Acta Cryst. (1998). D**54**, 735–749

The Refined Structure of dUTPase from *Escherichia coli*

ZBIGNIEW DAUTER,[a] KEITH S. WILSON,[a]* GUNILLA LARSSON,[b] PER OLOF NYMAN[b] AND EILA S. CEDERGREN-ZEPPEZAUER[c]

[a]*Department of Chemistry, University of York, Heslington, York YO1 5DD, England,* [b]*Department of Biochemistry, Center for Chemistry and Chemical Engineering, POB 124, Lund University, S-22100 Lund, Sweden, and* [c]*Department of Structural Chemistry, Arrhenius Laboratories for Natural Sciences, Stockholm University, S-10691 Stockholm, Sweden. E-mail: keith@yorvic.york.ac.uk*

(*Received 11 June 1997; accepted 10 November 1997*)

Abstract

Deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase, E.C. 3.6.1.23) catalyzes the hydrolysis of dUTP to dUMP and pyrophosphate and is involved in nucleotide metabolism and DNA synthesis. A crystal of the recombinant *E. coli* enzyme, precipitated from polyethylene glycol mixtures in the presence of succinate at pH 4.2, was used to collect synchrotron diffraction data to 1.9 Å resolution, in space group $R3$, $a = b = 86.62$, $c = 62.23$ Å. Mercury and platinum derivative data were collected at wavelengths to optimize the anomalous contribution. The resulting 2.2 Å MIRAS phases differed from the final set by 40° on average and produced an excellent map which was easy to interpret. The model contains 132 water molecules and refined to an R value of 13.7%. 136 residues have clear electron density out of 152 expected from the gene sequence. The 16 C-terminal residues are presumably disordered in the crystal lattice. The monomer is a 'jelly-roll' type, containing mostly β-sheet and only one short helix. The molecule is a tight trimer. A long C-terminal arm extends from one subunit and encompasses the next one within the trimer contributing to its β-sheet. Conserved sequence motifs common among dUTPases, previously suggested to compose the active site and confirmed in a recent study of the dUDP complex, are located at subunit–subunit interfaces along the threefold axis, in parts of the β-sheet and in loop regions. A similar molecular architecture has recently been found in two other trimeric dUTPases.

1. Abbreviations

dUTPase, deoxyuridine 5′-triphosphate nucleotidohydrolase; dUTP, deoxyuridine 5′-triphosphate; dUDP, deoxyuridine 5′-diphosphate; dUMP, deoxyuridine 5′-monophosphate; dTMP, deoxythymidine 5′-monophosphate; MIRAS, multiple isomorphous replacement/anomalous scattering.

2. Introduction

dUTPase, a key enzyme in pyrimidine nucleotide metabolism, specifically hydrolyzes deoxyuridine-5′-triphosphate (dUTP) to dUMP and inorganic pyrophosphate (Kornberg & Baker, 1991). The enzyme influences DNA synthesis in a crucial way and the existence of dUTPase has been shown to be important for the cell cycle. Attempts to completely inactivate the dUTPase gene in *Escherichia coli (E. coli)* and *Saccharomyces cerevisiae* have been unsuccessful (El-Hajj *et al.*, 1988; Gadsden *et al.*, 1993). dUTPase keeps the intracellular concentration of dUTP at a low level thus suppressing the incorporation of uracil into DNA (Tye *et al.*, 1977). Decreased dUTPase activity causes elevated recombination frequency (Konrad, 1977), abnormal mutation rate and accumulation of short DNA fragments as intermediates in DNA replication (Tye *et al.*, 1977). In some organisms, the dUTPase reaction is the major route delivering precursor molecules (dUMP) for the production of dTMP. dUTPase from *E. coli* shows a strict specificity towards dUTP (Bertani *et al.*, 1963; Larsson, Nyman *et al.*, 1996) and the active site discriminates between nucleotides with respect to the sugar moiety as well as the pyrimidine part.

The enzyme from *E. coli* is the best documented dUTPase among all species and was first described by Bertani *et al.* (1961). It was later purified and characterized by Shlomai & Kornberg (1978). The structural gene for the enzyme was isolated (Taylor *et al.*, 1980; Lundberg, Karlström *et al.*, 1983) and the nucleotide sequence determined (Lundberg, Thoresson *et al.*, 1983). It predicts a polypeptide of 152 amino-acid residues for the dUTPase subunit (Cedergren-Zeppezauer *et al.*, 1988). Using an expression vector, Hoffmann *et al.* (1987) obtained an overproducing system that provides about 600 times more dUTPase than wild-type bacterial strains and a simplified procedure was developed for the purification of the enzyme in large amounts. In contrast to the report by Shlomai & Kornberg (1978) about dUTPase being a zinc metalloenzyme, Hoffmann (1988) found that zinc is not necessary for activity; instead Mg^{2+} is essential (Larsson, Nyman *et al.*, 1996; G. A. Rogers,

© 1998 International Union of Crystallography
Printed in Great Britain – all rights reserved

G. Larsson & P. O. Nyman, unpublished work). The *E. coli* dUTPase was first shown to be a trimer in the crystal (Cedergren-Zeppezauer *et al.*, 1992). This is proven by gel-filtration experiments (this work) to be true in solution also.

A growing number of dUTPases from various sources have been reported and several sequences have been published (Björnberg *et al.*, 1993; Bergman *et al.*, 1994; Elder *et al.*, 1992; Broyles, 1993; Roseman *et al.*, 1996; Giroir & Deutsch, 1987; Climic *et al.*, 1994; McIntosh *et al.*, 1992, 1994; Pri-Hadash *et al.*, 1992; Wagaman *et al.*, 1993). Protein sequence comparisons between the *E. coli* enzyme (Lundberg, Thoresson *et al.*, 1983) and viral dUTPases (McGeoch, 1990) revealed five conserved motifs scattered along the sequence, indicating that those parts might be important for catalytic function and/or stability. The conserved motifs have been used to screen for genes encoding dUTPase in other genomes and such genes have been found among herpes viruses, retroviruses, pox viruses (McGeoch, 1990) and parasites (*Trypanosoma, Leishmania*) (Camacho *et al.*, 1998).

We previously published a brief report of the three-dimensional structure of the native *E. coli* enzyme (Cedergren-Zeppezauer *et al.*, 1992). Here, a detailed description of the refined structure at 1.9 Å resolution is given. The structure of *E. coli* dUTPase in complex with dUDP has been reported recently (Larsson, Svensson *et al.*, 1996). In addition the structures of feline immunodeficiency virus dUTPase (Sridhar Prasad *et al.*, 1996) and the human enzyme with its nucleotide complexes (Mol *et al.*, 1996) were determined: both are trimeric enzymes with overall organisation similar to that of *E. coli*. The *E. coli* and human complexes confirmed that in both species the conserved motifs do form the active centres of the enzymes.

3. Purification and crystallization

Recombinant dUTPase was isolated from an overproducing *E. coli* strain (MC 1000 containing plasmids pHW1 and pRK248cIts) as described by Hoffmann *et al.* (1987). The enzyme was stored in 25 m*M* Bis–Tris buffer, pH 6.2, 0.06% NaN_3, at 277 K for several months before crystallization. Alternatively, freshly prepared enzyme was incubated at 375 K for a few days. These treatments were needed to obtain reproducibly the crystal form suitable for diffraction experiments.

The enzyme was crystallized at 293 K using sitting drops. An enzyme solution in Bis–Tris buffer at pH 6.2 was mixed with precipitant and additives but not equilibrated against an outer solution. Crystallization mixtures of 40 µl, containing 2–3 mg protein ml^{-1}, were placed in small (100 µl) depressions in plastic cell-culture plates and sealed. Crystals grown in 3–5% polyethylene glycol (PEG) 8000, 50 m*M* succinate, 50 µ*M* $MgCl_2$ and 450 µ*M* sodium pyrophosphate ($NaPP_i$) at pH 4.2 were used for data collection on the native enzyme and for one of the heavy-atom derivatives by soaking in 75 µ*M* ethyl mercury phosphate for 4 h. A second derivative was obtained by soaking crystals in a solution containing 1.2 m*M* K_2PtCl_4 for 7 h. These crystals were grown in a mixture containing 6.5 m*M* succinate, 5 m*M* $MgCl_2$, 0.5 m*M* $NaPP_i$, 1.5% PEG 8000 at pH 4.2.

Crystalline dUTPase was analysed for full-length protein by tryptic degradation and sequence analysis of isolated peptides. Rhombohedral crystals (700 µg) were thoroughly washed before analysis. The redissolved protein was denatured in 0.1 *M* NH_4HCO_3, pH 8.1, 8 *M* urea at 293 K for 1.5 h. The sample was lyophilized and redissolved in 0.1 *M* NH_4HCO_3. The sample was again denatured for 5 min at 373 K, trypsin (2%) was added, the resulting solution incubated at 310 K for 4 h, lyophilized and redissolved in 0.1 *M* NH_4HCO_3, pH 7.1. The fragments were separated by reverse-phase HPLC on a proRPC 5/10 column (Pharmacia), buffer system *A*, 50 m*M* NH_4HCO_3, pH 7.2; and *B*, 50 m*M* NH_4HCO_3, pH 7.2, 70% CH_3CN. Selected fractions were lyophilized and sequenced by Edman degradation using an Applied Biosystems 470A Protein Sequencer. Five fragments, corresponding to amino-acid residues 14–34, 35–71, 105–116, 117–141, 142–152 were identified which confirmed that the 16 C-terminal residues (137–152) are present in the crystals.

