*Acta Cryst.* (1993). D**49**, 403–417

# Structure of Triosephosphate Isomerase from *Escherichia coli* Determined at 2.6 Å Resolution

BY M. E. M. NOBLE,* J. PH. ZEELEN AND R. K. WIERENGA†

*European Molecular Biology Laboratory, Meyerhofstrasse* 1, *D*-6900 *Heidelberg, Germany*

AND V. MAINFROID, K. GORAJ, A.-C. GOHIMONT AND J. A. MARTIAL

*Laboratoire de Biologie Moléculaire et de Génie Génétique, Université de Liège, Institut de Chimie, B*6, *B*-4000 *Sart-Tilman, Belgium*

(*Received* 16 *October* 1992; *accepted* 12 *March* 1993)

## Abstract

The structure of triosephosphate isomerase (TIM) from the organism *Escherichia coli* has been determined at a resolution of 2.6 Å. The structure was solved by the molecular replacement method, first at 2.8 Å resolution with a crystal grown by the technique of hanging-drop crystallization from a mother liquor containing the transition-state analogue 2-phosphoglycolate (2PG). As a search model in the molecular replacement calculations, the refined structure of TIM from *Trypanosoma brucei*, which has a sequence identity of 46% compared to the enzyme from *E. coli*, was used. An *E. coli* TIM crystal grown in the absence of 2PG, diffracting to 2.6 Å resolution, was later obtained by application of the technique of macro-seeding using a seed crystal grown from a mother liquor without 2PG. The final 2.6 Å model has a crystallographic *R* factor of 11.9%, and agrees well with standard stereochemical parameters. The structure of *E. coli* TIM suggests the importance of residues which favour helix initiation for the formation of the TIM fold. In addition, TIM from *E. coli* shows peculiarities in its dimer interface, and in the packing of core residues within the $\beta$-barrel.

## Introduction

The enzyme triosephosphate isomerase (TIM) is an object of interest because of both its functional and structural properties. The mechanism by which the dimeric enzyme catalyses the interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (GAP) (Fig. 1) has been the subject of study by biochemical techniques, techniques of molecular biology, and biophysical techniques such as nuclear magnetic resonance (NMR) and X-ray crystallography. Biochemical experiments have served to elucidate the free-energy profile of the catalytic pathway of the enzyme (Knowles & Albery, 1977); molecular biology has been used to examine the tools available to evolution at a molecular level (Blacklow & Knowles, 1990), and NMR and X-ray crystallography have been used to try to relate the resulting wealth of experimental results to structural details (Alber *et al.*, 1981; Davenport *et al.*, 1991; Knowles, 1991; Lolis & Petsko, 1990; Noble *et al.*, 1991; Wierenga, Noble, Postma *et al.*, 1991; Wierenga, Noble, Vriend, Nauche & Hol, 1991).

By chance, the determination of the structure of TIM has also revealed one of the protein folds used repeatedly by nature to provide a framework upon which a diverse range of catalytic functions has been built. This fold, consisting of an eight-fold repeat of a loop-$\beta$–loop-$\alpha$ motif, forms a regular eight-stranded parallel $\beta$-barrel (with strands termed $\beta$1–$\beta$8), surrounded by a coat of eight $\alpha$-helices (termed $\alpha$1–$\alpha$8) (Fig. 2). To date, 20 different functions have been seen to be performed by proteins having such a 'TIM barrel' fold (Brändén, 1991). Most of the proteins involved are enzymes, with the common feature of having their active site at the C-terminal end of the $\beta$-barrel. Most catalytic residues of TIM barrel proteins are contributed by the

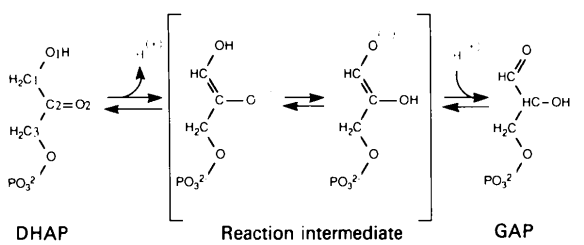


Fig. 1. The reaction catalysed by triosephosphate isomerase. Stereospecific interconversion of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (GAP) proceeds *via* a *cis* enediol(ate) intermediate. Atomic labels are given for DHAP atoms.

* Current address: Laboratory of Molecular Biophysics, University of Oxford, South Parks Road, Oxford OX1 3QU, England.
† Author to whom correspondence should be addressed.

$\beta$-strands and the loops connecting the $\beta$-strands with the following $\alpha$-helices (termed loops 1–8). The diversity of function of TIM barrel proteins lends considerable interest to understanding the structural properties and sequence determinants of this fold. Such understanding is not only relevant to the principles governing protein structure in general, but might also be applied in the design and construction of enzymes with novel structural and/or functional properties.

To date, the structure of TIM from three different sources has been reported; chicken, yeast and *Trypanosoma brucei*. Chicken TIM was the first to be solved (Banner *et al.*, 1975), although the quality of those crystals limited the resolution of the available data. Crystallographic studies on trypanosomal TIM (Wierenga, Kalk & Hol, 1987) and yeast TIM (Lolis, Alber *et al.*, 1990) extended the detail of the knowledge available about the TIM structure. Trypanosomal TIM has been studied to analyse the atomic interactions and structural changes involved in ligand binding (Noble *et al.*, 1991; Verlinde *et al.*, 1991; Wierenga, Noble, Postma *et al.*, 1991). The resulting information has been used in attempts to design specific inhibitors of the enzyme from *T. brucei* (Verlinde, Rudenko & Hol, 1992), and also to rationalize the binding of fortuitously discovered specific inhibitors thereof (Hol *et al.*, 1991), with the ultimate aim of structure-based drug design. The structure of yeast TIM either free (Lolis, Alber *et al.*, 1990), or in complex with transition-state and intermediate analogues (Davenport *et al.*, 1991; Lolis & Petsko, 1990) has since been the start point of quantum and molecular mechanical studies aimed at a further understanding of the reaction mechanism (Bash *et al.*, 1991; Davenport *et al.*, 1991).

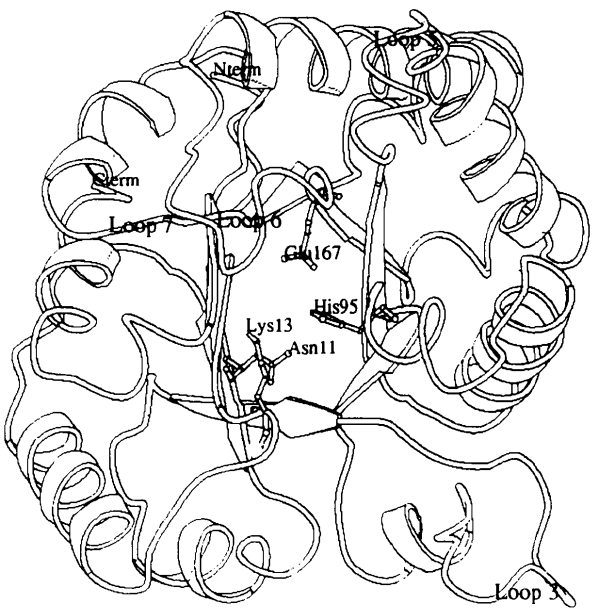


Fig. 2. Cartoon representation of one monomer of triosephosphate isomerase (trypanosomal TIM numbering). The eightfold repeat of a (loop-$\beta$–loop-$\alpha$) motif is viewed along the barrel axis from the C-terminal end of the $\beta$-strands. The explicitly drawn and labelled residues (Asn11, Lys13, His95 and Glu167) all form hydrogen bonds with bound substrate analogues. Also labelled are loops 5 (residues 129–139), 6 (residues 167–180) and 7 (residues 210–214). These loops, which are known to undergo conformational changes upon ligand binding, are depicted in their 'closed' conformation. The last labelled loop (loop 3, residues 70–80) is responsible for many of the contacts between monomers of the dimeric enzyme. This picture was generated using the program *XRENDER* (Noble, unpublished work).

TIM is a homodimeric enzyme with subunits consisting of around 250 amino acids (Fig. 3). *E. coli* TIM has 255 residues (Pichersky, Gottlieb & Hess, 1984), and the numbering scheme used throughout this paper will be that of *E. coli* TIM, unless otherwise stated. The alignment of the *E. coli* TIM sequence with those of other TIMs of known structure is given in Fig. 3. The dimer interface is formed principally from loops 1, 2 and 3 from each subunit. Of these, the most intimate association is formed by loop 3, the 'interface loop', which extends from the bulk of its own subunit into a complementary pocket in its partner subunit. The threonine at the tip of this loop (Thr77), forms a long hydrogen bond with one of the catalytically important residues of the partner subunit (Asn11), and an indirect interaction, *via* the carboxyl O atoms of Glu97, with another (Lys13). Other residues of loop 3 also form van der Waals contacts with active-site residues of their partner subunit. This probably explains the observations that, whilst no allostery exists between the two active sites, TIM is only measurably active as a dimer (Zabori, Rudolph & Jaenicke, 1980).

Binding of substrate or substrate analogues is accompanied by changes in conformation of loops 5, 6 and 7 (Wierenga, Noble, Postma *et al.*, 1991). This is most marked in the case of loop 6, the 'flexible loop,' the tip of which undergoes a C$\alpha$ movement of some 7 Å in response to ligand binding, and which closes off the active site from solvent. Closure of the flexible loop is important to prevent an unwanted phosphate elimination reaction which may otherwise occur from the reaction intermediate (Pompliano, Peyman & Knowles, 1990). The isomerization reaction itself proceeds *via* the abstraction of a proton from C1 of DHAP to produce a *cis*-enediolate intermediate (Fig. 1) (Rieder & Rose, 1959). This step is followed by the transfer of a proton from O1 to O2, and finally the readdition of a proton to C2.

A glutamate (Glu169) is implicated in the transfer of a proton between the two substrate C atoms. The side chain of this glutamate can occupy two different positions, depending on the contents of the active site. In ligand-free TIM, it is found in a 'swung out' conformation, interacting with residues His97 and Ser98, whilst in TIM complexed to substrate analogues, it occupies a 'swung in' conformation, interacting closely with the ligand (Lolis & Petsko, 1990;

Noble *et al.*, 1991). An uncharged histidine (His97) is thought to mediate transfer of a proton between the two O atoms, and to act as an electrophile to polarize the corresponding C—O bonds (Komives *et al.*, 1991). Residues Asn11 and Lys13 interact with O1 and O2 of the substrate respectively.

