated for the 10% of data not used in refinement.

tropic refinements and its 2.9% drop when protein atom ADPs were refined suggest that the introduction of anisotropy was sensible. An *ORTEP* (Burnett & Johnson, 1996) diagram of residues Val36–Lys39 of molecule *B* (localized in a loop), representative of most parts of the structure, is given in Fig. 2(*a*); the main-chain atoms, stabilized by a hydrogen bond between the carbonyl O atom of Val36 and the amide N atom of Lys39, show relatively isotropic small displacements, whereas side-chain atoms of Lys39 exposed to the solvent exhibit more pronounced anisotropic motion. In Fig. 2(*b*), residues Asp44–Pro47 of molecule *C* (localized at the end of the third helix) illustrate a less well ordered part of the structure with high anisotropic motion; the rigid-bond restraints on ADPs here result in an approximate rigid-body motion, perpendicular to the axis of the helix, towards the solvent.

The final $R$ factor is 0.115 for the 48118 unique reflections in the range 8–1.5 Å and 0.112 for the 44595 reflections with $F_o > 4\sigma(F_o)$ and the goodness of fit is 1.09. The final difference Fourier map was essentially flat with a standard deviation of 0.05 e Å$^{-3}$; the highest and lowest values were 0.29 and −0.28 e Å$^{-3}$, respectively. Significant positive peaks in the final $(F_o - F_c)$ map are located near the N-terminal extremity, the disorder of which could not be modelled; the negative peaks are located in the vicinity of disordered residues (Lys19*B* and Asn22*A*). A difference electron-density map was calculated using structure factors calculated from a model in which five well ordered water molecules were removed from the coordinate list. The deleted water molecules all appeared as peaks of at least 20$\sigma$ in the difference map. Deviations from ideal stereochemical parameters are 0.013 Å for bond lengths and 1.76° for bond angles. The high accuracy of this structure is confirmed by the Ramachandran plot for the three molecules (Fig. 3) generated using *PROCHECK* (Laskowski *et al.*, 1993), which shows no residues in the non-allowed $\varphi,\psi$ regions. The high accuracy is also confirmed by the estimated average coordinate error of 0.05–0.10 Å from the Luzzati plot (Luzzati, 1952) and of 0.045 Å from the $\sigma_A$ method of Read (1986). After conjugate-gradient refinement of all parameters had converged, a cycle of full-matrix blocked least-squares refinement was carried out in order to calculate standard uncertainties for the positional parameters. For this additional cycle, all restraints were turned off during the calculation to give unbiased estimates; three blocks were used, each block containing all positional parameters of one barnase molecule in the asymmetric unit. The overall three-dimensional positional standard uncertainties are 0.058 Å for protein main-chain atoms and 0.073 Å for all protein atoms, in agreement with values derived from the Luzzati plot.

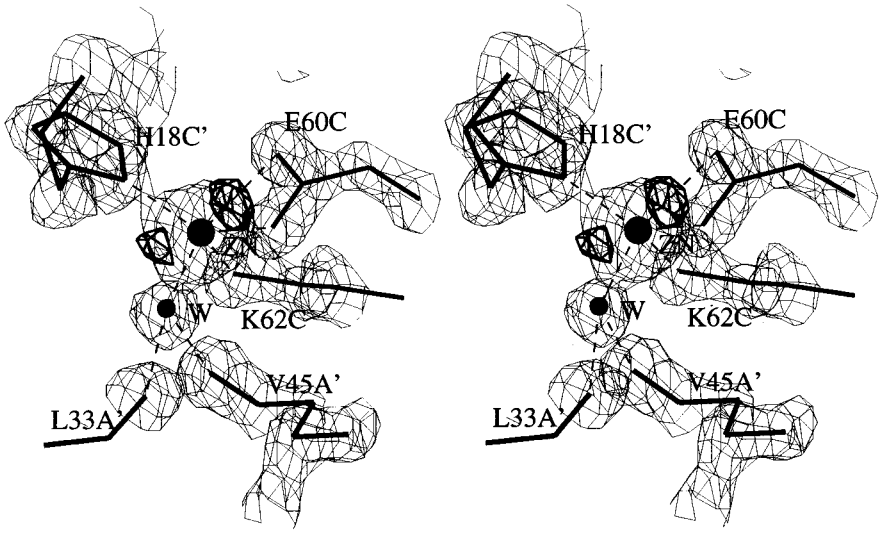


**Figure 1**
Zinc site of molecule *C*: zinc ion in ($3F_o - 2F_c$, $\alpha$c) electron-density map, contour level 2$\sigma$; regions of positive difference electron density contoured at 4$\sigma$ (bold line) show the anisotropy of the zinc thermal motion.

The final model is complete with the exception of the N-terminal Ala1 and Gln2 residues for which there is a lack of continuous density in the three molecules. This N-terminal disorder is a common feature of most barnase structures solved to date. Moreover, there are several poorly defined residues for which electron density for at least one side-chain atom in the ($3F_o - 2F_c$) map is less than 0.5$\sigma$ and/or the $B_{eq}$ value is higher than 45 Å$^2$, such as Tyr17*A*, Lys19*B*, Asn22*A*, Asn22*B*, Lys39*C*,

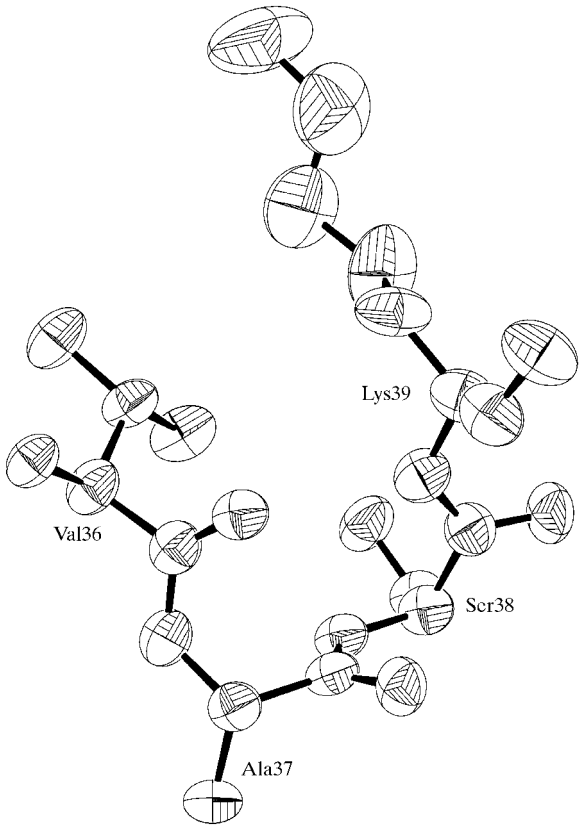


**Figure 2**
*ORTEP* (Burnett & Johnson, 1996) views of two groups of atoms. The thermal ellipsoids are drawn at the 20% probability level. (*a*) Residues Val36–Lys39 of molecule *B*, (*b*) residues Asp44–Pro47 of molecule *C*.

Arg59*A*, Arg59*B*, Arg59*C* and Ser67*B*. Most of these side-chain extremities have been modelled with partial occupancy factors. For the Arg59 side-chains, there is hardly any density in molecules *A* and *B*, in contrast to molecule *C* where mobility of this residue is limited by the close proximity of a symmetry-related molecule *C* (symmetry operation: $-x + y + 1, -x + 1, z + 1/3$). The Arg59*A* and Arg59*B* side chains are included in the model with zero occupancy values.

