Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution structure of a plasmid-encoded dihydrofolate reductase: pentagonal network of water molecules in the D_2-symmetric active site

Narendra Narayana

Department of Biochemistry, Case Western Reserve University, Cleveland, OH 44106, USA

Correspondence e-mail: nxn17@case.edu

Received 19 March 2006
Accepted 24 April 2006

PDB Reference: dihydrofolate reductase, 2gqv, r2gqvsf.

R67 plasmid-encoded dihydrofolate reductase (R67 DHFR) is an NADPH-dependent homotetrameric enzyme that catalyzes the reduction of dihydrofolate to tetrahydrofolate. The amino-acid sequence and molecular architecture of R67 DHFR and its inhibitory properties toward folate analogues are different from those of chromosomal DHFR. Here, the crystal structure of R67 DHFR refined using 1.1 Å resolution data is presented. Blocked full-matrix least-squares refinement without restraints resulted in a final R factor of 11.4%. The anisotropic atomic displacement parameters analyzed by Rosenfield matrices and translation–libration–screw validation suggested four quasi-rigid domains. A total of ten $C^{\alpha}-H\cdots O$ hydrogen bonds were identified between the β-strands. There is reasonable structural evidence that His62 is not protonated in the tetramer, which is in accord with previous pH-profile studies. The side chain of Gln67 that protrudes into the active site exhibits dual conformation, a feature noticed for the first time owing to the availability of atomic resolution data. The R67 DHFR active site is unique: it has D_2 symmetry and is a large active site with a pentagonal network of water molecules and exposure of backbone atoms to solvent; the central pore is favorable for planar ring-stacking interactions. The geometrical shape, overall symmetry, local asymmetry and waters appear to dominate the binding of ligands, catalysis and inhibition.

1. Introduction

Dihydrofolate reductase (DHFR; EC 1.5.1.3) plays an important role in folate metabolism. It catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor (Greenberg, 1954; Kraut & Matthews, 1987). Chromosomal DHFR is proposed to be a highly evolved enzyme, with a catalytic efficiency of 0.15 compared with 0.6 for triose phosphate isomerase, which is considered to be a perfect enzyme (Fierke *et al.*, 1987; Albery & Knowles, 1976). DHFR is a monomeric protein (MW $\simeq$ 18 kDa; 159 amino-acid residues in *Escherichia coli*) that folds into an eight-stranded β-sheet core with four α-helices packing against the central β-sheet (Matthews *et al.*, 1977). DHFR is essential for folate metabolism in both eukaryotic and prokaryotic cells, as the generation of THF, a carrier of one-carbon units, is required for the synthesis of thymidylate, purine nucleosides, methionine and other metabolic inter-

mediates. These metabolites are required for DNA, RNA and protein synthesis; blockage of DHFR-catalyzed reduction of DHF to THF thus results in cessation of DNA synthesis and consequent cell death. Furthermore, DHFR exhibits high-level species selectivity for inhibition by synthetic folate analogues (antifolates) such as methotrexate (MTX), trimethoprim (TMP) and pyrimethamine (Kuyper *et al.*, 1996; Roth & Cheng, 1982). Therefore, DHFR is of considerable interest to both enzymologists and drug designers (Morrison, 1991; Jones & Matthews, 1995; Falzone *et al.*, 1994).

Recent clinical studies have indicated that certain strains of *Escherichia coli* and *Klebsiella aerogenes* carry R factors that confer high levels of resistance to TMP (Fleming *et al.*, 1972). Subsequent work has shown that these R factors specify DHFRs (Pattishall *et al.*, 1977) and they were initially classified into two groups depending on their insensitivity and complete resistance to TMP. The most recent classification of plasmid-encoded DHFR genes indicates two families (dfrA and dfrB) with 17 different types, I–XVII (White & Rawlinson, 2001). In general, these resistant enzymes have amino-acid sequences that are closely related to those of bacterial chromosomal DHFRs. The type II enzyme is 1000-fold less sensitive to TMP compared with other R plasmid-encoded DHFRs and is weakly inhibited by methotrexate (MTX; $K_i = 0.9 \times 10^{-4}\,M$), a stoichiometric inhibitor of chromosomal DHFRs (Pattishall *et al.*, 1977; Zolg & Hanggi, 1981; Hitchings & Smith, 1980). R67 plasmid-encoded DHFR (R67 DHFR), a type II enzyme, is of particular interest as it is unrelated genetically and structurally to chromosomal DHFR. At physiological pH the functional R67 DHFR is a homotetramer (MW $\simeq$ 8.5 kDa; 78 amino-acid residues per subunit; Stone & Smith, 1979; Smith *et al.*, 1979) and at low pH native tetrameric R67 DHFR dissociates to form inactive dimers (Nichols *et al.*, 1993). Equilibrium folding studies of R67 DHFR indicate the absence of a folded monomer (Reece *et al.*, 1991; Zhuang *et al.*, 1994). The crystal structure of native R67 DHFR displays four identical subunits related by 222 symmetry with a central active-site pore (Narayana *et al.*, 1995; Figs. 1*c* and 1*d*). The kinetic parameters for R67 DHFR are as follows: the K_m values for DHF and NADPH are 5.8 and 3.0 μM, respectively, and K_{cat} (pH 7) for hydride transfer is 1.3 s^{-1}. For comparison, the corresponding kinetic parameters for chromosomal DHFR are 1.2 and 0.94 μM and 240 s^{-1}, respectively. The K_i (TMP) for R67 DHFR is 0.15 mM, compared with 20 pM for chromosomal DHFR. Although R67 DHFR catalyzes the same reaction as does chromosomal DHFR, it is quite distinct from the chromosomal, bacterial and mammalian DHFRs in amino-acid sequence, structure, catalytic and inhibitory mechanisms.

In the original report on the crystal structure of dimeric R67 DHFR (Matthews *et al.*, 1986; Fig. 1*b*; green/yellow or pink/cyan subunits in Fig. 1*c*), weak diffuse electron density was observed for the N-terminal 17 amino-acid residues. In an effort to produce a form of R67 DHFR that could be crystallized as the active tetramer, chymotrypsin was used to selectively cleave the full-length protein at Phe16. This yielded a shortened protein with 62 amino acids that is fully active as a tetramer (Reece *et al.*, 1991). Crystal structures of this truncated form of R67 DHFR were determined in the active tetrameric form as an apoenzyme (1.7 Å resolution) and as a binary complex with folate (1.8 Å resolution) (Narayana *et al.*, 1995). The tetrameric R67 DHFR is a dimer of dimers with an unusual pore 24 Å in length, 18 Å wide and 12 Å thick that passes through the middle of the molecule (Figs. 1*c* and 1*d*). NADPH and folate both bind to the central pore and all four subunits contribute to the active site. While a shared active site between protomers is not uncommon, only one active site per tetramer is unusual.

Atomic resolution structures that are refined at or better than 1.0 Å resolution are rare. Of a total of 33 931 coordinate sets deposited in the PDB, only 186 protein structures (including several variants of the same protein; Berman *et al.*, 2000) are better than or equal to 1 Å resolution. At this resolution, features such as hydrogen positions, anisotropic motions of atoms, discrete disorder and quasi-rigid bodies can be modeled (Sheldrick, 1990; Schomaker & Trueblood, 1968). In addition, atomic resolution structures yield valuable information about solvent organization and various geometrical properties of proteins and serve to improve the database of stereochemical target values for use in validating structures determined at lower resolution (MacArthur & Thornton, 1999). Here, we report the structure of truncated native R67 DHFR (residues 17–78) at 1.1 Å resolution determined from diffraction data collected under cryocooling conditions (100 K).

2. Experimental

2.1. Crystallization, data collection and processing

Recombinant R67 DHFR was purified as previously described (Reece *et al.*, 1991). The truncated enzyme (residues 17–78) was crystallized from hanging drops according to the protocol described previously (Narayana *et al.*, 1995). Prior to X-ray data collection, crystals were flash-cooled using 50% MPD in 50 mM Tris–HCl buffer as cryoprotectant. Diffraction data to 1.1 Å resolution were collected from a single flash-cooled R67 DHFR crystal using an ADSC Quantum 1 CCD detector at a synchrotron-radiation source. The R67 DHFR crystal belongs to space group $I4_122$, with unit-cell parameters $a = b = 68.05$, $c = 52.24$ Å, and has one molecule in the asymmetric unit ($V_M = 2.25$ Å^3 Da^{-1} and 45% solvent content). An overall B factor of 9.7 Å^2 was estimated from the Wilson plot. The data were processed using *DENZO* (Otwinowski & Minor, 1997). A summary of the data-collection statistics is given in Table 1.