Here we present the determination of the structure of TIM from *E. coli*. This structure broadens the species database of known TIM structures to include a prokaryotic representative. Since an effective expression and mutation system exists for *E. coli* TIM, this structure also provides the basis for the design of a range of mutants of TIM aimed at probing both the structural properties of the TIM barrel, and the dimer-interface interactions of this enzyme.

## Materials and methods

### (1) *Purification, crystallization and data collection*

*E. coli* triosephosphate isomerase was purified from an over-expressing strain (Mainfroid *et al.*, manuscript in preparation). Crystals were grown by the vapour-diffusion method using protein dissolved at a concentration of 5 mg $ml^{-1}$ in 25 m*M* Tris/HCl pH 7.6, containing 20 m*M* NaCl, 1 m*M* ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), $NaN_3$ and 1 m*M* 2-phosphoglycolate, and a well solution of 22% PEG 6000 in 200 m*M* 2-[*N*-morpholino]ethanesulfonic acid (MES)/NaOH pH 6.4, also containing 1 m*M* of each DTT, EDTA and $NaN_3$. In most crystallization experiments, needle-like crystals grew over a period of 2 d but these were not suitable for X-ray analysis. Only in a few wells, in which no needle-like crystals grew, did larger crystals grow over a period of several weeks, the best of which (0.6 × 0.2 × 0.2 mm) could be used to collect a three-dimensional data set.

Data were collected on an Enraf–Nonius FAST area detector using radiation from an Elliot GX21 rotating-anode generator equipped with a graphite monochromator, in 0.1° rotation frames exposed for 120 s each. 91° of data were collected around an arbitrary axis, and a further 47° about another axis 90° away from the first. Statistics describing the quality of the data are presented in Table 1. The *AUTI* option in *MADNESS* (Messerschmidt & Pflugrath, 1987) was used to determine that the cell was primitive orthorhombic. Inspection of axial

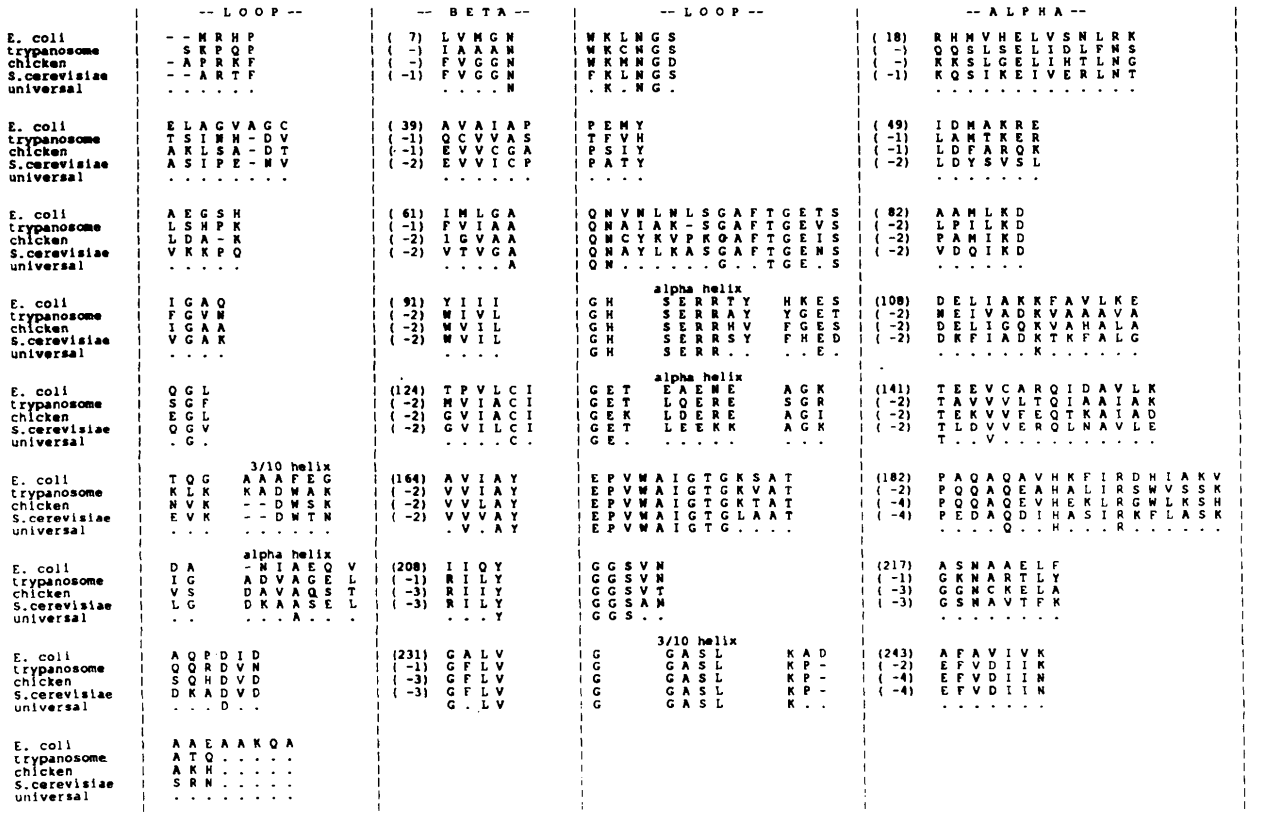


Fig. 3. Alignment of sequences of triosephosphate isomerases of known three-dimensional structure. The bottom line has entries in sequence positions which are conserved in 13 TIM sequences (according to Wierenga, Noble & Davenport, 1992). The sequence alignment is broken up into eight loop-$\beta$–loop-$\alpha$ units. Additional secondary structural elements within the loop regions are also labelled. The alignment here is based upon the structural superposition of the four known structures.

Table 1. *Statistics of the data collected*

| | FAST | IP |
|---|---|---|
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell edges (Å) | 68.4 47.1 152.1 | 67.7 46.8 151.3 |
| Maximum resolution (Å) | 2.8 | 2.6 |
| Observations | 28658 | 52487 (51869 fully recorded, 618 partial) |
| Overall merging $R$* (%) | 7.4 (25550 observations = 10836 unique reflections used in merging) | 8.3 (51476 observations = 12801 unique reflections used in merging) |
| 2.89–2.8 Å merging $R$ (%) | 21.7 | 19.0 |
| Unique reflections | 10852 | 13812 |
| Completeness (%) | | |
| 40.0–2.80 Å | 85.4 | 89.4 |
| 2.89–2.80 Å | 43.3 | 63.6 |

* Merging $R = \{[\Sigma_h\Sigma_i(\bar{I}_h - I_{h,i})]/(\Sigma_h\Sigma_i|\bar{I}_h|)\} \times 100\%$.

reflections indicated that the three twofold axes of this crystal class were all screw axes, identifying the space group as $P2_12_12_1$. Data collection and initial reduction was performed by the *MADNES* package, profile fitting by *XDS* (Kabsch, 1988), and subsequent batch scaling with programs of the Groningen *BIOMOL* suite.

(2) *Structure solution by molecular replacement*

The search model used for molecular replacement was the structure of trypanosomal TIM, which has been refined at 1.83 Å resolution (Wierenga, Noble, Vriend, Nauche & Hol, 1991), and which has 46% sequence identity to TIM from *E. coli*. Loops known to be variable in conformation were deleted from the model, *i.e.* the flexible loops of subunits 1 and 2 (residues 167–180), and residues of the neighbouring loop (loop 7, residues 211–214). Also deleted were residues flanking the locations of insertions or deletions in the *E. coli* sequence relative to the *T. brucei* sequence as predicted by sequence alignment according to the program *BESTFIT* of the *GCG* package (residues 31–32, 68–69, 194–196 and 240–241, trypanosomal TIM numbering) (Devereux, Haeberli & Smithies, 1984). The resulting dimeric model, consisting of 2 × 216 residues, was rotated so as to align the local twofold axis with the crystallographic $z$ axis. Structure factors for space group $P1$ were calculated for this molecule in a cell of dimensions 128 × 128 × 128 Å by the program *GENSFC* of the *CCP*4 suite (SERC Daresbury Laboratory, 1979). The cross-rotation function between this model and the *E. coli* TIM data was calculated with the fast-rotation-function algorithm of Crowther (Crowther, 1972) as implemented in the program *ALMN*. The limits of Patterson integration used were 3.0–25.0 Å, with resolution limits for included reflections of 15.0–4.0 Å. The highest two peaks in the resulting function were of magnitude 6.6 and 6.3$\sigma$, respectively. For comparision, the next highest peak was only 3.6$\sigma$ in height. The top two peaks were found at $\alpha = 68.5$, $\beta = 25.1$, $\gamma = 316.5°$, and $\alpha = 73.9$, $\beta = 23.4$, $\gamma = 131.5°$, respectively, with angles defined according to the eulerian angle convention used by the fast-rotation function. The existence of two peaks of approximately equal height with similar values for the angles $\alpha$ and $\beta$, but approximately 180° apart in $\gamma$, is what would be expected for the two alternative rotations which would be required to correctly orient a dimeric search model which had been pre-aligned with its local twofold axis parallel to the $z$ axis.

As a further confirmation of the correctness of this solution, the relation between the two peaks, corresponding to a polar rotation with $\kappa = 180$, $\omega = 278.2$ and $\varphi = 67.2°$, resembles the values found for the self-rotation function using the *E. coli* data ($\kappa = 180$, $\omega = 278.1$ and $\varphi = 65.4°$). After application of this rotation, the $T_2$ translation function (Crowther & Blow, 1967) was calculated using the *CCP*4 programs *GENSFC*, *CADLCF*, *TFSGEN* and *FFTKW*. Normalized structure factors ('$E$' values) for data between 20.0 and 3.0 Å were used in the calculation of the translation function. The highest peak had a height of 11.9$\sigma$, equal to 7.8$\sigma$ above the average map value. The corresponding section of the translation function is plotted in Fig. 4. For comparison, the second highest distinct peak had a height of only

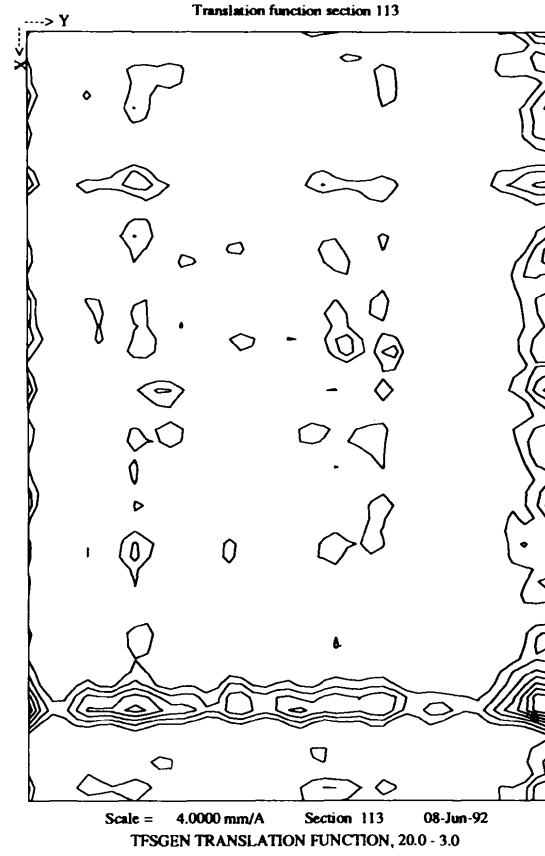


Fig. 4. Section containing the highest peak of the $T_2$ translation function (*TFSGEN*), calculated after applying the appropriate rotation from *ALMN* to the trypanosomal TIM search model. The section corresponds to the *ab* plane at $c = 0.38$, and is contoured starting from 1$\sigma$ above the mean in steps of $\sigma$.