### 3.2. The 1.5 Å barnase structure

The 1.5 Å model of barnase is essentially the same as in previously reported structures. In order to facilitate the following discussion, the main features of the barnase structure are reiterated briefly here. Barnase has been described as an $\alpha+\beta$ protein composed of three $\alpha$ helices (residues 6–18, 26–33 and 41–46) and a five-stranded antiparallel $\beta$-sheet (residues 50–56, 70–76, 86–91, 95–99 and 107–109) which, by packing against the $\alpha1$ helix, forms the major hydrophobic core of the protein (Serrano *et al.*, 1992). These secondary-structure elements are linked by five loops (residues 19–25, 34–40, 57–69, 77–85 and 100–106) and two $\beta$-turns (residues 46–49 and 91–94). Residues 23–25 of loop 1 connecting the $\alpha1$ and $\alpha2$ helices form a short strand parallel and hydrogen bonded to the first strand of the $\beta$-sheet, while residues 36–40 of loop 2 between the $\alpha2$ and $\alpha3$ helices can be described as a $3_{10}$ helix. Loop 3 is the longest and includes the residues involved in RNA-base recognition (Sevcik *et al.*, 1990). Loops 4 and 5 link strands II and III, and IV and V of the $\beta$-sheet, respectively.

There is little overall difference between the initial model determined at 2 Å resolution and that at 1.5 Å resolution; the r.m.s. difference between the two refined structures is 0.18 Å

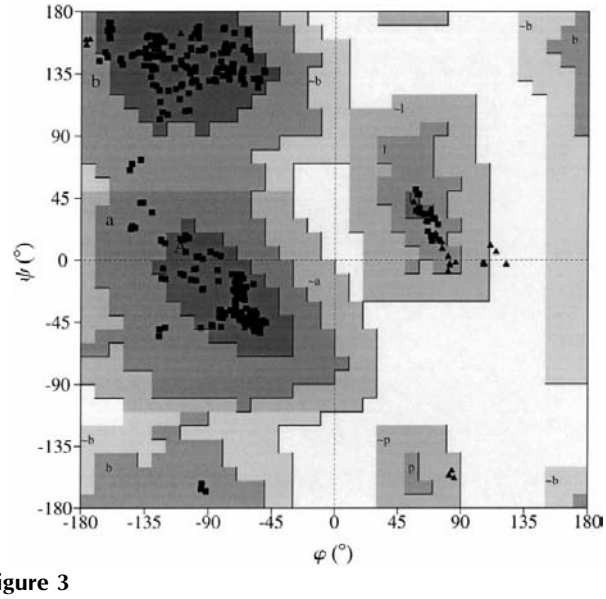


**Figure 3**
Ramachandran plot for the barnase structure drawn with the *PROCHECK* program. 89% of the residues are in the most favoured regions. Glycine residues are shown as triangles. The different regions defined by borderlines are labelled A, B and L (most favoured), a, b, l and p (allowed), and ~a, ~b, ~l and ~p (generously allowed) (Laskowski *et al.*, 1993).

for backbone atoms and 0.40 Å for all atoms common to both sets of coordinates. However, the electron-density maps calculated from the high-resolution data provide considerable details and reveal new structural information in the form of multiple side-chain conformers and additional solvent molecules. Improvement of the electron density is illustrated in Fig. 4 which shows the same residues 34*B*–36*B* taken from the structure previously refined at 2.0 Å resolution and from the present model.

As frequently observed in structures refined at sufficiently high resolution, several side chains on the surface of the protein adopt more than one discrete conformation (Smith *et al.*, 1986). The following residues have been interpreted as double conformers: Ser28*A*, Ser28*B*, Gln31*A*, Ser38*A*, Ser38*B*, Lys49*C*, Ser85*A*, Ile96*A*, Ile96*B* and Ile96*C*. Hydrogen-bond interactions with solvent and protein, as shown in Fig. 5 for Ser38*A*, explain the alternate positions of polar side chains and support the interpretation. All of these residues are exposed to solvent with the exception of Ile96, which is completely buried in the hydrophobic core of the protein. Discrete disorder of side chains has been described previously in the structure of the barnase–d(CGAC) complex at 1.76 Å resolution. Interestingly, the side chain of Ile96 in one molecule of the asymmetric unit also adopts two conformations close to those observed in free barnase. This indicates that the core in barnase is not perfectly well packed in the vicinity of Ile96 and that even in the hydrophobic interior the protein still has some flexibility. Contrary to Ile96, none of the residues lying on the surface of the free protein are disordered in the three molecules and thus it is probable that they do not reflect a general structural feature.

### 3.3. Crystal packing

The three barnase molecules (*A*, *B* and *C*) in the asymmetric unit are not related by a simple non-crystallographic symmetry axis. Each independent molecule and its crystallographically symmetry-related molecules stack in columns along the threefold screw axis of the cell (Fig. 6). The columns of molecules *B* and the columns of molecules *C* have the same direction whilst the columns of molecules *A* are in the opposite direction. Crystal contacts have been analyzed using a 4 Å distance cutoff. Contacts for each molecule, together with information on the symmetry and components of each contact, are given in Table 3. More contacts are made between molecules in the same column than between two molecules in different columns. A molecule *C* forms 14 hydrogen bonds

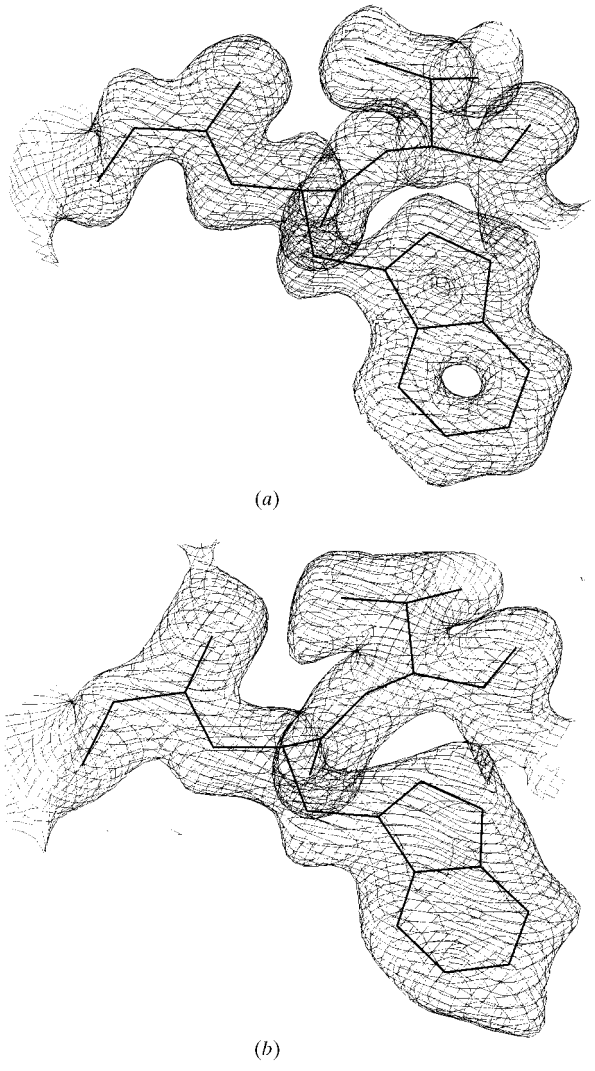


**Figure 4**
(*a*) Residues 34–36 in ($3F_o - 2F_c$, $\alpha c$) electron-density map contoured at 0.05 e Å$^{-3}$ (1$\sigma$). (*b*) Residues 34–36 in ($3F_o - 2F_c$, $\alpha c$) electron-density map from the 2.0 Å resolution structure contoured at 1$\sigma$.

