PROTEINS: Structure, Function, and Genetics 10:50–69 (1991)

The Adaptability of the Active Site of Trypanosomal Triosephosphate Isomerase as Observed in the Crystal Structures of Three Different Complexes

Martin E.M. Noble,[1] Rik K. Wierenga,[1] Anne-Marie Lambeir,[3] Fred R. Opperdoes,[3] Andy-Mark W.H. Thunnissen,[2] Kor H. Kalk,[2] Hillie Groendijk,[2] and Wim G.J. Hol[2]

[1]European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany, [2]Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, NL-9747 AG Groningen, The Netherlands; and [3]Research Unit for Tropical Diseases, International Institute of Cellular and Molecular Pathology, Avenue Hippocrate 74, B-1200, Brussels, Belgium

ABSTRACT **Crystals of triosephosphate isomerase from *Trypanosoma brucei brucei* have been used in binding studies with three competitive inhibitors of the enzyme's activity. Highly refined structures have been deduced for the complexes between trypanosomal triosephosphate isomerase and a substrate analogue (glycerol-3-phosphate to 2.2 Å), a transition state analogue (3-phosphonopropionic acid to 2.6 Å), and a compound structurally related to both (3-phosphoglycerate to 2.2 Å). The active site structures of these complexes were compared with each other, and with two previously determined structures of triosephosphate isomerase either free from inhibitor or complexed with sulfate. The comparison reveals three conformations available to the "flexible loop" near the active site of triosephosphate isomerase: open (no ligand), almost closed (sulfate), and fully closed (phosphate/phosphonate complexes). Also seen to be sensitive to the nature of the active site ligand is the catalytic residue Glu-167. The side chain of this residue occupies one of two discrete conformations in each of the structures so far observed. A "swung out" conformation unsuitable for catalysis is observed when sulfate, 3-phosphoglycerate, or no ligand is bound, while a "swung in" conformation ideal for catalysis is observed in the complexes with glycerol-3-phosphate or 3-phosphonopropionate. The water structure of the active site is different in all five structures. The results are discussed with respect to the triosephosphate isomerase structure function relationship, and with respect to an on-going drug design project aimed at the selective inhibition of glycolytic enzymes of *T. brucei*.**

Key words: TIM, protein–ligand complexes, water involvement in binding, drug design, active site structure, sleeping sickness

INTRODUCTION

The glycolytic enzyme triosephosphate isomerase (TIM) catalyzes one of the simplest reactions observed in biochemistry, i.e., the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) (Fig. 1). The simplicity of this conversion reaction, the absence of cofactors, and the high catalytic efficiency of the enzyme[1] makes this an ideal case for the study of a biological structure–function relationship.

For this reason, a large amount of spectroscopic and kinetic information has been amassed to dissect the function of TIM. This work has determined the free energy profile of the individual steps in the reaction pathway between substrate binding and product release,[2] and has demonstrated the encounter limited nature of the reaction rate.[1] The mechanism is thought to involve a single protein group transferring a proton between carbon atoms via a *cis* enediol or enediolate intermediate,[3] with other residues providing an additional electrophilic contribution to catalysis.[4]

In addition, efforts have been made to understand the functional characteristics of the enzyme in terms of its three dimensional structure. This work has lead to atomic resolution structures for TIM and TIM complexes derived from three different sources; chicken,[5] yeast,[6] and *Trypanosoma brucei*.[7] Chicken TIM, the first of these structures to be solved, revealed the first "TIM barrel" fold, which has now been observed in enzymes catalyzing a wide range of

Received June 20, 1990; revision accepted October 10, 1990.

Address reprint requests to Rik K. Wierenga, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany.

Present address of A.-M. Lambeir: Plant Genetic System NV, Plateaustraat 22, B-9000 Ghent, Belgium.

Abbreviations: TIM, triosephosphate isomerase (EC 5.3.1.1); GAP, glyceraldehyde-3-phosphate; DHAP, dihydroxyacetonephosphate; 3-PGA, 3-phosphoglycerate; 3-PP, 3-phosphonopropionate; G3P, glycerol-3-phosphate; EDTA, ethylenediaminetetraacetic acid; MOPS, 3-(*N*-morpholino)-propanesulfonic acid; RMS, root mean square; PEG, polyethyleneglycol.

