Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Atomic resolution structure of *Erwinia chrysanthemi* L-asparaginase

Jacek Lubkowski,[a]* Miroslawa Dauter,[a,b]† Khosrow Aghaiypour,[a]‡ Alexander Wlodawer[a] and Zbigniew Dauter[a]†

[a]Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, MD 21702, USA, and [b]Intramural Research Support Program, SAIC Frederick, National Cancer Institute at Frederick, Frederick, MD 21702, USA

† Mailing address: Brookhaven National Laboratory, Building 725A-X9, Upton, NY 11973, USA.
‡ Current address: Department of Biochemistry, Tehran Medical Sciences University, Tehran, Iran.

Correspondence e-mail: jacek@ncifcrf.gov

Received 19 July 2002
Accepted 22 October 2002

PDB Reference: L-asparaginase, 1o7j, r1o7jsf.

An X-ray structure of L-asparaginase from *Erwinia chrysanthemi* (ErA) has been refined at 1 Å resolution to an *R* factor of below 0.1, using data collected on a synchrotron source. With four molecules of the enzyme consisting of 327 amino acids each, this crystal contains one of the largest asymmetric units of a protein refined to date at atomic resolution. Previously, structures of ErA and of related enzymes from other bacterial sources have been refined at resolutions not exceeding 1.7 Å; thus, the present structure represents a very significant improvement in the quality of the available models of these proteins and should provide a good basis for future studies of the conformational variability of proteins, identification of subtle conformational features and corroboration of the stereochemical libraries, amongst other things. L-Asparaginases, which are enzymes that catalyze the hydrolysis of L-asparagine to aspartic acid, have been used for over 30 y as therapeutic agents in the treatment of acute childhood lymphoblastic leukemia, although the details of the enzymatic reaction and substrate specificity have not yet been completely elucidated. This atomic resolution structure is a step in that direction.

1. Introduction

L-Asparaginase (L-asparagine amidohydrolase; EC 3.5.1.1) is an enzyme that primarily catalyzes the conversion of L-asparagine to L-aspartic acid and ammonia. The anti-leukemic activity of guinea pig serum L-asparaginase, which was discovered over 40 y ago (Broome, 1961, 1968), led to clinical utilization of members of this family of enzymes and resulted in considerable interest in evaluating their biochemical properties, especially those of the enzyme isolated from *Escherichia coli* (Mashburn & Wriston, 1964; Roberts *et al.*, 1966). Since the early 1970s, type II *E. coli* L-asparaginase (EcA) and a closely related enzyme from *Erwinia chrysanthemi* (ErA) have been used as drugs in the treatment of acute childhood lymphoblastic leukemia, although the exact mechanism of their activity has not yet been completely verified. Similar enzymes, some of which could also efficiently utilize L-glutamine as a substrate and were therefore called glutaminase–asparaginases, have also been isolated and characterized from a number of other bacterial species, such as *Wolinella succinogenes* asparaginase (WsA), *Pseudomonas* 7A glutaminase–asparaginase (PGA) and *Acinetobacter glutaminasificans* glutaminase–asparaginase (AGA; Beacham & Jennings, 1990). Moreover, highly homologous L-asparaginases have been identified in nearly all members of the animal kingdom (Borek & Jaskólski, 2001), in contrast to plants, which utilize an unrelated enzyme with sequence similarity to mammalian glycosylasparaginases (Lough *et al.*, 1992).

Table 1
Data-collection statistics for the atomic resolution structure of ErA.

Data were collected in two passes, differing in the placement of the detector and the speed of data collection. Values in parentheses are for the final resolution shell.

	High resolution	Low resolution
Wavelength (Å)	0.98	0.98
Crystal-to-detector distance (mm)	60	130
Resolution (Å)	4.0–1.0	25.0–1.6
No. of images	230	180
Exposure time per frame (s)	100	10
Rotation range per frame (°)	0.5	1.0
Merging		
Unit-cell parameters (Å, °)	a = 106.38, b = 90.35, c = 127.59, β = 91.4	
Resolution (Å)	25.0–1.0 (1.04–1.00)	
Measured reflections	1780326	
Unique reflections	559625	
Completeness (%)	87.6 (75.1)	
$R_{\rm merge}$ (%)	5.0 (30.5)	
$I/\sigma(I)$	24.2 (3.0)	

Structural studies of L-asparaginases have been undertaken in the last 30 y with the aim of at least describing their enzymatic mechanism, if not the reasons for their biological activity. Although crystals were reported as early as 1969 (North *et al.*, 1969; Epp *et al.*, 1971; Wlodawer *et al.*, 1975), the first detailed structure of any member of this class of enzymes was only published in 1993. Crystal structures have been reported since then for five bacterial L-asparaginases and L-glutaminase–asparaginases, both as apoenzymes and in complexes with substrates and inhibitors (Swain *et al.*, 1993; Miller *et al.*, 1993; Lubkowski, Wlodawer, Ammon *et al.*, 1994; Lubkowski, Wlodawer, Housset *et al.*, 1994; Lubkowski *et al.*, 1996; Palm *et al.*, 1996; Jakob *et al.*, 1997; Ortlund *et al.*, 2000; Jaskólski *et al.*, 2001; Aghaiypour *et al.*, 2001*a*,*b*). The amino-acid sequences of these enzymes contain ∼40% identical residues in a single chain consisting of ∼330 amino acids, resulting in considerable similarity in their structures. These structures were refined at medium to high resolution (mostly in the range 1.8–2.4 Å), with the maximum resolution of the published structures being 1.7 Å for PGA (Jakob *et al.*, 1997) and ErA (Aghaiypour *et al.*, 2001*b*). All of these enzymes are active as homotetramers with a molecular weight in the range 140–150 kDa, although the four individual active sites are not reported to be cooperative.

The first structures of proteins refined at a resolution of 1 Å or better became available almost 20 y ago (Wlodawer *et al.*, 1984; Teeter *et al.*, 1993), but until recently their number has been quite small. However, improvements in the methods of data collection using synchrotron radiation and cryo-crystallography and in structure refinement with programs such as *SHELXL* (Sheldrick & Schneider, 1997) and *REFMAC* (Murshudov *et al.*, 1997) have made the elucidation of very high-resolution X-ray structures much easier. Atomic resolution structures have proven to be indispensable for the description of alternate conformations of amino-acid residues, for detailed description of the solvent, for providing unbiased data on the stereochemistry of proteins and for development and verification of new models restraining the geometry and stereochemistry of refined structures (Dauter, in preparation). However, the vast majority of such structures are of comparatively small proteins, although it has become clear that some crystals with quite large unit-cell parameters are also capable of diffracting to atomic resolution. Our long-term extensive studies of L-asparaginases from various bacterial sources resulted in the growth of excellent quality crystals of ErA that diffracted to a resolution limit of 1 Å. In this paper, we describe the atomic resolution structure of this enzyme, the crystals of which contain, to the best of our knowledge, the largest asymmetric unit that has been investigated to date at such a high resolution.