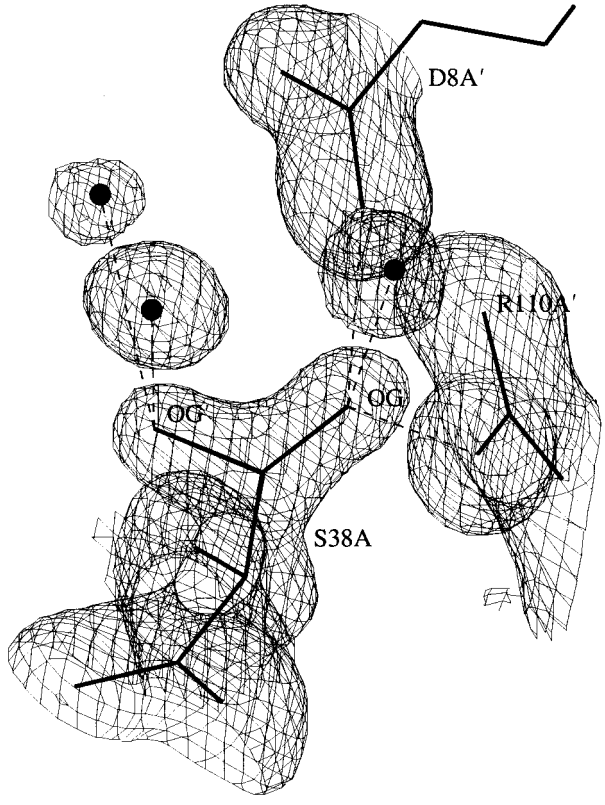


**Figure 5**
Ser38*A* with two alternative conformations in ($3F_o - 2F_c$, $\alpha c$) electron-density map contoured at 0.05 e Å$^{-3}$ (1$\sigma$).

**Table 3**
Molecular contacts in barnase crystals, space group $P3_2$.

Hydrogen-bond cutoff distance 3.3 Å; contact cutoff distance 4.0 Å.

| | Contacting molecule | Symmetry relation to contacting molecule | Contacting atoms ($n_1$:$n_2$)† | Contacting residues ($m_1$:$m_2$)‡ | Number of hydrogen bonds |
|---|---|---|---|---|---|
| Molecule *C* | *A* | $x-1, y-1, z$ | 3:3 | 2:2 | 0 |
| | *B* | $x, y, z$ | 7:10 | 5:5 | 3 |
| | *B* | $-y+1, x-y+1, z-1/3$ | 13:10 | 6:5 | 2 |
| | *B* | $-y+2, x-y+1, z-1/3$ | 4:8 | 2:3 | 2 |
| | *C* | $-y+1, x-y, z-1/3$ | 37:38 | 9:11 | 7 |
| | *A* | $-x+y+1, -x+1, z+1/3$ | 12:12 | 3:5 | 2 |
| | *C* | $-x+y+1, -x+1, z+1/3$ | 38:37 | 11:9 | 7 |
| Molecule *B* | *A* | $-y+1, x-y+1, z+2/3$ | 10:10 | 6:4 | 1 |
| | *B* | $-y+1, x-y+1, z-1/3$ | 28:32 | 10:8 | 6 |
| | *A* | $-x+y, -x+2, z+1/3$ | 9:12 | 4:6 | 2 |
| | *B* | $-x+y, -x+1, z+1/3$ | 32:28 | 10:8 | 6 |
| Molecule *A* | *A* | $-y+2, x-y+1, z-1/3$ | 28:40 | 8:9 | 6 |
| | *A* | $-x+y+1, -x+2, z+1/3$ | 40:28 | 9:8 | 6 |

† Number of atoms of the first ($n_1$) and of the second ($n_2$) molecule involved in the molecular contact. ‡ Number of residues of the first ($m_1$) and of the second ($m_2$) molecule involved in the molecular contact.

with the two neighbouring symmetry-related *C* molecules (in the *C* column) but only nine hydrogen-bonds with the contacting *A* and *B* molecules. Molecule *C* is the most stabilized molecule with seven contacting neighbours. The relative positions of two symmetry-related *A* or two symmetry-related *B* molecules are very similar to those of two symmetry-related *C* molecules, with hydrogen bonds stabilizing the columns in the same way for each of the three columns.

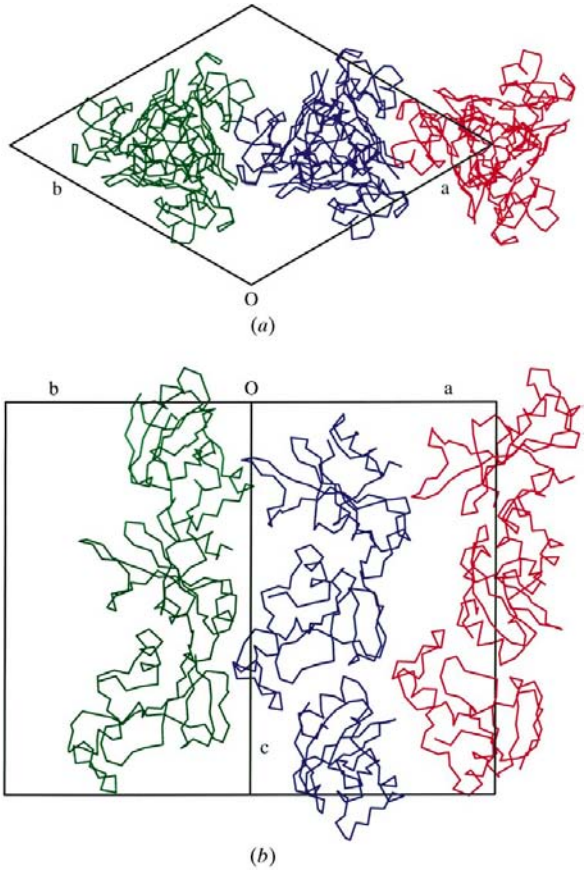


**Figure 6**
Crystal packing. Projection of the nine molecules of the unit cell: (*a*) along the *c* axis, (*b*) along the direction $a + b$ (molecule *A*, red; molecule *B*, green and molecule *C*, blue).

In each column, a zinc ion mediates the contacts between two symmetry-related molecules. Only the zinc ion bound to molecule *C* was identified in previously determined wild-type and mutant barnase structures. The identification of this site as a zinc ion was initially based on (i) the requirement for zinc sulfate in the crystallization medium, (ii) the height of the peak in difference Fourier maps calculated during the 2 Å refinement, (iii) the low isotropic *U* value of the zinc ion when introduced into the refinement with full occupancy, (iv) the disappearance of this peak in the crystal structure of the iodobarnase derivative obtained in a mother liquor without zinc sulfate (Hill, 1986) and (v) the chemical environment of the zinc ion. As shown in Fig. 1, this ion is bound to Glu60*C* (both to OE1 and OE2), Lys62*C* and His18*C*′ (symmetry-related *C* molecule) and to a water molecule which not only coordinates the zinc ion but also makes good hydrogen bonds with the carbonyl O atom of residues 33*A*′ and 45*A*′ (neighbouring symmetry-related molecule *A*). This zinc ion and its bound water molecule are completely surrounded by ordered protein atoms and do not make contact with any other solvent molecule. This proximity of a molecule *A*′ to the *C* zinc site is the only significant difference to the other *A* and *B* zinc sites. The zinc sites of molecules *A* and *B* are exposed to solvent and at 1.5 Å resolution there is density in the $(F_o - F_c)$ map at distances from protein atoms expected for a zinc ion. The occupancy factor was determined to be 0.30 by examining $(F_o - F_c)$ maps and was further refined with *SHELXL*93 to 0.30 (3) and 0.35 (3) for the zinc ions bound to molecule *A* and *B*, respectively. For molecules *A* and *B*, the zinc ion is bound to an exposed water molecule which was given the same occupancy factor as that of the cation. The standard uncertainties of the zinc occupancy factors were calculated at the end of the refinement by performing a cycle of full-matrix block refinement, each block containing positional parameters and ADPs of the zinc ions and all the surrounding residues.

