Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Atomic resolution structure of *Escherichia coli* dUTPase determined *ab initio*

A. González,[a]*† G. Larsson,[b] R. Persson[b]‡ and E. Cedergren-Zeppezauer[b]

[a] European Molecular Biology Laboratory (EMBL), Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany, and [b]Department of Biochemistry, Center for Chemistry and Chemical Engineering, Lund University, PO Box 124, 22100 Lund, Sweden

† Current address: Stanford Synchrotron Radiation Laboratory (SSRL), 2575 Sand Hill Road, MS 99, Menlo Park, CA 94025, USA.
‡ Current address: York Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, England.

Correspondence e-mail: ana@slac.stanford.edu

Received 20 November 2000
Accepted 8 March 2001

PDB References: *E. coli* dUTPase, 1eu5; *E. coli* dUTPase, Hg derivative, 1euw.

Cryocooled crystals of a mercury complex of *Escherichia coli* dUTPase diffract to atomic resolution. Data to 1.05 Å resolution were collected from a derivative crystal and the structure model was derived from a Fourier map with phases calculated from the coordinates of the Hg atom (one site per subunit of the trimeric enzyme) using the program *ARP/wARP*. After refinement with anisotropic temperature factors a highly accurate model of the bacterial dUTPase was obtained. Data to 1.45 Å from a native crystal were also collected and the 100 K structures were compared. Inspection of the refined models reveals that a large part of the dUTPase remains rather mobile upon freezing, with 14% of the main chain being totally disordered and with numerous side chains containing disordered atoms in multiple discrete conformations. A large number of those residues surround the active-site cavity. Two glycerol molecules (the cryosolvent) occupy the deoxyribose-binding site. Comparison between the native enzyme and the mercury complex shows that the active site is not adversely affected by the binding of mercury. An unexpected effect seems to be a stabilization of the crystal lattice by means of long-range interactions, making derivatization a potentially useful tool for further studies of inhibitor–substrate-analogue complexes of this protein at very high resolution.

1. Introduction

The essential enzyme 2′-deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase) catalyses the hydrolysis of dUTP to dUMP and pyrophosphate (Bertani *et al.*, 1961; Greenberg & Somerville, 1962). It maintains a central role in nucleotide metabolism and in the regulation of DNA synthesis (Kornberg & Baker, 1991). Various techniques have previously been used to characterize different members of the large dUTPase family, including molecular and structural biology approaches as well as physico-chemical studies, inhibitor screenings and cell biology. Studies have been initiated within a European Consortium for dUTPase research to explore the possibility to utilize dUTPase as a potential drug target. Viral, parasitic and bacterial enzymes are being investigated and compared with the human form of dUTPase and species-specific potent inhibitors are being developed.

Three different subunit organizations of dUTPases have been found. dUTPases are either monomers, dimers or trimers (Persson, 1998) and in the majority of them five conserved sequence motifs (McGeoch, 1990) have been identified. Monomeric and trimeric dUTPases are highly specific for dUTP (Larsson, Nyman *et al.*, 1996; Bergman *et al.*, 1998). The protozoan dimeric enzyme from *Leishmania major* can, in contrast, also utilize dUDP, a potent inhibitor for monomeric

and trimeric dUTPases (Nord *et al.*, 1997), as a substrate (Camacho *et al.*, 2000). The dUTPase-catalyzed reaction requires the presence of Mg^{2+} ions (Larsson, Nyman *et al.*, 1996) and water participates in the cleavage of the phosphor–ester bond (G. A. Rogers, personal communication). X-ray structures of several trimeric dUTPases are known to moderately high resolution (Cedergren-Zeppezauer *et al.*, 1992; Mol *et al.*, 1996; Prasad *et al.*, 1996, 2000; Dauter *et al.*, 1999; Larsson, Svensson *et al.*, 1996). To describe the active site of dUTPases with any precision, the accuracy of the structural data needs to be high. Small ions such as Mg^{2+} are difficult to detect at low resolution and the actual water-binding pattern in the active site is not well defined at a resolution around 2 Å, to which all but one of the known structures have been solved.

Most structures cited above were solved with data from crystals at room temperature or 277 K. In this work, we have collected data from frozen crystals of the trimeric *E. coli* dUTPase at 100 K and extended the resolution of the structure to the atomic level. This was achieved through stabilization of the enzyme by addition of ethylmercurythiosalicylate (Thimerosal). The structure of this derivative was solved *ab initio* using a single data set at 1.05 Å resolution and iteratively building up the model using a map calculated initially from a single mercury site per subunit of the trimer. This approach has been successful for small molecules (Perrakis, Antoniadou-Vyza *et al.*, 1999) and metalloproteins (Benini *et al.*, 2000; Perrakis *et al.*, 1997). Here, we demonstrate that it is also useful in the case of a heavy-atom derivative of a protein.

The refined model of *E. coli* dUTPase to 1.05 Å resolution will be used in the detailed description of the structure and will be utilized as a structural guide for future design of inhibitory compounds. To ensure that the heavy-metal interactions did not distort the structure, the mercury-bound enzyme was also compared with the native enzyme structure refined against data collected to 1.45 Å under identical conditions.