2.2. Structure refinement

Initial crystallographic refinement was carried out against 95% of the measured data using *CNS*. The remaining 5% of the observed data, which was randomly selected, was used in R_{free} calculations to monitor the progress of refinement. The starting model for the refinement, containing only the protein atoms, was obtained from the structure of R67 DHFR deter-

Table 1
Crystal and X-ray data-collection details.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	0.9495
Resolution (Å)	50–1.1 (1.14–1.10)
Space group	$I4_122$
Unit-cell parameters (Å)	$a = b = 68.05$, $c = 52.24$
Detector type	ADSC Quantum 1 CCD
No. of measured reflections	190479
Unique reflections	22453
R_{sym}†	0.042 (0.34)
Completeness (%)	90 (58)
Mean $I/\sigma(I)$	21.4 (3.7)
Matthews coefficient (Å^3 Da^{-1})	2.25
Solvent content (%)	45
Wilson B factor‡ (Å^2)	9.7
Redundancy of data set	8.5
Mosaicity (°)	0.8

† $R_{sym} = \sum_h \sum_i [|I_i(h) - \langle I(h)\rangle|/\sum_h \sum_i I_i(h)]$, where I_i is the ith measurement and $\langle I(h)\rangle$ is the weighted mean of all measurements of I(h). ‡ The Wilson B factor was obtained using *TRUNCATE* (Collaborative Computational Project, Number 4, 1994).

mined at 1.7 Å resolution (PDB code 1vie). An initial 40 cycles of rigid-body refinement were followed by consecutive cycles of simulated annealing and isotropic B-factor refinement interspersed with manual model building with *O* (Jones, 2004). The refinement in *CNS* was started at 1.7 Å resolution and gradually increased in subsequent rounds of refinement to a final resolution of 1.2 Å, resulting in an R factor and R_{free} of 26.8 and 27.6%, respectively. At this stage, water molecules were incorporated into the structure in several rounds of successive refinement. The water molecules were included if they had peaks greater than 3σ in $F_o - F_c$ electron-density maps and were within hydrogen-bonding distance of the polar protein atoms. Higher hydration shells were filled in the later rounds of refinement if the water peaks were at the 3σ level in $F_o - F_c$ maps and were involved in hydrogen-bonding interactions with the neighboring water molecules. The R factor, R_{free} and the mean figure of merit (FOM) converged at 19.7%, 21.1% and 0.91, respectively, for protein atoms (residues 20–78) and 80 water molecules. This model was further subjected to refinement with *TNT*. Two rounds of least-squares refinement with 30 cycles of conjugate-gradient least-squares (CGLS) refinement in *TNT* yielded an R factor of 17.9% for data between 20 and 1.1 Å resolution.

This refined model was used to perform CGLS refinement using *SHELXL*97, where X-ray intensities were used in refinement calculations. The bond distances and angles (1,2 and 1,3 distances) were constrained using standard Engh and Huber restraints. Each round of refinement with *SHELXL* consisted of ten cycles of restrained CGLS refinement followed by manual model adjustments using $2F_o - F_c$ and $F_o - F_c$ electron-density maps. Water peaks predicted by the *SHELXL* program were included if they were positioned at a hydrogen-bonding distance from other water molecules or protein atoms. Initial rounds of restrained CGLS refinement were carried out by keeping all atomic displacement parameters (ADPs) isotropic. Subsequently, the ADPs were converted to anisotropic values, leading to improved Fourier maps. At this stage, alternative conformations visible in the

Table 2
Structure-refinement statistics and geometrical parameters of the final model.

(*a*) Refinement statistics.

Resolution range (Å)	10.0–1.1
No. of protein residues	59
R factor/R_{free} for all data† (%)	11.1/17.2
R factor/R_{free} for $F > 4\sigma$ (%)	9.9/15.3
wR^2‡ (%)	25.7
Final R factor	11.4
No. of non-H atoms	
Protein§	455 (450)
Water	191
No. of observations/parameter¶	3.9
Goodness of fit (GooF)	2.3

(*b*) Deviations from ideal geometry. Values as reported by *SHELXL*.

	R.m.s.d.	Target σ
Bond distances (Å)	0.023	0.030
Angle distances (Å)	0.035	0.060
Chiral volumes (Å^3)	0.117	0.200

(*c*) Structure quality. Values obtained by *PROCHECK*.

Core Ramachandran (%)	87.2
Hydrogen bonds sd (Å)	0.6
χ_1 pooled sd (°)	9.8
χ_2 sd (°)	10.3
G factor (dihedral angles)	−0.21

(*d*) Anisotropic temperature-factor (U_{ij}) restraints. For the definition of DELU, SIMU and ISOR restraints, see §2.

	R.m.s.d.	Target σ
DELU (Å^2)	0.012	0.020
SIMU (Å^2)	0.034	0.050
ISOR (Å^2)	0.103	0.100

(*e*) Mean isotropic equivalent B factors (Å^2).

Overall structure	24.3
All protein atoms	14.9
Main-chain atoms	12.9
Side-chain atoms	17.1
Water molecules	46.2

† $R_{cryst} = 100 \times \sum_h |F_o - F_c|/\sum_h F_o$, where F_o and F_c are the observed and calculated structure-factor amplitudes of reflection h. R_{free} is the R factor corresponding to a randomly selected 5% subset of reflections not used in refinement. The final R factor was computed using all data, including reflections that were omitted for R_{free} calculations. The final R factor corresponds to the final round of unrestrained blocked least-squares refinement. ‡ $wR^2 = \{\sum[w(F_o^2 - F_c^2)^2]/\sum[w(F_o^2)^2]\}^{1/2}$. § The number of unique atoms excluding alternative conformers is given in parentheses. ¶ Assuming anisotropic atomic B factors.

electron-density maps were modeled and refined isotropically before being refined anisotropically. The site-occupation factors (SOFs) of alternative conformations were refined such that their sum was unity. The occupancy of water molecules was fixed at 1.0 or 0.5 depending on the electron-density peak height. The bulk solvent was modeled based on Babinet's principle, as implemented in the SWAT option in the *SHELXL* program. At the beginning of *SHELXL* refinement, default values were used for all distances, planarity, chiral and antibumping restraints (standard deviations were 0.02 Å for 1,2 distances, 0.04 Å for 1,3 distances, 0.2 Å^3 for planes and

chiral volumes and 0.02 Å for anti-bumping restraints). These restraints were relaxed by increasing the standard deviations of the 1,2 distances, 1,3 distances and antibumping restraints to 0.03, 0.06 and 0.04 Å, respectively, in the later stages of the refinement. The initial restraint applied to the planarity for the peptide bond (0.5 Å^3) was removed during final cycles of refinement. Both the parameters DELU and SIMU that define the restraints for the U_{ij} components of the bonded atoms and the restraints for the corresponding U_{ij} components of spatially adjacent atoms, respectively, were weakly restrained with a standard deviation of 0.01 Å^2. DELU and SIMU restraints on the ADPs were softened to final standard deviations of 0.02 and 0.05 Å^2, respectively, in the concluding cycles of refinement. The ISOR restraint of 0.1 Å^2 was applied to the ADPs of solvent atoms to approximate an isotropic behavior. The *SHELXL* instruction WGHT was set at 0.1 throughout the refinement.

The final round of refinement was carried out with the inclusion of riding H atoms for protein residues. The positions of H atoms assigned based on the known geometrical criteria were not refined. Structure refinement was concluded by performing blocked full-matrix refinement with and without restraints. The protein atoms and water molecules were grouped into blocks with overlapping residues. The final R factor for the structure using all data between 10 and 1.1 Å resolution with no stereochemical restraints was 11.4%. Structure-refinement details are given in Table 2.

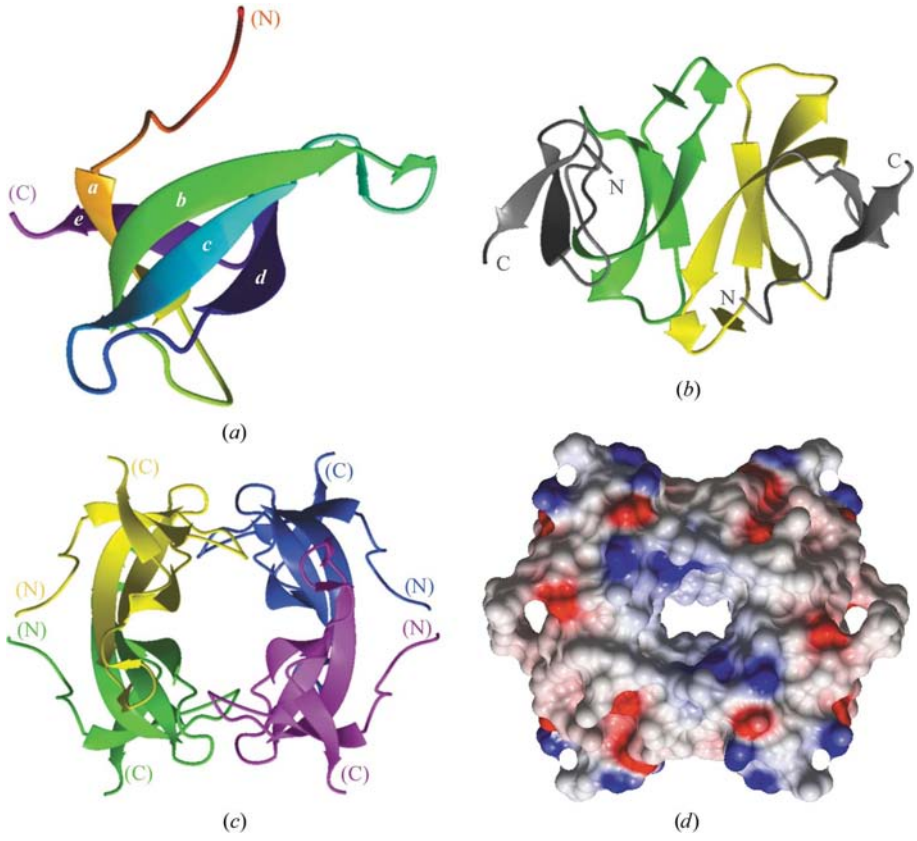

Figure 1
Ribbon representation of R67 DHFR. (*a*) An R67 DHFR monomer (extracted from the tetramer) forms a compact β-barrel comprised of β-strands *a*–*e*. It may be noted that monomeric R67 DHFR is not detected in solution or in crystals. (*b*) Dimeric R67 DHFR with another β-barrel at the interface formed by β-strands *b*–*d* (green) and symmetry-related strands (yellow). The remaining strands in both subunits are shown in gray. This form occurs at low pH. (*c*) The active tetrameric form of R67 DHFR is a dimer of dimers at physiological pH. Subunits *A*–*D* are colored green, yellow, cyan and pink. (*d*) Electrostatic surface of the R67 DHFR tetramer. The negative, neutral and positive surfaces are displayed in red, white and blue colors, respectively. The active-site pore bears a positive character owing to Lys32 and Lys33, which are positioned at the outer edge of the pore. This figure was produced using *CCP4MG* (Potterton *et al.*, 2004).