4.6$\sigma$ above the mean value of the map. The *R* factor for data between 10.0 and 6.0 Å for the model after application of this translation was 53.0%.

(3) *Structure refinement*

Structure refinement was begun with rigid-body refinement of the model against data between 10.0 and 6.0 Å, using the *TNT* refinement package (Tronrud, Ten Eyck & Matthews, 1987). Each subunit was allowed to move independently during this refinement. This lead to an r.m.s. coordinate shift of 1.3 Å in the first refinement cycle, and a drop in *R* factor from 53.0 to 48.5%. Further rigid-body refinement against data between 6.0 and 3.0 Å lead to a drop in *R* factor from 51.5 to 48.5% for this resolution range. At this point, $\sigma_A$-weighted (Read, 1986) $2F_o - F_c$ and $F_o - F_c$ maps were calculated, and the model was assessed against the density. The correctness of the solution was confirmed by the quality of the fit of the model to the apparent electron density, with several sequence differences being clearly indicated. One such example was the density for atoms beyond C*B* of Met9 of *E. coli* TIM, an alanine in the search model (Fig. 5).

The structure was refined using a combination of conventional least-squares refinement (*TNT* package), and molecular dynamics X-ray refinement (*X-PLOR*) (Brünger, Kuriyan & Karplus, 1987), interspersed with manual fitting of the model to electron-density maps using the programs *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) and *FRODO* (Jones, 1985) running on an ESV graphics workstation. In the first round of interactive model building, the sequence was mutated from that of *T. brucei* TIM to that of *E. coli* TIM for subunit 1, and the fit of the model to its density was optimized. At this stage, atoms from residues not common to *E. coli* and trypanosomal TIMs were assigned a *B* factor of 25 Å$^2$, whereas common atoms kept the *B* factor of the trypanosomal enzyme. In regions where the tracing of the chain became ambiguous, considerable sections of sequence were deleted, to avoid model bias at a later stage in the refinement procedure. This lead to the deletion of residues 20–21, 30–35, 69–74, 154–163, 196–200 and 239–243, as well as residues 169–182 (loop 6) and residues 211–215 (loop 7). After completing mutation and rebuilding of subunit 1, the coordinates of subunit 2 were generated by application of the local twofold axis. From this point on, the two subunits were refined independently of each other. The first step in this refinement was a further rigid-body refinement using *TNT*, followed by an *X-PLOR* simulated-annealing run. This and subsequent *X-PLOR* runs followed a standard protocol summarized in Table 2.

During the course of refinement, the maps steadily improved, allowing the cautious readdition of the loops which had been initially deleted. Where appropriate, this was performed with the *LEGO* options of the program *O*, with side-chain conformations for the new residues being selected from the rotamer

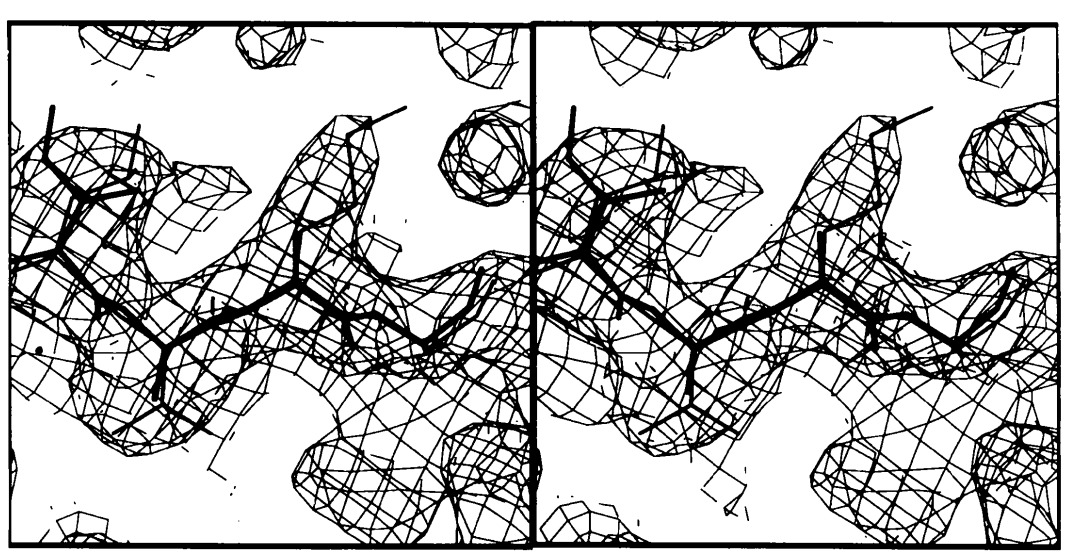


Fig. 5. Electron density consistent with the sequence difference at position 9. $\sigma_A$-weighted $2mF_o - DF_c\ \alpha_{calc}$ electron density, calculated with data between 10 and 3 Å, is shown contoured at a level of 1$\sigma$. The calculated structure factors and phases are derived from the trypanosomal-search model after application of the translation-function solution, and rigid-body refinement with the *TNT* package. At this point, the *R* factor was 49%. The model drawn in thick lines is the trypanosomal TIM model used in phasing (alanine at position 9), and the model in thin lines represents the final refined *E. coli* TIM model (methionine at position 9). Clear electron density extends beyond C*B* of the phasing model.

Table 2. *Details of the X-PLOR simulated-annealing protocol*

| | |
|---|---|
| Non-bonded parameters | Dielectric constant = 1, maximum distance = 7.0 Å, switching function 6.0–6.5 Å, full atomic charges set to zero (Lys, Arg, Glu, Asp) |
| Relaxed stage | |
| Tolerance (Å) | 0.05 |
| VDW repulsion turned on | 50 cycles |
| VDW repulsion turned off | 50 cycles |
| Annealing stage | |
| Initial temperature (K) | 1500 |
| Temperature step (K) | 25 |
| Cycles at each step | 25 |
| Tolerance (Å) | 0.2 |
| Final temperature (K) | 300 |
| Regularization stage | 100 steps of conjugate-gradient refinement |

database associated with the program. For minor adjustments of fit in essentially correct regions, however, the program *FRODO* was preferred. When the *R* factor of the model had reached a value below 20%, limited atomic *B*-factor refinement was allowed, such that atomic *B* factors were constrained to lie between 10 and 50 Å$^2$. This limitation was applied because of the restricted resolution of the available data.

Refinement was stopped after a total of ten sessions of interactive graphics, between which four simulated-annealing *X-PLOR* runs and 100 cycles of *TNT* conjugate-direction refinement had been performed, including the above mentioned constrained *B*-factor refinement. At the end of the refinement protocol, a complete chain tracing was achieved everywhere except in the region of the two flexible loops, and at the C termini, of which four residues from each subunit could not be built. In the region of the flexible loops, only broken density was visible at the 0.75$\sigma$ level. This density did not allow reliable chain tracing. Refinement statistics of the model at this point are given in the 'intermediate' column of Table 3.

### (4) *Higher resolution data set*

To improve the quality of available crystals, the technique of macro-seeding was used (Thaller *et al.*, 1981). For this purpose, a tray of hanging-drop crystallization trials without 2PG was chosen in which, after several months of equilibration, most droplets contained the usual needle-like crystals, and some droplets contained neither crystals nor precipitated protein. A seed crystal of dimensions 0.05 × 0.05 × 0.5 mm was chosen from amongst the needle-like crystals, and transferred to one of the droplets in which no nucleation had occurred. This droplet was then replaced over its original well solution, and allowed to equilibrate for a further month. At the end of this period, a good quality crystal of dimensions 0.15 × 0.15 × 0.5 mm was found in the droplet which could be successfully mounted in a siliconized quartz capillary. The conditions for the macro-seeding experiment are summarized in Table 4.

Table 3. *Refinement statistics of the intermediate and final models*

| | Intermediate | Final |
|---|---|---|
| Data set | FAST | IP |
| Protein atoms | 3744 | 3772 |
| Solvent atoms | 14 (1 $PO_4^{3-}$ + 1 2PG) | 172 (waters) |
| Crystallographic parameters | | |
| Resolution (Å) | 25.0–2.8 | 25.0–2.6 |
| Reflections in refinement | 10824 | 13812 |
| *R* factor* (%) | 14.6 | 11.9 |
| Luzzati plot estimated error (Å) | 0.18–0.20 | 0.15–0.18 |
| Geometric parameters† | | |
| R.m.s. bond-length deviation (Å) | 0.024 | 0.018 |
| R.m.s. bond-angle deviation (°) | 3.1 | 2.4 |
| $\chi$-imperfection value‡ (°) | 46 | 36 |
| Residues with deviating $\varphi,\psi$§ | 36 | 9 |
| R.m.s. $\Delta B$ for bonded atoms (Å$^2$) | 25 | 8 |

* *R* factor = $\{[\sum_h|(F_{obs,h} - F_{calc,h})|]/(\sum_h F_{obs,h}|)\} \times 100\%$.
† Unless otherwise stated, calculated with the *TNT* refinement package.
‡ R.m.s. deviation from Ponder & Richards $\chi_1\chi_2$ torsion-angle database.
§ Identified by the programs *PCAP* and *PHIPSI* of the Groningen *BIOMOL* suite.