2. Materials and methods

2.1. Crystallization of ErA

The sample of ErA used for all crystallization studies described here was a generous gift from the Intramural Research Support Program, SAIC Frederick Inc., where it was isolated and purified; we used it without any further purification. All crystals used for the diffraction experiments were grown using the protocol described previously by Miller *et al.* (1993). Monoclinic crystals appeared after 2–3 d and grew to their final size (deformed prisms, 0.5 × 0.5 × 0.5 mm) within a few weeks.

2.2. Data collection and processing, structure solution and refinement

X-ray data were collected from a single crystal of ErA at liquid-nitrogen temperature. Before flash-freezing the crystal in a stream of N_2 (100 K), it was transferred to mother liquor containing 15% glycerol. Diffraction data were collected on beamline X9B (National Synchrotron Light Source, Brookhaven National Laboratory, Upton, NY, USA) using an ADSC Quantum 4 CCD detector. Data were collected in two passes to allow accurate measurement of both the intense low-resolution data as well as the weaker high-resolution reflections. Data were processed and scaled using the *HKL*2000 suite of programs (Otwinowski & Minor, 1997). The completeness of the data was only 87.6%, mainly owing to the comparatively low symmetry of the monoclinic crystals, although it approached 98% for the lower resolution shells. The completeness of the outermost shell was 75%. Other statistics are shown in Table 1.

The initial model of ErA was based on the 1.85 Å structure of this enzyme reported previously (Miller *et al.*, 1993). Since the crystals were completely isomorphous, the structure could be directly refined with *CNS-XPLOR* (Brünger *et al.*, 1998) to convergence at R = 18.4% ($R_{\rm free}$ = 20.3%). It contained four molecules of ErA, each containing all 327 residues, and 657 water molecules. From this point the program *SHELXL* (Sheldrick & Schneider, 1997) was used for further refinement, in the conjugate-gradient mode (CGLS) against F^2, in a resolution range subsequently extended to 10.0–1.0 Å. Rounds of ten cycles of CGLS refinements were interspersed with inspection of the model and Fourier maps ($2F_o - F_c$ and $F_o - F_c$) using *QUANTA* (Accelrys Inc., San Diego, USA)

Table 2
Refinement statistics.

No. of reflections [>4$\sigma(F)$ in parentheses]	
Working set	553402 (456093)
Test set	5592 (4600)
Resolution range (Å)	25.0–1.0
R factor	0.1098 (0.0981)
R_{free}	0.1283 (0.1162)
Total No. of non-H atoms	
With occupancy 1.0	10842
With partial occupancy	724
No. of water molecules	1360
Average B factor (Å^2)	
All non-H atoms	14.50
Protein atoms	13.19
Water molecules	23.18
Non-water ligands	28.30
R.m.s. deviations from ideality	
Bonds (Å)	0.015
Angle distances (Å)	0.031
Chiral volumes (Å^3)	0.099

and the manual modeling of various features, such as alternative conformations and solvent molecules.

Except for the first round of refinement, all non-H atoms were refined anisotropically. The *SHELXL* default restraints and their weights were used for protein geometry (based on the library of Engh & Huber, 1991) and for anisotropic displacement parameters (ADPs). The H atoms were treated as 'riding' rigidly on their parent atoms, with their coordinates recalculated after each cycle and the isotropic displacement parameters set to 20% (or 50% for methyl groups) higher than the equivalent value of the parent atom. The H atoms were not placed on atoms in disordered regions of the protein, *i.e.* not on atoms having partial occupancies. Water molecules were placed according to the difference Fourier map. The refinement process converged at R_1 = 9.81% for 456 093 reflections >4$\sigma(F_o)$ and 10.98% for all 553 402 reflections and 104 166 refined parameters. The corresponding R_{free} values are 11.62% (4600 reflections) and 12.83% (5592 reflections), respectively. The final statistics from the structural refinement and the basic characteristics of the models are shown in Table 2.

3. Results and discussion

3.1. Atomic resolution structure of ErA

The overall fold of ErA observed after atomic resolution refinement was identical to the fold previously described for this enzyme (Miller *et al.*, 1993; Aghaiypour *et al.*, 2001*a*,*b*) and for other L-asparaginases (Fig. 1). All such enzymes are active as homotetramers with 222 symmetry. Each monomer consists of about 330 amino-acid residues that form 14 β-strands and eight α-helices arranged into two easily identifiable domains, the larger N-terminal domain and the smaller C-terminal domain, connected by a linker consisting of ∼20 residues. The crossover between the fourth and the fifth β-strands of the N-terminal domain is left-handed (Miller *et al.*, 1993). This type of motif is rarely observed in proteins (Richardson, 1981), but when it is seen it always serves an important determinant of their activity.

The first two N-terminal amino acids could not be modeled in any of the four independent monomers of ErA in the asymmetric unit. By contrast, the C-terminal residues are all well ordered, resulting in models for residues 3–327 in all peptide chains. A number of residues could be modeled in two conformations. These residues include Gln25, Met65, Leu71, Met133, Leu136, Ile199, Val208, Ser216, Tyr232, Gln239, Ser254, Val257, Arg264, Arg278, Leu298, His302, Met308 and Arg31 in chain *A*, Leu71, Ser74, Gln75, Val89, Met133, Ile199, Phe209, Asp210, Asp221, Glu231, Tyr232, Ser254, Ile260, His302, Met308, Ser315 and Asp316 in chain *B*, Asn51, Ser74, Glu79, Val89, Ile91, Arg178, Arg198, Ile199, Arg206, Lys219, Asp221, Leu233, Val257, Arg264, Leu298, His302 and Arg313 in chain *C* and Val46, Asp68, Leu71, Ser74, Glu79, Lys110, Leu136, Arg192, Arg198, Ile199, Arg212, Gly213, Ser216, Lys219, Glu231, Tyr232, Ser254, Val257, His302, Pro317 and Glu322 in chain *D*. For a few residues, such as Ser254 and His302, discrete disorder was detected in all of the molecules and seems to be an intrinsic property of their side chains. For many other residues listed above, disorder results from lack of interactions owing to their location on the surface of the protein. While some of the residues observed in multiple orientations have also been described as such in the lower resolution structures of ErA (Aghaiypour *et al.*, 2001*b*), this agreement holds for fewer than half of them.