2. Crystal preparation, data collection and processing

The purification procedure of *E. coli* dUTPase has been described by Hoffmann *et al.* (1987). The mercury-derivative crystals were obtained from a solution containing 5 m*M* succinate at pH 5.5, 100 m*M* sodium ethylmercurythiosalicylate (Thimerosal, Hampton Research), 3 mg ml^{-1} dUTPase in 10 m*M* bis-tris buffer at pH 6.2 and 1.5% polyethylene glycol 8000. Native enzyme crystals were prepared under similar conditions in the absence of heavy-atom salt. For data collection at 100 K the crystals were cryoprotected by a short immersion into a solution of mother liquor containing 15% glycerol.

The mercury-derivative crystals diffracted to atomic resolution. A data set to 1.05 Å was collected from a frozen crystal at EMBL Hamburg beamline X11 (λ = 0.91 Å) using a fast image-plate scanner (MAR X-ray Research GmbH, Hamburg). The crystals belonged to space group *R*3 and the unit-cell parameters were $a = b = 85.742$, $c = 61.749$ Å. Data from a native enzyme crystal was collected at EMBL Hamburg beamline BW7B at a wavelength of 0.83 Å. The diffraction limit for the native crystals was 1.45 Å at 100 K; the space group was the same as for the mercury derivative but the unit-cell parameters were slightly different ($a = b = 85.821$, $c = 60.476$ Å), with a maximum difference in the *c* axis of 2.1%.

Table 1
Data collection and scaling statistics.

The values in parentheses correspond to the highest resolution bin (1.11–1.05 Å for the derivative and 1.53–1.45 Å for the native data).

Data set	Hg derivative	Native
Resolution range (Å)	31.782–1.050	47.140–1.450
Unique reflections	76450	29458
Reflections with $I > 3\sigma$	58141	24300
Completeness (%)	99.5 (100.0)	99.3 (99.9)
Multiplicity	2.9 (2.5)	3.7 (3.1)
$I/\sigma(I)$	12.3 (3.0)	9.1 (4.6)
Wilson plot B (Å^2)	7.4	14.4
$R_{\rm merge}$†	0.059 (0.280)	0.043 (0.150)
$R_{\rm meas}$‡	0.072 (0.358)	0.050 (0.182)

† $\sum|I - \langle I\rangle|/\sum\langle I\rangle$ over all reflections. ‡ Multiplicity-weighted $R_{\rm merge}$ (Diederichs & Karplus, 1997).

Image integration was carried out with the program *MOSFLM* (Leslie, 1991). The data were scaled and merged with *SCALA* (Collaborative Computational Project, Number 4, 1994). The merging statistics are shown in Table 1. The structure-factor amplitudes were derived from the averaged intensities with the program *TRUNCATE* (Collaborative Computational Project, Number 4, 1994; French & Wilson, 1978). The overall temperature factor was estimated from Wilson plots (Wilson, 1942) using all the reflections at a resolution higher than 2 Å.

3. *Ab initio* structure solution of the mercury derivative

In the derivative structure, the position of the Hg atom was obtained from an isomorphous Fourier difference map $F_H - F_N$. F_H are the observed structure factors for the mercury-derivative data and F_N are the amplitudes and phases calculated from the native room-temperature structure of *E. coli* dUTPase (ID code 1dup; Dauter *et al.*, 1998) available from the PDB (Bernstein *et al.*, 1977). A 22σ level peak was also found in the Harker section of the anomalous difference Patterson map with coefficients $F_H(+)$ and $F_H(-)$, corresponding to the same site. Although the heavy-atom position could have thus been derived from the derivative data only, making use of the existing native data made it possible to select the correct hand of the site coordinates without an additional trial. The procedure proposed by Perrakis, Antoniadou-Vyza *et al.* (1999) was then used to solve the derivative structure *ab initio*. The program *ARP*/*wARP* (Perrakis, Morris *et al.*, 1999; Lamzin et al., 1999) was combined with refinement with the program *REFMAC*, which uses a maximum-likelihood target (Murshudov *et al.*, 1997) to automatically build and refine O-atom positions in a density map initially calculated from the position of the single Hg atom.

In the two initial model building and refinement cycles atoms were added one by one (at distances between 1.0 and 3.3 Å from an existing atom) based on electron-density peaks in the $mF_o - DF_c$ Fourier map calculated after maximum-likelihood unrestrained refinement. F_o and F_c are the scaled observed and calculated structure factors, m is the figure of merit and D is the Fourier transform of the probability distribution of the coordinate error (Luzzati, 1952). These coefficients are calculated using the free reflection set in order to minimize bias to the model (Murshudov *et al.*, 1997). Atoms were merged when too close to each other or when appearing in too low density (below 0.3σ) in a $3mF_o - 2DF_c$ density map. In subsequent cycles, the σ cutoff to find and reject atoms was increased successively. The maximum number of new atoms allowed was also increased. The total number of O atoms searched was based on the number of residues in the crystallographic asymmetric unit. At the end of the fourth cycle the *wARP* model, consisting of 1925 O atoms, was considered to be complete. The *R* factor for this model was 20%.

After rejection of badly placed atoms and a further *REFMAC*/*ARP* building cycle, automatic chain tracing was attempted by searching the O-atom model for groups of atoms fulfilling the distance and angle criteria expected for peptide chains. 108 residues scattered along ten polypeptide stretches were found in the first cycle. Each main-chain tracing cycle was followed by further addition and removal of atoms based on $mF_o - DF_c$ and $3mF_o - 2DF_c$ Fourier maps, respectively. Automatic tracing of 132 of a total of 136 ordered residues was achieved after 40 cycles of autobuilding, substantially less than the number recommended in the program manual. The optional automatic side-chain docking procedure correctly assigned 49 side chains for the ordered residues. Residues 137–152 could not be traced either automatically or manually.