### 3.4. Displacement parameters

The average $B_{eq}$ value is 19.9 Å$^2$ for the protein atoms (excluding H atoms), 39.2 Å$^2$ for the solvent molecules and 22.6 Å$^2$ for the 3042 atoms included in the model. Average protein $B_{eq}$ values for the three molecules, 20.8, 20.1 and 18.8 Å$^2$ for molecules *A*, *B* and *C*, respectively, are very similar. Plots of the average $B_{eq}$ values for the main-chain atoms of each residue in the three independent barnase molecules are given in Fig. 7; they are useful indicators of the inherent flexibility of the polypeptide-chain segments. Regions with $B_{eq}$ values greatly exceeding 30 Å$^2$ are mainly localized at the N- and C- termini of the chains and in loop 3 (residues 57–69), which contains all RNA-base recognition residues; as expected, lower $B_{eq}$ values are associated with secondary-structure elements. The $B_{eq}$ value plots for the three molecules show a significant correlation; this is true for molecules *A* and *B* along the whole chain and for the three molecules in the C-terminal half corresponding to the $\beta$-sheet. As a result of crystal packing, the $B_{eq}$ values of molecule *C* are on average smaller than those of the other molecules and their variations in the N-terminal half have a distinct behaviour. Crystal contacts involving residues 25–50 (the $\alpha$2 helix, loop 2 and the $\alpha$3 helix) are different for the three molecules and allow more freedom for molecule *C*. For the $\alpha$1 helix most of the crystal contacts are very similar, but the C-terminal end of this helix in molecule *C* is stabilized by interactions with a neighbouring molecule *A* *via* a zinc ion.

### 3.5. The three independent barnase molecules

The three copies of the barnase molecule in the asymmetric unit are essentially identical: a systematic superposition of two independent molecules gives r.m.s. differences of 0.23, 0.29 and 0.31 Å for C$\alpha$ atoms and 0.49, 0.37 and 0.50 Å for all non-H atoms. Least-squares superposition (Fig. 8) and inspection of the graphics shows that the major differences are limited to surface atoms involved in intermolecular contacts and that there are no major changes in conformation. The largest r.m.s. differences are localized in a chain segment including loop 2 and the $\alpha$3 helix, in loop 3 and, to a lesser extent, in loops 4 and 5 and the last strand of the $\beta$-sheet. Least-squares superposition of the main-chain atoms of the hydrophobic core ($\alpha$1 helix, $\beta$-I, II, III, IV strands) gives r.m.s. differences of 0.08, 0.09 and 0.11 Å, which are close to the average coordinate error. These figures and the variations of $B_{eq}$ values along the polypeptide chain show that barnase is composed of a rather rigid core, but that a relatively high proportion of the protein has some flexibility. This picture is in agreement with the solution structure of barnase determined by NMR (Bycroft *et al.*, 1991). The catalytic site, located in a depression at the surface of the $\beta$-sheet, is extremely conserved, except for the last few atoms in the Lys27 side chain, which project into the solvent, and the side chain of His102 which is involved in intermolecular contacts. A large number of hydrogen bonds, discussed below, are formed between the charged catalytic site residues and between these residues and ordered conserved solvent molecules. All these interactions stabilize the catalytic site conformation.

### 3.6. Water structure

This barnase model contains 415 solvent molecules with an average $B_{eq}$ value of 39.2 Å$^2$. Statistics of the hydrogen-bond interactions of solvent molecules with protein atoms are described in Table 4. The trend in $B_{eq}$ values correlates well with the number of hydrogen-bond interactions, decreasing from 49 Å$^2$ for water molecules making only one hydrogen-bond interaction to 32 Å$^2$ for those making four hydrogen-bond interactions. Protein–water interactions may be grouped

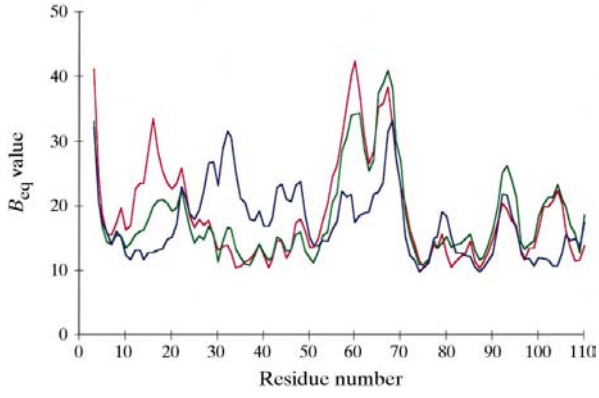


**Figure 7**
Average $B_{eq}$ values for main-chain atoms as a function of residue number for the three independent barnase molecules: molecule *A* (red), molecule *B* (green) and molecule *C* (blue).

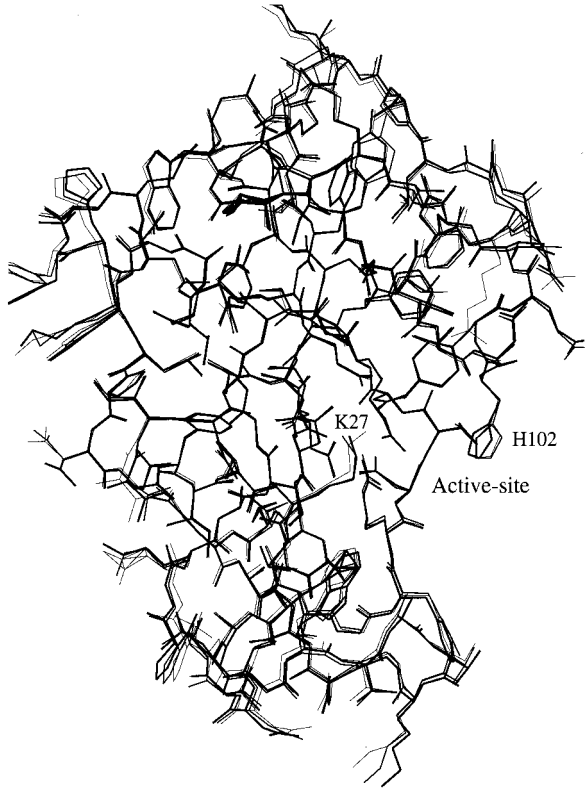


**Figure 8**
Comparison of the three independent barnase molecules after least-squares superposition.

**Table 4**
Solvent-structure statistics.

A hydrogen-bond cutoff distance of 3.5 Å was used with angular cutoff values of 120 and 90° (as minimum acceptable values) at the H and O atoms, respectively. Program *CONTACT* (Collaborative Computational Project, Number 4, 1994). $B_{av}$ is the averaged $B_{eq}$ value.