The molecular mass of freshly prepared enzyme was determined by gel filtration on a Sephadex G-75 column (1.6 × 65 cm), equilibrated with 25 m*M* Bis–Tris buffer, pH 7, 0.1 m*M* NaCl, Fig. 1. The value obtained, 52 kDa, corresponds to an assembly of three subunits since the

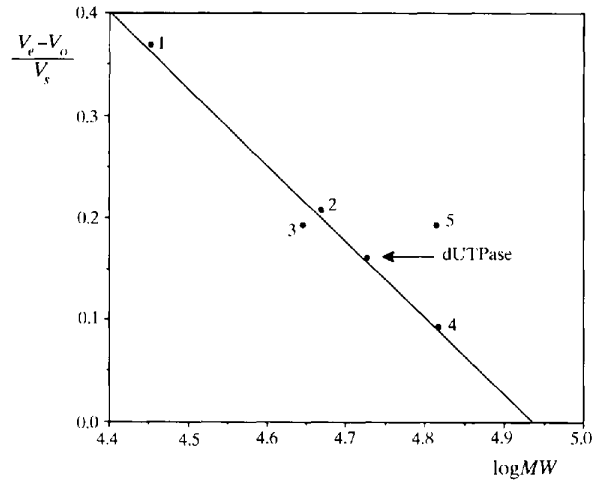

Fig. 1. Molecular-mass determination of *E. coli* dUTPase by gel filtration on Sephadex G-75. Calibration proteins were: (1) carbonic anhydrase, 29 kDa; (2) ovalbumin, 45 kDa; (3) 3-phosphoglycerate kinase, 47 kDa; (4) bovine serum albumin, 66 kDa and (5) hemoglobin, 64.5 kDa. The position of dUTPase is indicated by an arrow and a value of 52 kDa is estimated.

Table 1. *Summary of data collection and processing*

	Native		Mercury	Platinum		
	High	Low	High	Low	High	Low
Resolution (Å)	1.9	2.6	2.0	2.5	2.1	3.0
Wavelength (Å)	0.995		0.995		1.05	
Total reflections	48670		37295		30385	
Unique reflections	13640		11737		9910	
$R(I)$† (%)	6.9		5.9		5.7	
$R(I)$ outer shell (%)	15.0		24.1		18.1	
$I/\sigma(I)$	18.6		11.9		11.0	
$I/\sigma(I)$ outer shell	8.3		3.6		5.1	
Completeness (%)	98.9		94.1		96.9	
Completeness outer shell (%)	92.1		89.6		98.0	

† The merging $R(I)$ factor is defined as $\sum |I - \langle I \rangle| / \sum I$.

molecular mass of the trimer calculated from the amino-acid sequence is 48.9 kDa.

4. Data collection and processing

The diffraction data for the native crystal as well as for the heavy-atom derivatives were collected on the X31 EMBL beamline at DESY, Hamburg. A prototype imaging-plate scanner built in EMBL – the forerunner of the MAR Research device (Hendrix & Lentfer, unpublished work) – was used. All data, for native enzyme and derivatives, were collected at room temperature in two passes at high and low resolution, Table 1. Two crystals were used for the native, single samples for both derivatives. The wavelength for the mercury and platinum derivatives was adjusted to be slightly shorter than the corresponding metal L_{III} absorption edges (1.009 and 1.072 Å, respectively) to maximize the anomalous scattering. The orientation of the rhombohedral crystals was such that the threefold axis was close to perpendicular to the spindle axis, and hence each crystal was rotated through about 90° in total.

MOSFLM (Leslie *et al.*, 1986) was used for the integration of the images. All subsequent programs used were from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Data collection and processing are summarized in Table 1. Detailed results are presented for the native data only; the characteristics of the derivative data are similar. Fig. 2 shows the merging R in resolution ranges and Fig. 3 the completeness together with the proportion of reflections greater than $3\sigma(I)$. Fig. 4 shows the Wilson plot of the native data (Wilson, 1942) which indicates an average B value of 15.6 Å^2. The Bijvoet differences were measured for 85% of the Hg and 92% of the Pt derivative data.

The mean fractional isomorphous difference was 15.4% for the mercury and 15.0% for the platinum. The mean empirical ratio between the anomalous and isomorphous differences, K_{emp} defined as the average of $[F_{PH}^{+} - F_{PH}^{-}]/[F_{PH} - F_{P}]$, was 3.2 for the mercury and 3.4 for the platinum data. The theoretical values of K_{emp} for

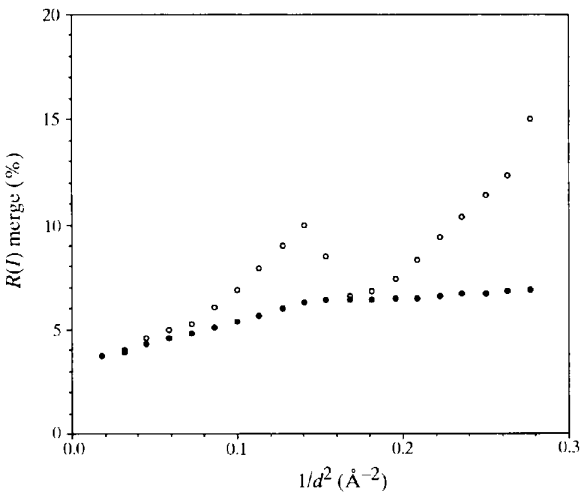

Fig. 2. The merging R factor ($\sum |I - \langle I \rangle| \sum I$). The open circles correspond to values in the individual resolution ranges, and the black circles to the cumulative values to that resolution.

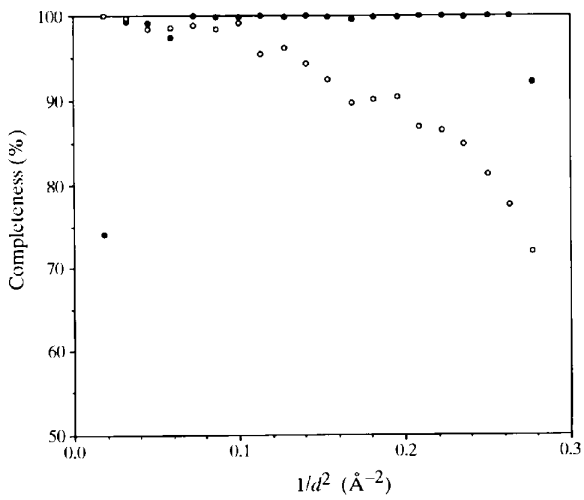

Fig. 3. Percentage completeness of the native data as a function of resolution (black circles) and the percentage of the measured X-ray intensities grater than 3σ (open circles) for the native data.

Table 2. *Heavy-atom parameters and phasing statistics*

	x	y	z	B	occ	Aocc†
Hg	0.206	0.329	0.452	25	0.84	0.84
Pt	0.145	0.391	0.446	52	1.21	1.18

Resolution range (Å)	N reflections	FOM‡	Hg R_{iso}§	Hg K_{emp}¶	Hg PhP††	Hg R_{Cullis}‡‡	Pt R_{iso}§	Pt K_{emp}¶	Pt PhP††	Pt R_{Cullis}‡‡
20.0–8.48	99	0.82	17.2	6.44	4.33	0.37	19.7	6.70	2.39	0.61
8.48–5.92	299	0.85	16.2	5.93	4.67	0.33	21.0	6.74	2.83	0.47
5.92–4.54	542	0.75	12.3	5.82	3.02	0.47	15.3	6.08	1.99	0.63
4.54–3.68	880	0.66	12.4	5.46	2.14	0.59	14.3	6.11	1.56	0.73
3.68–3.10	1273	0.60	13.2	4.83	1.98	0.63	13.9	5.26	1.11	0.84
3.10–2.68	1748	0.53	16.0	3.81	1.88	0.65	14.1	3.91	0.97	0.86
2.68–2.35	2313	0.43	17.0	2.76	1.56	0.74	14.9	2.56	0.66	0.92
2.35–2.10	2933	0.30	17.0	2.14	1.17	0.83	18.6	1.83	0.33	0.99
Overall	10087	0.49	15.3	3.3	1.84	0.66	15.5	3.5	1.02	0.82

† Anomalous occupancy. ‡ FOM, figure of merit. § $R_{iso} = 100\sum\left||F_{PH}| - |F_P|\right|/\sum|F_P|$. ¶ $K_{emp} = \sum\left||F^{+}_{PH}| - |F^{-}_{PH}|\right|/\sum\left||F_{PH}| - |F_P|\right|$. †† Phasing power, PhP = r.m.s. $(|F_H|/E)$. ‡‡ $R_{Cullis} = 100\sum E/\sum\left||F_{PH}| - |F_P|\right|$. E, residual lack-of-closure error = $\left||F_P + F_H| - |F_{PH}|\right|$. $|F_P|$, native protein structure-factor amplitude; $|F_{PH}|$, heavy-atom derivative structure-factor amplitude; $|F_H|$, heavy-atom structure-factor amplitude.

both derivatives are 6.9 (Cromer & Liberman, 1970). There was no attempt to optimize the anomalous signal in the data collection and reduction, other than using the appropriate wavelength. Firstly, it is impossible in space group $R3$ to record Bijvoet related pairs of reflections simultaneously on the same image. Secondly, no 'local' scaling was performed.