The whole automated building procedure took approximately 28 CPU hours on an SGI O2 R10000 workstation (Silicon Graphics, Mountain View, CA, USA). The automatically traced model was close to the final refined model, with an r.m.s. difference between them of only 0.066 Å for the main chain. However, some side-chain positions had to be corrected and multiple conformations in both the main chain and side chains had to be modelled manually.

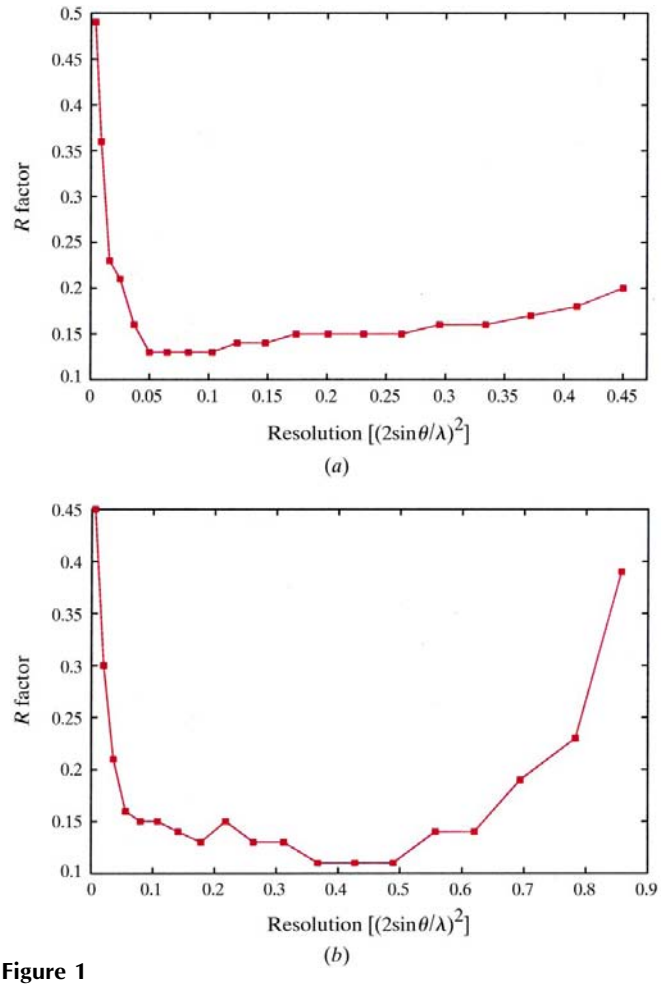

Figure 1
The *R* factor as a function of resolution for the refined structures: (*a*) native structure, (*b*) derivative structure.

4. Refinement and validation

4.1. The dUTPase–Hg complex

The model, consisting of the polypeptide chain and water molecules, was subjected to several further cycles of auto-

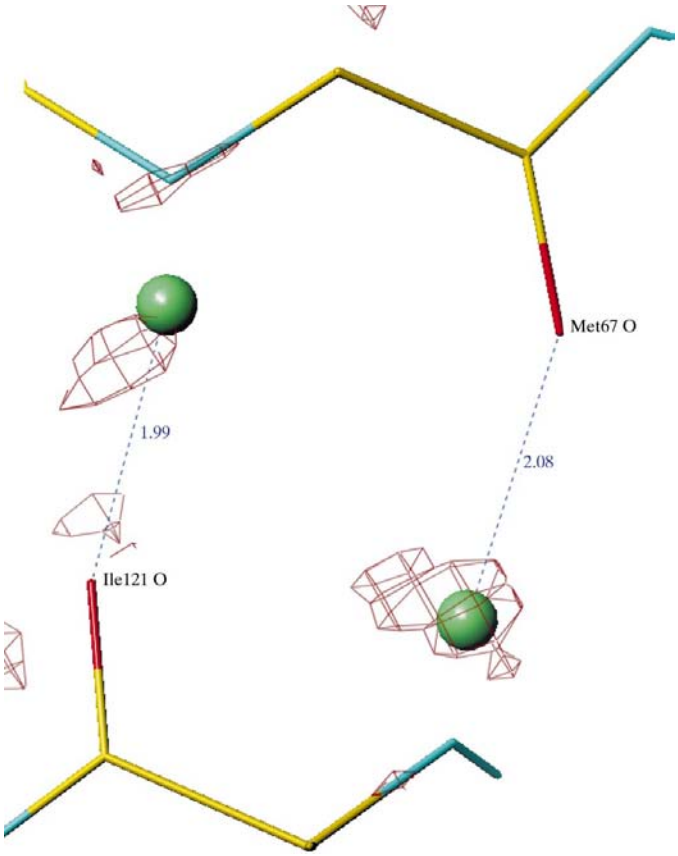

Figure 2
A difference electron-density map $mF_o - DF_c$ contoured at 2σ (red) shows peaks close to the predicted positions of H atoms (in green) involved in N—O bonds between β-chains. The distances between the carboxyl O atoms and the predicted H atoms are also shown.

mated refinement using the program *ARP* (Lamzin & Wilson, 1997) and conjugate-direction refinement (Tronrud, 1992) as implemented in *REFMAC*. The geometry was restrained to the standard Engh and Huber values (Engh & Huber, 1991). Isotropic temperature factors were refined for individual atoms. At this stage, visual inspection of the Fourier difference maps showed residual electron density near the active site and the Hg atom. Glycerol was found to fit the density at the active site, while a strong peak in the $2mF_o - DF_c$ map (25σ) near the mercury main peak (50σ) was interpreted as an alternative position for the heavy atom.