| | | $B_{av}$ (Å$^2$) |
|---|---|---|
| Number of water molecules | | |
| in the model | 415 | — |
| forming no hydrogen bonds | 0 | — |
| forming one hydrogen bond | 53 | 49 |
| forming two hydrogen bonds | 105 | 44 |
| forming three hydrogen bonds | 138 | 38 |
| forming four hydrogen bonds | 119 | 32 |
| forming no hydrogen bonds to protein | 61 | 50 |
| forming one hydrogen bond to protein | 175 | 42 |
| forming two hydrogen bonds to protein | 134 | 35 |
| forming three hydrogen bonds to protein | 42 | 28 |
| forming four hydrogen bonds to protein | 3 | 23 |

| Hydrogen-bond type | Number | $d_{av}$ (Å) | ⟨Angle⟩ | $B_{av}$ (Å$^2$) |
|---|---|---|---|---|
| Main-chain O···$H_2O$ | 208 | 2.94 | 130.6 | 36.2 |
| Main-chain N···$H_2O$ | 97 | 3.01 | 158.3 | 29.2 |
| Side-chain O···$H_2O$ | 192 | 2.90 | 119.6 | 35.3 |
| Side-chain N···$H_2O$ | 84 | 3.04 | 150.4 | 40.2 |

into the first hydration shell for those water molecules making at least one hydrogen-bond interaction with a protein atom and the second or higher hydration shells for those water molecules making hydrogen-bond interactions only with other water molecules. 354 water molecules were found in the first hydration shell and 61 were distributed in higher hydration shells. We note the strong preference of water molecules to bind to protein O atoms rather than to N atoms, using their protons rather than their lone pairs of electrons in hydrogen bonds. This accounts for the greater geometrical flexibility of hydrogen bonding to O atoms compared with that to N atoms (Frey, 1994). Bound water molecules are an integral part of the protein structure. Analysis of solvent structure in the three crystallographically independent barnase molecules showed a very similar number of water molecules associated with each chain (136, 142 and 137 for molecules *A*, *B* and *C*, respectively); 67 of these are 'equivalent' water molecules. The solvent molecules were qualified as structurally equivalent if they lay, after least-squares superposition of the molecules to which they are hydrogen-bonded, within 1.1 Å of each other (about half of the *ARP* merging distance). Among these, 16 (listed in Table 5 along with their hydrogen-bond interactions) have a bulk solvent-accessible area less than or equal to 10 Å$^2$. Calculations of solvent-accessible surface areas were performed using a probe radius of 1.4 Å (programs *AERAIMOL*, *RESAERA*; Collaborative Computational Project, Number 4, 1994). These 16 solvent molecules – with the exception of W13 – coincide within 0.45 Å when their positions in molecules *A*, *B* and *C* are compared. Two molecules (W7 and W20) are each located in an invagination at the surface of barnase. The others (Fig. 9), on the basis of their structural role, can be assigned to the following classes: (i) belonging to the catalytic site; (ii) stabilizing the recognition loop and (iii) mediating hydrogen bonds between loop 4 and either the turn connecting the $\alpha$3 helix to the $\beta$I strand or loop 2 and the N-terminal part of the $\alpha$3 helix.

**3.6.1. Class (i)**. The water molecules located in the catalytic site (Fig. 10) form well defined hydrogen bonds. W1 interacts both with the main chain of Asp75 and Tyr78 and with the $\beta$-strand 50–55. Two additional solvent molecules W3 and W9 fix the side-chain OD2 atom of Asp75. These three water molecules are located in deep invaginations at the surface of barnase and make no contact with any other water molecule. Furthermore, the Asp75 OD1 atom is bound to Glu73 and to Asp54 *via* two water molecules, W2 and W11. Each water molecule in this first class has a zero solvent-accessible area and is conserved within 0.15 Å in the three barnase molecules. Thus, the conformation of the catalytic site and, in particular, the appropriate mutual distances of the catalytically important residues are regulated not only by intraprotein interactions, such as the salt bridge formed between the Asp75 and Arg83 side chains, but also by solvent–protein interactions.

**3.6.2. Class (ii)**. The N-terminal part of loop 3 (residues 56–69) contains all guanine base liganding residues. This recognition site does not adopt a defined secondary structure, but its conformation is stabilized by main-chain to main-chain and side-chain to side-chain hydrogen bonds as well as by conserved solvent–protein interactions. W19 is bound to the main-chain N atom of residue Lys62 as well as W13 and to the Asn58 OD1 atom. W19 is extremely conserved in the three barnase molecules while W13 (or equivalent), with a surface accessibility of 10, 6 and 3 Å$^2$ in molecules *A*, *B* and *C*, respectively, is conserved within 0.80 Å among the three molecules.

**3.6.3. Class (iii)**. Many water molecules stabilize loop 4

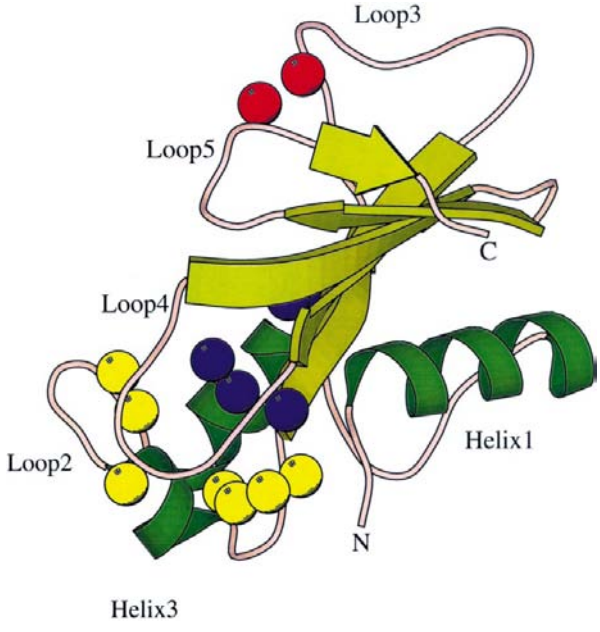


**Figure 9**
*MOLSCRIPT* (Kraulis, 1991) diagram of barnase with the buried water molecules grouped in three classes distinguished by colour: blue in the catalytic site, red in the recognition loop and yellow in loop 4.

**Table 5**
Equivalent buried water molecules in molecules *A*, *B* and *C* and their hydrogen-bond interactions.

Water molecules involved in these interactions are given in parentheses.