3. Results

3.1. Atomic resolution structure of R67 DHFR

The structure was successively refined using *CNS* (Brünger *et al.*, 1998), *TNT* (Tronrud *et al.*, 1987) and *SHELXL*97 (Sheldrick & Schneider, 1997). 30 omit maps were computed by omitting three residues at a time and the model was inspected for small adjustments or alternative conformations. In the final round of CGLS refinements, an $F_o - F_c$ omit map revealed an alternative side-chain conformation for Gln67 (Fig. 2*c*). The penultimate cycle of refinement was performed using the blocked least-squares full-matrix option in order to obtain reliable estimates of standard deviations as reported in Table 2. The protein molecule plus the water molecules were grouped into eight blocks. The positional parameters as well as ADPs for atoms were simultaneously refined in the same block. Protein residues 22–78 were refined anisotropically, whereas residues 20 and 21 were refined isotropically at half occupancy because they exhibited high mobility, as indicated by diffuse density and high thermal factors (mean B is 44.6 Å^2). Most water O atoms that were added in the later stages of refinement were assigned half occupancy owing to their relatively low peak strength. A difference Fourier map calculated in the final stages of refinement revealed several H atoms (Fig. 3). In the concluding steps of refinement, riding H atoms were added to the protein molecule but the H atoms were not refined. The final round of blocked full-matrix least-squares refinement without restraints resulted in a final R factor of 11.4% for all data between 10 and 1.1 Å resolution. The final model consists of protein residues 20–78 and water molecules 79–269. There are 148 water molecules with unit occupancy and 43 water molecules with half occupancy. The largest peak and hole in the final difference Fourier synthesis

are 0.35 and −0.28 e Å^{-3}, respectively. A summary of the refinement statistics is given in Table 2.

3.2. Quality of the structure

The refined model of R67 DHFR includes amino-acid residues 20–78, the side chain of residue Gln67 in dual conformations with SOFs of 0.46 and 0.54 and 191 water molecules with either unit or half occupancy. The amino-acid residues are designated by standard three-letter or one-letter codes. The backbone of the protein molecule is well ordered. There were no non-glycine residues in the disallowed region of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965). The quality of the electron-density maps was uniformly excellent as shown in Fig. 2. There was no discernable electron density for residues 17–19, implying flexibility, and density was weak for residues 20 and 21. The N-terminal portion (17– 21) was also found to be disordered in the original 1.7 Å resolution structure of the apoenzyme as well as in the binary complex of R67 DHFR bound to folate (Narayana *et al.*, 1995; data sets were at 277 K). Thus, the conformation of the stretch of residues 17–21 appears to be independent of the temperature at which the data were collected, suggesting intrinsic flexibility. *PROCHECK* (Laskowski *et al.*, 1993) was used for geometric analysis of the final model. *CCP*4 (Collaborative Computational Project, Number 4, 1994) was used to analyze the solvent network and hydrogen-bonding interactions and to perform anisotropic and translation–libration–screw (TLS) analyses. The program *PARVATI* (Merritt, 1999) was used to analyze *B*-factor anisotropy in the structure. Least-squares superposition of equivalent C$^{\alpha}$ atoms (20–78) in the present structure determined at 100 K and the previously determined apoenzyme structure at 277 K yields a root-mean-square deviation (r.m.s.d.) of 0.57 Å. In this overlay of molecules, a few side chains (Ser20, Asn21, Arg29, Arg31, Lys32, Gln41, Trp45, Gln67, Leu74, Glu75 and Asn78) display significantly different conformations (Fig. 4). Additionally, the backbone atoms superpose well except in the loop region (33–37), where there is a shift of about 0.4 Å (Fig. 4).

3.3. Overall enzyme structure

Four identical subunits of R67 DHFR assemble with a dihedral symmetry axes passing through the tetramer. Each subunit is a compact and elongated β-barrel comprised of five antiparallel β-strands, labelled *a*, *b*, *c*, *d* and *e* (residues 28–32, 38–47, 54–59, 65–70 and 74–76, respectively; Fig. 1*a*). A total of 191 solvent molecules are found associated with each subunit (Fig. 5*a*). The water-filled central hour-glass-shaped pore has a total of 168 water molecules inclusive of symmetry-related water molecules (Fig. 5*b*). Each subunit is exposed to the front and back portions of the pore. The loop region between strands *a* and *b* (30–40), the loop between strands *b* and *c* (44–55) and the region comprised of β-strand *d* (63–74) and their symmetry-related residues form the surface of the pore (Fig. 5*b*). The side chain of residue Gln67 and its symmetry equivalent are positioned nearest to the intersection of the dyad axes and form the upper and lower surfaces of the central pore. The overall average isotropic equivalent *B* factor is 24.3 Å^2. The mean anisotropy of protein atoms is 0.59 and that for solvent atoms is 0.49. About 30% of the protein atoms and 50% of the solvent molecules have an anisotropy value of 0.5 or less. The electrostatic surface of the tetramer shows a positive potential at the entrance (or exit) and in the interior of the pore, owing to residues Lys32 and Lys33. Patches of negative and positive potential are seen at the periphery of the molecule (Fig. 1*d*). The overall structure is similar to the previous 1.7 Å resolution structure of R67 DHFR (PDB code 1vie). The r.m.s.d.s of superposition of equivalent C$^{\alpha}$ atoms, main-chain atoms and all non-H protein atoms of the two structures are 0.57, 0.66 and 1.11 Å, respectively.

3.4. Disordered regions

In the present structure, a single residue (Gln67) out of a total of 62 amino-acid residues exhibits alternative side-chain conformations. This side chain of Gln67 protrudes into the middle of the active-site pore. The omit map corresponding to Gln67 showed electron density for the main-chain atoms (C$^{\alpha}$, C, N and O) and the side-chain atom C$^{\beta}$ at about the 9σ level, whereas the disordered portion C$^{\gamma}$, C$^{\delta}$, O$^{\varepsilon1}$ and N$^{\varepsilon2}$ were seen at about the 3σ level. The dual conformational states of the side chain of Gln67 are designated as *A* and *B* (Fig. 2*c*). The N-terminal amino-acid residues 17–19 were invisible in the electron-density maps, implying disorder. Residues 20–23 exhibit diffuse and weak electron density with high mean *B* values of 44.7, 39.8, 28.1 and 19.0 Å^2, respectively. The side chains of Arg31 and Trp45 display high mobility, as indicated by high mean *B* values of 27.9 and 30.3 Å^2, respectively. The C-terminal residues Ile77 and Asn78 have large thermal motions, with mean *B* values of 18.6 and 39.3 Å^2, respectively. Comparison of this structure with the previously determined structure at 277 K indicates differences at a few residues (polar or charged) distributed over the surface of molecule, suggesting the possibility of a certain degree of flexibility at those locations (Fig. 4).

3.5. Observation of H atoms

A $F_o - F_c$ map was computed in the final round of refinement to locate peaks representing H atoms. Several H atoms of the protein molecule were visible in the electron-density map (Fig. 3). Most of the observed isolated H-atom peaks were at the 3σ level and some of them at around the 2.5σ level. A total of 115 H atoms, which is 26% of the total H atoms included in the refinement with riding H atoms option, could be visualized at the 2.5σ level and above in the final difference Fourier map. Almost all H atoms attached to the main-chain C$^{\alpha}$ and N atoms were visible. This observation is in agreement with the fact that peaks corresponding to H atoms are better seen in the ordered and rigid regions of the structure.