Table 4. *Details of crystal growth conditions and macro-seeding*

| | FAST | IP | |
|---|---|---|---|
| Crystal | | Donor well | Acceptor well |
| Buffer | 200 m*M* MES/NaOH | 200 m*M* MES/NaOH | 200 m*M* MES/NaOH |
| pH | 6.4 | 5.75 | 6.0 |
| 2PG (m*M*) | 1 | — | |
| Additives | 1 m*M* (EDTA, DTT, $NaN_3$) | 1 m*M* (EDTA, DTT, $NaN_3$) | 1 m*M* (EDTA, DTT, $NaN_3$) |
| Precipitant | 22% PEG6000 | 22% PEG6000 | 22% PEG6000 |
| Crystal size (mm) | 0.6 × 0.2 × 0.2 | 0.5 × 0.05 × 0.05 | 0.5 × 0.15 × 0.15 |
| Diffraction limit (Å) | 2.8 | — | 2.6 |

A three-dimensional data set was collected from this crystal on a MAR research image-plate device, in 100 frames of 1° oscillation per frame. An exposure of 900 s per frame was chosen, using X-rays from a Siemens rotating-anode generator equipped with a graphite monochromator. Although the crystal used in the image-plate data collection was smaller than that used in the FAST data collection, it was visibly more regular and a larger collimator was used for data collection. This is probably the reason for the higher resolution diffraction. Data collection was controlled with the program *XIPS*, and subsequent batch scaling was carried out through the programs *ABSCALE, ROTAVATA, AGROVATA* and *TRUNCATE* of the *CCP*4 package. Merging and completeness statistics for this data set are given in the image-plate (IP) column of Table 1. Since this second crystal was grown from a mother liquor without substrate analogue present, no attempt was made to merge the two data sets. The cell dimensions of the second crystal were also somewhat different. Instead, the intermediate structure was used for rigid-body refinement against the new data between resolutions of 6.0 and 4.0 Å. The structure was then further refined by a combination of molecular dynamics (*X-PLOR*), conventional least-squares refinement (*TNT* package), and interactive model building (*FRODO*).

With the inclusion of higher resolution data in the refinement, clear chain tracings could be deduced for

the four remaining carboxy-terminal residues of subunit 1, and two of the carboxy-terminal residues of subunit 2. After inclusion of these residues in the refinement, as well as several rounds of refinement and interactive building, density became apparent in the region of the two 'missing' flexible loops. These could be reliably traced in a conformation resembling the 'open' conformation as discussed previously (Wierenga, Noble, Postma *et al.*, 1991). The constraints on *B* factors were removed, and replaced by restraints on the *B*-factor difference for covalently bonded atoms. Water molecules were added to the model, chosen from the highest peaks in an $F_o - F_c$ difference electron-density map calculated after completion of the refinement of the protein model. These peaks were accepted as water molecules only if they had a suitable protein environment (maximum distance to hydrogen-bonding partner < 3.1 Å, minimum distance to hydrogen-bonding partner > 2.5 Å, and minimum distance to a C atom > 3.3 Å). Final refinement statistics for the model are given in Table 3.

(5) *Free R-factor calculations*

Calculation of the free *R* factor (Brünger, 1992) was performed with data between 10.0 and 2.6 Å, after completion of the model refinement. Although the original free *R*-factor calculations have been tested by removing a fraction of the data at the beginning of the refinement, the procedure tested here was also suggested (Brünger, 1992). As a first step, water molecules were deleted from the final model. This model, which had been refined principally using the *TNT* package, was then 'equilibrated' against the *X-PLOR* force field using ten cycles of *B*-factor refinement (B), followed by one cycle of simulated annealing (SA), followed by ten cycles of *B*-factor refinement (B). The resulting model was then subjected to a further round of B–SA–B refinement against a randomly selected 90% subset of the full IP data set. The *R* factor against the remaining 10% of the data (the 'free *R* factor') was calculated before and after this last round of refinement.

To provide a point of comparison, a similar protocol was performed using the intermediate model obtained from the 2.8 Å FAST data set, after applying rigid-body refinement with the IP data set to correctly position the model. The same sets of structure factors of the IP data set (see above) were used for these calculations. The free *R* factor of this model, which by several other criteria was inferior, and that of the final model are summarized in Table 5.

(6) *Structure analysis*

Analysis of the structure was performed chiefly on an Evans and Sutherland ESV10 workstation. Cavity and accessible surface-area calculations were carried out with the *MSP* package (Connolly, 1985), and hydrogen bonding was analysed with *WHAT IF* (Vriend, 1990) using the same criteria as previously described (Wierenga, Noble, Vriend, Nauche & Hol, 1991). Before comparison of the structures of different monomers of TIM, superposition was performed on the basis of 104 C*A* atoms common to the structural $\alpha$-helices and $\beta$-strands common to all of the studied TIMs. The $\chi$-imperfection values, calculated with the program *TORSIONS* (Noble, unpublished) are the r.m.s. differences between observed ($\chi_1\chi_2$) values and the nearest preferred-cluster values of ($\chi_1$ and $\chi_2$), observed by Ponder & Richards (1987) in a database of well refined protein structures.

Table 5. *Results of the free-R calculations*

$\chi_1\chi_2$ represents the r.m.s. deviation from the Ponder & Richards database of side-chain torsion angles. *R* represents the conventional *R* value for data from 10 to 2.6 Å resolution against the entire IP data set. $R_{free}$ represents the *R* factor against a randomly chosen 10% subset of the IP data set.

| | Final | | | Intermediate | | |
|---|---|---|---|---|---|---|
| | $\chi_1\chi_2$ (°) | *R* (%) | $R_{free}$ (%) | $\chi_1\chi_2$ (°) | *R* (%) | $R_{free}$ (%) |
| Before rigid body* | — | | | 47.6 | 33.87 | 34.87 |
| Start | 41.2 | 22.74 | 23.43 | 47.6 | 26.67 | 29.70 |
| Equilibrated† | 38.3 | 17.83 | 18.28 | 43.9 | 19.86 | 20.08 |
| After free *R*‡ | 37.9 | 18.25 | 27.91 | 42.7 | 19.73 | 31.07 |

* To account for slight differences in cell dimensions, the intermediate structure, but not the final structure, was refined as a rigid body against the IP data set prior to the free-*R* protocol.

† The equilibriated model was calculated after the equilibration round of *B*-factor annealing-*B*-factor (BAB) refinement against the entire IP data set.

‡ The after free-*R* model was calculated after the free-*R* BAB refinement cycle against the IP data set from which the data used in the free-*R* calculation were omitted.

## Results

(1) *Model quality*

The structure of *E. coli* TIM was determined and refined to a low *R* factor with good agreement to standard geometry values (Table 3). The real-space *R* factor for the structure, as calculated by the program *O–RSF* (Jones, Zou, Cowan & Kjeldgaard, 1991), is shown as a function of residue number in Fig. 6. From this figure it can be seen that most regions of the structure agree well with the $2F_o - F_c$ electron density used in the calculation of the real-space *R* values. The estimated coordinate error according to Luzzati (Luzzati, 1952) is in the range 0.15–0.20 Å (data not shown). This value is, however, too low, since the intermediate 2.8 Å structure gave an estimated coordinate error of approximately the same size, despite an r.m.s. coordinate difference of 0.35 and 0.80 Å between the two structures, calculated for the subunit 1 C$\alpha$ atoms of the 104 core residues or all protein atoms (not including residues with *B* >

50 Å$^2$, loop 6, loop 7, regions of insertions, deletions and the carboxy termini), respectively.

The Ramachandran plot (Ramachandran & Sasisekharan, 1968) for the entire structure is shown in Fig. 7. The two residues lying in the lower right quadrant of the diagram correspond to Lys 13 (one of the catalytic residues) from subunits 1 and 2, respectively. This aberrant $\varphi,\psi$ combination has been observed in all TIMs to date, and probably represents a necessary strain to provide optimal active-site architecture. A comparison of the coordinate difference between the two independent monomers gives an r.m.s. value of 0.55 and 0.43 Å for the 1008 common main-chain atoms and the 104 core C$\alpha$ atoms, respectively; this represents an upper error for the combined coordinate error in the two subunits. The correlation between the atomic *B* factors of subunit 1 and subunit 2 is 0.66.

(2) *Model quality evaluated by free R factors and $\chi$-imperfection value calculations*

The free *R*-factor protocol described here tests the ability of this statistical parameter to identify the better of two essentially correct models, *i.e.* the 'final' and 'intermediate' models. These models differ in details of conformation, and model completeness [the intermediate model contained 3744 (non-H) protein atoms, whereas the final model contained 3772 (non-H) protein atoms]. The results presented in Table 5 show that after equilibration against the higher resolution IP data set, the final model had a conventional *R* factor of 17.8%, while the intermediate model had a conventional *R* factor of 19.9%. The *R* factor of the final model (17.8%) is high compared to the quoted *R* factor in Table 3 (11.9%). This difference is because of the absence of water molecules from the model used in free-*R* calculations, as well as differences in the *TNT* and *X-PLOR* packages (*e.g.* no bulk solvent modelling was included in the *X-PLOR* calculations). Free-*R* refinement of each structure gave rise to a free *R* factor of 27.9% for the final model, and 31.1% for the intermediate model. The discrimination of the conventional *R* factor [(19.9 − 17.8)/17.8 = 11.8%] is similar to the discrimination of the free *R* value [(31.1 − 27.9)/27.9 = 11.5%].

The free *R* factor is 10.1 and 11.2% higher than the normal *R* factor for the final and intermediate structures respectively (discrimination of 10.9%). These differences are similar for free *R*-factor refinements using the standard protocol (Brünger, 1992). The discrimination of the $\chi$-imperfection values after the free *R*-factor refinement is 12.7%. Interestingly, the lower quality of the intermediate model is more strongly highlighted by the respective $\chi$-imperfection values of the original models discussed in Table 3; the discriminating value is (46 − 36)/36 = 28%.

(3) *Structure of E. coli TIM*

The C*A* trace of subunit 1 of *E. coli* TIM is compared to that of subunit 1 of trypanosomal TIM in Fig. 8. The similarity in fold extends to the C*A* positions of residues of the flexible loop, which are clearly defined in the 'open' conformation (Wierenga, Noble, Postma *et al.*, 1991). The fit of the flexible-loop residues of subunit 2 into their omit electron density is shown in Fig. 9. This fit is good,

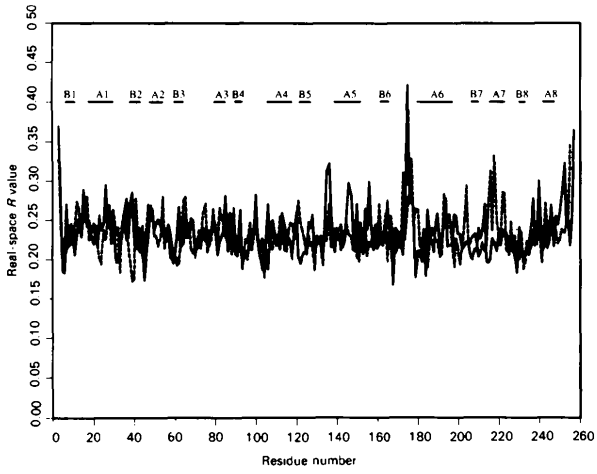


Fig. 6. Main-chain real-space *R* value. The real-space *R* value is plotted as a function of residue number for subunit 1 (solid line) and subunit 2 (broken line).