| | | Molecule *A* | | Molecule *B* | | Molecule *C* | |
|---|---|---|---|---|---|---|---|
| Atom | Residue | | Distance (Å) | | Distance (Å) | | Distance (Å) |
| N | Tyr78 | W1*A* | 2.94 | W1*B* | 2.97 | W1*C* | 2.99 |
| OG | Ser50 | | 2.84 | | 2.89 | | 2.80 |
| O | Asp75 | | 2.99 | | 2.99 | | 2.94 |
| OD1 | Asp75 | W2*A* | 2.89 | W2*B* | 2.90 | W2*C* | 2.85 |
| OE1 | Glu73 | | 2.90 | | 2.99 | | 2.93 |
| | (W11*A*) | | 2.90 | (W11*B*) | 2.85 | (W11*C*) | 2.91 |
| OD2 | Asp75 | W3*A* | 2.85 | W3*B* | 2.84 | W3*C* | 2.81 |
| N | Asn84 | | 2.92 | | 2.88 | | 2.99 |
| OD1 | Asn84 | | 2.81 | | 2.91 | | 2.86 |
| N | Gly81 | W4*A* | 2.95 | W4*B* | 2.96 | W4*C* | 2.89 |
| O | Gly40 | | 2.78 | | 2.82 | | 2.86 |
| N | Ala43 | | 3.11 | | 3.08 | | 3.10 |
| | (W85*A*) | | 2.80 | (W55*B*) | 2.71 | (W53*C*) | 2.77 |
| O | Gly81 | W5*A* | 2.88 | W5*B* | 2.92 | W5*C* | 2.87 |
| NE1 | Trp35 | | 3.11 | | 2.93 | | 3.22 |
| | (W6*A*) | | 2.70 | (W6*B*) | 2.80 | (W6*C*) | 2.82 |
| N | Leu42 | | 3.33 | | 3.44 | | 3.18 |
| O | Ala37 | W6*A* | 2.76 | W6*B* | 2.71 | W6*C* | 2.73 |
| | (W12*A*) | | 2.77 | (W18*B*) | 2.71 | (W45*C*) | 2.70 |
| | (W5*A*) | | 2.70 | (W5*B*) | 2.80 | (W5*C*) | 2.82 |
| N | Thr100 | W7*A* | 2.96 | W7*B* | 2.99 | W7*C* | 2.99 |
| O | Thr105 | | 3.01 | | 3.07 | | 3.09 |
| OG1 | Thr105 | | 3.13 | | 2.96 | | 2.99 |
| | (W57*A*) | | 2.83 | (W45*B*) | 2.85 | (W66*C*) | 2.96 |
| O | Gly48 | W8*A* | 2.89 | W8*B* | 2.92 | W8*C* | 2.78 |
| OG | Ser50 | | 2.82 | | 2.74 | | 2.78 |
| OD2 | Asp23 | | 3.16 | | 2.98 | | 3.12 |
| | (W23*A*) | | 2.71 | (W23*B*) | 2.82 | (W23*C*) | 2.73 |
| OD2 | Asp75 | W9*A* | 2.78 | W9*B* | 2.81 | W9*C* | 2.80 |
| N | Ile51 | | 2.82 | | 2.92 | | 2.91 |
| N | Asp54 | W11*A* | 3.14 | W11*B* | 3.06 | W11*C* | 3.10 |
| OD2 | Asp54 | | 2.84 | | 2.75 | | 2.95 |
| O | Glu73 | | 3.43 | | 3.49 | | 3.31 |
| | (W2*A*) | | 2.90 | (W2*B*) | 2.85 | (W2*C*) | 2.91 |
| OE1 | Glu60 | W13*A* | 2.68 | W13*B* | 2.43 | W13*C* | 2.98 |
| O | Tyr103 | | 2.93 | | 2.92 | | |
| | (W19*A*) | | 3.32 | (W19*B*) | 3.02 | (W19*C*) | 3.01 |
| | | | | (W36*B*) | 2.52 | (W36*C*) | 2.62 |
| N | Lys62 | W19*A* | 3.11 | W19*B* | 2.96 | W19*C* | 3.10 |
| OD1 | Asn58 | | 2.97 | | 2.88 | | 2.67 |
| | (W13*A*) | | 3.32 | (W13*B*) | 3.02 | (W13*C*) | 3.01 |
| | | | | (W83*B*) | 3.19 | (W40*C*) | 2.67 |
| O | Tyr90 | W20*A* | 2.88 | W20*B* | 2.99 | W20*C* | 2.94 |
| O | Thr70 | | 3.23 | | 3.34 | | 3.20 |
| | (W127*A*) | | 2.74 | (W138*B*) | 2.79 | (W76*C*) | 2.63 |
| | | | | (W70*B*) | 2.75 | (W97*C*) | 2.81 |
| O | Tyr78 | W23*A* | 2.77 | W23*B* | 2.64 | W23*C* | 2.80 |
| | (W8*A*) | | 2.71 | (W8*B*) | 2.82 | (W8*C*) | 2.73 |
| | (W41*A*) | | 3.05 | (W41*B*) | 3.20 | (W41*C*) | 2.85 |
| | (W78*A*) | | 2.65 | | | | |
| O | Ala46 | W33*A* | 2.91 | W33*B* | 2.86 | W33*C* | 2.80 |
| O | Lys49 | | 2.52 | | 2.70 | | 2.65 |
| | (W41*A*) | | 2.66 | (W41*B*) | 2.60 | (W41*C*) | 2.66 |
| O | Thr79 | W41*A* | 2.88 | W41*B* | 2.93 | W41*C* | 2.79 |
| O | Lys49 | | 3.25 | | 3.24 | | 3.23 |
| | (W23*A*) | | 3.06 | (W23*B*) | 3.20 | (W23*C*) | 2.85 |
| | (W33*A*) | | 2.66 | (W33*B*) | 2.60 | (W33*C*) | 2.66 |

(residues 77–85): one first subgroup constituted of W8, W23, W33 and W41 mediates hydrogen bonds between this loop and turn 1 connecting the $\alpha$3 helix to the $\beta$I strand. W4, W5 and W6 form the second subgroup and make hydrogen bonds between loop 4 and loop 2 and the N-terminal part of the $\alpha$3 helix (Fig. 9).

It is worth pointing out that these buried water molecules analyzed in the free barnase all have equivalents in all the complexed barnase structures so far studied, despite their different crystal forms and crystallization media. Thus, it should be emphasized that these water molecules are a constitutive part of barnase. They play a crucial role in its

**Table 6**
Least-squares superposition of molecules *A*, *B* and *C* of free barnase with molecules *A*, *B* and *C* of barnase linked to barstar and with molecules *L*, *M* of barnase linked to tetranucleotide d(CGAC).

Values given are r.m.s. (Å) for C$\alpha$ atoms and, in parentheses, for all non-H atoms. Program *LSQKAB* (Collaborative Computational Project, Number 4 1994).

| | Complexed with barstar | | | Complexed with d(CGAC) | |
|---|---|---|---|---|---|
| | *A* | *B* | *C* | *L* | *M* |
| Barnase *A* | 0.46 (0.70) | 0.48 (0.76) | 0.53 (0.72) | 0.45 (0.84) | 0.47 (0.83) |
| Barnase *B* | 0.49 (0.70) | 0.55 (0.76) | 0.53 (0.72) | 0.52 (0.86) | 0.54 (0.82) |
| Barnase *C* | 0.45 (0.72) | 0.51 (0.78) | 0.49 (0.75) | 0.43 (0.87) | 0.44 (0.85) |

structural and functional properties by stabilizing the loops and so immobilizing the catalytic site.

### 3.7. Comparison with bound barnase structures (linked to barstar or to nucleotides)

The high-resolution refinement reported here for the free enzyme and the availability of the refined structure of the enzyme complexed with barstar (at 2.0 Å resolution) and with the nucleotides 3′-GMP (at 2.2 Å resolution), d(GpC) (at 1.9 Å resolution) and d(CGAC) (at 1.76 Å resolution) allow us to compare details of the free and the complexed structures of the same protein. In these structures, with the exception of the barnase–d(GpC) complex, there is more than one molecule in the asymmetric unit. This allows more accurate conclusions to be drawn.

The free barnase structure and the complexed structures in the barnase–tetranucleotide complex (the nucleotide-complexed structure that has been determined to highest resolution) and in the barnase–barstar complex were compared after least-squares fitting of residues 4–110. The r.m.s. deviations are shown in Table 6. Barnase binding produces only localized changes in the barnase structure, which are higher on average in the barnase–d(CGAC) complex than in the barnase–barstar complex. Structural differences between free and complexed barnase structures and between the three independent molecules of the free barnase structure are found in the same parts of the polypeptide chain. They are confined to the flexible regions of the protein which have changed their crystal environment, such as the His18 side chain which moves by up to 2 Å, or are involved in ligand binding.