3.6. Thermal vibrations

The mean B_{eq} values for overall structure, protein atoms, main-chain atoms, side-chain atoms and water molecules are 24.3, 14.9, 12.9, 17.1 and 46.2 Å^2, respectively. The mean

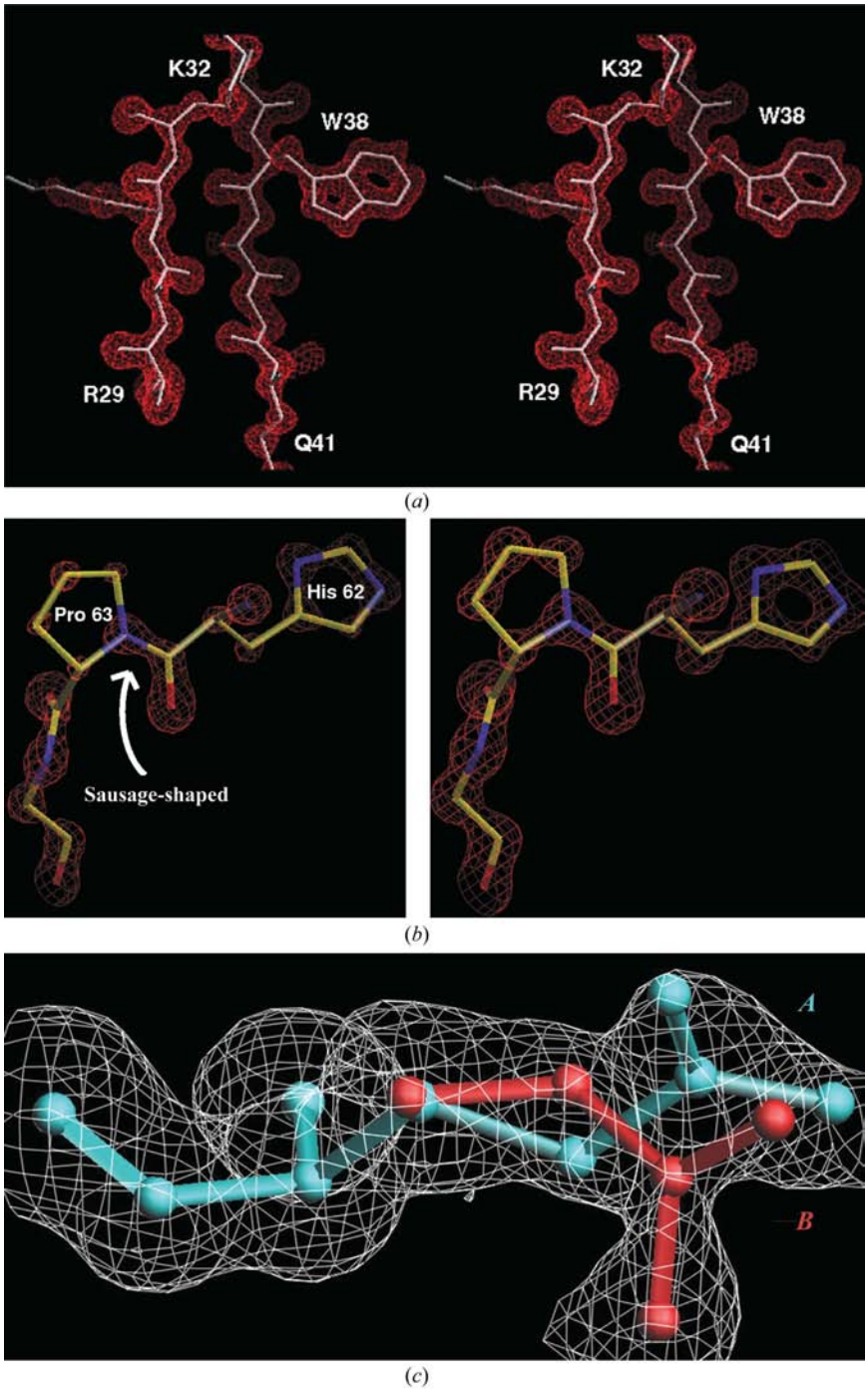

Figure 2
Omit electron-density maps. (*a*) Stereoview of portions of β-strands *a* (29–32) and *b* (37–41) superimposed on the difference map computed at the final stages of refinement by omitting these residues. The map is contoured at the 3.5σ level. (*b*) A short stretch of an omit map for residues 62–64. The left panel shows electron density contoured at the 10σ level. Blobs of electron density correspond to the atomic positions, a feature seen in high-resolution refinement. A sausage-shaped density at the same contour level is seen at the peptide linkage indicated by an arrow. This feature was seen at almost all of the peptide bonds in the molecule. The right panel shows the same segment contoured at a slightly lower level (7σ). Here, the boundaries of the residues are seen with a conspicuous hole in the middle of the pentagonal rings (His62 and Pro63), a feature normally observed in atomic resolution structures. (*c*) The side chain of residue Gln67 exhibits dual conformation as shown. The Gln67 residue is superimposed on the omit map (contour level of 3σ) calculated without its coordinates. The backbone atoms are seen at the 9σ level, whereas the side chain is only seen at about the 3σ level. The two conformations for the tip of the side chain (C^{γ}, C^{δ}, $O^{\varepsilon1}$, and $N^{\varepsilon2}$ atoms) designated as *A* and *B*, are shown in cyan and red. Figs. 2 and 3 were produced using *O*.

anisotropy (Merritt, 1999) for protein atoms and water molecules are 0.59 and 0.49, with standard deviations of 0.16 and 0.18, respectively. The thermal ellipsoids are displayed at 50% probability for C^{α} (Fig. 6*a*) and main-chain atoms (Fig. 6*b*).

The equivalent isotropic displacement parameter $\langle U^2 \rangle$ (the mean-square displacement; MSD) of an atom along a specific direction can be calculated from the anisotropic parameters. If the direction of interest is the interatomic vector between atoms *A* and *B* (bonded or not) and if the two atoms share a rigid-body, then the difference in the MSD (Δ_{AB}) along the interatomic vector is zero. A Δ_{AB} value of 0 is a necessary condition for the pair of atoms to move as a rigid body. The correlation decreases with the magnitude of the delta value. Rosenfield analysis (Rosenfield & Trueblood, 1978) of $|\Delta_{AB}|$ values for all possible pairwise combinations of C^{α} atoms within the R67 DHFR structure was computed and plotted as a delta matrix (Fig. 7) using the program *ANISOANL* (Collaborative Computational Project, Number 4, 1994).

A Δ_{AB} value of 0 is a necessary condition for rigid-body motion but does not confirm the presence of rigid domain by itself. Rigid-body domains derived by Δ_{AB} analysis were evaluated in terms of a TLS model for rigid-body motion (Schomaker & Trueblood, 1968; Winn *et al.*, 2001). TLS analysis was performed using *ANISOANL* and the calculated ADPs from the TLS model were fitted to the observed ADPs obtained from anisotropic refinement. The agreement between the calculated and the observed ADPs was evaluated using an *R* factor ($R_{\rm TLS}$) and goodness-of-fit ($\rm GOF_{TLS}$).

The difference values (Δ_{AB}) were computed for main-chain atoms and C^{α} atoms. The matrix for main-chain atoms was noisy compared with that obtained from using C^{α} atoms. Therefore, we have used the delta matrix of C^{α} atoms for identification of quasi-rigid domains (Fig. 7). The delta differences were averaged in bins of three residues. 12 TLS groups were generated using the delta matrix as a guide and were evaluated for $R_{\rm TLS}$ factor and $\rm GOF_{TLS}$ values. The $R_{\rm TLS}$ for various trials ranged between 0.20 and 0.33 and $\rm GOF_{TLS}$ values varied between 24.0 and 39.0. The TLS group comprised of the whole molecule has the highest *R*-factor and GOF values. Partitioning of the molecule as indicated in

Fig. 7 was validated to have the best agreement index. This TLS group with R_{TLS} factor and GOF_{TLS} values of 0.20 and 24.4, respectively, is comprised of residues 22–32, 33–37, 38–70 and 71–78 that move as quasi-rigid bodies.

3.7. Solvent organization

A total of 191 water molecules were included in the R67 DHFR structure, of which 148 were refined with unit occupancy and 43 were refined with half occupancy on the basis of the quality of electron density. There are 168 water molecules (including the symmetry-related molecules) in the active-site pore. The crystal has a solvent content of 45%(*v*/*v*), which corresponds to a volume of 108 860 Å^3 being occupied by water molecules in the unit cell. This value corresponds to a theoretical number of 4354 water molecules in the unit cell or 272 water molecules per asymmetric unit, assuming an average volume of 25 Å^3 per water molecule (Hubbard & Argos, 1994). Therefore, the 191 solvent molecules observed in the present structure account for 70% of the total water content in the unit cell. The solvent molecules are well ordered, with an average isotropic *B* factor of 46.2 Å^2. The trend in *B* factors correlates well with the number of hydrogen bonds to protein. The mean *B* values decrease with increases in the number of contacts, with 40 Å^2 for a single contact, 33 Å^2 for two contacts and 17 Å^2 for three contacts.

The high resolution of the structure facilitates the identification of solvent molecules located in the second coordination shell and beyond from the protein surface. Contact analysis was applied to cluster water molecules into different hydration shells (Collaborative Computational Project, Number 4, 1994). Water molecules were assigned to the first hydration shell if they were involved in at least one hydrogen-bonding interaction with a protein atom. Water molecules with hydrogen bonding only to other water molecules were classified into the second hydration shell. By using these criteria and a hydrogen-bond cutoff distance of 3.4 Å, 85 water molecules were identified in the first hydration shell and 106 in higher shells. The mean *B* values for water molecules in the first and higher shells are 37 and 53 Å^2, respectively. Analysis of the distribution of water sites as a function of distance from the nearest protein atom indicates that the first shell is clustered around 2.8 Å, with a second cluster between 3.8 and 4.2 Å.