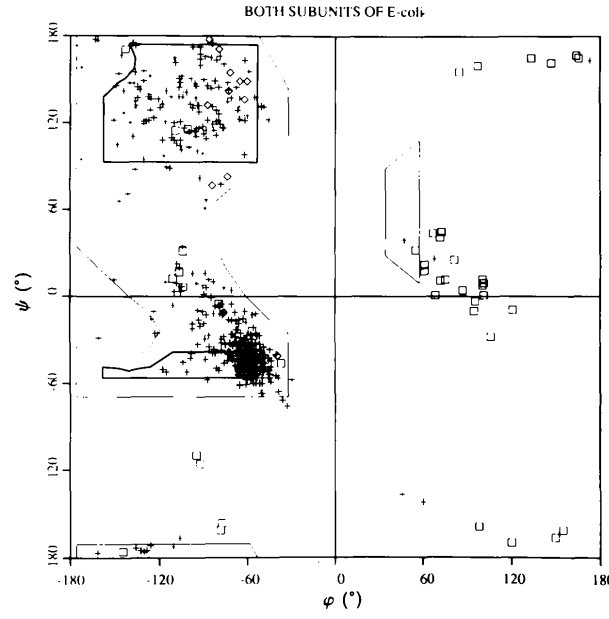


Fig. 7. Ramachandran plot of *E. coli* TIM. The scatter of main-chain torsion angles observed in the final model is presented. Thick lines surround the fully allowed regions (Ramachandran & Sasisekharan, 1968), and the thin lines surround regions allowed with a relaxed van der Waals constraint. Plusses are used for the observed $\varphi/\psi$ combinations, except in the case of proline (diamonds) and glycine (squares).

despite being the least well defined part of the model, as judged from the real-space *R*-factor plot in Fig. 6. Residues of the flexible loop have also the highest atomic *B* factors (Fig. 10). The r.m.s. difference for the C*A* atoms of residues 169–182 between subunit 1 trypanosomal and *E. coli* models is 0.86 and 0.98 Å for subunits 1 and 2, respectively. This similarity is surprisingly good for the most flexible part of the protein model (Fig. 10). Consistent with the open state of the flexible loop is the 'swung out' conformation of the catalytic glutamate (Glu169). This conformation, unsuitable for catalysis, has always previously been observed in the absence of a ligand bound at the active site. The flexible loop of subunit 1 is involved in crystal contacts, while that of subunit 2 is not.

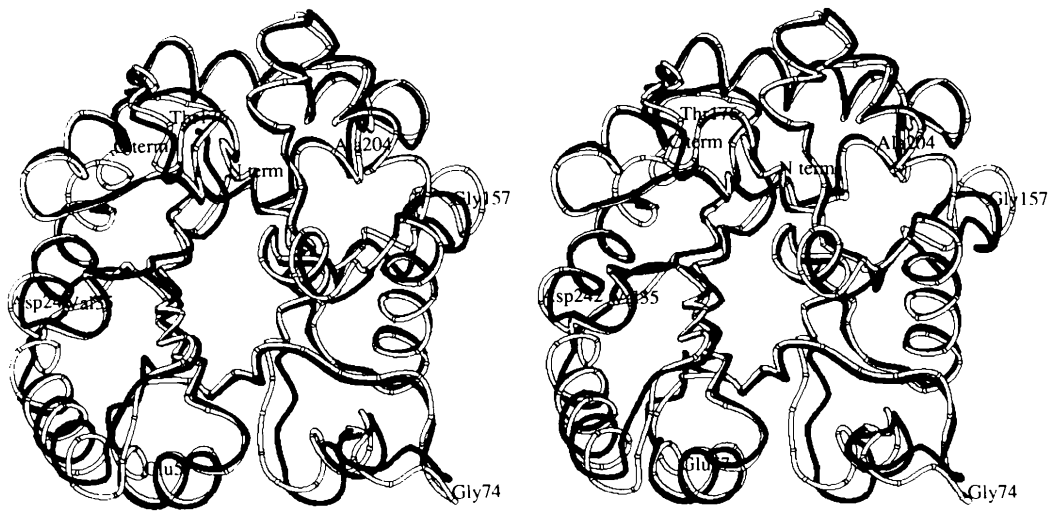


Fig. 8. Stereo picture of the comparison of the C*A* trace of subunit 1 of *E. coli* TIM (white) with subunit 1 of *T. brucei* TIM (black). The plot is made after superposition of the two monomers on the basis of 104 C$\alpha$ atoms common to the structural $\alpha$-helices and $\beta$-sheets of the two molecules. Regions of difference discussed in the text are labelled: Val35, Glu57, Gly74, Gly157, Thr176, Ala204 and Asp242. The equivalent comparison for subunit 2 of *E. coli* TIM gives similar results (data not shown). For reference, the N and C termini are also labelled, as is the flexible loop (loop 6) near Thr176.

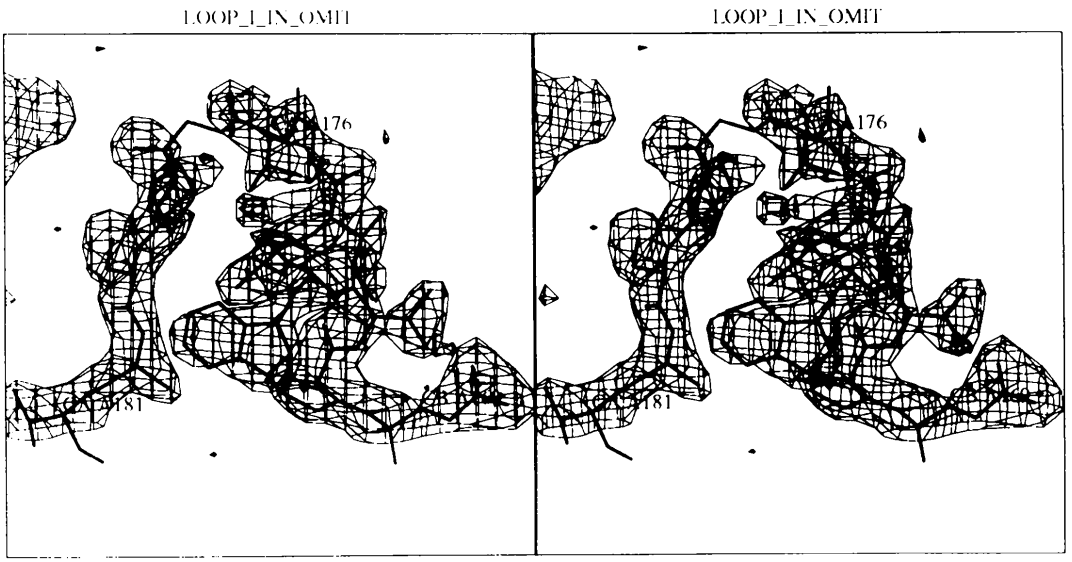


Fig. 9. The fit of the flexible loop of subunit 2 into 'omit' electron density. After completion of the refinement, residues 169–182 of subunit 2 were deleted from the model, which was then subjected to five cycles of combined positional and *B*-factor refinement by the *TNT* package. After this, the truncated model was used to provide calculated structure factors and phases for the calculation of a $\sigma_A$-weighted $F_o - F_c$ $\alpha_{calc}$ map. This map is shown in thin lines, contoured at a level of 1$\sigma$. The atomic model (thick lines) in the background is the interpreted structure for the flexible loop.

(4) *Crystals grown in the presence of 2PG*

The FAST data set, collected from a crystal grown in the presence of 2PG, showed only weak electron density in the two active sites. This electron density could be interpreted as a phosphate bound at the active site of subunit 1, and a 2PG molecule at the active site of subunit 2. A difference Fourier calculated using data from the FAST data set and phases from the final model after rigid-body refinement against the FAST data set shows this weak density in the active site of subunit 2 (Fig. 11). This map also reveals electron density next to the 'swung out' catalytic glutamate indicating that this residue has partial occupancy of the 'swung in' conformation. A 2PG molecule was built into this electron density, and a phosphate molecule into the corresponding electron density in the active site of subunit 1. On refinement, the thermal factors for the ligand molecules tended to extremely high values ($B > 100$ Å$^2$ data now shown), consistent with low occupancy of the ligands. Because of this fact, no further analysis of this structure was carried out.

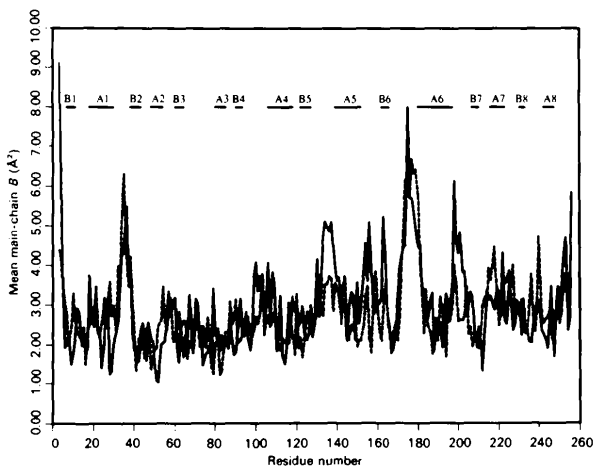


Fig. 10. *B*-factor plot of *E. coli* TIM. The mean main-chain *B* factors of subunit 1 (solid line) and subunit 2 (broken line) of the final model are plotted as a function of residue number. The secondary structure assignment is plotted above to indicate regions of $\alpha$-helix and $\beta$-sheet. The regions of highest *B* factor are the flexible loops (loop 6) of subunit 1 and subunit 2.

(5) *Insertions and deletions in E. coli TIM*

The sequence of *E. coli* TIM has 255 residues compared to the 249 residues of trypanosomal TIM. Electron density for trypanosomal TIM indicates that the crystallized species begins with residue Ser2, without the N-terminal methionine residue, whereas in *E. coli* TIM, electron density exists for all of the coded residues, from the first methionine (Met3). The sequence of *E. coli* TIM contains three insertions compared to trypanosomal TIM, residues Ala36, Leu72 and Asp242. These insertions map respectively to the loop before $\beta$2, loop 3 and the start of $\alpha$8. In addition, *E. coli* TIM has a five-residue C-terminal extension compared to trypanosomal TIM. *E. coli* TIM is one residue shorter than trypanosomal TIM at the N terminus. The only other deletion in *E. coli* TIM relative to trypanosomal TIM is in the short helix connecting $\alpha$6 with $\beta$7, between residues Ala201 and Asn202.