Further refinement of the structure was carried out with the program *SHELXL*97 (Sheldrick & Schneider, 1997). The model was refined against experimental intensities using the conjugate-gradient algorithm and a diffuse solvent correction was applied (Moews & Kretsinger, 1975). The occupancies for the Hg atoms and the glycerol molecules were refined. The relative occupancies of multiple side-chain conformations were refined restraining the sum of the occupancies to unity. After restrained anisotropic refinement of all atoms, more multiple conformations were introduced in the model (main-chain and side-chain atoms, water atoms and a third mercury position). The occupancy of waters with a temperature factor larger than 40 Å^2 was set to 0.5. For waters near disordered side chains the occupancy was set to the same value as the conformer at the optimal distance for hydrogen bonding. H atoms generated from fixed stereochemistry were included at the end of the refinement.

4.2. The native enzyme

The initial model used for the refinement of the native dUTPase structure was the atomic resolution model of dUTPase (polypeptide chain and waters) before refinement with *SHELXL*. The native structure model was refined against the native data with *SHELXL*97 using a protocol similar to that used for the mercury complex. As the ratio number of observations:number of atoms (4.6) differs owing to the lower resolution of the native data set (1.45 Å) only the S atoms were refined anisotropically. There are seven S atoms in total in the subunit: one in a cysteine residue (which is the mercury-binding site) and six in methionine side chains.

4.3. Accuracy of the models

After the refinement, a block-matrix least-squares minimization was carried out to estimate errors in coordinates and temperature factors. The statistics for both refined models are given in Table 2. Visual inspection of the models was performed on Silicon Graphics workstations using the program *TURBO-FRODO* (Roussel & Cambillau, 1991) and the stereochemical quality of the final models was checked with *PROCHECK* (Laskowski *et al.*, 1993).

The free *R* factor (Brünger, 1993) was used as a validation test throughout the refinement. In the last cycle all the reflections were included, but this did not cause a drop in the *R* factor. This has also been observed in the refinement at atomic resolution of *Fusarium solani* cutinase (Longhi *et al.*, 1997), while in the case of *Streptomyces aureofaciens* RNase (Sevcik *et al.*, 1996) inclusion of the $R_{\rm free}$ set of reflections in the refinement caused a drop in the *R* factor of 0.3%. Longhi and coworkers suggested that the different behaviour of the refinement might be related to the lower data-to-parameter ratio (5.5 for cutinase, 3.6 for RNase) affecting the convergence of the refinement. In the case of the dUTPase mercury derivative, the data-to-parameter ratio is 5.7. Thus, the behaviour of the refinement is consistent with the above hypothesis.

Table 2
Summary of the crystallographic analysis and refinement for *E. coli* dUTPase.

The statistics apply to the 136 residues present in the model.

	Native	Hg derivative
Disordered residues	35	38
Solvent atoms	308	298
Glycerol†	2 (1.53)	2 (1.63)
Hg atom†	—	3 (0.79)
R.m.s. on final $F_o - F_c$ map (e Å^{-3})	0.05	0.06
Maximum value of final $F_o - F_c$ map (e Å^{-3})	0.41	0.48
Minimum value of final $F_o - F_c$ map (e Å^{-3})	−0.25	−0.36
Ramachandran most favoured region (%)	93.7	92.9
Ramachandran allowed region (%)	4.5	5.4
Ramachandran generously allowed region (%)	0	0
Ramachandran disallowed region (%)	1.8	1.8
E.s.u. for C (Å)	0.122	0.048
E.s.u. for N (Å)	0.093	0.036
E.s.u. for O (Å)	0.075	0.030
R.m.s deviation from target values (Å)		
Bond lengths	0.012	0.015
Angle lengths	0.032	0.037
Distances from restraint planes	0.029	0.030
Chiral volumes	0.083	0.137
R factor ($R_{\rm free}$)‡	0.14 (0.19)	0.14 (0.16)

† The value in parentheses is the sum of all the occupancies. ‡ $R = \sum |I_o| - |I_c| / \sum |I_o|$. The *R* factor was calculated using all reflections and $R_{\rm free}$ was calculated with the free reflection set (5% of the total).

The final *R*-factor value for the mercury-derivative structure of the enzyme (14.1%) is slightly higher than expected for atomic resolution structures (Fig. 1), where the *R* factor often lies within the range 8–12% (Dauter *et al.*, 1995). The high *R*-factor values could partly be a consequence of the 14 totally disordered residues at the C-terminal tail of the protein. Although they are present in the crystal (Dauter *et al.*, 1998), the absence of a model for these residues is likely to affect correct modelling of the solvent and to be a source of disagreement with the data. The highest resolution weaker data also contribute to the high overall *R*-factor value, as can be seen in Fig. 1. The overall *R* factor to 1.1 Å is 13.3%.

In the following description we will mainly refer to the most accurate of the structures, the dUTPase–Hg complex, unless otherwise stated.