An analysis of solvent interactions with protein atoms shows that there are 2.1 times more hydrogen bonds to main-chain O atoms (42) than there are to main-chain N atoms (20). In comparison, the number of hydrogen bonds made to side-chain O atoms (35) is only around 1.2 times the number made to side-chain N atoms (29). These results are in accord with previous statistical and simulation studies, which suggest that main-chain O atoms have a greater ability than side-chain O atoms to form hydrogen bonds to water molecules (Thanki *et al.*, 1988).

3.8. Pentagonal array of water molecules

The active-site pore is filled with water molecules. Each subunit has 42 water molecules associated with the surface facing the pore. Because all four subunits contribute to the active-site pore, there are a total of 168 water molecules in the pore, the volume of which is approximately 2600 Å^3. The hour-glass-shaped pore is densely populated with well ordered water molecules (Fig. 5*b*) possessing a mean *B* value of 46 Å^2, which is the same as the mean value obtained for all water molecules. It was interesting to note that these waters form a network of hydrogen bonds that resembles fused pentagonal

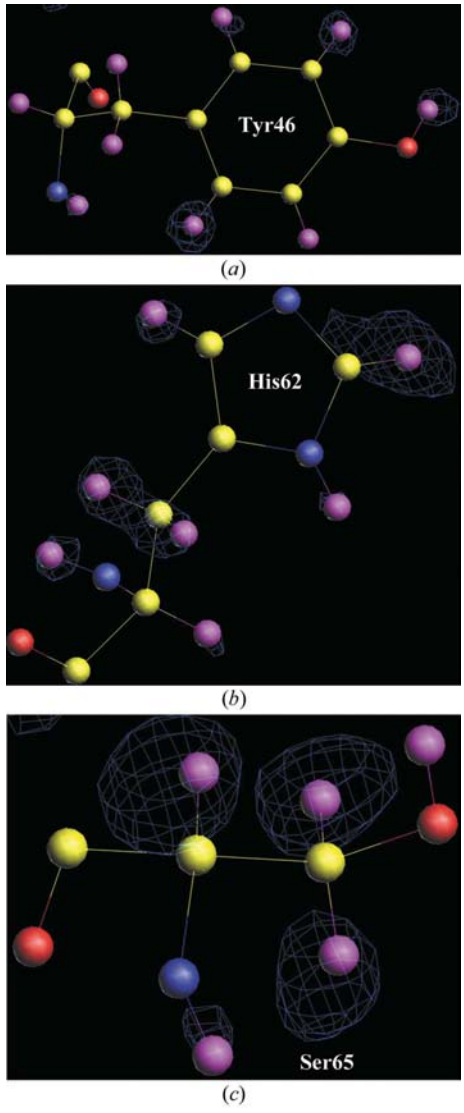

Figure 3
Electron density for H atoms. A difference Fourier map (blue) computed in the final stages of refinement indicates H-atom positions. Representative portions of the protein molecule are shown in the figure. The N, O, C and H atoms are displayed as blue, red, yellow, and purple spheres of the same size, respectively. (*a*) The H atoms attached to Tyr46 are superimposed on the electron-density map contoured at the 2.5σ level. (*b*) The H atoms of His62 are shown superposed on the difference Fourier map. Clearly, there is no feature for the presence of a H atom at $N^{\varepsilon 2}$ as expected for an unprotonated histidine. There is density at about the 3σ level for all the other H atoms associated with His62. The H atom bonded to $C^{\varepsilon 1}$ exhibits a much higher peak strength (4.5σ) compared with other H atoms, so the density is smeared at $C^{\varepsilon 1}$ when contoured at the 2.5σ level. (*c*) Density for H atoms of Ser65 contoured at the 2.5σ level. It may be noted that the H atoms attached to C^{α} and C^{β} atoms are seen at a much higher level (~4–5σ) than the H atom attached to the backbone N atom.

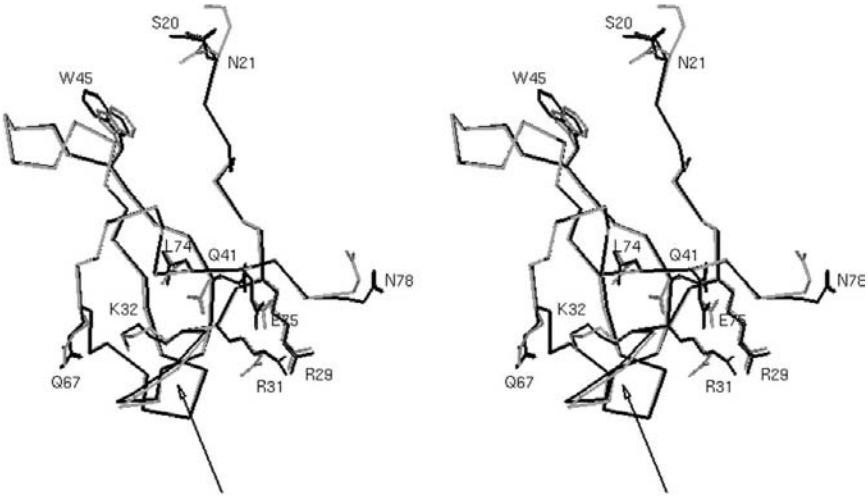

Figure 4
Comparison of cryogenic (100 K) and low (277 K) temperature structures. Stereoview of the least-squares superposition of C^{α} atoms of one subunit of the present structure (black) with the equivalent atoms of the structure determined at 277 K (gray; r.m.s.d. of 0.57 Å). The arrow indicates a significant shift in the loop comprised of residues 33–37 (∼0.4 Å). Side chains of several residues that exhibit considerable deviation are displayed.

rings (Fig. 5*c*). There are eight pentagonal rings linked to each other, forming a sheet of water molecules. Three of the eight pentagonal rings are slightly distorted. The water molecules belonging to one subunit are colored green and the symmetry-related water molecules are shown in yellow in Fig. 5. Each subunit contributes 28 water molecules to the pentagonal array of hydrogen bonds. This sheet of water molecules extends from the center to the outer pore. Owing to the 222 symmetry, there are four such sheets covering the respective subunits in the active-site pore.

3.9. C—H···O hydrogen bonds

The presence of C—H···O hydrogen bonds has been experimentally established by spectroscopic techniques and in crystals of organic compounds (Sutor, 1962; Desiraju, 1991). The occurrence of these hydrogen bonds in proteins have been described recently by several groups (Bella & Berman, 1996; Derewenda *et al.*, 1995; Fabiola *et al.*, 1997). The C—H···O bonds are weak attractive forces (∼2–4 kJ mol^{-1}; Desiraju, 1991) compared with familiar N—H···O or O—H···O hydrogen bonds (17–21 kJ mol^{-1}). The stereochemical preferences of C—H···O hydrogen bonds are similar to those of standard hydrogen bonds. Hydrogen bonds found between main-chain C^{α}—H and the carbonyl O atom in adjacent strands of β-sheets are the most common. We have investigated the occurrence of C—H···O bonds in the present structure, excluding water molecules, with a cutoff distance of 3.3 Å between C and O atoms. Almost all the bonds observed using this criterion were of the type C^{α}—H···O=C, in agreement with Derewenda *et al.* (1995). We have included in our analysis only those bonds for which the H atom attached to the C^{α} atom was seen in the electron-density map. Therefore, we were able to measure the actual distances and angles of the C^{α}—H···O bonds between the interstrand atoms. The average values of bond lengths $H^{\alpha}\cdots O$ and $C^{\alpha}\cdots O$ and bond angles C^{α}—H···O for a total of ten C^{α}—H···O hydrogen bonds observed in the structure are 2.4 and 3.2 Å and 142°, respectively, which is in excellent agreement with previous analysis of C—H···O hydrogen bonds in β-sheets (Fabiola *et al.*, 1997).

4. Discussion

This is the first high-resolution structure of R67 DHFR. The previous apoenzyme structure was refined at 1.7 Å resolution using data collected at 277 K (Narayana *et al.*, 1995). The structure of a binary complex with folate determined at 1.8 Å resolution revealed that the active site is the central pore (Narayana *et al.*, 1995). The crystallographic 222 symmetry axes passing through the middle of the pore result in overlapping of electron density owing to multiple binding sites for the ligands. From these previous structures, it was not clear whether the residues lining the pore exhibited alternative conformations. R67 DHFR does not possess a proton donor (Park, Zhuang *et al.*, 1997) and has no active-site carboxylate to interact with a protonated N1 of DHF. However, the active site has easy access to solvent molecules and may contribute to ligand binding, catalysis and inhibition. Kinetic studies and numerous mutations in the active site have been constructed and evaluated (Howell, 2005; Schmitzer *et al.*, 2004; Dam *et al.*, 2000; Bodenreider *et al.*, 2002; Mejean *et al.*, 2001). NMR studies of the dynamic behavior of the protein indicate a rigid structure (Pitcher *et al.*, 2003). Given that the complexity of the binding mode and the minor changes that occur at the active site are amplified owing to symmetry, it is necessary to better understand the structural details at atomic resolution. Such high-resolution data will enable us to perform refinement with anisotropic ADPs and provide an estimate of the dynamics of the molecule in the crystal. The atomic resolution structure of R67 DHFR described in this paper has fine details of alternative conformations, solvent organization, molecular motions in the crystal, positions of H atoms and comparative analysis with the previous structure determined at 1.7 Å resolution.