With two notable exceptions, all main-chain torsion angles are found in either the most favored region of the Ramachandran plot (90.9%) or in the additionally allowed regions

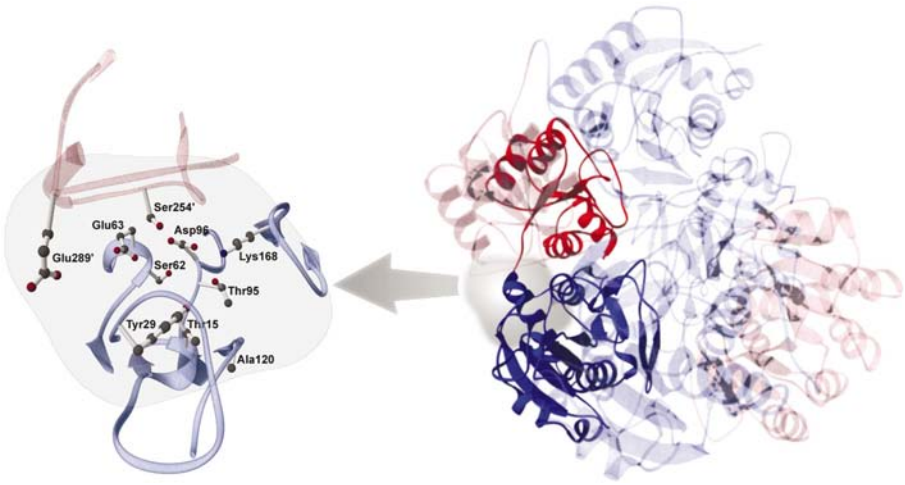

Figure 1
Schematic representation of the ErA molecule. In the tetrameric enzyme shown on the right-hand side of this figure, one of the monomers is accented with solid colors. In all four monomers the N-terminal domains are drawn in blue and the C-terminal domains in red. The location of one of the four equivalent active sites is indicated by a semi-transparent gray blob. The same active site is shown in more detail on the left-hand side of the figure. The active-site pocket is formed from the fragments of two domains contributed by different monomers. The figure was prepared with the program *RIBBONS* (Carson, 1991), followed by rendering with the program *POV-RAY* (http://www.povray.org).

(8.4%) (Fig. 2). The two exceptions are Asp296, which is found in the generously allowed region, and Thr204, which is found in the disallowed region. The torsion angles for these residues are virtually identical in all four molecules; the temperature factors are low (∼10 Å^2) and the electron-density maps are very clear (Fig. 3). The former residue is fixed in its orientation by hydrogen bonds to the main-chain N atom of residue 297 (through OD2) and with the ND1 atom of His325. The OG atom of the latter residue makes hydrogen bonds to the amide N atom of Thr205 and to ND1 of His203. It appears that the hydrogen bonds to the amide N atoms of the adjacent residues may be responsible for fixing the main-chain torsion angles away from the most common values and for stabilizing these strained structures.

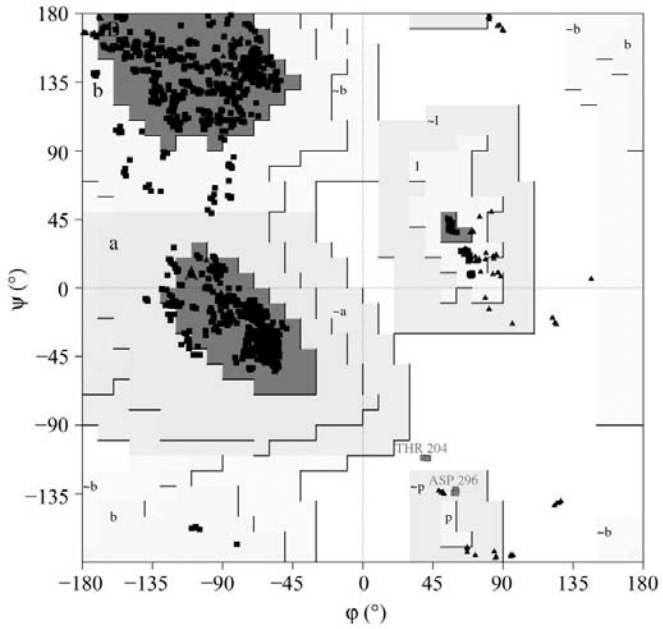

Figure 2
Ramachandran plot for ErA, prepared with *PROCHECK* (Laskowski *et al.*, 1993). The values are marked as squares for all non-glycine and non-proline residues for all four molecules and as triangles for glycines and prolines. The values for two residues, Asp296 and Thr204, are outside the most favored regions, although they are highly consistent between different molecules (see text for discussion).

3.2. Conformation of Thr204

As discussed above, analysis of the Ramachandran plot revealed Thr204 to be in a disallowed conformation in all four monomers of the enzyme. Such a conformation was also previously observed for this residue in L-asparaginases from other bacterial sources (Swain *et al.*, 1993; Lubkowski, Wlodawer, Ammon *et al.*, 1994; Lubkowski, Wlodawer, Housset *et al.*, 1994; Lubkowski *et al.*, 1996). In the case of this particular structure, however, excellent electron density for this residue assures the correctness of the conformational assignment. Thr204 is located at the beginning of the inter-domain linker, a fragment that is characterized by increased flexibility compared with the core of either domain. The immediate environment of Thr204 is shown in Fig. 4, which also presents the network of hydrogen bonds. Excluding the side chains of Lys110 and Arg206, all other atoms (and also water molecules) have very low *B* factors (≤15 Å^2), indicating that the site of Thr204 is quite rigid. Moreover, when all the atoms presented in Fig. 4 are compared across different monomers, the r.m.s. deviations vary between 0.10 and 0.15 Å, confirming the high structural conservation of this site. All nine water molecules identified within this pocket are also conserved in all four monomers. Analysis of the hydrogen-bond network shows that only three atoms with the potential to form hydrogen bonds, Thr205 N, Val109 O and Ser111 N, are not part of the network. Thr204 participates in five hydrogen bonds; without substantial reorientation of several surrounding residues, such a strong stabilization is possible only for the conformation of this residue observed in the structure. Although the inter-domain linkers are located on the surface of the L-asparaginase tetramer, nine water molecules located in the vicinity of Thr204 are involved in multiple short hydrogen bonds and are likely to be an integral part of the protein, assuring its correct fold. The stabilization of Thr204 arising from the hydrogen-bond interactions is clearly sufficient to compensate for the energy gain caused by its unfavorable conformation. Had Thr204 assumed a different topology, one of the resulting conformational changes would take place in the region Gly149-Arg150-Gly151, a stretch of residues that belongs to the left-handed crossover discussed earlier. It is thus possible that the conformation of Thr204 and the presence of the crossover are correlated.