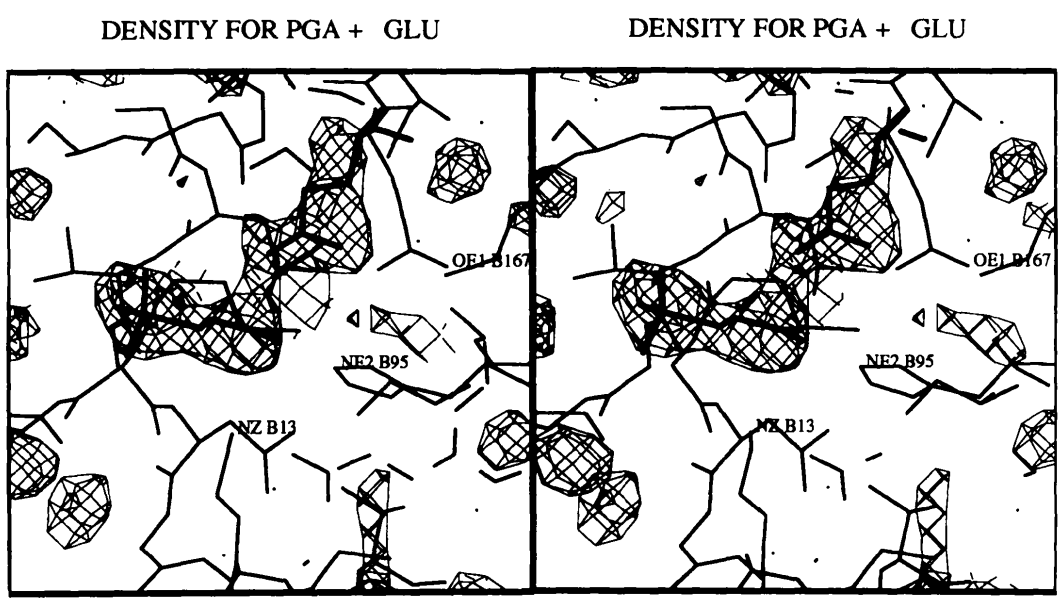


Fig. 11. Difference Fourier of the FAST data set. The final model from refinement against the image-plate data set was refined as a rigid body against the lower resolution FAST data set. The resulting model was used to provide calculated structure factors and phases for a $\sigma_A$-weighted $F_o - F_c$ $\alpha_{calc}$ map, contoured here at a level of $3\sigma$. Thick lines are used to show the phasing model. This feature, at the active site of subunit 2, was the highest in the map (with a magnitude of $8\sigma$).

Of the insertions in the *E. coli* sequence, Asp242 is an exclusive feature of *E. coli* TIM compared to other TIMs of published sequence (Wierenga, Noble & Davenport, 1992). In other sequences, the dipeptide Ala241-Asp242 of *E. coli* TIM is replaced by a single proline residue. The local structures of *E. coli* and trypanosomal TIM in this region are shown in Fig. 12. Whereas in trypanosomal TIM, and other TIMs, helix initiation is facilitated by the presence of the proline residue, in *E. coli* TIM, stabilization of the N terminus of the helix is performed by 'N-capping' hydrogen bonds to the side chain of Asp242.

The C-terminal extension of *E. coli* TIM adopts a helical conformation for all five residues where they can be observed in subunit 1, and for the three which can be seen in subunit 2 (Fig. 8).

(6) *Other differences between trypanosomal search model and refined E. coli TIM structure*

The overall fold of trypanosomal TIM and *E. coli* TIM has been compared in Fig. 8. The r.m.s. C$A$ coordinate differences between the trypanosomal search model and the refined *E. coli* model are shown in Fig. 13. For both subunit 1 and subunit 2 of *E. coli* TIM, the comparison is with subunit 1 of trypanosomal TIM, after superposition on the basis of the 104 equivalent C$A$ atoms comprising the eight $\alpha$-helices and eight $\beta$-strands. The r.m.s. difference for these C$A$ positions is 0.8 Å for *E. coli* subunit 1 and 0.9 Å for *E. coli* subunit 2. It can be seen that most of the large differences occur at the above described positions of sequence insertions or deletions.

The other regions of difference are around residues 58 and 157. The difference around residue 58 maps to the loop before $\beta$3, and seems to be just a different loop conformation. The difference around residue 157 maps to the short $3_{10}$ helix preceding $\beta$6. The connection between $\alpha$5 and $\beta$6 is somewhat longer in trypanosomal and *E. coli* TIMs compared to other published TIM sequences. In the case of trypanosomal TIM, this connection forms a three-residue loop, followed by six residues in $3_{10}$ helical conformation, followed by $\beta$6. In *E. coli* TIM on the other hand, there is only one residue after the end of $\alpha$5 before the start of the $3_{10}$ helix. This helix is then only four residues long, followed by three residues of random coil before the start of $\beta$6 [with secondary structure defined by the *DSSP* program (Kabsch & Sander, 1983)].

(7) *Differences in monomer–monomer interactions*

As for trypanosomal TIM, the constituent monomers of *E. coli* TIM are related by a near exact twofold axis ($\kappa = 178.2°$, shift along translation axis = 0.16 Å). Several differences are, however, apparent between the monomer–monomer interactions in trypanosomal TIM and those in *E. coli* TIM. Notably, the side chains of residues Arg100 from one monomer and Glu79 from the other monomer, which form a salt bridge across the dimer interface in trypanosomal, yeast and chicken TIMs, are reoriented in *E. coli* TIM so that they no longer interact (Fig. 14) (C$D$ of Arg 100 differs by 1.5 Å from trypanosomal TIM, and C$D$ of Glu79 differs by 4.5 Å). In *E. coli* TIM a water molecule is sitting at the site equivalent to the position of the carboxylate moiety of this glutamate in trypanosomal, yeast and chicken TIM. In each monomer of *E. coli* TIM, Glu79 now interacts with a histidine residue of the other monomer (His104). As in trypanosomal TIM, Arg100 interacts with Glu106, which is in turn compensating the charge of Lys114. A further consequence of the loss of the inter-monomer salt bridge is therefore the introduction of an uncompensated charge in *E. coli* TIM. Despite the apparent variability in their structural roles, Glu79 and Arg100 are seen to be amongst the 56 from 250 residues conserved in all known TIM sequences (Fig. 3).

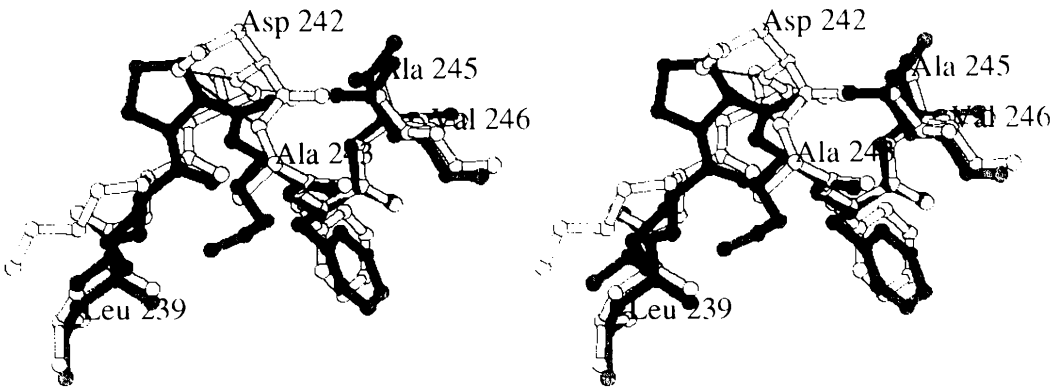


Fig. 12. Stereographic comparison of the models of trypanosomal TIM (in white) and *E. coli* TIM (in grey) at the start of helix 8. In *E. coli* TIM, the O$D$1 of Asp242 forms hydrogen bonds to its own main-chain N atom, and to that of the following residue. Trypanosomal TIM, by contrast, has a proline residue at the start of this helix.

A count of the inter-subunit hydrogen bonds in *E. coli* TIM yields a total of 26, compared to 20 inter-subunit hydrogen bonds in trypanosomal TIM (Wierenga, Noble, Vriend, Nauche & Hol, 1991). *E. coli* TIM has a total of four or six inter-subunit salt bridges (2 × Arg54–Asp50′ + 2 × Arg15–Asp87′ + possibly 2 × Glu79–His104′), compared to only two observed across the trypanosomal TIM subunit interface (2 × Glu77–Arg98′). In trypanosomal TIM (Wierenga, Noble, Vriend, Nauche & Hol, 1991), there is a cavity at the interface near to residues Glu77 and Arg98, which is filled by five ordered water molecules. In *E. coli* TIM by contrast, the surface of the equivalent volume is continuous with the outer surface of the protein [as calculated with the MSP package (Connolly, 1985)], as a result of missing monomer–monomer contacts which close one end of the cavity in trypanosomal TIM. This region is, therefore, a cleft in *E. coli* TIM, rather than a cavity. At the end of this cleft is found another interesting difference between the interfaces of trypanosomal TIM and *E. coli* TIM. Residue 45, at the tip of loop 2, is a phenylalanine in trypanosomal TIM, which forms a hydrophobic contact across the local dyad with the equivalent residue on the other monomer. In *E. coli* TIM, this residue is replaced by a glutamate residue (Glu46) (see also Fig. 14). This residue also interacts with the symmetry-related copy of itself across the local dyad, forming extensive van der Waal's contacts along the aliphatic part of the side chain. The charge of Glu46 is not compensated by a positively charged protein group, but rather is solvated by several water molecules, and a hydrogen bond to the side chain of Thr80 of the same subunit. The interaction with Thr80 requires that loop 3 should have a slightly different conformation in *E. coli* TIM than it does in trypanosomal TIM. This subtle difference would not allow Glu79 to form the salt bridge with Arg100 from the other monomer, which is observed in other TIM structures.

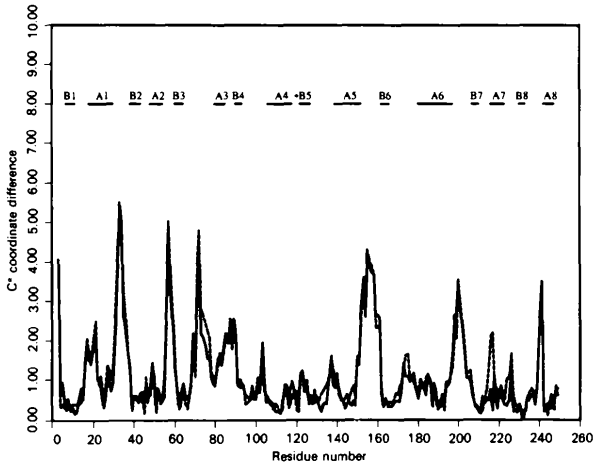


Fig. 13. Analysis of the distribution of coordinate difference between trypanosomal TIM subunit 1 and either *E. coli* subunit 1 (solid line) or *E. coli* subunit 2 (broken line).