Amino-acid residue Gln67 has a long linear side chain and is positioned in the middle of the pore. In the 1.7 Å resolution structure, the terminal atoms of the side-chain C^{γ}, C^{δ}, $O^{\varepsilon1}$ and $N^{\varepsilon2}$ have high *B* factors and the corresponding electron density was weak in the Fourier maps. The atoms $O^{\varepsilon1}$ and $N^{\varepsilon2}$ were interpreted as forming hydrogen bonds with their symmetry-related atoms. In the present study, we observe electron density that is compatible with dual conformations for the four terminal side-chain atoms (Fig. 2*c*). The two side-chain conformations *A* and *B* are mutually exclusive. $N^{\varepsilon2}$ and $O^{\varepsilon1}$ form hydrogen bonds with water molecules in both conformations *A* and *B*. Only $N^{\varepsilon2}$ in conformation *A* can form

a hydrogen bond with $O^{\varepsilon 1}$ in conformation *B* of a symmetry-related molecule. The occurrence of conformation *A* for the Gln67 side chain in one subunit necessitates conformation *B* in a symmetry-related molecule to avoid steric clashes. This scenario breaks the 222 symmetry in a subtle way as a consequence of the discrete disorder. Although the overall symmetry of the molecule is dihedral, this minor variation in the active-site pore results in local asymmetry. However, all subunits in the tetramer can have the side chain of Gln67 in conformation *B* without steric clash, thus not violating exact 222 symmetry. Apart from Gln67, all the other residues lining the active site do not display alternative conformations, suggesting overall rigidity of the pore surface. This feature clearly represents the inherent flexibility of Gln67. Both conformations *A* and *B* are stabilized by hydrogen bonds to solvent and protein atoms. The averaged electron density produced by dual occupancy at low 1.7 Å resolution resulting in an averaged conformation is now resolved into its individual components at high 1.1 Å resolution. In the native enzyme, owing to a slight shift in the side chain of Gln67, the double hydrogen bonding with its symmetry-related side chain is disrupted, presumably resulting in an appreciable loss of favourable stacking capacity over the ring structures of the ligands. In contrast, the Q67H mutant of R67 DHFR displays stronger binding to both cofactor (110-fold) and substrate (36-fold) (Park, Bradrick *et al.*, 1997). Thus, it appears that the presence of a side chain with a flat ring (histidine in the Q67H mutant) in the center of the pore favors the tight binding of both the substrate and the cofactor.

Comparison of this structure with the previously determined 277 K structure indicates significant movements in the side chains of several residues (Arg31, Lys32, Gln41, Trp45, Gln67, Leu74, Glu75 and Asn78) as well as a stretch of main-chain atoms of residues 33–37. These residues are scattered over the surface with no correlated side-chain disorder. Instead, the side-chain disorder is isolated, unlike in crambin (Yamano & Teeter, 1994; Teeter *et al.*, 1993), repressor of primer (Vlassi *et al.*, 1998) and eosinophil-derived neurotoxin (Swaminathan *et al.*, 2002). The residues Arg31, Lys32, Lys33 and Glu75 are conserved in all R plasmid-encoded variants, suggesting that they play a functional role (Hicks *et al.*, 2003). Residue Gln67 lies in the active site and plays a role in binding and catalysis (Park, Bradrick *et al.*, 1997; Strader *et al.*, 2001). Trp45 occurs at the monomer–monomer interface with a relatively high mean *B* value for its side chain (30.3 $Å^2$) and weak and diffuse density in electron-density maps. The W45F

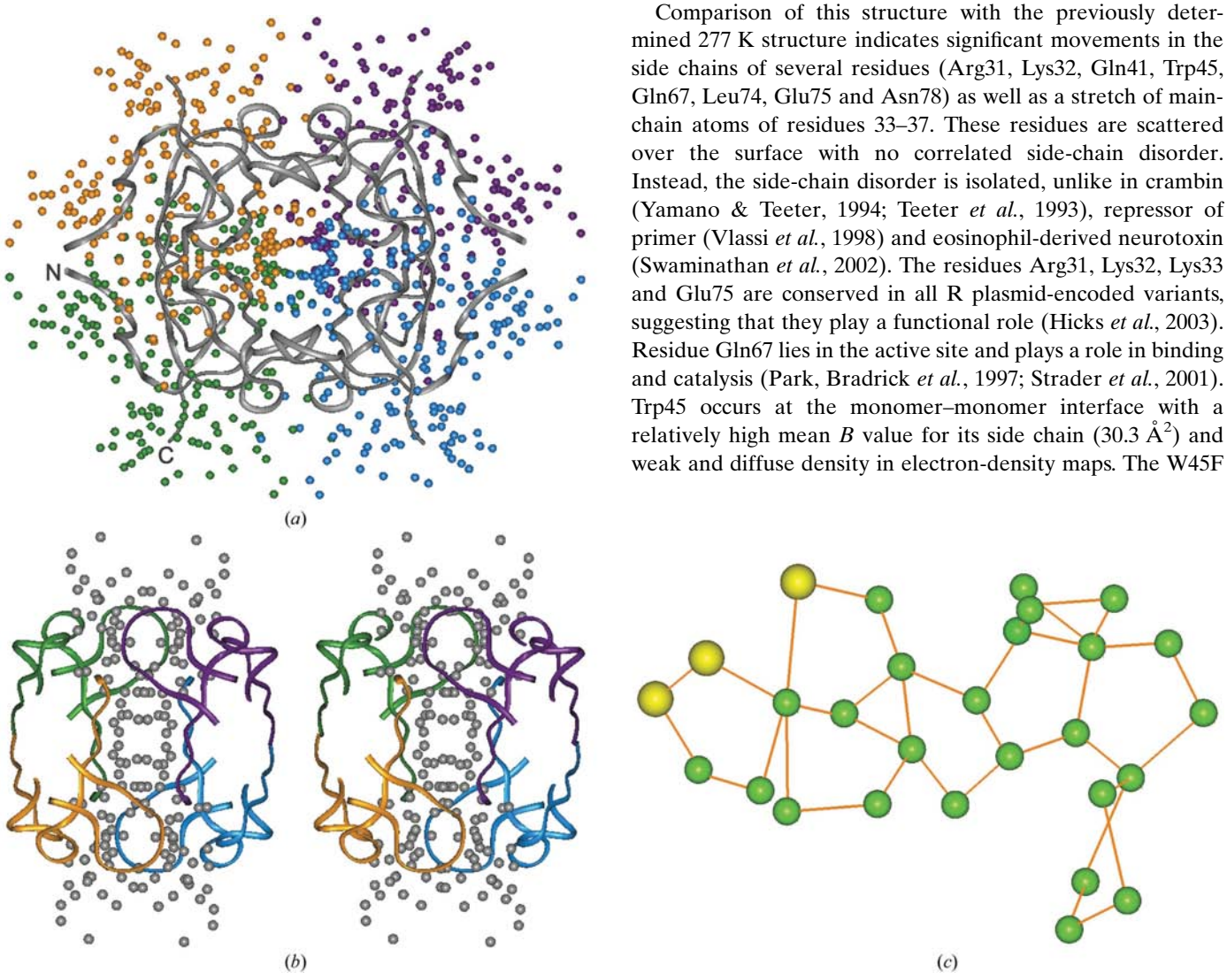

Figure 5
Solvent organization. (*a*) R67 DHFR tetramer (gray) is surrounded by water molecules. The water molecules belonging to different subunits are shown in different colors. (*b*) Stereoview of an hour-glass-shaped pore filled with 168 well ordered water molecules (including symmetry-related water molecules; grey spheres) is shown. This view is rotated by 90° about the horizontal axis to the view in (*a*). Portions of protein molecule that cover the water surface in the pore are displayed as ribbons. The ribbons belonging to the four subunits are in different colors. (*c*) A pentagonal network of water molecules is seen at the surface of subunits lining the active-site pore. Fused pentagons formed by hydrogen-bonding interactions among water molecules are shown for one subunit. Such fused pentagonal rings are likewise associated with other subunits of the tetramer. The green spheres represent water molecules associated with one subunit. The yellow spheres belong to a symmetry-related molecule. The orange lines denote hydrogen bonds. Figs. 4 and 5 were prepared using *INSIGHT II*.