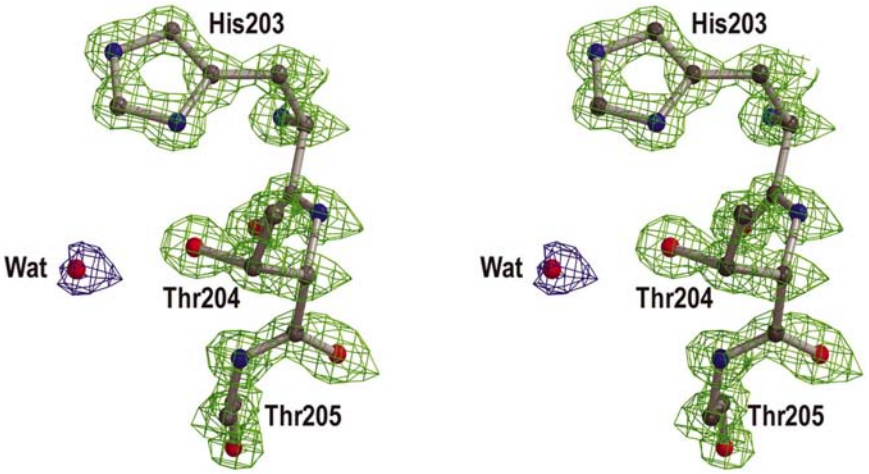

Figure 3
Atomic coordinates and the superimposed $2F_o - F_c$ electron-density map in the vicinity of Thr204, an outlier residue in the Ramachandran plot. The green contour is at the 1σ level and the blue contour at the 0.6σ level. Different contour levels are needed to simultaneously show the well ordered residue with low temperature factors and the solvent, which is significantly more mobile. Breaks in the electron density are a consequence of the high resolution of the data used to calculate the map, such that the individual atoms become visible.

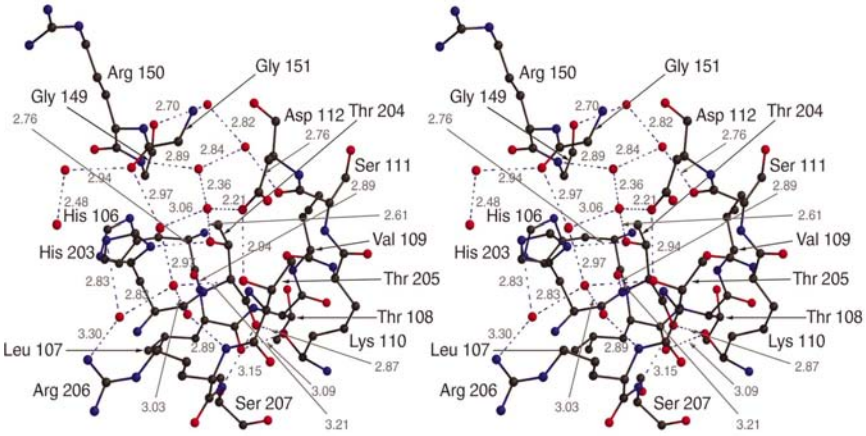

Figure 4
Conformation of Thr204. This figure represents the environment of Thr204 as determined in monomer *C* of the ErA complex with SO_4^{2-}. Except for the side chains of Arg206 and Lys110, which are characterized by high *B* factors, all other residues as well as nine water molecules shown in this figure have virtually identical conformations in other monomers of the enzyme. Hydrogen bonds are marked by dashed lines, with the interatomic distances shown as found in monomer *C*. The figure was prepared with the program *BOBSCRIPT* (Esnouf, 1997), followed by rendering with the program *POV-RAY* (http://www.povray.org).

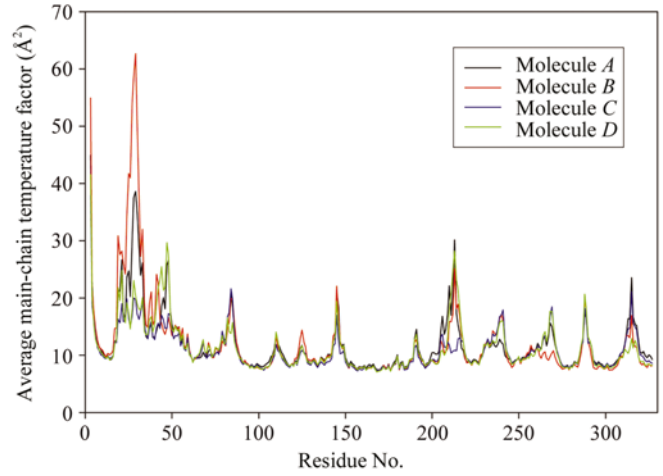

Figure 5
Temperature factors for the four crystallographically independent ErA molecules. The average values for the main-chain atoms were calculated with the program *MOLEMAN* (Kleywegt & Jones, 1997). The values for each molecule are shown in different colors, as indicated on the plot.

3.3. Comparison of the independent molecules

The quaternary structure of ErA comprises a non-crystallographic tetramer displaying 222 symmetry, with the active site created by parts of two molecules described as an intimate dimer. These molecules are designated *A*, *B*, *C* and *D*, with the intimate dimer pairs being *AC* and *BD*. The four independent molecules found in the asymmetric unit of these monoclinic crystals are extremely similar and, indeed, this atomic resolution structure can be used to analyze the limits of the crystal-imposed variability of proteins. The r.m.s. deviation between the positions of C^{α} atoms of any two subunits was of the order of 0.24 Å for all 325 common atom pairs (ranging from a low of 0.210 Å between molecules *B* and *C* to 0.265 Å for *B* and *D*), while the more commonly quoted deviations that excluded a few of the largest outliers ranged from 0.139 Å for molecules *A* and *C* to 0.222 Å for *B* and *D* for 306–317 atom pairs. These differences are small but significant, since the accuracy of atomic positions can be expected to lie in the range 0.02–0.05 Å, as for other structures refined at 1.0 Å resolution. The most visible differences were limited to only two regions of the structure. One of these regions consisted of residues 43–47 and the other of residues 210–217. The former part of the structure is virtually identical in molecules *A* and *D* and in molecules *B* and *C*, but differs by almost 1 Å between these two pairs. In molecules *B* and *C* this area is stabilized by a hydrogen bond between Asn41 and Lys47, while no similar interaction is observed in molecules *A* and *D*. Some residues in the stretch including residues 210–217 have alternative conformations (including their C^{α} atoms) and all four molecules differ by as much as 2 Å in that area. In all other parts of ErA, the conformation of the main chain is extremely similar, although the conformation of some side chains may vary.