Active-site residues in free barnase (three molecules), in the barnase–barstar complex (three molecules) and in the barnase–d(GCAC) complex (two molecules) are shown in Fig. 11 after least-squares superposition onto molecule *C* of free barnase. Most of the side-chain conformations appear as clearly distinct for each of the three structures. Even when movements induced by ligand binding are limited, they are significant compared to coordinate errors. While residues of the catalytic site on the $\beta$-sheet (Glu73, Arg87) are well conserved, larger displacements are observed for those located on loops (Arg83, His102, Tyr103) and for Lys27. The most striking difference is at His102 which plays a major role in the activity of the enzyme and in barstar binding. An important change in the side-chain conformation enables it to fit in a preformed pocket on the barstar surface in the complex with the inhibitor and to form a hydrogen bond to the phosphate group $P_1$ in the complex with the tetranucleotide. The movement of His102 is correlated to that of Tyr103 *via* a shift in the main-chain and side-chain interactions and to that of Glu73 which is hydrogen bonded to the Tyr103 hydroxyl. Large differences are also found at the guanine-binding site (residues 57–60) in loop 3, despite the maintenance of the hydrogen-bond network between residues and between solvent molecules and protein. In the unligated barnase structure the Arg59 side chain is highly disordered and that of Glu60, stabilized by interactions with a neighbouring molecule, is located in weak electron density for molecules *A* and *B*. In the complexed barnase structures, both side chains interact with the ligands and adopt well defined conformations. It is worth noting that the recognition loop has high thermal motion both for the main-chain and the side-chain atoms in the free barnase structure as well as in the barnase–barstar complex. In the tetranucleotide-complexed barnase structure these residues are stabilized by binding the guanosine base and their $B_{eq}$ values decrease accordingly.

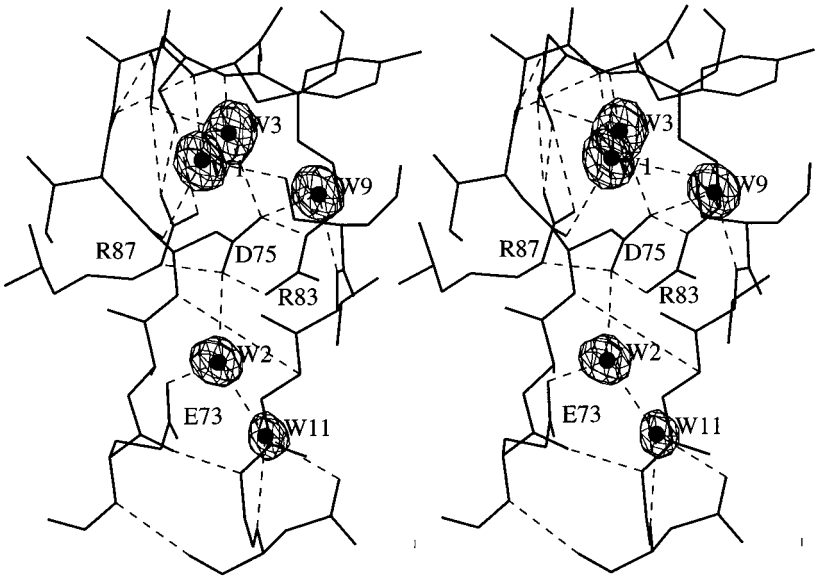


**Figure 10**
Stereoview of buried water molecules located in the catalytic site as seen in a $(F_o - F_c)$ map, contoured at $2\sigma$, with waters omitted from the structure-factor calculation. Dashed lines show hydrogen bonds.

### 3.8. Role of solvent in the barnase–barstar interaction

The important role played by water molecules in the barnase–barstar interaction has already been emphasized on the basis of comparison

with the free barnase structure refined at 2 Å (Guillet, Lapthorn, Hartley *et al.*, 1993; Buckle *et al.*, 1994). The analysis given by Buckle *et al.* has shown that at the barnase–barstar interface, 12 out of the 35 water molecules within 4.5 Å of both protein molecules have a solvent accessibility less than 10 Å$^2$ and 23 molecules with a solvent accessibility ranging from 11 to 56 Å$^2$ form a rim at the interface. It was found that five water molecules of the first group have equivalent molecules in free barnase structure, and none of the solvent molecules at the interface rim are conserved in the free barnase structure. The high resolution of this present model allows a more accurate analysis of these water molecules. Water molecules were considered as conserved if their positions in the free barnase structure and in the complex deviated by less than 1.8 Å and if their hydrogen-bonding pattern to protein atoms was retained (Fig. 12). Indeed, a side-chain conformational change between two structures induces a displacement of the bound water molecule. For the first 12 buried waters, eight water molecules are conserved in each free barnase molecule and another one is conserved in two molecules. Among the three non-conserved water molecules, one makes three barstar hydrogen bonds and should participate in the free barstar structure, another is bound to Arg59, which is highly disordered in free barnase, and the last one has a symmetric barnase molecule near to its position. Of the conserved totally buried water molecules at the barnase–barstar interface, two are already similarly buried in free barnase: those stabilizing the recognition loop. Thus, among the buried water molecules at the barnase–barstar interface, nine are seen in the structure of free barnase and at most three are present in the uncomplexed barstar structure, despite both interacting surfaces having a similar number of charged residues. The difference could be related to the concave shape of the barnase active site, which facilitates multiple hydrogen bonding of water molecules to protein atoms. This set of water molecules form an invariant structural component of the barnase active site which contributes significantly to the complementarity of the interacting surfaces.

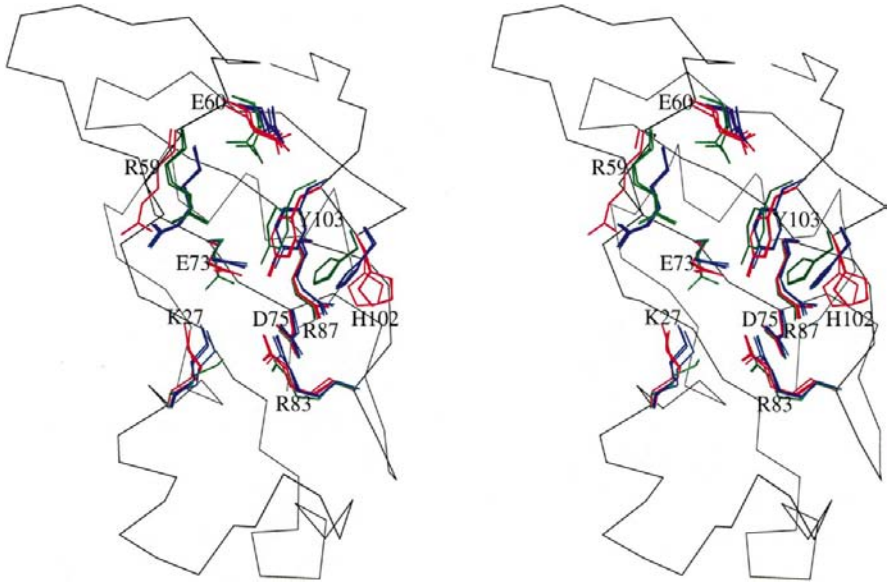


**Figure 11**
Stereoview of the active-site residues after least-squares superposition with molecule C of free barnase, shown as the C$\alpha$ trace, of the three molecules of free barnase (red), of the three molecules of barnase–barstar complex (blue) and of the two molecules of the barnase–d(GCAC) complex (green).