5. Structure solution and refinement

Both mercury and platinum derivatives produced clear isomorphous difference Patterson maps (Figs. 5*a* and 5*b*). The Hg derivative gave also a clearly interpretable anomalous Patterson (Fig. 5*c*) whereas the Pt derivative did not (Fig. 5*d*). In both derivatives only a single site

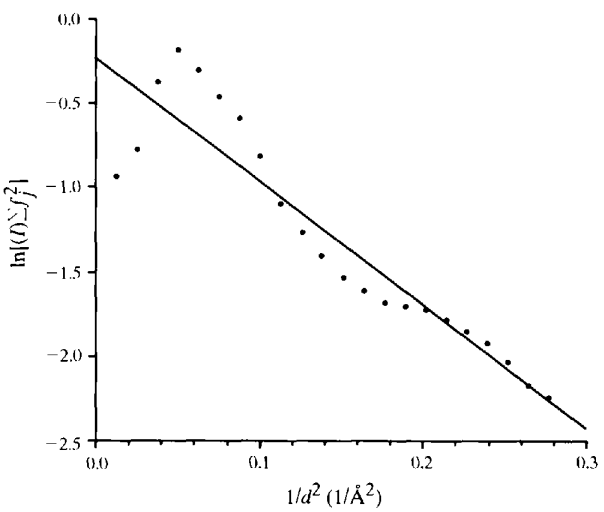

Fig. 4. The Wilson plot for the native data.

was identified. The relative position of the platinum to the Hg atom was found from the difference Fourier synthesis with phases calculated from the two enantiomeric Hg-atom positions. In the synthesis computed with phases from the correct mercury enantiomer there was a single prominent peak corresponding to the maximum in the platinum Patterson synthesis. In the other synthesis, there were no significant peaks. This provided independent confirmation of the choice of the enantiomer and showed the powerful contribution of the anomalous measurements to the phasing.

The occupancy, positional and isotropic temperature parameters for the Hg and Pt atoms were refined using both the isomorphous and anomalous signal. The phases were originally determined with earlier programs from the *CCP4* suite. They have been recently re-estimated using the current version of *MLPHARE* to provide more comprehensive statistics and these are presented in Table 2. In spite of the featureless Pt anomalous Patterson, the contribution of the Pt anomalous signal to the phasing was substantial. The real and anomalous occupancies of Pt refined to a value higher than 1.0 reflecting either imperfect scaling or perhaps the contribution of the coordinating Cl atoms. The reason for the lack of significant features in the Pt anomalous Patterson Harker section compared to the Hg, and to the isomorphous syntheses of both, is not clear. By restricting the resolution and using σ cutoffs, the Pt vectors could be enhanced, but were still much lower than in the other syntheses. The maps shown in Fig. 5 were all calculated with comparable data.

The isomorphous and anomalous contributions of the two heavy-atom derivatives were used to obtain the initial phase set. The overall figure of merit was 0.49 for 10 087 reflections at 2.1 Å resolution. The resulting density map was subjected to solvent flattening

Table 3. *Target values (σ) and standard deviations for the restrained parameters in the refinement*

The weight corresponds to $1/\sigma^2$.

Standard	σ	Deviation	Number of parameters
Distances (Å)			
Bond lengths (1–2 neighbours)	0.020	0.010	1048
Bond angles (1–3 neighbours)	0.040	0.035	1424
Dihedral angles (1–4 neighbours)	0.050	0.042	345
Planar groups	0.020	0.011	177
Chiral volumes (Å^3)	0.200	0.180	166
Non-bonded contacts (Å)			
Single torsion contacts	0.500	0.180	347
Multiple torsion contacts	0.500	0.263	361
Hydrogen bonding contacts	0.500	0.201	109
Torsion angles ( )			
Peptide plane (ω)	5.0	2.5	140
Staggered (c_1)	15.0	16.2	182
Orthonormal (c_2)	20.0	28.2	11

according to Wang (1985) and Leslie (1987). Data were included to a limiting resolution of 2.2 Å and the solvent content assumed to be 50%. The solvent content estimated from the packing density, 2.8 Å^3 Da^{-1} (Matthews, 1968), is 56%. Six cycles of density modification were performed, each followed by combination of the isomorphous replacement and calculated phase distribution. The overall figure of merit after solvent flattening was 0.78.

The resulting map was easily interpretable and the polypeptide chain with the known sequence of amino acids (Lundberg, Thoresson *et al.*, 1983) was built into the electron density using *FRODO* (Jones, 1978). There was no density corresponding to the C-terminal 16 residues, and the initial model consisted of 136 amino acids out of the expected 152. A small number of side-chain atoms were omitted from the model, as there was no clear electron density.

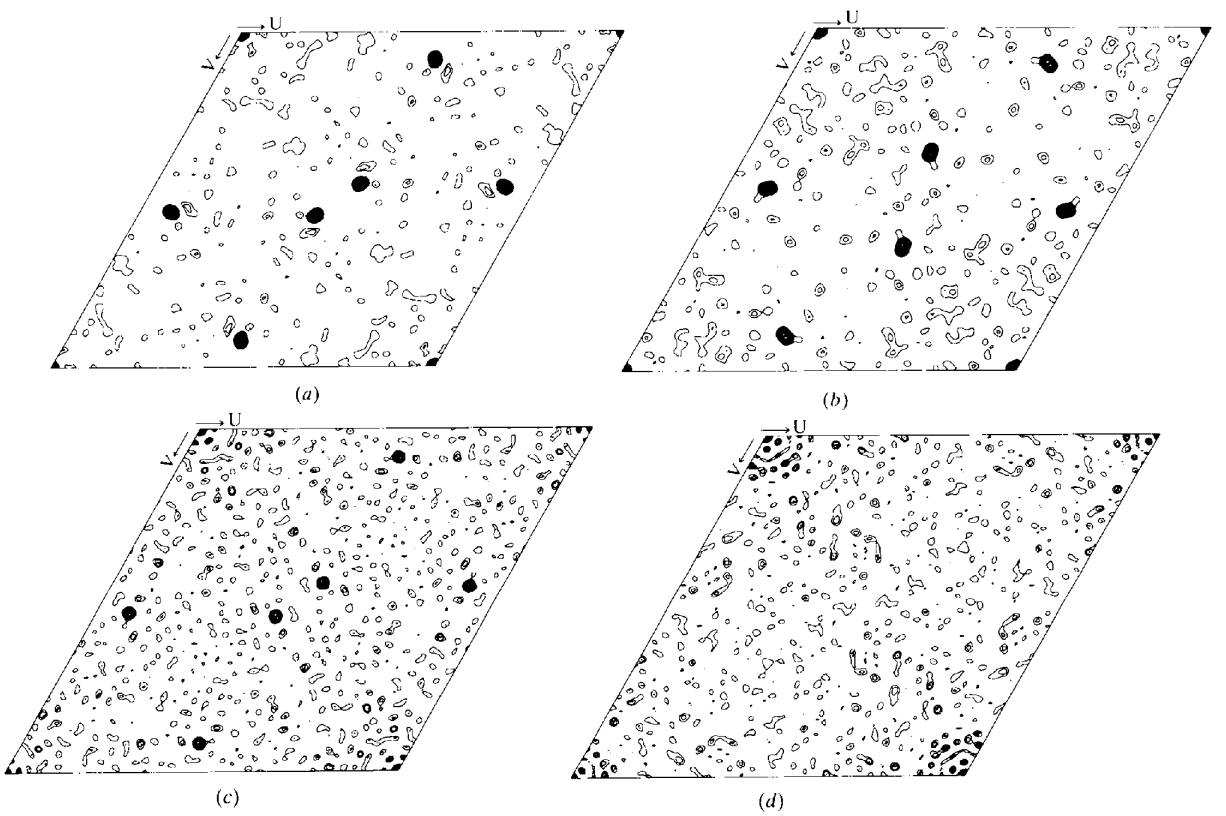

Fig. 5. (*a*) and (*b*) $w = 0$ Harker sections of the isomorphous and (*c*) and (*d*) anomalous difference Patterson syntheses for the mercury and platinum derivative, respectively, contoured at every 1σ.

The initial model was refined for a single round by molecular dynamics with the X-ray terms treated as an additional potential using *X-PLOR* (Brünger, 1988). The full resolution range of X-ray data to 1.9 Å was included with the default parameters set within the program, and the R factor fell from 39.2 to 30.1%.