3.4. Temperature factors

As was expected from the atomic resolution structure, the temperature factors observed here are rather low, with the average isotropic *B* for all protein atoms being only 13.2 $Å^2$ (Table 2). Although non-crystallographic symmetry was not enforced during any stages of the refinement, the agreement between the temperature factors averaged over the main-chain atoms in the four molecules is remarkably good (Fig. 5). The average *B* factors are of the order of 10 $Å^2$ for the parts of the molecules belonging to well defined secondary structure and increase to about 20 $Å^2$ for the tips of the loops. The highest temperature factors are found in the flexible loop covering the active site, particularly for molecules *A* and *B*, although the *B* factors are much lower in the other two molecules. Since the exact crystallographic environment of each molecule is different, the distribution of the highest *B* factors is different for each of them. To illustrate the detailed distribution of displacement parameters in the active-site flexible loop, the atoms from this fragment are represented in Fig. 6 in the form of thermal ellipsoids. The overall increase in the volume of the ellipsoids observed for the middle part of the loop (residues 18–34) can be seen clearly compared with those for the terminal regions (Thr15, Leu34). The active-site flexible loop contains two residues (Thr15 and Tyr29) shown to be crucial for the activity of the enzyme. The low *B* factors of the nucleophilic Thr15 (average *B* factor = 10.2 $Å^2$) are the result of conformational stabilization of this residue by the hydrogen bonds with the ligand (see Fig. 6). The other residue, Tyr29, which is much less restricted structurally, only participates in two hydrogen bonds, including one with a water

molecule. Consequently, the atoms of Tyr29 are characterized by higher *B* factors (average *B* factor = 30.2 Å^2). The values of the *B* factors for Tyr29 are particularly high in monomer *B* (average *B* factor = 50.1 Å^2), while in monomers *C* and *D* their values are only slightly higher than the average for the entire molecule at 17.8 and 19.6 Å^2, respectively. These differences in the active-site loop flexibility can be attributed to two factors. The first is the different environment of the four loops in the crystal, resulting in different stabilization through interactions with symmetry-related ErA molecules, although the structural analysis does not delineate any particular stabilizing contacts differentiating between the four active-site loops. The other possibility is the interaction bridged by the water molecule between the carbonyl O atoms of Ala21 and Gly22 and the SO_4^{2-} anion in the vicinity of the active site, which is present only in monomers *A*, *C* and *D*.

3.5. Solvent structure

Inspection of the electron density allowed us to identify 1451 water molecules in the asymmetric unit. Based on the distance and structural environment criteria, 366, 345, 354 and 346 waters were assigned to chains *A*, *B*, *C* and *D* of the enzyme, respectively. The remaining 43 water molecules occupy the interfaces between monomers and interact to a comparable extent with residues from different protein chains. An analysis of the structurally conserved water molecules revealed that 132 of them occupy virtually identical positions with respect to each protein chain of ErA. The discrepancies in the positions of these water molecules from the average, calculated over four monomers for each water molecule, is no larger than 0.25 Å, *i.e.* it is comparable to the r.m.s. deviations between the positions of C$^\alpha$ atoms of any two subunits. Whereas the average *B* factor for all water molecules is ~25 Å^2, the values for structurally conserved and variable water molecules are 14.3 and 31.4 Å^2, respectively. Visual inspection shows that the distribution of conserved water molecules is bipolar, with a significantly larger number located near the interfaces between the protein chains compared with the exterior of the enzyme molecule.

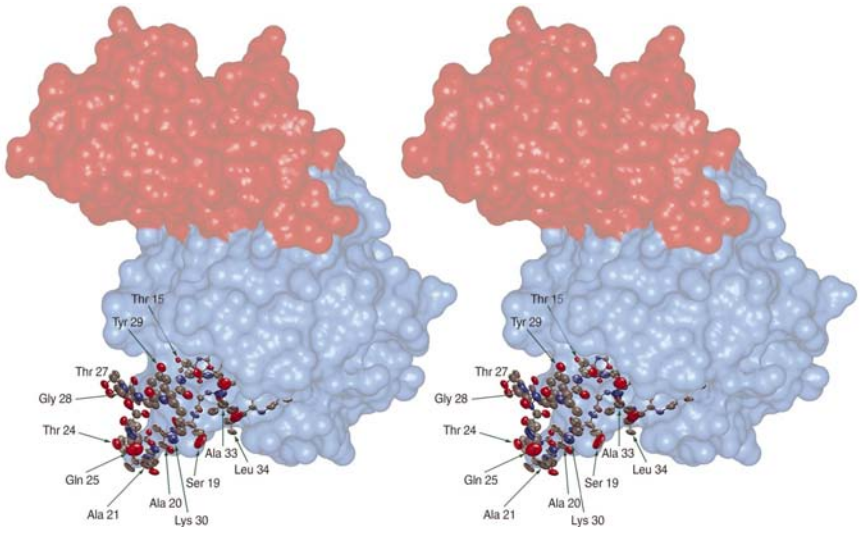

Figure 6
The flexible loop covering the active site of ErA. The solvent-accessible surface is shown for all residues of monomer *B*, except for the fragment Gly13–Val36. For this fragment, referred to as the active-site flexible loop, the individual non-H atoms are represented by the thermal ellipsoids calculated from anisotropic displacement parameters. The ellipsoids are drawn at the probability level of 30%. The accessible surface, corresponding to the N-terminal domain of the ErA monomer, is shown in blue, while the red surface represents the C-terminal domain. Selected residues of the flexible loop are labeled. This figure was prepared using the programs *ORTEX* (McArdle, 1995), *RASTER3D* (Merritt & Murphy, 1994) and *INSIGHT II* (Accelrys, San Diego, CA, USA).