5. The *E. coli* dUTPase structure at 100 K

Freezing the crystal does not appear to alter significantly the tertiary structure of the protein, although the unit-cell para-

meters shrink by about 1% in the derivative structure and by up to 3% on the *c* axis of the low-temperature native crystal compared with the unit-cell parameters at higher temperature. Despite these unit-cell changes, the fold and overall hydrogen-bonding pattern are very similar to the room-temperature structure previously described (Cedergren-Zeppezauer *et al.*, 1992; Dauter *et al.*, 1998). Difference electron-density maps showed some peaks which could be assigned to H atoms in the well ordered buried areas of the atomic resolution structure (Fig. 2), but most H atoms were not visible in the maps.

5.1. High degree of disorder

Both the native and the derivative 100 K structures have many disordered parts, such as the C-terminal tail comprising residues 137–152, which were not visible in difference Fourier maps of this enzyme and are presumably disordered and not visible in any crystallographic electron-density map (Cedergren-Zeppezauer *et al.*, 1992; Larsson, Svensson *et al.*, 1996; Dauter *et al.*, 1998). While the N-terminal methionine side chain is poorly refined in the low-temperature structure, with *B* factors taking extremely large values as seen in Fig. 3(*c*), some other exposed side chains which were not found in density in the room-temperature structure (Lys3, Lys4, Lys16, Lys77, Gln106 and Glu129) could be modelled. At 100 K these side chains could be totally or partly defined in double conformations, albeit with temperature factors above the average (17.8 Å^2). The exception is Gln106, where all the atoms are clearly defined in the electron density at a level of 1.8σ. At low temperature this residue interacts with the neighbouring monomer (Gly46), possibly increasing the stability of the trimer.

A total of 37 residues, mostly in exposed areas of the monomer, were found with several atoms in alternate positions at 100 K, including two residues at the monomer–monomer hydrophobic interface area (Met67 and Pro124). In the loop region (see Table 4 in Dauter *et al.*, 1998) comprising residues 15–17 the whole polypeptide chain is in a double conformation. This shows that in these crystals, while some mobile side chains tend to become 'immobilized' in two preferred conformations, freezing of the crystal does not increase the order. In general, there is a high degree of dynamic disorder which does not decrease as the resolution of the data is increased (Table 2). This might further add to the relatively high *R* values.

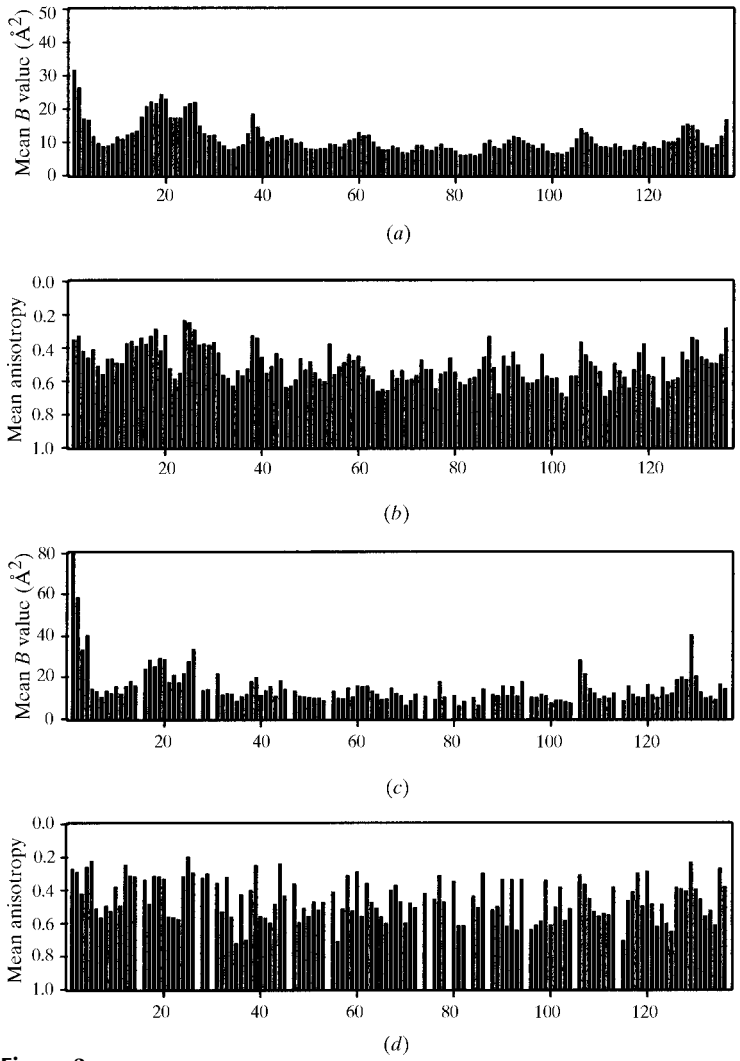

Figure 3
Temperature-factor statistics for the polypeptide chain of the Hg–dUTPase complex at 1.05 Å resolution. (*a*) Mean temperature factor of main-chain atoms. (*b*) Mean anisotropy of main-chain atoms. (*c*) Mean temperature factor of side-chain atoms. (*d*) Mean anisotropy of side-chain atoms. Here, the anisotropy is defined as the smaller eigenvalue divided by the largest eigenvalue of the anisotropic temperature-factor components.