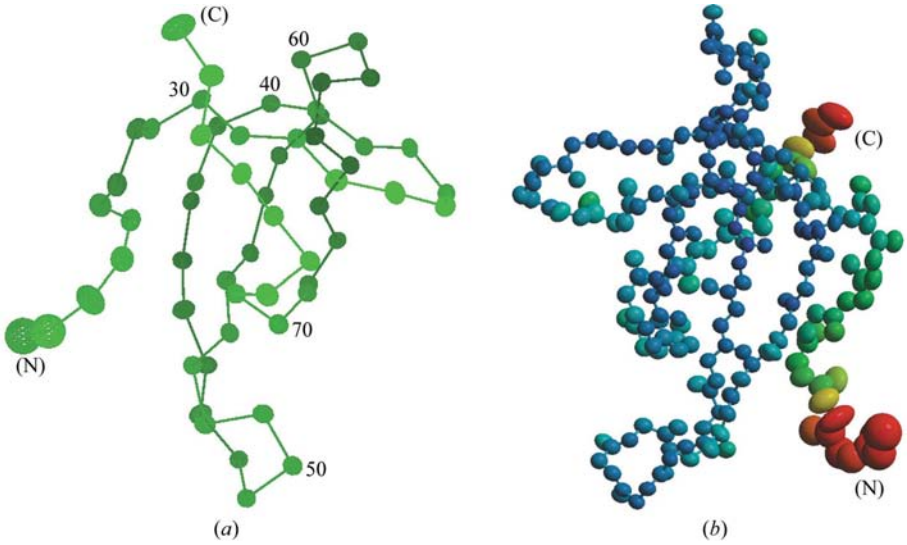

Figure 6
Representation of anisotropic thermal motions. (*a*) Thermal ellipsoids for C^{α} atoms are drawn with 50% probability using *XtalView* (McRee, 1999). The ADPs for C^{α} atoms were calculated from the experimental U_{ij}s. (*b*) Thermal ellipsoid diagram of the main-chain atoms in R67 DHFR drawn using *RASTEP* (Merritt, 1999). The atoms are color-ramped according to the magnitude of the displacements, ranging from blue (lowest) through cyan, green and yellow (medium) to red (highest).

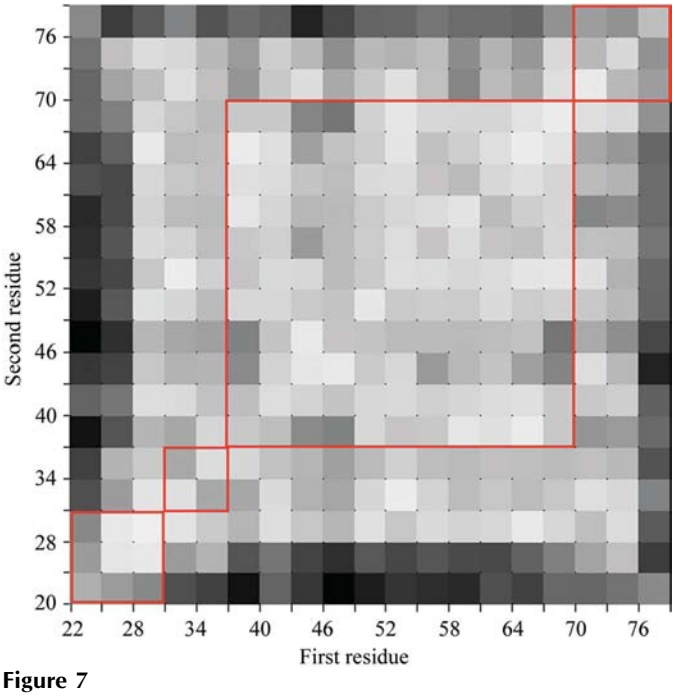

Figure 7
Partitioning the structure of R67 DHFR into multiple TLS groups. The Rosenfield $|\Delta_{AB}|$ matrix for the refined anisotropic positional displacement along the $C^{\alpha}-C^{\alpha}$ interatomic vectors of R67 DHFR is shown. The difference values are averaged in bins of three residues. The anisotropic atoms used in the calculation are from residues 22 to 78. The $|\Delta_{AB}|$ values are shaded where regions of uniformly white and light shading correspond to possible quasi-rigid bodies in the protein. The larger difference values correspond to darker boxes, implying flexible regions in the molecule. The plot suggests an optimal partition of the subunit into four TLS groups as indicated by red squares. Combining the results from delta-matrix plot and by consideration of the β-strand boundaries, four TLS groups (quasi-rigid units; residues 22–32, 33–37, 38–70 and 71–78) were constructed and validated. This figure was produced using *ANISOANL*.

mutant had a marginal influence in maintaining the monomer–monomer interface (West *et al.*, 2000). Interestingly, the main-chain movement of residues 33–37 observed in the comparative analysis matches one of the quasi-rigid groups obtained by Rosenfield and TLS analyses. Therefore, we predict that the loop connecting β-strands *a* and *b* facing the active-site pore to have some level of flexibility that may have functional implications.

Water molecules play an important role in maintaining the structural stability of proteins. The present structure has revealed extensive water-mediated hydrogen-bond networks. There are a total of 191 water molecules associated with each subunit (Fig. 5*a*), of which 85 bind directly to the protein. The average hydrogen-bond distance decreases from 2.96 Å for protein–water contacts to 2.90 Å for water–water contacts. This trend is similar to that seen in other atomic resolution protein structures (Kumaraswamy *et al.*, 1996).

It was interesting to note the presence of a layer of a pentagonal ring pattern for water molecules in the active-site pore covering the surface of each subunit (Fig. 5*c*). The presence of such highly ordered water molecules in the active site may provide clues to its importance in catalysis. Solvent reorganization may be involved in the effects of different ligands and mutant enzymes on ΔH and $T\Delta S$ (Grunwald & Steel, 1995). There are only a few cases in which water molecules associated with the surface of a protein adopt a certain well defined geometrical pattern. The pentagonal pattern of the water network seen here is reminiscent of ice clathrates in the crystal structure of the plant protein crambin (Teeter, 1992). In crambin, five water pentagons are fused together covering a hydrophobic patch on the protein. Water pentagons have also been found in crystal structures of DNA fragments (Kennard, 1985; Neidle *et al.*, 1980). A recent ultrahigh-resolution crystal structure of ribonuclease A has shown ordered water molecules in several fused rings forming a semiclathrate hydrate structure (Esposito *et al.*, 2000). The occurrence of a regular pattern of water molecules mimic a central core with somewhat rigid characteristics. As noted in previous studies, specific water molecules in the active site play a role in catalysis and inhibition and as recognition features for the binding of ligands (Narayana *et al.*, 1995). Recent work using combinatorial mutagenesis has shown the compensatory mutations at the active site preserve the enzyme's activity (Schmitzer *et al.*, 2004). We predict that active-site water molecules are resilient and may display context-dependent rearrangement of the water network and help facilitate such compensatory mutations. These solvent molecules may have a characteristic pore or channel effect

contributing to the dielectric constant in the active-site pore. One active site accommodates many different ligands: NADPH, folate, Congo red and novobiocin, among others (Howell, 2005). It is likely that both protein residues and the associated water molecules are partnered together to bind such diverse ligands. Each subunit is covered by a sheet of eight fused water pentagons made up of 28 water molecules. There are four such sheets in the pore and to our knowledge this is the first report of such large number of ordered water molecules (112) clustered in a symmetrical fashion in the active site.

Atomic resolution data has the virtue of revealing H-atom positions in a protein molecule. The identification of H atoms in specific residues is especially important to understand catalytic mechanism and stability, among other characteristics of the molecule. For example, in the present case, His62 is a key residue implicated in pH-dependent tetramer–dimer equilibrium (Nichols *et al.*, 1993; Dam *et al.*, 2000; Mejean *et al.*, 2001). Electron density for H atoms (contoured at the 2.5σ–4.5σ level) associated with His62 is seen unambiguously (Fig. 3*b*). The H atom bonded to $C^{\varepsilon1}$ displays the strongest density (4.5σ) among the H atoms attached to His62. The H atoms associated with the main-chain N and C^{α} atoms and the side-chain C^{β}, C^{γ}, $N^{\delta1}$ and $C^{\delta2}$ atoms were seen at about the 2.5σ level. The $N^{\varepsilon2}$ atom of His62 is hydrogen bonded to O^{γ} of the symmetry-related Ser59 (2.64 Å). The difference Fourier maps do not decisively indicate the location of a H atom attached to O^{γ} of Ser59 because the nearby density is diffuse and the strongest isolated peak (2.1σ) is at a distance of about 1.4 Å. This peak is about 2 Å from $N^{\varepsilon2}$ of the symmetry-related His62. This led us to infer that $N^{\varepsilon2}$ is not protonated, which is in agreement with kinetic pH-profile studies (Nichols *et al.*, 1993). Protonation of His62 (pK_a of 6.84) in the low pH range 5–8 is thought to be responsible for dissociation of a tetramer to a dimer (Nichols *et al.*, 1993; Park, Zhuang *et al.*, 1997; Dam *et al.*, 2000; Mejean *et al.*, 2001). This dissociation is presumably a consequence of electrostatic repulsion between the two symmetry-related His62 residues and steric clash between the side chain of His62 and the side-chain hydroxyl of the symmetry-related Ser59. Thus, interaction between the symmetry-related pairs His62 and Ser59 plays an important role in subunit–subunit association (Dam *et al.*, 2000). This study confirms the likely molecular basis for pH-dependent tetramer to dimer dissociation.

We were able to identify at least ten $C^{\alpha}-H\cdots O{=}C$ hydrogen bonds between residues belonging to adjacent β-strands. $C-H\cdots O$ interactions in biological systems have recently received attention owing to their structural and functional role. At present, the specific role of these bonds in R67 DHFR is not clear. It may be that these unconventional hydrogen bonds may cooperatively help to contribute to the stability of the protein.