(8) *Packing of the barrel core*

A unique feature of *E. coli* TIM when compared to other TIMs is the presence of a methionine residue (Met 9) in $\beta$1, with its side chain oriented into the interior of the barrel. In all other known TIM structures, this sequence locus is occupied by a small residue – either a glycine or an alanine (Fig. 3). A comparison of the local structures of TIM from trypanosome and *E. coli* in this region is shown in Fig. 15. Space for the large methionine residue, as compared to the alanine residue in trypanosomal TIM is made by a number of complementary changes of other residues forming the hydrophobic interior. Notably, Val40 of trypanosomal TIM is replaced by Ala41 in *E. coli* TIM, Ala63 of trypanosomal TIM is replaced by Gly64 in *E. coli* TIM, and Ile124 in trypanosomal TIM is replaced by Val126 in *E. coli* TIM. In addition, the methionine residue of *E. coli* TIM serves to fill a cavity found within the hydrophobic interior of trypanosomal TIM, also shown in Fig. 15.

(9) *The water structure*

In the refined *E. coli* TIM model there are 172 water molecules. The average *B* factor of these waters is 47 Å$^2$ (compared to 31 Å$^2$ for the protein atoms). At the dimer interface there are 15 water molecules within 3.5 Å of both subunits. Some of these water molecules have equivalents in the structure of trypanosomal TIM. Close to the dimer interface, two water molecules are observed in trypanosomal TIM to fill a cavity between O*E*1 Glu104 and N His95 (Wat601, Wat578). It was postulated (Wierenga, Noble, Vriend, Nauche & Hol, 1991) that these water molecules are important for the stability of TIM, because the human TIM variant E104D is thermally labile. The importance of these water molecules is supported by the observation that well defined water molecules are also located in this region at equivalent positions in *E. coli* TIM. The positions of two prominent water molecules in the ligand-free active-site pocket of trypanosomal TIM have been discussed (Wierenga, Noble, Vriend, Nauche & Hol, 1991; Wierenga, Borchert & Noble, 1992): Wat600 (between N*D*2 Asn11 and N*E*2 His95) and Wat610 (between O*E*2 Glu167 and N*E* Arg99). In *E. coli* TIM, electron density for water molecules is observed at equivalent positions to each of these water molecules in both subunits.

## Discussion

The structure of TIM from *E. coli* has been determined and refined to a low *R* factor with good

agreement with standard stereochemical parameters. Further evidence of the quality of the structure comes from the close agreement of atomic coordinates (r.m.s. = 0.55 Å for main-chain atoms) and atomic *B* factors (correlation coefficient = 0.66) between the two subunits. This close agreement was achieved without the imposition of non-crystallographic restraints during the refinement. The low *R* factor achieved probably reflects good data and a good starting model for molecular replacement,

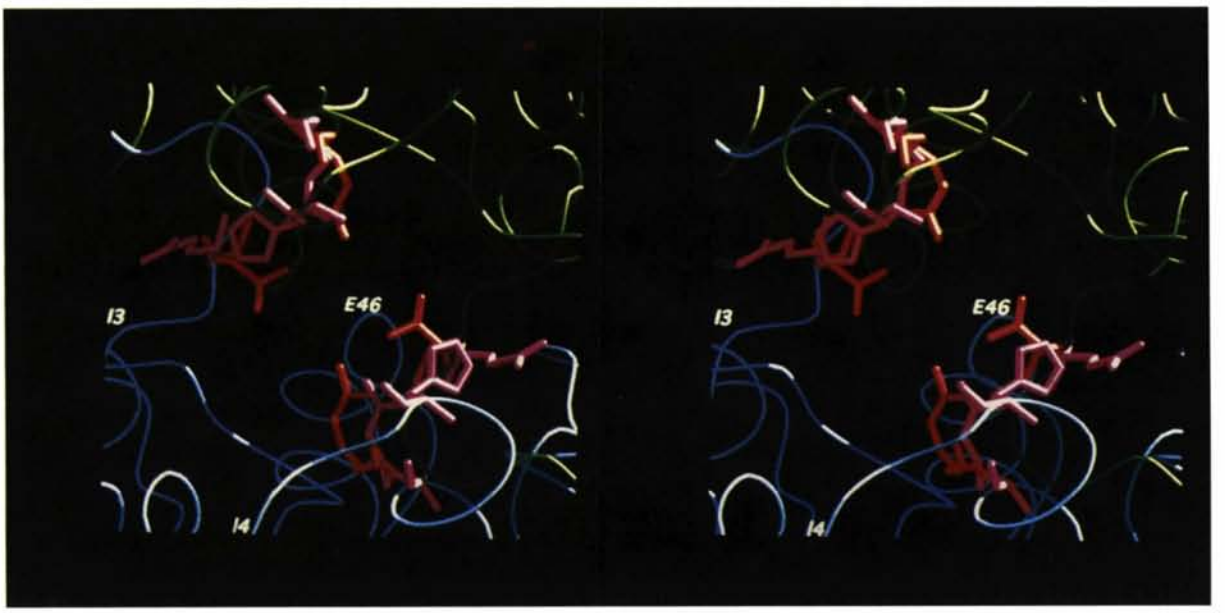


Fig. 14. Difference in the intermolecular interactions of residues Glu79 of both subunits. A smooth worm representation shows the C*A* trace of the *E. coli* TIM molecule viewed along the local twofold axis. One monomer is shown in blue, and the other in green. Residues Glu79 interact with residues His104 (of the other subunit) in *E.* coli TIM (residues shown in magenta). In trypanosomal TIM, as in yeast and chicken TIM, the equivalent glutamate residues (Glu77) form a salt bridge with arginines (Arg98 of the other subunit) (residues shown in red). Label *E*46 indicates the approximate position of Glu46 [at the tip of loop 2; Glu46 is at the bottom of a deep, solvent accessible, cleft (see text)]; 13 is near the N terminus of loop 3; 14 is near the C terminus of loop 4. This figure was produced with the program *XRENDER* (Noble, unpublished work).

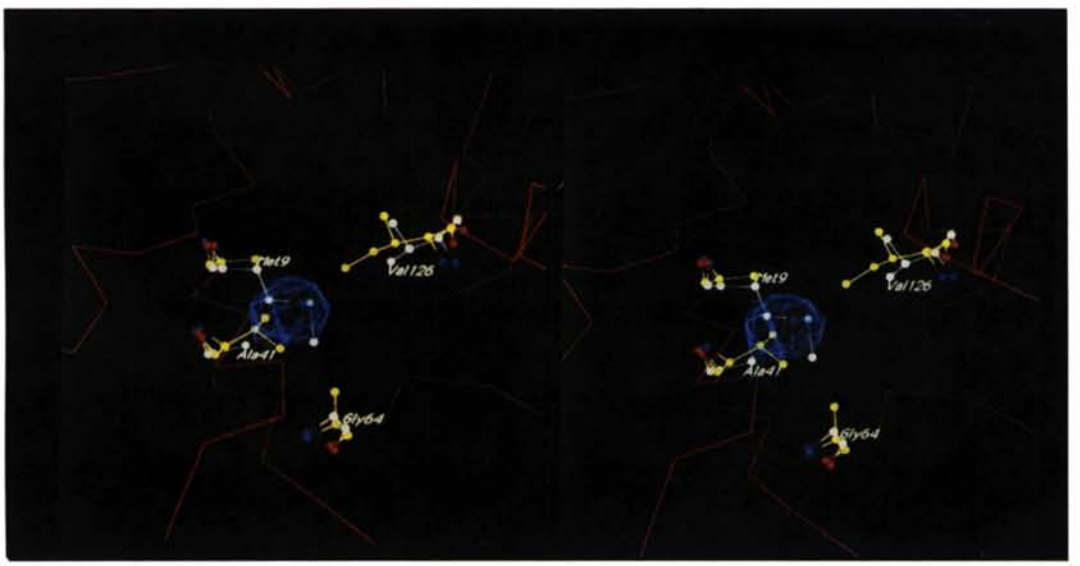


Fig. 15. Comparison of the packing of the core of *E. coli* TIM (labelled, white side chains) and trypanosomal TIM (yellow side chains). A view along the barrel axis towards Met9. Residues of trypanosomal TIM which would contact a methionine residue at sequence position 9 are shown, along with the equivalent residues of *E. coli* TIM. The shorter residue (Ala9) in trypanosomal TIM is partially complemented by larger surrounding side chains, but a cavity remains (blue wire frame). The cavity volume is 23 Å$^3$.

which had been refined at high resolution. Another contributing factor is the relatively weak restraint on the correlation of covalently bonded $B$ factors. Tighter restraints give rise to a decrease in r.m.s. $\Delta B$ from 8.5 to 5.1 Å$^2$, at the expense of an increase in $R$ factor from 11.9 to 14.5%. It is difficult to identify which of these sets of $B$ factors is most physically meaningful, since both have approximately the same atom by atom correlation of the equivalent temperature factors in subunit 1 and subunit 2 (0.66 for the model with $R = 11.9\%$ and 0.69 for the model with $R = 14.5\%$).

The structure was first determined at 2.8 Å and then improved at 2.6 Å resolution. In going from the 2.8 Å 'intermediate' structure to the 2.6 Å 'final' structure, the quality of the model increased considerably. This improvement could be seen in terms of a cleaner Ramachandran plot (Table 3), and an improved agreement with database side-chain torsion angles. Since the Ponder and Richards torsion-angle database is not used explicitly in structure refinement, agreement with this database represents an unbiased assessment of the quality of a model, similar in principle to the use of free $R$-factor calculations. The free $R$-factor protocol (Table 5), as implemented here, gives a discrimination for the free $R$ factor (11.5%) and the $\chi$-imperfection value (12.7%) which is lower than for the $\chi$-imperfection value (28%) between the original models, before the free $R$-factor protocol was carried out (Table 3). In our experience, the $\chi$-imperfection value of 45° (Table 3) is rather high. For example, the refinement of the 2.4 MAS trypanosomal TIM structure at 1.83 Å (Wierenga, Noble, Vriend, Nauche & Hol, 1991) decreased the $\chi$-imperfection value from 44 to 31°. Although an exhaustive survey of the Protein Data Bank (Bernstein *et al.*, 1977) has not been undertaken, the lowest $\chi$-imperfection value which we have yet seen is 20.4°, for the 2APR aspartic protease molecule solved at 1.8 Å resolution (Suguna *et al.*, 1987).