5.2. The active site

The 100 K and room-temperature models differ in the active-site cavity by the presence of the cryoprotectant glycerol (Fig. 4). Two glycerol molecules displace water molecules and two of the O atoms of each glycerol take the place of the expelled waters maintaining the general hydrogen-bonding pattern of the active site. One of the glycerol molecules, Glyc139, is stacked with a tyrosine side chain (Tyr93) in a position similar to that of the deoxyribose in the structure of the *E. coli* enzyme in complex with dUDP (Larsson, Svensson *et al.*, 1996). The other glycerol, Glyc138, with a somewhat lower occupancy, is located deeper in the active-site crevice and interacts with the side chain of Asn84, a residue known to be important in substrate and water interactions. There may be a hydrogen bond between O3 of Glyc138 and O1 of Glyc139 and between the O2 atoms of both glycerol molecules.

All the water molecules present in the room-temperature model were also found in the derivative low-temperature structure, except for the three of them displaced by the glycerol molecules. Because the glycerol molecules do not bind with full occupancy, it is possible that these water sites are still partially occupied in the low-temperature structure. This might explain the peak height of 5σ found at the O3 position

in Glyc139 shown in Fig. 4. The O3 atom is occupying the position of one of the displaced water molecules. Two additional water molecules are observed in the active site in the low-temperature structure. Water 445 has low occupancy and is buried. It might have been pushed aside by the presence of Glyc139 and could be responsible for the adoption of one of the alternative conformations of Met68. Water 401 may have been displaced by Glyc138 and positioned closer to the side chain of Lys77, causing it to adopt a different conformation compared with the unliganded room-temperature structure. The r.m.s. distance between the room-temperature (where glycerol is not present) and the low-temperature derivative structure is 0.73 Å for the atoms in residues 61–93 (0.17 Å for the main-chain atoms). The residues in this stretch are all located in the active site. The small difference between the models around this area indicates that apart from the changes discussed above, the binding of glycerol does not grossly alter the *E. coli* dUTPase structure.

Residues Ser72, Asp90 (conserved among dUTPases) and Asp92 (part of the tyrosine hairpin loop in the active site) could be modelled in multiple conformations. This may be a consequence of the better quality of the electron-density maps at 1.05 Å rather than being a feature of the structure at 100 K, because the multiple side-chain conformations are within hydrogen-bonding distance of water molecules which are also present in the room-temperature structure.

5.3. Heavy-atom site

The Hg atom in the atomic resolution structure can be seen in three different positions (Fig. 5) with varying bond distances to the S^{γ} atom of Cys36. This is the only cysteine residue in the sequence and the side chain is located in a small hydrophobic cavity. Sites *A* and *B* at 2.452 and 2.284 Å (e.s.u. 0.006 and 0.007, respectively) are the major sites. Site *C* at 2.370 Å (e.s.u. 0.014) has a much lower occupancy. These bond distances were left unrestrained during the refinement. The variety of bond lengths and positions for the Hg atom can be explained by the flexibility of the polypeptide chain around the binding cavity, with an abundance of mobile side chains (Asp11, Arg13, Asn38) surrounding the cysteine groove. Typical cysteine–mercury bond lengths reported for structures in the PDB range between 2.28 and 2.36 Å (Nair *et al.*, 1991; Church *et al.*, 1986; Nagar *et al.*, 1996; Sinha *et al.*, 1999). Weak elongated density near the most occupied metal site (*A*) could be assigned to the ethyl moiety of sodium ethylmercurythiosalicylate. The distance between the closest ethyl C atom and the highest occupied mercury site is 2.066 Å and the S^{γ}—Hg—C angle is 171°. Sinha *et al.* (1999) report an angle of 165° for two-coordinated Hg. The close agreement in bond angles and distances with this structure implies that the thiosalicylate group leaves upon derivatization.

In the 100 K native structure Arg13 points towards the solvated Glu17 and its N^{γ} forms a hydrogen bond to a water molecule in the cysteine cavity. In the derivative model this water is absent and the arginine side chain is in a clear double conformation with the amine groups in both conformers directed towards the mercury site and residues Asp11 and Asn38, forming hydrogen bonds to both of these residues

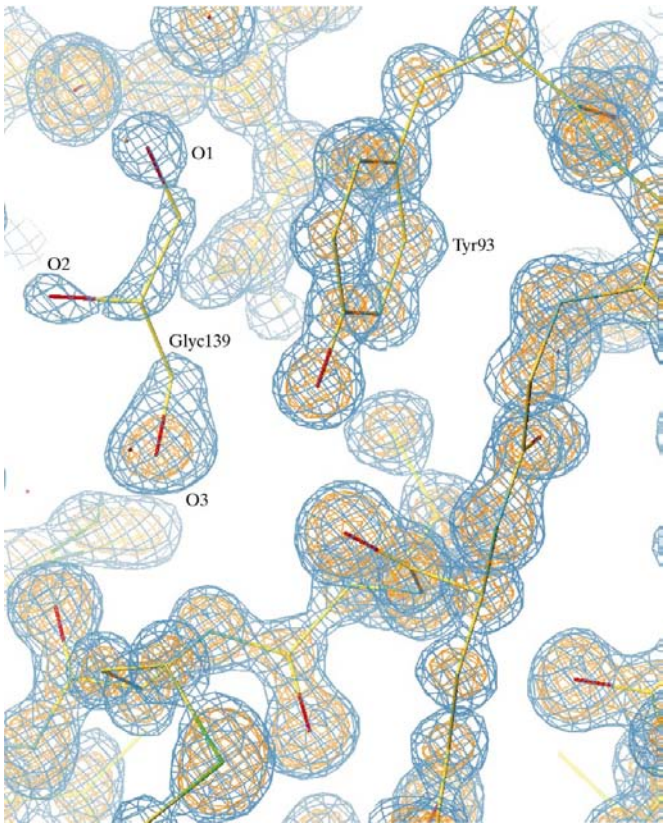

Figure 4
Detail of the 1.05 Å structure, showing parts of the polypeptide chain at the active site, including Tyr93 and one molecule of glycerol (Glyc139). The $2mF_o - DF_c$ electron-density map is contoured at 2σ (in blue) and 4σ (coral).