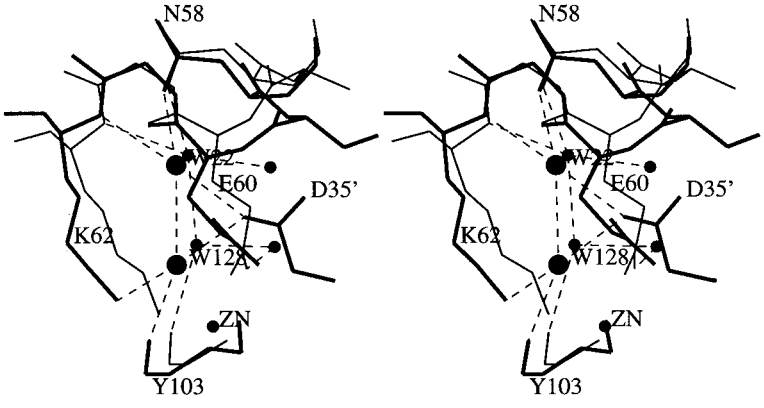


**Figure 12**
Stereoscopic view of conserved solvent molecules, filling the barnase–barstar interface and forming a conserved hydrogen-bonding network: bold lines show the barnase–barstar molecules with buried solvent molecules (large spheres) and thin lines show the free barnase molecule *C* after least-squares superposition with associated solvent molecules (smaller spheres). Barstar residues are labelled as D35′.

Although the 23 water molecules which form the rim are less well ordered and none of them was considered as conserved in the 2 Å barnase structure, the present analysis shows that they are not all recruited from bulk solvent. They can be grouped according to the way they are bound in the interface: 15 are bound to barnase by at least one hydrogen bond, while the others are bound to barstar or to another water molecule. Nine of these 15 water molecules are conserved in at least one free barnase molecule; the others have not been located, partly as a consequence of crystal contacts. This shows that a significant proportion of the solvent molecules around the interface can also be considered as conserved or partially conserved; they contribute towards reducing the solvent entropic cost from the recruitment of bulk-solvent molecules for the hydration of the complex. Thus, the formation of an extensive network of ordered solvent molecules linking barnase and barstar through a large number of solvent-mediated hydrogen bonds

suggests that, in addition to protein–protein interactions, hydration enthalpy is an important factor in the complex stabilization, while the hydrophobic contribution to binding is restricted by the retention of water molecules bound to the barnase binding site.

We are grateful to J. Fourniat and E. Vallet for help and support during this work. We thank M. Knossow and J. Navaza for their constant interest and critical reading of the manuscript. We acknowledge the staff of LURE (Orsay) for making station W32 on the wiggler line of LURE-DCI available to us. Part of the calculations were performed on the CRAY computers at IDRIS, Orsay, France.

## References

Baudet, S. & Janin, J. (1991). *J. Mol. Biol.* **219**, 123–132.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Buckle, A., Cramer, P. & Fersht, A. (1996). *Biochemistry*, **35**, 4298–4305.
Buckle, A. & Fersht, A. (1994). *Biochemistry*, **33**, 1644–1653.
Buckle, A., Henrick, K. & Fersht, A. (1993). *J. Mol. Biol.* **234**, 847–860.
Buckle, A., Schreiber, G. & Fersht, A. (1994). *Biochemistry*, **33**, 8878–8889.
Burnett, N. M. & Johnson, K. C. (1996). *ORTEP III.* Report ORNL-6895. Oak Ridge National Laboratory, Tennesse, USA.
Bycroft, M., Ludvigsen, S., Fersht, A., Poulsen, F. (1991). *Biochemistry*, **30**, 8697–8701.
Chen, Y., Fersht, A. & Henrick, K. (1993). *J. Mol. Biol.* **234**, 1158–1170.
Chen, Y., Fersht, A. & Henrick, K. (1995). *Acta Cryst.* D**51**, 220–231.
Clarke, J., Henrick, K. & Fersht, A. (1995). *J. Mol. Biol.* **253**, 493–513.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Day, A. G., Parsonage, D., Ebel, S., Brown, T. & Fersht, A. (1992). *Biochemistry*, **31**, 6390–6395.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Fersht, A. R. (1993). *FEBS Lett.* **325**, 5–16.
Frey, M. (1994). *Acta Cryst.* D**50**, 663–666.
Frisch, C., Schreiber, G., Johnson, C. M. & Fersht, A. R. (1997). *J. Mol. Biol.* **267**, 696–706.
Guillet, V., Lapthorn, A., Hartley, R. W. & Mauguen, Y. (1993). *Structure*, **3**, 165–176.
Guillet, V., Lapthorn, A. & Mauguen, Y. (1993). *FEBS Lett.* **330**, 137–140.
Hartley, R. W. (1980). *J. Mol. Evol.* **15**, 355–358.
Hartley, R. W. (1993). *Biochemistry*, **32**, 5978–5984.
Hartley, R. W. & Rogerson, D. L. (1972). *Prep. Biochem.* **2**(3), 229–242.
Hill, C., Dodson, G., Heinemann, Y., Saenger, W., Mitsui, Y., Nakamura, K., Borisov, S., Tischenko, G., Polykov, K. & Pavlovsky, S. (1983). *Trends Biochem. Sci.* **8**, 364–369.
Hill, C. P. (1986). PhD thesis, University of York, England.
Jack, A. (1973). PhD thesis, University of Cambridge, England.
Jones, T. A. & Kjeldgaard, M. (1993). *O*, Version 5.9. Uppsala University, Sweden.
Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Lees, C. W. & Hartley, R. W. (1966). *Biochemistry*, **5**, 3951–3960.
Leslie, A. G. W. (1992). *CCP4 ESF–EACMB Newslett. Protein Crystallogr.* **26**, 27–33.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Mauguen, Y., Hartley, R. W., Dodson, E., Dodson, G., Bricogne, G., Chothia, C. & Jack, A. (1982). *Nature (London)*, **297**, 162–154.
Meiering, E., Bycroft, M. & Fersht, A. R. (1991). *Biochemistry*, **30**, 11348–11356.
Meiering, E., Bycroft, M., Lubienski M. J. & Fersht, A. R. (1993). *Biochemistry*, **32**, 10975–10987.
Nishimura, S. (1960). *Biochem. Biophys. Acta*, **45**, 15–27.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Roussel, A. & Cambillau, C. (1991). *TURBO-FRODO*, Silicon Graphics Geometry Partner Directory, Mountain View, CA, USA.
Rushizky, G. W., Greco, A. E., Hartley, R. W. & Sober, H. A. (1963). *Biochemistry*, **2**(4), 787–793.
Schreiber, G. & Fersht, A. (1993). *Biochemistry*, **32**, 5145–5150.
Serrano, L., Kellis, J. T., Cann, P., Matousheck, A. & Fersht, A. R. (1992). *J. Mol. Biol.* **224**, 783–804.
Sevcik, J., Sanishvili, R. G., Pavlosky, A. G. & Polyakov, K. M. (1990). *Trends Biochem. Sci.* **15**, 158–162.
Sheldrick, G. M. (1993). *SHELXL*93. University of Göttingen, Germany.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1986). *Biochemistry*, **25**, 5018–5027.
Takahashi, K. & Moore, S. (1982). *The Enzymes*, Vol. 15, 3rd ed., edited by P. D. Boyer, pp. 435–468. New York: Academic Press.
Trueblood, K. N., Burgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Cryst.* A**52**, 770–781.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
Zelinkova, E., Bacova, M. & Zelinka, J. (1971). *Biochim. Biophys. Acta*, **235**, 342–344.