Refinement was continued using the procedure of Konnert & Hendrickson (1980), covering the full resolution range to 1.9 Å and no σ cutoff on amplitudes. All reflections were used in both refinement and map calculation in this and all subsequent steps. Solvent molecules were identified by the *ARP* procedure (Lamzin & Wilson, 1993). The refinement consisted of several sets each of several cycles, which were interspersed with rebuilding of the model into $3F_o - 2F_c$ and $F_o - F_c$ electron density. 'Omit maps' were used to identify and rebuild poorly defined regions of the model. More recently, the model was refined by the maximum-likelihood procedure using *REFMAC* (Murshudov *et al.*, 1997). R_{free} cross validation was only applied during this final *REFMAC* stage to allow unbiased scaling. The initial refinements were carried out some time ago before R_{free} was in general use. They would be extremely hard to repeat and given the fact that only accepted protocols were used, we do not believe this affects the quality of the model.

The atomic coordinates and native and two derivative sets of amplitudes have been deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).†

6. Accuracy of the model

The final model consisted of 136 residues and 132 waters. All atoms were assigned unit occupancy although some protein atoms have weak electron density and high B factors. Table 3 shows the weights of the stereochemical restraints used during the refinement together with the final deviations of the geometrical parameters from the 'ideal' values. In spite of the fact that over 10% of the residues appear to be totally disordered in the crystal, the final R factor is 13.7% for the 13 597 reflections between 10.0 and 1.9 Å. In the final difference Fourier synthesis, the maximum density was 0.44, the minimum −0.34 and the root-mean-square density 0.07 e Å^{-3}. There is no electron density for the C-terminal residues, 137–152, shown to be present in the crystalline material by biochemical analysis, §3. Presumably they are highly disordered. The Ramachandran plot (Ramakrishnan & Ramachandran, 1965), produced by *PROCHECK* (Laskowski *et al.*, 1993), is shown in Fig. 6. Two residues have (φ, ψ) conformational angles outside the allowed region, Ala60 (70, −35°) and Leu85 (71, −61°) and both are in turns. Their unusual conformation is confirmed unequivocally by clear electron density, Fig. 7, and they are discussed below.

The average B value for main-chain atoms is 15.9 Å^2, for side-chain atoms is 24.7 Å^2 and for the solvent it is 41.3 Å^2. The overall B from the Wilson plot is 15.6 Å^2. The mean B value for main-chain and side-chain atoms is plotted as a function of residue number in Fig. 8(*a*). 12 residues have two or more atoms with B values greater than 50 Å^2. Most lie at the N-terminus or on the surface of the molecule; eight of the 12 are charged residues exposed to solvent.

The solvent accessibility (*DSSP*, Kabsch & Sander, 1983) is shown as a function of residue number in Fig. 8(*b*) for the monomer isolated from the trimer by the total height of the histogram bars. The shaded area corresponds to differences between the accessibility of residues buried in the trimer and that in the free monomer. The black parts illustrate the area buried by crystal contacts between trimers, so that the lowest white parts of the histogram bars relate to the solvent accessibility of dUTPase in the crystal. The white part is most relevant to the flexible regions in Fig. 8(*a*). The black parts allow the residues involved in intertrimer contacts in the crystal to be identified, and the shaded parts show those residues involved in making up the trimer from monomers. In the isolated monomer the total accessible surface area is 9211 Å^2, which is reduced to 6027 Å^2 per monomer in the trimer, and to 5129 Å^2 in the crystal. More than one third of the monomer surface is covered by its neighbours in the trimer and an additional 10% by the crystal contacts.

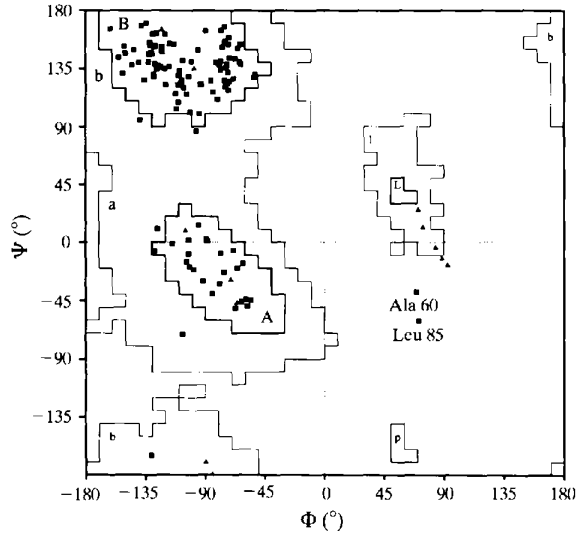

Fig. 6. The Ramachandran plot. The allowed regions for (φ, ψ) angles are marked for non-glycine residues in secondary-structural elements: A, L and B for right- and left-handed helix and β-sheet core regions, respectively. Triangles correspond to glycine residues.

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1DUP).

7. Accuracy of the experimental phases

Initial phases were calculated using the isomorphous and anomalous contributions of two heavy-atom derivatives and were improved by solvent flattening. In the light of the ease of interpretation of the resulting electron-density map, we compared the phases calculated from the refined model with those from experimental MIRAS, Fig. 9. The average phase change between the MIRAS and final refined phases was only 40° and at low resolution this value is about 20°.

8. Description of the model

8.1. *The monomer*

8.1.1. *General shape*. The general shape and topology of the monomer are irregular, Fig. 10. Disregarding the extended C-terminal portion which wraps around the neighbouring subunit, the monomer can be roughly described as a pyramid with a height of 41 Å and a width at half height of 15 Å. The intersubunit interfaces in the trimer are essentially planar and largely hydrophobic.

8.1.2. *Main-chain hydrogen bonds*. The main-chain hydrogen-bond pattern of the monomer is shown in Fig. 11, which also shows the secondary-structural elements and amino-acid sequence.

(*a*) *β-Strands*. The monomer consists predominantly of β-pleated sheets, Fig. 10, with only a single short α-helix. There are 11 β-strands, nine forming a barrel built from two highly twisted sheets. The strands in the barrel have been labelled β1 to β8, strands β1 to β7 coming from the same monomer. The last strand in the barrel, β8′, comprising residues Gln127 to Val133, is part of the neighbouring subunit, thus constituting an important link between monomers. β5 consists of two short consecutive strands, β5a and 5b, separated by a type I turn having one amino acid, Leu85, with a strained backbone conformation, Fig. 7(*b*). With one exception, all the strands are antiparallel, the exception being β8′ which forms the extended 'arm' in the C-terminal

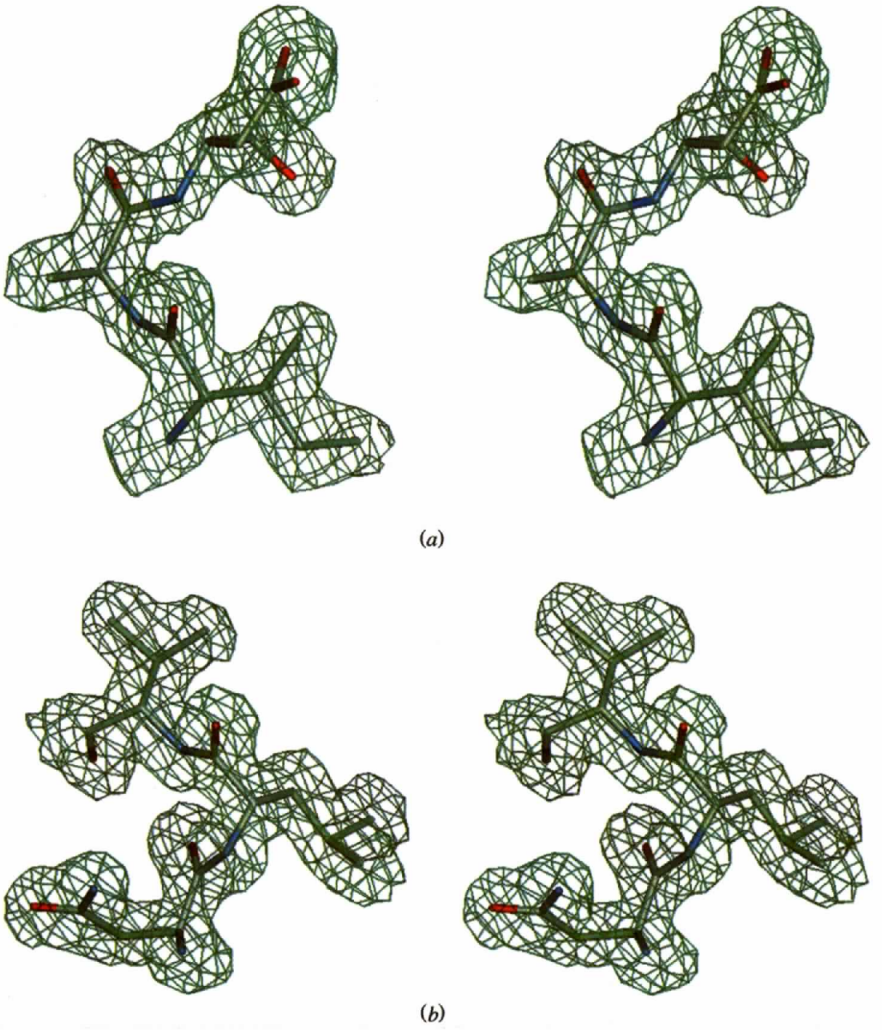

Fig. 7. The final $(2F_o - F_c)$ electron-density map for (*a*) residues around Ala60 (φ/ψ; 70/−35°) and (*b*) turn 5 with Leu85 (φ/ψ; 71/−61°) contoured at 1σ above the mean.

portion of the monomer, binding to $\beta 1$ of the neighbouring subunit.