3.6. Heterogen molecules

Analysis of the electron-density peaks allowed the identification of eight sulfate anions, two glycerol molecules and four polyethylene glycol molecules specifically bound to the enzyme. Four sulfate anions are bound in a strictly conserved mode in the active sites and were observed previously in crystals of ErA crystallized under the same or similar conditions (Miller *et al.*, 1993; Jaskólski *et al.*, 2001). The tight binding of these anions is indicated by very low *B* factors and by exceptionally clear electron-density peaks. An additional three SO_4^{2-} ions are also located close to the active sites of monomers *A*, *C* and *D*. Although these three ions are slightly less restricted by the interactions with the enzyme, their binding is likely to contribute to the conformational stabilization of the active-site loops (see §3.4). The fourth sulfate anion is bound to the C-terminal domain of monomer *B*, near residues Glu288, Leu290 and Gly292. The interactions of this ion with the side chain of Glu288 are mediated by the water molecules. The first of the two glycerol molecules is found almost exactly at the center of the tetrameric molecule of the enzyme. Its particular environment is shaped by the side chains of Asp175, Arg178 and Asn180 contributed by all four monomers, as well as by several water molecules which occupy the rest of the central cavity. The second molecule of glycerol is bound in the cleft formed by both domains of monomer *A* (residues 66–68, 216–220 and 308), as well as by the side chain of Asp228 from monomer *C*. Several water molecules complement this binding pocket. Glycerol molecules are characterized by relatively low *B* factors (20–25 Å^2) and the corresponding electron-density peaks were easily interpretable. Whereas both SO_4^{2-} ions and (with the advent of cryo-crystallography) glycerol molecules are frequently observed during structure determination to be bound in specific sites, identification of well ordered polyethylene glycol (PEG) molecules in protein crystals is quite rare. We were therefore surprised to identify three elongated clusters of electron density that could not be interpreted as water molecules, yet could be easily modeled with small PEG chains. Two of

those molecules are located near the tips of the extended side chains of Lys72 (monomer *D*) and Lys243 (monomer *A*) and are found in very similar ring-like conformations. The planes of those rings are approximately perpendicular to the axes of the extended side chains of the lysines. The PEG fragment interacting with Lys243 corresponds to the formula $HO-(CH_2-CH_2-O-)_5H$ with a molecular weight of 238 Da, while the PEG fragment binding to Lys72 is shorter by one CH_2-CH_2-O- protomer and its molecular weight is 194 Da. The third and shortest PEG fragment (molecular weight 150 Da) is located near the side chain of Arg212 in monomer *B*. Interestingly, all three ordered PEG fragments accompany side chains of positively charged residues.

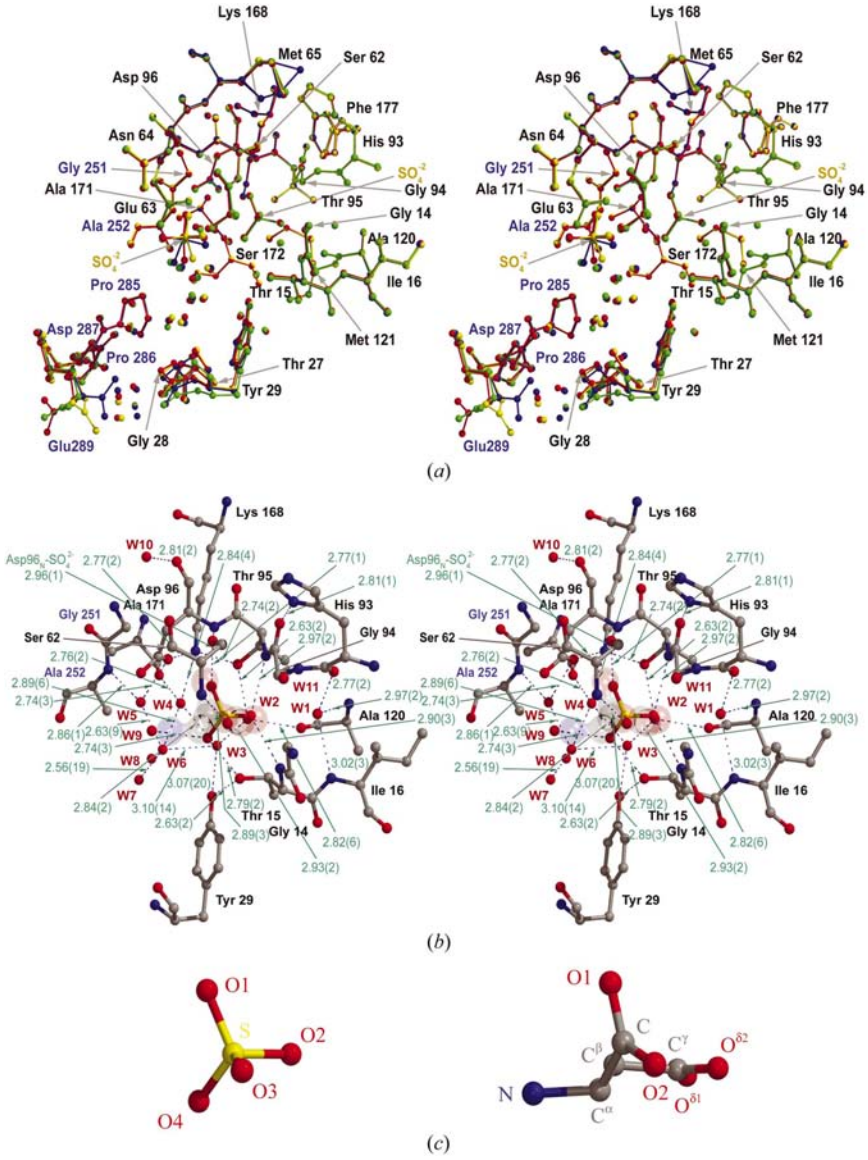

Figure 7
Active site of ErA. (*a*) Superposition of the four independent active sites present in ErA. Residues contributed by different monomers are annotated with black and blue labels. The active sites have been superimposed using the program *ALIGN* (Cohen, 1997), based on all atoms shown in this figure. The r.m.s. deviations between all aligned atoms are lower than 0.08 Å for any pair of aligned active sites. The sulfate anions, as well as solvent molecules, are conserved in all four active sites of the enzyme. Except for the side chain of Glu289, all other active-site atoms (including most of the solvent molecules) are located in remarkably conserved positions. (*b*) Close-up of the active site formed by monomers *A* and *C*, with the hydrogen bonds annotated and marked with dashed lines. The lengths of the hydrogen bonds are represented by the average values for four active sites. Numbers in parentheses represent the ranges of distances, *i.e.* 2.77 (1) indicates that a particular hydrogen bond varies between 2.77 − 0.01 Å and 2.77 + 0.01 Å for different active sites. The molecule of L-aspartate, as determined previously in the complex with ErA (Miller *et al.*, 1993), is also shown as a semi-transparent shape. (*c*) A sulfate (left) and aspartate (right), with their atoms labeled, in the orientation corresponding to (*b*). All atoms shown in (*b*) and (*c*) are colored by type. For clarity and consistency of discussion, water molecules are annotated with red labels. The figure was prepared with the program *BOBSCRIPT* (Esnouf, 1997), followed by rendering with the program *POV-RAY* (http://www.povray.org).