The availability of anisotropic thermal parameters enabled us to perform Rosenfield delta-matrix analysis (Rosenfield & Trueblood, 1978). The quasi-rigid-body groupings of residues based on the delta plot was validated by TLS analysis (Schomaker & Trueblood, 1968; Winn *et al.*, 2001; Howlin *et al.*, 1993). The TLS model was used to elucidate the characteristics of the rigid-body motions within a molecule in the crystal. The protein molecule was partitioned into four TLS groups labeled TLS1 (22–32), TLS2 (33–37), TLS3 (38–70) and TLS4 (71–78). TLS1, comprised of β-strand *a*, TLS2, consisting of the loop residues between β-strands *a* and *b*, TLS3, comprising β-strands *b*, *c* and *d*, and TLS4, consisting of β-strand *e*, exhibit quasi-rigid motions. Inclusion of TLS2 into either TLS1 or TLS3 or TLS4 results in higher R_{TLS} and GOF_{TLS} values, suggesting that this section is variable in comparison to the rest of the structure. The TLS2 is the shortest stretch among the four TLS groups and is located near the active-site pore. Residues Lys33, Ser34, Gly35, Ala36 and Ala37 are grouped in TLS2. Comparison of the R67 DHFR structure determined at 1.1 Å reported here and the previous 1.7 Å resolution structure indicate variation of main-chain atoms in the same loop (33–37) of the structure (Fig. 4). This indication of variability of the loop connecting β-strands *a* and *b* from different analyses suggest inherent flexibility of the loop and may have functional relevance to the enzyme activity or inhibition.

The present high-resolution structure of R67 DHFR represents a significant improvement in the structural model available for this protein. The active-site residues and the solvent organization in the pore are described in greater detail. The accurate description of the active-site geometry, ordered pentagonal network of water molecules, possible flexibility of specific residues and quasi-rigid domains will aid greatly in the design of selective and potent inhibitors as an alternative to TMP. The availability of fine details of the active site may help in modeling studies. A mix of rigid and flexible characteristics offered by solvent molecules in the core of the molecule may contribute to the plasticity of the R67 DHFR active site. The synthesis of knowledge from the present high-resolution structure and future chemomutagenesis studies may help to better understand the catalytic mechanism as well as providing insights into drug design. This work suggests that active-site water molecules could serve as potential molecular-recognition features in ligand binding, catalysis and inhibition.

NN thanks Drs Nguyen-Huu Xuong and David Matthews for encouragement and Dr Elizabeth Howell for the sample. NN thanks Dr Michael Weiss for support and the staff at APS beamline14-BM-D for help with X-ray data collection. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science under Contract No. W-31-109-Eng-38. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources under grant No. RR07707.

References

Albery, W. J. & Knowles, J. R. (1976). *Biochemistry*, **15**, 5631–5640.
Bella, J. & Berman, H. M. (1996). *J. Mol. Biol.* **264**, 734–742.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bodenreider, C., Kellershohn, N., Goldberg, M. E. & Mejean, A. (2002). *Biochemistry*, **41**, 14988–14999.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dam, J., Rose, T., Goldberg, M. E. & Blondel, A. (2000). *J. Mol. Biol.* **302**, 235–250.
Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.
Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.
Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.
Fabiola, G. F., Krishnaswamy, S., Nagarajan, V. & Pattabhi, V. (1997). *Acta Cryst.* D**53**, 316–320.
Falzone, C. J., Wright, P. E. & Benkovic, S. J. (1994). *Biochemistry*, **33**, 439–442.
Fierke, C. A., Kuchta, R. D., Johnson, K. A. & Benkovic, S. J. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52**, 631–638.
Fleming, M. P., Datta, N. & Gruneberg, R. N. (1972). *Br. Med. J.* **1**, 726–728.
Greenberg, G. R. (1954). *Fed. Proc.* **13**, 745–759.
Grunwald, E. & Steel, C. (1995). *J. Am. Chem. Soc.* **117**, 5687–5692.
Hicks, S. N., Smiley, R. D., Hamilton, J. B. & Howell, E. E. (2003). *Biochemistry*, **42**, 10569–10578.
Hitchings, G. H. & Smith, S. L. (1980). *Adv. Enzyme Regul.* **18**, 349–371.
Howell, E. E. (2005). *Chembiochem*, **6**, 590–600.
Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. & Driessen, H. P. C. (1993). *J. Appl. Cryst.* **26**, 622–624.
Hubbard, S. J. & Argos, P. (1994). *Protein Sci.* **3**, 2194–2206.
Jones, B. E. & Matthews, C. R. (1995). *Protein Sci.* **4**, 167–177.
Jones, T. A. (2004). *Acta Cryst.* D**60**, 2115–2125.
Kennard, O. (1985). *J. Biomol. Struct. Dyn.* **3**, 205–226.
Kraut, J. & Matthews, D. A. (1987). *Active Sites of Enzymes*, edited by F. A. Jurnak & A. McPherson, Vol. 3, pp. 1–71. New York: John Wiley & Sons.
Kumaraswamy, V. S., Lindley, P. F., Slingsby, C. & Glover, I. D. (1996). *Acta Cryst.* D**52**, 611–622.
Kuyper, L. F., Baccanari, D. P., Jones, M. L., Hunter, R. N., Tansik, R. L., Joyner, S. S., Boytos, C. M., Rudolph, S. K., Knick, V., Wilson, H. R., Caddell, J. M., Friedman, H. S., Comley, J. C. & Stables, J. N. (1996). *J. Med. Chem.* **39**, 892–903.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
MacArthur, M. W. & Thornton, J. M. (1999). *Acta Cryst.* D**55**, 994–1004.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. & Hoogsteen, K. (1977). *Science*, **197**, 452–455.
Matthews, D. A., Smith, S. L., Baccanari, D. P., Burchall, J. J., Oatley, S. J. & Kraut, J. (1986). *Biochemistry*, **26**, 8591–8598.
Mejean, A., Bodenreider, C., Schuerer, K. & Goldberg, M. E. (2001). *Biochemistry*, **40**, 8169–8179.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117**.**
Morrison, J. F. (1991). *A Study of Enzymes*, edited by S. A. Kuby, Vol. 2, pp. 193–226. Boca Raton, FL, USA: CRC Press.
Narayana, N., Matthews, D. A., Howell, E. E. & Xuong, N.-H. (1995). *Nature Struct. Biol.* **2**, 1018–1025.
Neidle, S., Berman, H. M. & Shieh, H. S. (1980). *Nature (London)*, **288**, 129–133.
Nichols, R., Weaver, C. D., Eisenstein, E., Blakley, R. L., Appleman, J., Huang, T. H., Huang, F. Y. & Howell, E. E. (1993). *Biochemistry*, **32**, 1695–1706.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, H., Bradrick, T. D. & Howell, E. E. (1997). *Protein Eng.* **10**, 1415–1424.
Park, H., Zhuang, P., Nichols, R. & Howell, E. E. (1997). *J. Biol. Chem.* **272**, 2252–2258.
Pattishall, K. H., Acar, J., Burchall, J. J., Goldstein, F. W. & Harvey, R. J. (1977). *J. Biol. Chem.* **252**, 2319–2323.
Pitcher, W. H. III., DeRose, E. F., Mueller, G. A., Howell, E. E. & London, R. E. (2003). *Biochemistry*, **42**, 11150–11160.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* D**60**, 2288–2294.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Reece, L. J., Nichols, R., Ogden, R. C. & Howell, E. E. (1991). *Biochemistry*, **30**, 10895–10904.
Rosenfield, R. E. & Trueblood, K. N. (1978). *Acta Cryst.* A**34**, 828–829.
Roth, B. & Cheng, C. C. (1982). *Prog. Med. Chem.* **19**, 269–331.
Schmitzer, A. R., Lepine, F. & Pelletier, J. N. (2004). *Protein Eng. Des. Sel.* **17**, 809–819.
Schomaker, V. & Trueblood, K. (1968). *Acta Cryst.* B**24**, 63–76.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Smith, S. L., Stone, D., Novak, P., Baccanari, D. P. & Burchall, J. J. (1979). *J. Biol. Chem.* **254**, 6222–6225.
Stone, D. & Smith, S. L. (1979). *J. Biol. Chem.* **254**, 10857–10861.
Strader, M. B., Smiley, R. D., Stinnett, L. G., VerBerkmoes, N. C. & Howell, E. E. (2001). *Biochemistry*, **40**, 11344–11352.
Sutor, D. J. (1962). *Nature (London)*, **195**, 68–69.
Swaminathan, G. J., Holloway, D. E., Veluraja, K. & Acharya, K. R. (2002). *Biochemistry*, **41**, 3341–3352.
Teeter, M. M. (1992). *Dev. Biol. Stand.* **74**, 63–72.
Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.
Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1988). *J. Mol. Biol.* **202**, 637–657.
Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Vlassi, M., Dauter, Z., Wilson, K. S. & Kokkinidis, M. (1998). *Acta Cryst.* D**54**, 1245–1260.
West, F. W., Seo, H. S., Bradrick, T. D. & Howell, E. E. (2000). *Biochemistry*, **39**, 3678–3689.
White, P. A. & Rawlinson, W. D. (2001). *J. Antimicrob. Chemother.* **47**, 495–496.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Yamano, A. & Teeter, M. M. (1994). *J. Biol. Chem.* **269**, 13956–13965.
Zhuang, P., Eisenstein, E. & Howell, E. E. (1994). *Biochemistry*, **33**, 4237–4244.
Zolg, J. W. & Hanggi, U. J. (1981). *Nucleic Acids Res.* **9**, 697–710.