The barrel structure has a 'jelly-roll' topology (Fig. 2*a* in Cedergren-Zeppezauer *et al.*, 1992). Sheet 1 comprises strands $\beta 8'$, $\beta 1$, $\beta 3$, $\beta 6$ and $\beta 5a$ arranged in four ladders, and sheet 2 of $\beta 2$, $\beta 7$, $\beta 4$ and $\beta 5b$ in the form of three ladders. The shortest ladder has three hydrogen-bond bridges and the longest has seven, Fig. 11. Besides the barrel structure, there are two short β-strands, βI (Val41 to Leu43) and βII (Phe109 to Ile111) linked together by four hydrogen bonds.

(*b*) *α-Helix*. The monomer contains only a single, short helix of one and a half turns comprising residues Ser71 to His79 and carrying a cluster of basic side chains (His-Lys-His). The torsion angles (φ, ψ) are close to ideal for α_R in the first half of the helix but are distorted at the end. The helix is situated at the surface of the trimer with its side chains solvated and has only one hydrophobic residue. The carbonyl O atoms of residues 71–75 make hydrogen bonds corresponding to regular i; $i + 4$ interactions and Leu74 O has a bifurcated bond. The remaining amide N atoms and the carbonyl O atoms of residues 76 and 78 make hydrogen-bond interactions with six water molecules.

(*c*) *Turns and loops*. The monomer contains a number of loops as listed in Table 4 and the majority of them can be classified into standard types of β-bend, predominantly types I and II. Several of the loops lie in the region between residues 9 and 31. Residues 25–28 form a non-classifiable loop structure of the 'subunit contact loop 1', L1. Turn 1 makes a main-chain hydrogen bond with strand $\beta 3$ at a point where that strand changes

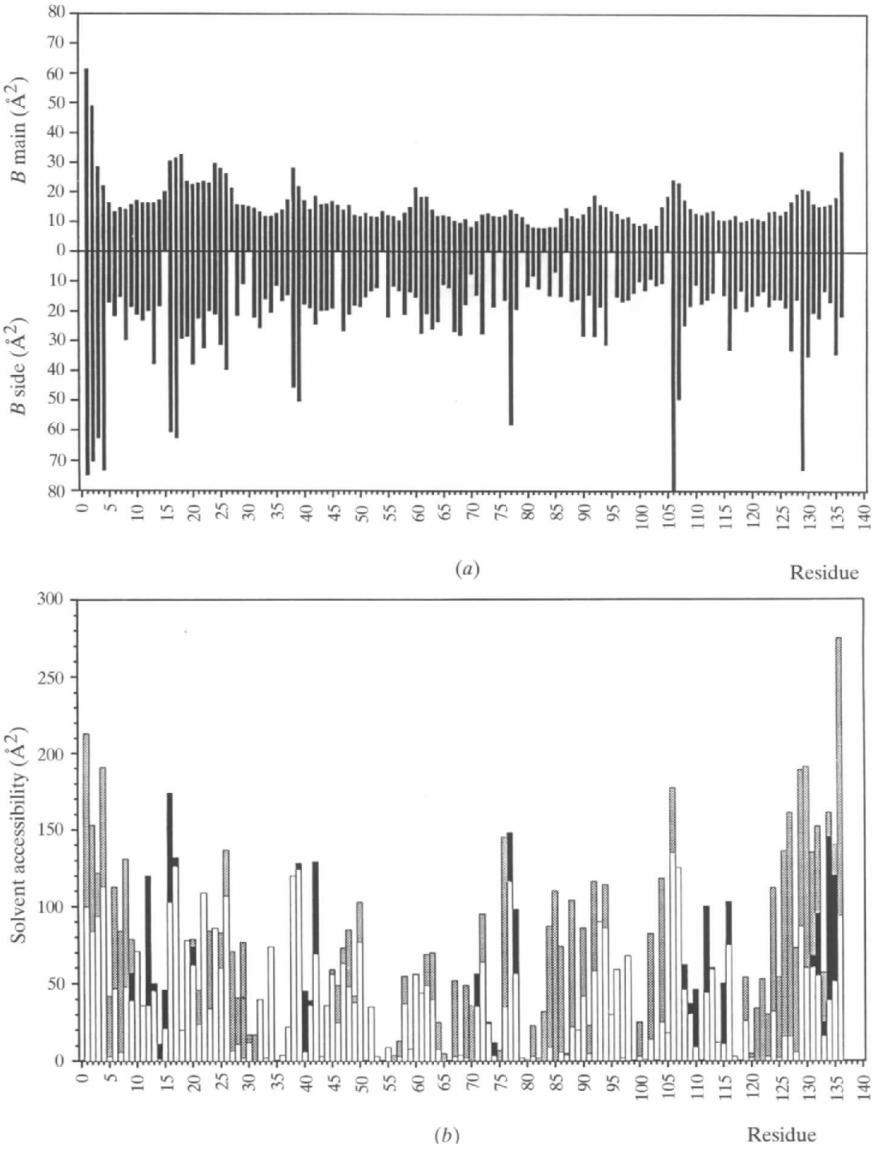

Fig. 8. As a function of residue number (*a*) the average *B* value for the main- and side-chain atoms and (*b*) histograms of the accessible surface area for dUTPase: total height of the bars, the isolated monomer; shaded parts, the area shielded by contacts within the trimer; black parts, intertrimer crystal contacts.

Table 4. *Turns in the dUTPase chain*

φ, ψ angles are given for residues $i+1$ and $i+2$ in classified reverse turns. The standard values of φ, ψ ( ) for turns are as follows: type I: −60, −30 and −90, 0, type II: −60, 120 and 80, 0, type III: −60, −30 and −60, −30. γ-Turn defined according to Baker & Hubbard (1984).

Turn	Type	Residue numbers	Sequence	Hydrogen bond Atom	Distance (Å)	φ_{i+1}, ψ_{i+1} (°)	φ_{i+2}, ψ_{i+2} ( )
T1	III	11–14	DPRV	O11–N14	3.1	−74, −24	−57, −18
T2	II	13–16	RVGKE	O13–N16	3.0	−67, 128	84, −6
				O13–N17	2.7		
		15–18	GKEF	O18–N15	2.6		
	γ-T	23–25	YAT	O23–N25	3.2		
L1		25–28	TSGS	OG28–N25	2.9		
				OG1 25–N28	3.2		
T3	II	44–47	APGN	O44–N47	3.0	−52, 126	92, −16
T4	I	61–64	DPSL	O61–N64	3.1	−71, −6	−93, 2
T5	I	82–85	LGNL	O82–N85	3.2	−72, −31	−95, 14
	γ-T	84–86	NLV	O84–N86	2.7		
T6	I	90–93	DSDY	O90–N93	3.2	−67, −21	−103, 0
L2		105–108	GQDS	OG1 103–N105	2.9		
				ND2 103–O107			
T7	II	112–115	EPGE	O112–N115	2.9	−54, 130	79, 0

direction, and together with the adjacent turn 2, ten water molecules surround the backbone atoms. Turn 3 seems to have a stabilizing effect on the conformation of the connecting loop between strand β6 and βII by making main-chain to main-chain hydrogen bonds to both the amide N atom and the carbonyl O atom of residue 103. Turn 3 is in a water-rich region situated at a subunit–subunit interface which is also the case for turn 4. Turn 4 is preceded by Ala60, which has a strained backbone conformation, and this turn has main-chain hydrogen bond contacts with active-site residues (Larsson, Svensson *et al.*, 1996) in addition to a loop connecting strand β7 and the subunit connecting arm, β8. Turn 5, which divides strand β5 into two halves, contains Leu85, with unusual Ramachandran angles, as first residue and contributes to the formation of the active site together with turn 6 and strand β6. Thus, the reverse turns 4, 5 and 6 form part of the active site. Turns 3 and 7 stabilize the conformation of two different six-residue loops connecting strands. All turns have some water molecules associated to their backbone atoms, mainly to the carbonyl O atoms, turn 5 being least accessible to solvent.

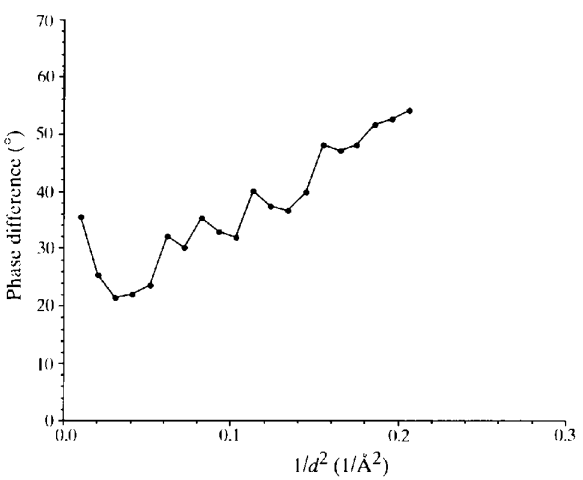

Fig. 9. The average difference between the initial MIRAS and final refined phases as a function of resolution.