3.7. Active site

The location of the active site in ErA has been previously elucidated in studies of the binding of ligands such as the reaction substrate/product L-aspartate (Miller *et al.*, 1993), the alternative substrate/products D-aspartate, L-glutamate and succinic acid monoamide (Aghaiypour *et al.*, 2001*a*), or suicide inhibitors such as the L and D stereoisomers of 6-diazo-5-oxy-norleucine (Aghaiypour *et al.*, 2001*b*). In the absence of any of these cofactors, however, the active site is usually marked by the presence of a sulfate molecule originating from the crystallization medium (Jaskólski *et al.*, 2001). This was indeed the case in the atomic resolution structure presented here, in which well ordered sulfates are seen in the active site of each molecule (Fig. 7). Detailed analysis of the four active sites shows their remarkable structural conservation. Although the structural conservation of the active sites is paralleled by low *B*-factor values (<10 Å^2) for most of the contributing atoms, the residues forming a lid that separates the active-site cavity from the solvent are significantly more flexible (*B* factors >30 Å^2). In Figs. 7(*a*) and 7(*b*), these flexible residues are shown at the bottom of the active sites (Thr15, Tyr29 and Glu289). 11 water molecules, labeled in Fig. 7(*b*) as W1–W11, are structurally conserved and are part of a well defined hydrogen-bond network. Most of those molecules can also be identified in the active sites of other bacterial L-asparaginases (Swain *et al.*, 1993; Lubkowski *et al.*, 1996) as well as in complexes between ErA and other ligands (Miller *et al.*, 1993; Aghaiypour *et al.*, 2001*a*,*b*). This observation indicates that such waters are an integral part of the

active-site architecture and that they possibly play a significant role in the catalytic properties of L-asparaginases. Water molecules located in the rigid part of the active site (top of the active site in Figs. 7*a* and 7*b*) are characterized by very low *B* factors (<15 Å^2). Particularly interesting observations are made by analysis of water W1. As seen in Fig. 7(*b*), this molecule is encapsulated in the cavity adjacent to the active site and forms three strong hydrogen bonds with the main-chain atoms of the enzyme. Within this interaction, one of the protons from this water is shared in a hydrogen bond with the ligand molecule. In the structure discussed here, the active sites are occupied by SO_4^{2-} anions; however, when the natural substrates are bound to the active site, the position of the O2 atom of the sulfate is occupied by the carbonyl O atom of an amide group (shown as a semi-transparent shape in Fig. 7*b*). Moreover, for enzymes with vacant active sites, the conformations of Thr15 and Ile16 (the latter residue interacting with W1) differ from those in the complexes (Lubkowski, Wlodawer, Housset *et al.*, 1994). Analysis of all high-resolution structures of bacterial L-asparaginases with well determined solvent molecules additionally shows that W1 is always present in its conserved site and is characterized by a very low *B*-factor value. Therefore, it is highly possible that W1 plays an instrumental role in stabilizing the active conformation of the flexible active-site region as well as in modifying the electronic structure of the substrate molecule *via* interaction through the hydrogen bond with the O atom of the amide group of the substrate. Other active-site water molecules, despite their roles in stabilizing particular conformations of surrounding residues, form a very well conserved network of hydrogen bonds which can be utilized for rapid proton/electron transfer during the enzymatic reaction.

Comparison of the binding modes of SO_4^{2-} and L-asparatate ions is of particular interest. Alignment of both these ligands, as shown in Fig. 7(*c*), indicates that all atoms of the sulfate anion are located in positions occupied by structurally equivalent atoms of the L-aspartate. These equivalences are S(SO_4^{2-})/C(L-Asp), O1(SO_4^{2-})/O1(L-Asp), O2(SO_4^{2-})/$O^{\delta1}$(L-Asp), O3(SO_4^{2-})/O2(L-Asp) and O4(SO_4^{2-})/C^{α}(L-Asp). Additionally, two of the L-aspartate atom sites (N and $O^{\delta2}$) are occupied in the complex with sulfate by water molecules W2 and W9. It was postulated previously that the protonation state of the substrate molecule can be directly correlated with its affinity for the active site of L-asparaginase (Lubkowski *et al.*, 1996). The primary difference between aspartate and asparagine is the presence of the negative charge on the carboxyl group of the former. The repulsive interaction between COO^- and the carbonyl O atom of Ala120 is believed to be responsible for the removal of the product molecule from the active site of L-asparaginase. For the same reason, a complex between L-asparaginase and L-aspartate can only be formed at acidic pH (Aghaiypour *et al.*, 2001*a*; Miller *et al.*, 1993). In the case of SO_4^{2-}, however, an even larger negative charge is carried by the ligand molecule and its binding to the active site of the enzyme does not seem to be pH dependent. The difference can be easily explained by the interaction with the hydroxyl group of Thr15 in both complexes. In the case of the enzyme–aspartate complex, the side chain of Thr15 presents the electron pair of its hydroxyl O atom towards the C^{γ} atom of the ligand, while the hydroxyl proton is contributed to the hydrogen bond with Tyr29. In complex with a sulfate anion, the hydroxyl group of Thr15 forms a strong hydrogen bond with the O4 atom of the ligand and provides additional stabilization of the complex. Consequently, the hydrogen bond between Thr15 and Tyr29 utilizes a proton donated by the hydroxyl group of the second residue. An additional factor contributing favorably to the binding of the SO_4^{2-} anion is its smaller size compared with the natural substrate.

An additional novel observation resulting from this atomic resolution study is the identification of a second sulfate molecule present near the active site (Fig. 7*a*). In contrast to the active-site ligand, however, this sulfate molecule is significantly more weakly bound. This sulfate anion was only identified in three monomers and its low-affinity binding is additionally reflected by only moderate structural conservation (Fig. 7*a*) and high *B* factors (>35 Å^2, compared with <10 Å^2 for the active-site SO_4^{2-}).

4. Conclusions

Crystallographic studies of the structures of bacterial L-asparaginases have been intensively pursued throughout the last decade both in our laboratory (Swain *et al.*, 1993; Miller *et al.*, 1993; Lubkowski, Wlodawer, Ammon *et al.*, 1994; Lubkowski, Wlodawer, Housset *et al.*, 1994; Lubkowski *et al.*, 1996; Palm *et al.*, 1996; Jaskólski *et al.*, 2001; Aghaiypour *et al.*, 2001*a*,*b*) as well as by others (Jakob *et al.*, 1997; Ortlund *et al.*, 2000). Despite the great amount of structural information about these enzymes that has been gained in this period, some details of the mechanism of enzymatic reaction are still not fully resolved. One of the more significant puzzles is the basis of the nucleophilicity of Thr15, as no basic residues could be identified in its close proximity. The details of the shape and size of the active-site cavity and its consequent selectivity of specifically sized and shaped ligands have also not been completely elucidated. It is generally accepted that the two natural ligands of L-asparaginases are L-asparagine and L-glutamine; however, some evidence indicates that somewhat larger substrates can also be hydrolyzed by these enzymes. Knowing the precise requirements set on the stereochemical characteristics of ligands will greatly help researchers understand the mechanism of catalysis and may shed new insights into the basis of the toxicity associated with L-asparaginase therapy. The atomic resolution structure of the complex of ErA with SO_4^{2-} presented in this work indirectly addresses some of these problems.