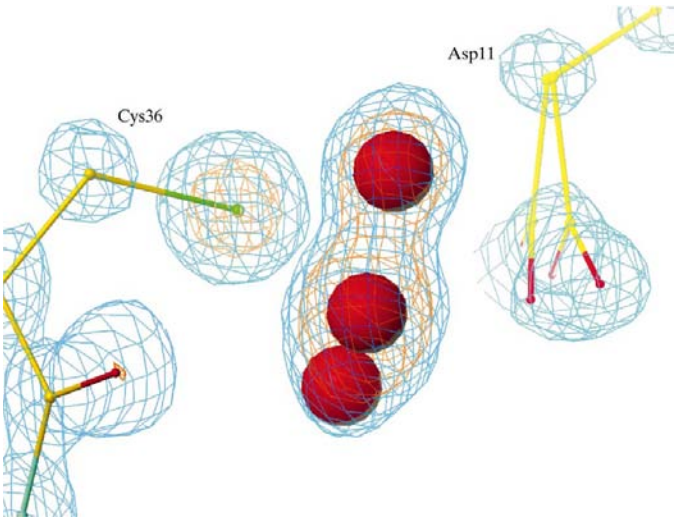

Figure 5
View of the mercury-binding site in a $2mF_o - DF_c$ electron-density map contoured at 3 and 10σ (in blue and orange, respectively) showing clearly the multiple conformations of the Hg atom (red spheres) bound to the S^{γ} of Cys36.

which do not exist in the native structure. Glu17 is also disordered, as is the whole peptide chain down to Gly15. This part of the structure is a loop situated close to strand β-I of a neighbouring subunit (for strand numbering, see Cedergren-Zeppezauer *et al.*, 1992). Lys16 N^{ζ} is involved in a direct hydrogen bond to Glu42 O^{γ} of a neighbouring monomer. These additional contacts may contribute to the stability of the crystal and explain the better diffraction properties of *E. coli* dUTPase after formation of the mercury derivative.

No further significant differences were found between the two structures. In particular, the active site seems to be unaltered by the covalent attachment of the heavy-atom compound to Cys36.

5.4. Solvent structure

A total of 298 water molecules were modelled in the atomic resolution derivative structure, many of which are absent from the room-temperature structure. There are approximately 50 waters with two or more protein ligands. About 85% of the well ordered waters in the first solvation shell, with temperature factors below 30 Å^2, are present in both the 100 K and room-temperature models. Some of the waters playing important structural roles in the 100 K model are, however, not present in the room-temperature model, such as water 248, which binds to the main chain at Leu137 N and Ala135 O, or waters 237, 250 and 279, involved in trimer–trimer interactions.

Nearly 60% of the waters have no protein ligands, binding to other water molecules instead. These waters tend to have temperature factors higher than 30 Å^2 and appear at lower levels in electron-density maps. They might be important in stabilizing the protein, since they often cluster at monomer–monomer interfaces, such as (i) the crevice formed by the helix Ser72–Gly79, the loop Arg104–Ser108 and the loop Ala44–Thr48 from the neighbouring monomer, (ii) the cavity above the hydrophobic cluster along the crystallographic threefold axis, (iii) the entry to the active-site cavity and (iv) the large area along the extended C-terminal arm (from Gln127) which wraps around the N-terminal part of another monomer.

6. Conclusions

The presence of mercury bound to the single free cysteine in *E. coli* dUTPase, in combination with cryocooling of the crystals to 100 K, made it possible to collect data to atomic resolution, allowing visualization of the structure to a finer detail than previously achieved. Data to 1.05 Å were collected in this study, but it might be possible to reach even higher resolution using a more intense synchrotron source or longer exposures. Higher resolution would increase the observation-to-parameter ratio, allowing an even more accurate refinement of the structure and definition of H-atom positions. The stabilizing effect of the mercury compound seems to take place through long-range interactions, possibly improving contacts within and between trimers along flexible parts of the polypeptide chain.

Despite the improvement of resolution, many exposed parts of the structure remain flexible and difficult to model adequately. Thus, structural details of the enzyme relevant to its mechanism still remain unknown, although a thorough kinetic analysis has been made (Larsson, Nyman *et al.*, 1996). As the active site is remote from the mercury-binding site and no significant structural differences have been found in this area when comparing the derivative and the native enzyme, we assume that the derivatization of the protein does not affect normal function. This assumption is strengthened by results obtained by G. Rogers (personal communication) who reacted *E. coli* dUTPase with Ellman's reagent (interacting with SH groups) without disturbing the activity of the enzyme. Therefore, the Hg–dUTPase complex could become important in further studies of the function of the enzyme, particularly in structure determinations of inhibitor complexes. Since glycerol could be a potential competitor for ligand binding, an alternative cryoprotectant should be found for these studies.