8.1.3. *Proline residues*. There are nine prolines all in the *trans* conformation. Four are the second residues in turns, Table 4. Three are located in the loop region 9–31, giving this part of the protein a characteristic shape and two are at the ends of strands β4 and β7. After Pro124 of β7 (conserved in several trimeric dUTPases) the chain changes direction to cross a subunit contact surface.

Pro52 is located in the middle of strand β3 contributing to the formation of a sharp kink. Pro21, 45, 62 and 70 are conserved (McIntosh *et al.*, 1994) among species as remotely related as *E. coli*, *Candida albicans* (a fungal pathogen), yeast, tomato and human. Their positions in the three-dimensional structure indicate a similarity in the subunit topology among the different enzymes since the conserved prolines are located at reverse turns connecting strands or adjacent to highly conserved sequence motifs (McGeoch, 1990). For the human enzyme and that from feline immunodeficiency virus (FIV), similarity of fold to the *E. coli* enzyme has been demonstrated (Mol *et al.*, 1996; Sridhar Prasad *et al.*, 1996).

8.1.4. *Hydrophobic residues*. Among the hydrophobic residues, 30 form an inner core packed inside and on top of the β-barrel surrounded by all strands or constitute the hydrophobic contacts between β-sheets around the threefold axis. One is in the α-helix and 15 are situated in loops and bends. Phe146 is in the highly conserved C-terminal sequence, invisible in the X-ray structure.

The residues in the core can be divided into three clusters. Inside the β-barrel, side chains from sheets 1 and 2 form two clusters. A third cluster is formed by

residues from strands βI (Val41, Leu43) and βII (Phe109, Ile111), Leu74 in the α-helix and the 'top' part of the β-sandwich (Leu37, Leu50, Val51, Ile80, Ile99, Val101, Ile117) which together make up a hydrophobic surface separating the closely packed side chains of the barrel.

8.1.5. *Cavities and internal waters*. There is an enclosed cavity in cluster 3, which is lined exclusively by hydrophobic side chains and contains no water. The two hydrophobic clusters in the core of the barrel are separated by a second cavity, part of a depression in the inter-subunit region, in which two water molecules are trapped making hydrogen bonds to the hydroxyl group of Tyr23. There are three phenylalanine residues (Phe122, Phe130 and Phe136) close by originating from two subunits. Val14 and Phe18 in the N-terminal loop region are near the surface of the monomer. Ile9, Leu10, Leu20, Leu50, Phe130, Leu132, Val133 are localized close to a water-rich region thus shielding the hydrophobic interior of the barrel from solvent.

8.1.6. *Charged groups*. There are 28 potentially charged residues visible in the maps of the monomer (five Lys, five Arg, ten Asp, five Glu and three His) plus one N-terminal NH_3^+. Another seven charges lie in the non-visible C-terminal part. Most charged side chains interact with water. There are three intramolecular ion pairs but three acidic side chains stick out of the surface lacking bonding partners within the protein. The aliphatic part of Glu115 covers the hydrophobic cluster 3, whereas the carboxyl group interacts with the main chain at position 112 in addition to Glu134 from a neighbouring trimer. This interaction requires that one of the two glutamates is protonated which is plausible since the pH of the crystals is around 4.

A positively charged cluster is formed by the helix and three arginine residues (34, 71 and 116) two of which are conserved. All these charged residues have well defined electron densities, the guanidinium groups of Arg34 and Arg71 being anchored through hydrogen bonds to the main-chain O atoms. The three arginines are in line, outstretched and parallel to one another at the surface of the substrate binding site. The diphosphate moiety of dUDP is directed towards these arginine side chains in the dUDP complex (Larsson, Svensson *et al.*, 1996).

A second charged area on the surface consists of residues Asp11, Arg13, Lys16 and Glu17. Asp11 and Arg13 form a salt bridge and Asp11 interacts with the backbone of turn 1. Lys16 and Glu17 are too far apart to make an ion pair and both are engaged in polar interactions with a neighbouring trimer. Arg104 is one of the few charged residues that participate in subunit contacts around the threefold axis and is partially buried making an aliphatic–aromatic interaction with Trp102 of a second subunit. The guanidinium group of Arg104 is hydrogen bonded to main-chain atoms of two subunits and only one atom is exposed to a water-rich region ('the pond') at the top of the molecule.

8.2. *Monomer–monomer interface*

8.2.1. *General features*. The interface between subunits can be divided into three interacting surfaces. The first is along the threefold axis where mostly hydrophobic residues are stacked. Second, there are extensive interactions between the extended part of the C-terminal 'arm' (residues 127–136) and a neighbouring subunit across the triangular 'bottom' face of the trimer. Finally, an interface between monomers is built up by loops, where the monomer–monomer contacts involve hydrogen bonds between main-chain as well as side-chain atoms. Several water molecules form an integral part of the subunit interface. There are 13 main-chain water hydrogen bonds and eight side-chain water interactions between monomers.

8.2.2. *Trimer interactions along the threefold axis*. The interactions along the threefold axis involve about 20 residues, and are totally hydrophobic in the central part of the molecule, Fig. 12. The length of the interface is about 40 Å, almost the total height of the monomer. At the 'top' of the trimer, three glutamine side chains form the rim of a wide 'pond' containing water. The walls of

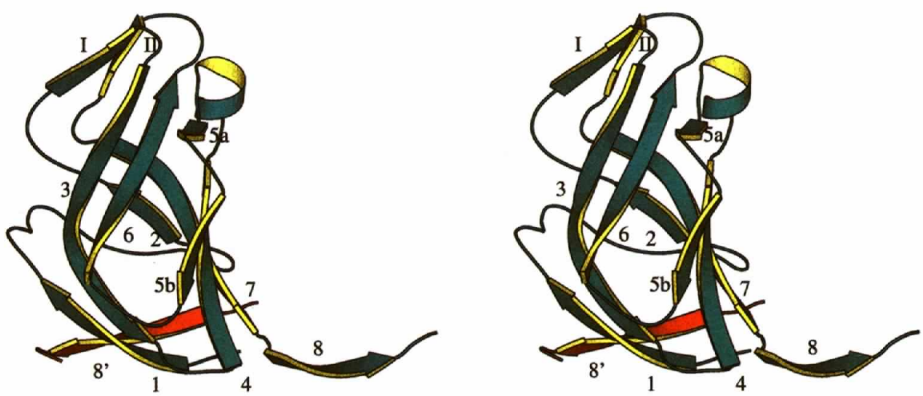

Fig. 10. Stereo ribbon diagram showing the fold of the dUTPase monomer and the strand numbering (program used: *MOLSCRIPT*, Kraulis, 1991).

the container, holding at least 20 water molecules, are made up by prolines and glycines. Further down the threefold axis, three tryptophan and three arginine side chains block access of water to the remaining part of the threefold interface. The three Trp102 side chains lie in van der Waals contact with one another around the threefold axis. From here on over a length of about 25 Å, hydrophobic residues form rings of closely packed side chains around the axis. Occasionally, there is an empty space between hydrophobic layers with a diameter of 10–13 Å. Thus, the contacts are tightened and widened periodically. On the bottom face of the trimer, at the point where crossing arms pass the surface, the distances between symmetry-related groups around the axis are 13 Å. Strands $\beta 4$, $\beta 5$ and $\beta 7$, Figs. 10 and 11, contribute most residues to the hydrophobic interactions around the axis, whereas residues from the loops 44–47 (connecting strands βI with sheet 2 of the barrel) and 104–108 (denoted subunit contact loop L2 in Table 4) form the polar contacts at the 'top' of the trimer.

8.2.3. *The crossing 'arm'*. The subunit contact involves a C-terminal crossing 'arm'. The main chain from strand $\beta 7$ of subunit *A* (passing close to the threefold axis) crosses above strand $\beta 7$ of subunit *B* and is covered by the N-terminus *B*, which in turn is stabilized by an interaction with Ser63 of subunit *A*. Strand *A* $\beta 8$ lies