The atomic resolution of the structure presented here proves that the four independent active sites of this protein present in the tetramer are truly identical within the experimental error and thus are fully biologically equivalent. Moreover, with the exception of the flexible loop, the active-site residues as well as a number of water molecules have very

low mobility, indicating that their catalytic properties result strictly from their relative topological arrangement and the interactions with the substrate. The relatively isotropic shape of the displacement parameters (as represented by their thermal ellipsoids) for individual atoms located within the active-site flexible loop (Fig. 6) indicates that their motions are not correlated in the absence of a ligand (Lubkowski *et al.*, 1996; Lubkowski, Wlodawer, Ammon *et al.*, 1994) and that the observed topology of this loop emerges subsequent to the ligand binding. By comparison, the atomic displacement parameters of a flexible loop in retroviral integrase, another enzyme refined at a comparable resolution, have a highly correlated anisotropic distribution (Lubkowski *et al.*, 1999) suggesting the nature of the movement of the loop.

The remarkable similarity of individual monomers indicates that the effects of the crystal packing in ErA are rather minor. However, the specific conformation of readily disordered fragments of the protein, such as the N-termini of L-asparaginase, cannot be determined even at the very highest resolution.

Bacterial type II L-asparaginases are active as tetramers; to our knowledge, no functional dimers of these enzymes have been reported to date. However, inspection of the intimate dimer in the atomic resolution structure of ErA shows that two active sites are formed entirely by the two monomers, as are the other two active sites in the second dimer. The role of tetramer formation in the proper folding of the enzyme can be studied in terms of detailed specific interactions between the two intimate dimers; the atomic resolution structure provides an excellent opportunity for studying the intermolecular contacts. The quality of X-ray data and subsequently of the atomic coordinates is particularly important during analysis of the solvent structure. In our studies, we were able to determine the positions of 1451 water molecules, corresponding to 1.1 solvent molecule per protein residue in the crystal, which contains 44% solvent. Using quite stringent distance criteria, we found that over 30% of identified solvent molecules occupy sites that are completely conserved in all four monomers.

Finally, the work presented here is one of only a very few examples of structural determinations of a relatively large protein (~140 kDa per tetramer) at a resolution as high as 1.0 Å. As such, it will provide a valuable source for analysis of the stereochemical properties of individual residues in proteins.

This work was sponsored in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000. The work of KA at NCI was sponsored in part by a short-term fellowship from the Ministry of Health and Medical Education of Iran. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*a*). *Biochemistry*, **40**, 5655–5664.
Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*b*). *Biochim. Biophys. Acta*, **1550**, 117–129.
Beacham, I. & Jennings, M. (1990). *Today's Life Sci.* **2**, 40–42.
Borek, D. & Jaskólski, M. (2001). *Acta Biochim. Pol.* **48**, 893–902.
Broome, J. D. (1961). *Nature (London)*, **191**, 1114–1115.
Broome, J. D. (1968) *J. Cancer*, **22,** 595–602.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Carson, M. (1991). *J. Appl. Cryst.* **24**, 958–961.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Engh, R. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Epp, O., Steigemann, W., Formanek, H. & Huber, R. (1971). *Eur. J. Biochem.* **20**, 432–437.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Jakob, C. G., Lewinski, K., LaCount, M. W., Roberts, J. & Lebioda, L. (1997). *Biochemistry*, **36**, 923–931.
Jaskólski, M., Kozak, M., Lubkowski, J., Palm, G. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 369–377.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 208–230.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lough, T. J., Reddington, B. D., Grant, M. R., Hill, D. F., Reynolds, P. H. & Farnden, K. J. (1992). *Plant Mol. Biol.* **19**, 391–399.
Lubkowski, J., Dauter, Z., Yang, F., Alexandratos, J., Merkel, G., Skalka, A. M. & Wlodawer, A. (1999). *Biochemistry*, **38**, 13512–13522.
Lubkowski, J., Palm, G. J., Gilliland, G. L., Derst, C., Röhm, K.-H. & Wlodawer, A. (1996). *Eur. J. Biochem.* **241**, 201–207.
Lubkowski, J., Wlodawer, A., Ammon, H. L., Copeland, T. D. & Swain, A. L. (1994). *Biochemistry*, **33**, 10257–10265.
Lubkowski, J., Wlodawer, A., Housset, D., Weber, I. T., Ammon, H. L., Murphy, K. C. & Swain, A. L. (1994). *Acta Cryst.* D**50**, 826–832.
McArdle, P. (1995). *J. Appl. Cryst.* **28**, 65–65.
Mashburn, L. T. & Wriston, J. C. (1964). *Arch. Biochem. Biophys.* **105**, 450–452.
Merritt, E. A. & Murphy, M. E. P. (1994). *Acta Cryst.* D**50**, 869–873.
Miller, M., Rao, J. K. M., Wlodawer, A. & Gribskov, M. R. (1993). *FEBS Lett.* **328**, 275–279.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
North, A. C., Wade, H. E. & Cammack, K. A. (1969). *Nature (London)*, **224**, 594–595.
Ortlund, E., LaCount, M. W., Lewinski, K. & Lebioda, L. (2000). *Biochemistry*, **39**, 1199–1204.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Palm, G. J., Lubkowski, J., Derst, C., Schleper, S., Röhm, K.-H. & Wlodawer, A. (1996). *FEBS Lett.* **390**, 211–216.
Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
Roberts, J., Prager, M. D. & Bachynsky, N. (1966). *Cancer Res.* **26**, 2213–2217.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Swain, A. L., Jaskólski, M., Housset, D., Rao, J. K. M. & Wlodawer, A. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1474–1478.
Teeter, M. M., Roe, S. M. & Heo, N. H. (1993). *J. Mol. Biol.* **230**, 292–311.
Wlodawer, A., Hodgson, K. O. & Bensch, K. (1975). *J. Mol. Biol.* **99**, 295–299.
Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984). *J. Mol. Biol.* **180**, 301–329.