From a purely crystallographic point of view, the contribution of mercury (or other heavy atoms) to high-resolution diffraction data offers the possibility of easily solving a structure *ab initio*, similarly to that presented here. The high completeness and accuracy of the automatically traced model shows the value of this approach for automated structure determination. As in the case of structures solved by multi-wavelength anomalous dispersion (MAD), this approach allows the characterization of small changes in the structure without bias to a pre-existing model. The small r.m.s. difference between the automatically traced and the final refined model shows that very accurate phases can be determined with totally objective criteria. The initial electron-density maps can thus be used to compare theoretical solvent models, as suggested by Burling *et al.* (1996) for high-resolution MAD maps. The advantage with respect to the MAD method would be the considerably smaller amount of data required for structure solution, which could be crucial when radiation damage becomes a serious problem.

This work was carried out within a European Commission Research Consortium, contract No. BMH4-CT97-2711 (DG12-SSMI). EC-Z received support from the EC TMR/LSF grant ERBFMGE-CT98-0134. RP was supported by an EU Research Training Grant (Biomedicine and Health Research Programme, contract No. ERBBMH4CT985105). The authors wish to thank the EMBL Hamburg Outstation for access to synchrotron beamlines. Financial support from the Swedish Cancer Society and Magnus Bergwalls Stiftelse is gratefully acknowledged.

References

Benini, S., González, A., Rypniewski, W. R., Wilson, K. S., Van Beeumen, J. J. & Ciurli, S. (2000). *Biochemistry*, **39**, 13115–13126.
Bergman, A. C., Nyman, P. O. & Larsson, G. (1998). *FEBS Lett.* **441**, 327–330.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bertani, L. E., Häggmark, A. & Reichard, P. (1961). *J. Biol. Chem.* **236**, Pc67–Pc68.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
Camacho, A., Hidalgo-Zarco, F., Bernier-Villamor, V., Ruiz-Pérez, L. M. & González-Pacanowska, D. (2000). *Biochem. J.* **346**, 163–168.
Cedergren-Zeppezauer, E., Larsson, G., Nyman, P. O., Dauter, Z. & Wilson, K. S. (1992). *Nature (London)*, **70**, 740–743.
Church, W. B., Guss, J. M., Potter, J. J. & Freeman, H. C. (1986). *J. Biol. Chem.* **261**, 234–237.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.
Dauter, Z., Persson, R., Rosengren, A. M., Nyman, P. O., Wilson, K. S. & Cedergren-Zeppezauer, E. S. (1999). *J. Mol. Biol.* **285**, 655–673.
Dauter, Z., Wilson, K. W., Larsson, G., Nyman, P. O. & Cedergren-Zeppezauer, E. (1998). *Acta Cryst.* D**54**, 735–749.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
French, G. S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–524.
Greenberg, G. R. & Somerville, R. L. (1962). *Proc. Natl Acad. Sci. USA*, **48**, 247–256.
Hoffmann, I., Widström, J., Zeppezauer, M. & Nyman, P. O. (1987). *Eur. J. Biochem.* **164**, 45–51.
Kornberg, A. & Baker, T. A. (1991). In *DNA Replication.* New York: Freeman Press.
Lamzin, V. S., Perrakis, A. & Morris, R. (1999). *ARP/wARP Manual*. EMBL, Hamburg, Germany.
Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
Larsson, G., Nyman, P. O. & Kvassman, J. O. (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. (1991). *Crystallographic Computing V*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 27–38. Oxford University Press.
Longhi, S., Czejek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997). *J. Mol. Biol.* **268**, 779–799.
Luzzati, V. (1952). *Acta Cryst.* **6**, 142–152.
McGeoch, D. J. (1990). *Nucleic Acids Res.* **18**, 4105–4110.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
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.* D**53**, 240–255.
Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. (1996). *Nature (London)*, **380**, 360–364.
Nair, S. K., Calderone, T. C., Christianson, D. W & Fierke, C. A. (1991). *J. Biol. Chem.* **266**, 17320–17325.
Nord, J., Larsson, G., Kvassman, J. O., Rosengren, A. M. & Nyman, P. O. (1997). *FEBS Lett.* **414**, 271–274.
Perrakis, A., Antoniadou-Vyza, E., Tsitsa, P., Lamzin, V. S., Wilson, K. S. & Hamodrakas, S. J. (1999). *Carbohydr. Res.* **317**, 19–28.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 87–91. Warrington: Daresbury Laboratory.
Persson, R. (1998). Thesis, Lund University, Sweden.
Prasad, G. S., Stura, E. A., Elder, J. H. & Stout, C. D. (2000). *Acta Cryst.* D**56**, 1100–1109.
Prasad, G. S., Stura, E. A., McRee, D. E., Laco, G. S., Hasselkus-Light, C., Elder, J. H. & Stout, C. D. (1996). *Protein Sci.* **5**, 2429–2437.
Roussel, A. & Cambillau, C. (1991). *Silicon Graphics Geometry Partners Directory*, p. 86. Silicon Graphics, Mountain View, CA, USA.
Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. (1996). *Acta Cryst.* D**52**, 327–344.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sinha, S., Rappu, P., Lange, S. C., Mantsala, P., Zalkin, H. & Smith, J. L. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13074–13079.
Tronrud, D. E. (1992). *Acta Cryst.* A**48**, 912–916.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.