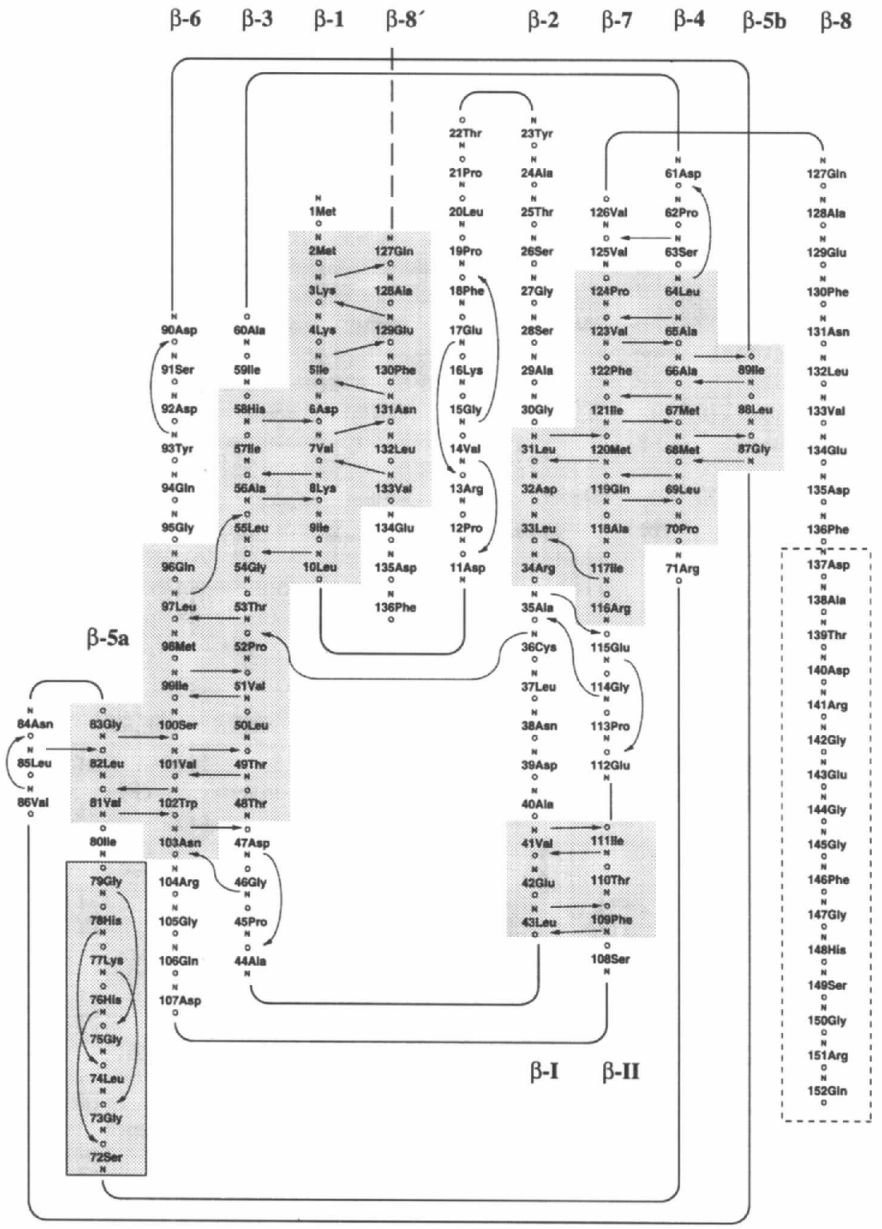

Fig. 11. Main-chain hydrogen-bonding scheme for the dUTPase subunit, outlining the secondary-structural elements, including strand $\beta 8'$ from a neighbouring subunit. The C-terminal region not seen in the maps is indicated.

parallel to strand $\beta 1$ of subunit *B* forming six hydrogen bonds, Fig. 11. The hydrophobic residues Ala128, Phe130, Leu132 and Val133 in the crossing 'arm' are partly embedded in a hydrophobic cluster inside the β-barrel of its neighbour. The four N-terminal residues of subunit *B* also contribute to the highly intertwined subunit contacts by bending over and covering the crossing 'arm' (*cf.* Fig. 3*a* in Cedergren-Zeppezauer *et al.*, 1992). Two acidic residues (Glu129 and Asp139) from the 'arm' and two basic residues (Lys4 and Lys8) from the other subunit also make contacts. The last three visible residues of the 'arm' pass over the edge of the *B* subunit approaching conserved loops which form the active-site surface cavity (Larsson, Svensson *et al.*, 1996). We expect the non-visible flexible part of the sequence (137–152) to participate in binding substrate and, thus, all three subunits to build up the active site in analogy with what is observed for human dUTPase–nucleotide complexes (Mol *et al.*, 1996).

8.2.4. *The loop contact region.* In the loop contact area, different parts of the sequence are brought together. Residues 23–30 (subunit contact loop L1), residues 67–79 (including the α-helix) and part of $\beta 4$ from one subunit make contacts with an adjacent subunit through residues 58–66 and loop 87–95 containing turn 6. Most of the hydrophobic residues in the loop region are directed into the barrel clusters. The tightest polar interactions between loops connecting subunits at this interface are those between residues 25–29 and 90–92. The polypeptide segment 14–32, including the subunit interaction loop L1, is particularly rich in bound waters (about 25 in total). An important part of the stabilization of the loop originates from binding of water in the interior of the loop facing strand $\beta 2$ of the β-barrel. Furthermore, main-chain–side-chain interactions *via* the threonine and serine hydroxyl groups within the loop (conserved side chains) tighten the structure.

8.3. *Crystal packing of trimers*

The trimer has the shape of a triangular prism, 45 Å high and 60 Å across the triangular face. The packing arrangement in the crystal is analogous to cubic close packing of spheres. One trimer interacts with three trimers from the layer above and three from the layer below. The packing is loose with 25 Å gaps between trimers along the *z* axis. Since the full-length protein is present in the crystals, the disordered C-termini presumably lie in the apparently empty space in this region. There are four surface patches that interact with neighbouring trimers. The residues involved can be identified from the histogram in Fig. 8(*b*) (see black bars).

9. Overall solvent structure

There are 132 water molecules per subunit, selected according to criteria set in the *ARP* program: a water molecule was assigned when the difference density was at the 3σ level, the *B* factor did not exceed 60 Å^2 and hydrogen-bond distances fell within the range 2.3–3.3 Å. There are few water molecules buried within the monomer with four protein ligands. About 50% of the waters make four or three hydrogen-bond contacts. There are six waters with three protein ligands. Waters making two hydrogen-bond interactions (one with a protein atom and the other to a water molecule) are most common. Among the water molecules displaying only a single interaction there is a tendency for association to a main-chain atom (11 cases) rather than to a side-chain atom (three) or a single water (five). Eight water molecules bind to the polypeptide backbone.

Several waters are bound exclusively to other waters, and these are often clustered, *e.g.* forming a 'pond' in the cavity on 'top' of the trimer above the hydrophobic cluster along the threefold axis. Another extensive cluster of 25 waters is located in the active site, some of which remain bound deeply inside the protein in the presence of dUDP (Larsson, Svensson *et al.*, 1996).

10. Relation to other dUTPases and the active site

Two other dUTPase structures are known, that of the feline immunodeficiency virus enzyme (FIV, Sridhar

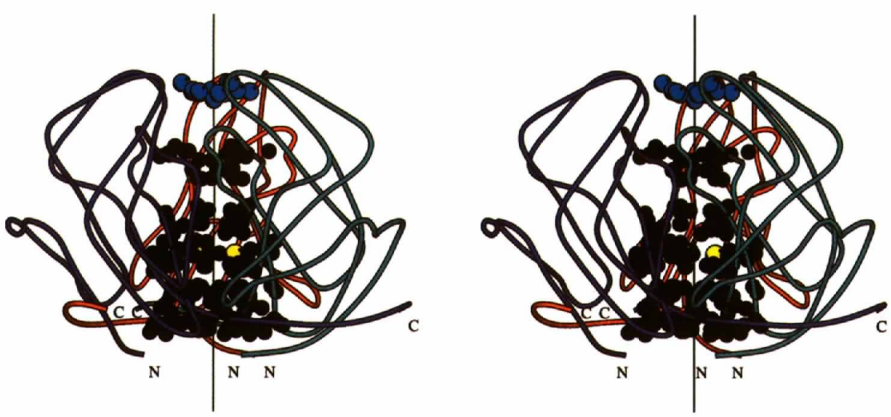

Fig. 12. The trimer of *E. coli* dUTPase with the hydrophobic core residues shown in black. The 'pond' of waters on top of the trimer is indicated in blue. The threefold axis is marked.

Prasad *et al.*, 1996) and the human enzyme in the native form and in complexes with dUMP, dUDP and dUTP (Mol *et al.*, 1996). In addition, the structure of the complex of the *E. coli* enzyme with the inhibitor dUDP has been published (Larsson, Svensson *et al.*, 1996). The position of the dUDP on the *E. coli* trimer is shown in Fig. 13 to indicate the active site. The similarities and differences of these structures to that of the *E. coli* enzyme is discussed in those papers and, therefore, only a brief account will be given here.

The overall fold and secondary structures of all three enzymes are similar and all form tight trimers with the long C-terminal arm of one subunit wrapped around the neighbouring subunits with its end totally disordered in the native structure. In spite of the fact that the sequence identity is only around 25%, there are five highly conserved sequence motifs located around the active site.

However, there are also marked differences between the three. The human structure was solved by molecular replacement using the preliminary model of the *E. coli* enzyme but it was impossible to solve the FIV enzyme structure the same way, in spite of the *E. coli* and FIV trimers forming a strictly hydrophobic core around a threefold axis whereas the core of the trimer in the human enzyme has a hydrophilic character.

The complex of *E. coli* enzyme with dUDP allowed the unequivocal identification of the active site in the depression at the interface between two subunits, which confirms our earlier (Cedergren-Zeppezauer *et al.*, 1992) proposal. Only in the complexes of the human enzyme (Mol *et al.*, 1996) could the extension of the C-terminal arm be located in the electron density. It contributes to the active site formed between the two other subunits of the trimer. Therefore, dUTPases have three active sites with all three subunits taking part in each of them.