A crystal grown in the presence of 1 m$M$ 2PG did not have the high occupancy of this transition-state analogue. This is surprising, since the $K_i$ value for 2PG is only 27 $\mu M$ (Lambeir, Opperdoes & Wierenga, 1987), so that a 40-fold excess was present in the crystallization medium. The absence of 2PG from the active site of the crystals must then result from one of two alternatives. The first possibility is that 2PG became hydrolyzed during the long crystallization period, releasing phosphate and 2-hydroxyethanoic acid. Evidence for this comes from the presence of electron density consistent with a low-occupancy phosphate molecule in the active site of subunit 1 of these crystals. Low occupancy for phosphate would be expected at the concentration of 1 m$M$, since the $K_i$ for phosphate of a range of TIMs is around 5 m$M$ (Lambeir, Opperdoes & Wierenga, 1987). The second alternative is that crystal contacts affect the structure and flexibility of TIM to the extent that ligand binding is severely disfavoured in this crystal form. This second alternative is made likely by the extent of structural changes which accompany ligand binding. The solution binding constant of around 27 $\mu M$ could be raised significantly if some of these structural changes are disfavoured by contacts of the crystal lattice.

The structure of TIM from *E. coli* provides a number of structural details which complement those from previous structures. An interesting example has been seen of two residues – Glu79 and Arg100 – entirely conserved across evolution, which had been seen in three TIM structures to form a salt bridge across the dimer interface, but which do not interact in TIM from *E. coli*. Such an observation should be taken as a cautionary point in the process of modelling by homology. Possibly, by further broadening the database of known homologous structures, a predictive understanding of such changes can be achieved.

Sequence peculiarities within the TIM primary structure are not always perfectly complemented by contacting residues. Imperfect complementation is observed in the case of Met9 of *E. coli* TIM, which is a small residue in all other TIM sequences. Trypanosomal TIM, for example, complements the absence of the large Met9 residue by having somewhat larger neighbouring residues, but fails to completely fill the equivalent space, leaving instead a hydrophobic cavity in the protein core. The presence of such a packing defect in some, but not all TIMs, suggests that there is some freedom for imperfection in protein folding. The cavity in question has been previously discussed in studies of the packing of the barrel core in TIM barrel proteins (Lesk, Brändén & Chothia, 1989).

An alignment of all known TIM sequences shows totally conserved 'N-cappping' residues [serine, threonine, asparagine, or aspartic acid (Richardson & Richardson, 1988)], at the start of $\alpha$3, $\alpha$4 and $\alpha$6 (see also Rentier-Delrue *et al.*, 1992). In *E. coli* TIM, the residues performing this role are Ser81, Ser107 and Thr181. $\alpha$1 is capped by one of this set of residues (Ser17 in *E. coli* TIM) in all but mammalian TIMs, where the equivalent sequence locus is occupied by an arginine. $\alpha$2 is capped by a conserved water-molecule position in trypanosomal and *E. coli* TIM structures. $\alpha$5 is also 'N-capped' by a conserved water molecule. The residue at the N-cap position of $\alpha$7 is Asn219; however, the side chain is not involved in N-capping, as it is pointing in a different direction. The start of $\alpha$8 displays one of two forms of stabilization of the N-terminus of a helix. In *E. coli* TIM, the dipeptide Ala241–Asp242 is present, with the aspartic acid serving to cap the helix (Fig. 12). In all other published TIM sequences, a proline replaces these two residues, which represents another postulated means of stabilizing the N terminus of a helix.

The good conservation of helix N-terminus-stabilizing residues suggests that their role is important in either the stability, or the pathway of folding of the TIM barrel. Protein stability is a small net result in the balance of large energetic and entropic terms, which can be altered significantly by a number of different minor sequence alterations. For this reason, it is tempting to speculate that good conservation of a sequence element involved in structure stabilization, such as the sequence constraints observed here and elsewhere for the N termini of $\alpha$-helices, reflects a determinant more of the folding pathway than of the final stability of a protein fold.

The authors would like to thank Dr Paul Tucker for help with the data-collection facilities at the EMBL in Heidelberg. The authors are also grateful to Professor Wim Hol for many stimulating discussions on TIM structures. This work was partly funded by EC grant BIOT-CT90-0182 to RKW. The coordinates and structure factors have been deposited at the Brookhaven Protein Data Bank.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TRE, R1TRESF). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Supplementary Publication No. SUP 37082). A list of deposited data is given at the end of this issue.

## References


ALBER, T., BANNER, D. W., BLOOMER, A. C., PETSKO, G. A., PHILLIPS, D. C., RIVERS, P. S. & WILSON, I. A. (1981). *Philos. Trans. R. Soc. London Ser. B*, **293**, 159–171.

BANNER, D. W., BLOOMER, A. C., PETSKO, G. A., PHILLIPS, D. C., POGSON, C. I., WILSON, I. A., CORRAN, P. H., FURTH, A. J., MILMAN, J. D., OFFORD, R. E., PRIDDLE, J. D. & WALEY, S. G. (1975). *Nature (London)*, **255**, 609–614.

BASH, P. A., FIELD, M. J., DAVENPORT, R. C., PETSKO, G. A., RINGE, D. & KARPLUS, M. (1991). *Biochemistry*, **30**, 5826–5832.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T., TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BLACKLOW, S. C. & KNOWLES, J. R. (1990). *Biochemistry*, **29**, 4099–4108.

BRÄNDÉN, C.-I. (1991). *Curr. Opin. Struct. Biol.* **1**, 978–983.

BRÜNGER, A. T. (1992). *Nature (London)*, **355**, 472–475.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

CONNOLLY, M. L. (1985). *J. Am. Chem. Soc.* **107**, 1118–1124.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

DAVENPORT, R. C., BASH, P. A., SEATON, B. A., KARPLUS, M., PETSKO, G. A. & RINGE, D. (1991). *Biochemistry*, **30**, 5821–5826.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). *Nucleic Acids Res.* **12**, 387–395.

HOL, W. G. J., VELLIEUX, F. M. D., VERLINDE, C. L. M. J., WIERENGA, R. K., NOBLE, M. E. M. & READ, R. J. (1991). *Molecular Conformations and Biological Interactions*, edited by P. BALARAM & S. RAMESESHAN, pp. 215–244. Bangalore: Indian Academy of Science.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.

KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KNOWLES, J. R. (1991). *Nature (London)*, **350**, 121–124.

KNOWLES, J. R. & ALBERY, W. J. (1977). *Acc. Chem. Res.* **10**, 105–111.

KOMIVES, E. A., CHANG, L. C., LOLIS, E., TILTON, R. F., PETSKO, G. A. & KNOWLES, J. R. (1991). *Biochemistry*, **30**, 3011–3019.

LAMBEIR, A.-M., OPPERDOES, F. R. & WIERENGA, R. K. (1987). *Eur. J. Biochem.* **168**, 69–74.

LESK, A. M., BRÄNDÉN, C.-I. & CHOTHIA, C. (1989). *Proteins*, **5**, 139–148.

LOLIS, E., ALBER, T., DAVENPORT, R. C., ROSE, D., HARTMANN, F. C. & PETSKO, G. A. (1990). *Biochemistry*, **29**, 6609–6618.

LOLIS, E. & PETSKO, G. A. (1990). *Biochemistry*, **29**, 6619–6625.

LUZZATI, P. V. (1952). *Acta Cryst.* **5**, 802–810.

MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

NOBLE, M. E. M., WIERENGA, R. K., LAMBEIR, A.-M., OPPERDOES, F. R., THUNNISSEN, A.-M. W. H., KALK, K. H., GROENDIJK, H. & HOL, W. G. J. (1991). *Proteins*, **10**, 50–69.

PICHERSKY, E., GOTTLIEB, L. D. & HESS, J. F. (1984). *Mol. Gen. Genet.* **195**, 314–320.

POMPLIANO, D. L., PEYMAN, A. & KNOWLES, J. R. (1990). *Biochemistry*, **29**, 3186–3194.

PONDER, J. W. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

RENTIER-DELRUE, F., MANDE, S. C., MOYENS, S., TERPSTRA, P., MAINFROID, V., GORAJ, K., LION, M., HOL, W. G. J. & MARTIAL, J. A. (1993). *J. Mol. Biol.* **229**, 85–93.

RICHARDSON, J. S. & RICHARDSON, D. C. (1988). *Science*, **240**, 1648–1652.

RIEDER, S. V. & ROSE, I. A. (1959). *J. Biol. Chem.* **234**, 1007–1010.

SERC Daresbury Laboratory (1979). *CCP4 Program Suite*. SERC Daresbury Laboratory, Warrington, England.

SUGUNA, K., BOTT, R. R., PADLAN, E. A., SUBRAMANIAN, E., SHERIFF, S., COHEN, G. H. & DAVIES, D. R. (1987). *J. Mol. Biol.* **196**, 877–900.

THALLER, C., WEAVER, L. H., EICHELE, G., WILSON, E., KARLSSON, R. & JANSONIUS, J. N. (1981). *J. Mol. Biol.* **147**, 465–469.

TRONRUD, D. E., TEN EYCK, L. F. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

VERLINDE, C. L. M. J., NOBLE, M. E. M., KALK, K. H., GROENDIJK, H., WIERENGA, R. K. & HOL, W. G. J. (1991). *Eur. J. Biochem.* **198**, 53–57.

VERLINDE, C. L. M. J., RUDENKO, G. & HOL, W. G. J. (1992). *J. Comput. Aided Mol. Des.* **6**, 131–147.

VRIEND, G. (1990). *J. Mol. Graph.* **8**, 52–56.

WIERENGA, R. K., BORCHERT, T. V. & NOBLE, M. E. M. (1992). *FEBS Lett.* **307**, 34–49.

WIERENGA, R. K., KALK, K. H. & HOL, W. G. J. (1987). *J. Mol. Biol.* **198**, 109–121.

WIERENGA, R. K., NOBLE, M. E. M. & DAVENPORT, R. C. (1992). *J. Mol. Biol.* **224**, 1115–1126.

WIERENGA, R. K., NOBLE, M. E. M., POSTMA, J. P. M., GROENDIJK, H., KALK, K. H., HOL, W. G. J. & OPPERDOES, F. R. (1991). *Proteins*, **10**, 33–49.

WIERENGA, R. K., NOBLE, M. E. M., VRIEND, G., NAUCHE, S. & HOL, W. G. J. (1991). *J. Mol. Biol.* **220**, 995–1015.

ZABORI, S., RUDOLPH, R. & JAENICKE, R. (1980). *Z. Naturforsch. Teil C*, **35**, 999–1004.