The uracil moiety of the nucleotide sits in the pocket, which is formed by a hairpin composed of strands $\beta5$ and $\beta6$, and whose shape is highly complementary to the base and discriminates again other bases. In addition the ribose is placed in a highly specific environment precluding the presence of a 2′O atom, strongly favouring the deoxyribonucleotide against the ribonucleotide. All residues crucial for nucleotide binding and active-site specificity are located in the five highly conserved sequence motifs within the diverse dUTPases (McGeoch, 1990).

dUTPase requires Mg^{2+} or other divalent metal ions for activity, but in the known structures no metal ions were observed in the active site in spite of being present in the crystallization medium. Thus, the location and the role of metal ions in the catalysis of dUTP remains to be established.

11. Conclusions

The structure of *E. coli* dUTPase was determined using two single-site heavy-atom derivatives. The experimental MIRAS phases with optimized wavelength for the anomalous scattering measurements, were of high quality and close to those of the final refined model. This resulted in an initial electron-density map which was easy to interpret.

Although dUTPase binds a nucleotide triphosphate substrate, the overall protein fold does not resemble the classical Rossmann fold and thus represents a new folding pattern among the diverse collection of proteins which bind NTP's or NAD. However, the C-terminal sequence 137-DATDRGEGGFGHSGRQ-152, highly conserved among all dUTPases, is similar to phosphate-binding loops in ATP- and GTP-binding proteins (Saraste *et al.*, 1990). It is only in the human dUTP structure in complex with dUMP, dUDP and dUTP that this flexible C-terminus has a defined conformation and is found to enclose the nucleotides within the active-site cavity (Mol *et al.*, 1996). To date, the bacterial enzyme is kinetically the best characterized dUTPase and based on such data (Larsson, Nyman *et al.*, 1996) and the structural evidence it is possible that the flexible 'arm' in the *E. coli* dUTPase, in contrast to the human enzyme, is only ordered after binding of the Mg^{2+}–substrate complex.

Fig. 13. Ribbon diagram of the dUTPase trimer with each monomer colour coded, viewed down the threefold axis. The position of the dUDP in the complex (Larsson, Svensson *et al.*, 1996) is shown.

In a review describing strained conformations in enzymes, Herzberg & Moult (1991) point out that such arrangements are often found in or in the neighbourhood of catalytic sites or other ligand-binding areas. Leu85 in *E. coli* dUTPase and probably Ala75 in the human enzyme (Mol *et al.*, 1996) have conformational angles in the disallowed region of the Ramachandran plot and they are indeed located in the uracil binding subsite of the tyrosine hairpin loop forming the active centre.

Sequences are known for many dUTPases and classified into trimeric and monomeric enzymes. A sequence alignment of trimeric enzymes has recently been reported (Sridhar Prasad *et al.*, 1996), where the *E. coli* enzyme and other bacterial dUTPases constitute a separate group with an insertion of three residues around positions 83–86 (*E. coli* numbering). An additional member of this group is represented by a recently identified gene in bacteriophage T5 presumably encoding a dUTPase (Kaliman, 1996). One of the residues inserted is Asn84 which in the *E. coli* enzyme contributes to substrate binding by hydrogen bonding of its side chain to O4 in the uracil ring. The next amino acid in the insertion is hydrophobic, Leu85 in *E. coli* and has a strained backbone conformation, Fig. 7(*b*). The human enzyme has only an alanine inserted in the corresponding region (with a strained conformation) and in FIV dUTPase there is no insertion. Thus, the detailed structural features of the substrate binding sites for the bacterial, human and the viral dUTPases differ, which may reflect differences in the catalytic properties of the enzymes. Unfortunately, neither the human nor the FIV enzymes have been characterized kinetically. However, the *E. coli* dUTPase was found to have a higher specificity towards dUTP (Larsson, Nyman *et al.*, 1996) compared with two viral dUTPases (Björnberg & Nyman, 1996).

This work was supported by the Swedish Natural Science Council, the Swedish Cancer Society, Magnus Bergwalls stiftelse, Craafords stiftelse and Kungliga Fysiografiska Sällskapet, Lund. We also wish to thank Dr Anna-Maria Rosengren, Department of Biochemistry, Lund University, for kindly providing the FPLC facility, and Mirka Dauter for skilful assistance.

References

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

Bergman, A. C., Björnberg, O., Nord, J., Nyman, P. O. & Rosengren, A. M. (1994). *Virology*, **204**, 420–424.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Mayer, E. F. Jr, Bryce, M. D., Rodgers, J. R., Kennard, O., Simanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bertani, L. E., Häggmark, A. & Reichard, P. J. (1961). *J. Biol. Chem.* **236**, 67–68.

Bertani, L. E., Häggmark, A. & Reichard, P. J. (1963). *J. Biol. Chem.* **238**, 3407–3413.

Björnberg, O., Bergman, A. C., Rosengren, A. M., Persson, R., Lehman, I. R. & Nyman, P. O. (1993). *Protein Expres. Purif.* **4**, 149–159.

Björnberg, O. & Nyman, P. O. (1996). *J. Gen. Virol.* **77**, 3107–3111.

Broyles, S. S. (1993). *Virology*, **195**, 863–865.

Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

Camacho, A., Arrebola, R., Péna-Diaz, J., Ruiz-Perés, L. M. & González-Pacanowska, D. (1998). *Biochem. J.* In the press.

Cedergren-Zeppezauer, E. S., Larsson, G., Hoffmann, I., Törnroos, K. W., Al-Karadaghi, S. & Nyman, P. O. (1988). *Proteins Struct. Funct. Genet.* **4**, 71–75.

Cedergren-Zeppezauer, E. S., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **355**, 740–743.

Climie, S., Lutz, T., Radul, J., Sumner-Smith, M., Vandenberg, E. & McIntosh, E. (1994). *Protein Expres. Purif.* **5**, 254–258.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cromer, D. T. & Liberman, D. (1970). *J. Chem. Phys.* **53**, 1891–1898.

Elder, J. H., Lerner, D. L., Hasselkus-Light, C. S., Fontenot, D. J., Hunter, E., Luciw, P. A., Montelaro, R. C. & Phillips, T. R. (1992). *J. Virol.* **66**, 1791–1794.

El-Hajj, H. H., Zhang, H. & Weiss, B. (1988). *J. Bacteriol.* **170**, 1069–1075.

Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. (1993). *EMBO J.* **12**, 4425–4431.

Giroir, L. E. & Deutsch, W. A. (1987). *J. Biol. Chem.* **262**, 130–134.

Herzberg, O. & Moult, J. (1991). *Proteins Struct. Funct. Genet.* **11**, 223–229.

Hoffmann, I. (1988). Doctoral thesis. University of Saarbrücken, Germany.

Hoffmann, I., Widström, J., Zeppezauer, M. & Nyman, P. O. (1987). *Eur. J. Biochem.* **164**, 45–51.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kaliman, A. V. (1996). *DNA Seq. J. Seq. Mapp.* **6**, 347–350.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

Konrad, E. B. (1977). *J. Bacteriol.* **130**, 167–172.

Kornberg, A. & Baker, T. (1991). In *DNA Replication*. New York: Freeman Press.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Larsson, G., Nyman, P. O. & Kvassman, J. (1996). *J. Biol. Chem.* **271**, 24010–24016.

Larsson, G., Svensson, L. A. & Nyman, P.-O. (1996). *Nature Struct. Biol.* **3**, 532–538.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Leslie, A. G. W. (1987). *Computational Aspects of Protein Crystal Analysis*, edited by J. R. Helliwell, P. A. Machin & M. Z. Papiz, pp. 35–90. Warrington: Daresbury Laboratory.

Leslie, A. G. W., Brick, P. & Wonacott, A. J. (1986). *CCP4 Newslett.* **18**, 33–39.

Lundberg, L. G., Karlström, O. H., Nyman, P. O. & Neuhard, J. (1983). *Gene*, **22**, 115–126.

Lundberg, L. G., Thoresson, H.-O., Karlström, O. H. & Nyman, P. O. (1983). *EMBO J.* **2**, 967–971.

McGeoch, D. J. (1990). *Nucleic Acids Res.* **18**, 4105–4110.

McIntosh, E. M., Ager, D. D., Gadsden, M. H. & Haynes, R. H. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 8020–8024.

McIntosh, M. E., Looser, J., Haynes, R. H. & Pearlman, R. E. (1994). *Curr. Genet.* **24**, 415–421.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Mol, C. D., Harris, J. M., McIntosh, E. M. & Tainer, J. A. (1996). *Structure*, **4**, 1077–1092.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.

Pri-Hadash, A., Hareven, D. & Lifschitz, E. (1992). *Plant Cell*, **4**, 149–159.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Roseman, N. A., Evans, R. K., Mayer, E. L., Rossi, M. A. & Slabaugh, M. B. (1996). *J. Biol. Chem.* **271**, 23506–23511.

Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990). *Trends Biochem. Sci.* **15**, 430–434.

Shlomai, J. & Kornberg, A. (1978). *J. Biol. Chem.* **253**, 3305–3312.

Sridhar Prasad, G., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H. & Stout, C. D. (1996). *Protein Sci.* **5**, 2429–2437.

Taylor, A. F., Siliciano, P. G. & Weiss, B. (1980). *Gene*, **9**, 321–336.

Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S. & Weiss, B. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 154–157.

Wagaman, P. C., Hasselkus-Light, C. S., Henson, M., Lerner, D. L., Phillips, T. R. & Elder, J. H. (1993). *Virology*, **196**, 451–457.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.