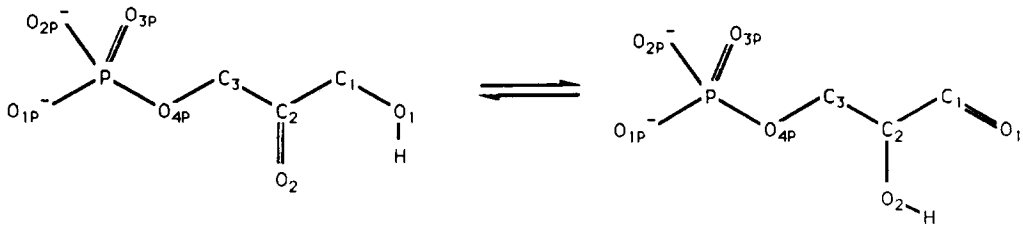

Fig. 1. The reaction catalysed by triosephosphate isomerase. The product is the D enantiomer of glyceraldehyde-3-phosphate.

biochemical reactions.[8] The TIM barrel is made up of an 8-fold repeat of strand-turn–helix-turn units, forming a parallel eight-stranded β-barrel on the inside of the protein surrounded by a coat of eight α-helices.

Knowledge of the structures of the active site of chicken TIM, situated at the C-terminal end of the β-barrel, leads to the development of ideas to explain the catalytic mechanism of TIM. Analysis of the active site in a chicken TIM DHAP complex identified three residues particularly important for the reaction mechanism[9]: a lysine residue (corresponding to residue 13 in trypanosomal TIM), a histidine residue (residue 95 in trypanosomal TIM), and a glutamic acid residue (167 in trypanosomal TIM). The side chains of these amino acids are thought to be involved in electrophilic catalysis, shuttling of protons between substrate oxygen atoms, and shuttling of protons between substrate carbon atoms respectively. The overall structure of TIM is shown schematically in Figure 2, with these catalytic residues shown to indicate the location of the active site.

Also apparent from crystallographic binding studies with chicken TIM and either DHAP or phosphate ions was a large conformational change involving the protein loop connecting strand β6 with helix α6, hereafter known as the flexible loop.[9] Atoms of this loop were seen to move by up to 10 Å to close off the active site from solvent and form a hydrogen bond to a terminal phosphate oxygen. One of the mechanistic functions of the flexible loop has recently been shown to be stabilisation of the reaction intermediate.[10] This serves to improve the overall ratio of the desired isomerization reaction to the unwanted phosphate elimination side reaction.

Details of substrate binding were inaccessible at the low resolution (6 Å) of the chicken TIM–substrate structure. Later work with yeast TIM has provided higher resolution information about the binding of substrate analogues. Structures of yeast TIM both free (to 1.9 Å resolution)[44], and complexed with either phosphoglycolohydroxamate (to 1.9 Å)[11] or phosphoglycollic acid (to 2.6 Å)[45] have been solved. Mutant TIMs produced on the basis of yeast and chicken TIM structures have been characterized in order to test hypotheses about the catalyic residues and the flexible loop.[10,12,13]

The structure of the TIM used in the binding studies described here, derived from the organism *Trypanosoma brucei brucei,* had been solved to 1.83 Å resolution (Wierenga et al., manuscript in preparation). The "native" crystals are grown from a solution containing 2.4 M ammonium sulfate, and contain a dimer of covalently identical subunits, each of 249 amino acids, in the crystallographic asymmetric unit.[7] Subunit 1, consisting of residues 2–250, has no sulfate bound at its active site and has a flexible loop (residues 167–180) constrained by crystal contacts in the "open" conformation. By contrast, subunit 2 (residues 302–550) has a bound sulfate ion at the active site, inducing a "closed" conformation for the flexible loop. Transfer of native crystals to a sulfate-free medium causes dissociation of the bound sulfate, and allows the subunit 2 flexible loop to adopt a conformation similar to that observed in subunit 1 (Wierenga et al., in press, 1991).

Described here is the structure of trypanosomal TIM in complex with three active site inhibitors, which are 3-phosphoglyceric acid (3-PGA), glycerol-3-phosphate (G3P), and 3-phosphonopropionic acid (3-PP). The structures of these inhibitors and the nomenclature used in this paper are shown in Figure 3. The ligands studied here contain either one less nonhydrogen atom than the actual substrate (3-PP), the same number as the substrate (G3P), or one more than the substrate (3-PGA). The overall charge of G3P is determined only by the phosphate group. In contrast, 3-PP and 3-PGA each has an additional carboxyl group, so that these ligands may exist as either dianionic or trianionic species. The differences in binding mode and protein adaptation resulting from the differences in size and chemical nature of the ligands have been analyzed. Also studied is the binding mode of the first phosphonate compound complexed to TIM (i.e., 3-PP).

The organism *Trypanosoma brucei* is a pathogen of great human and economic importance in tropical Africa.[14] In this region it is responsible for diseases such as sleeping sickness in man and nagana in cat-

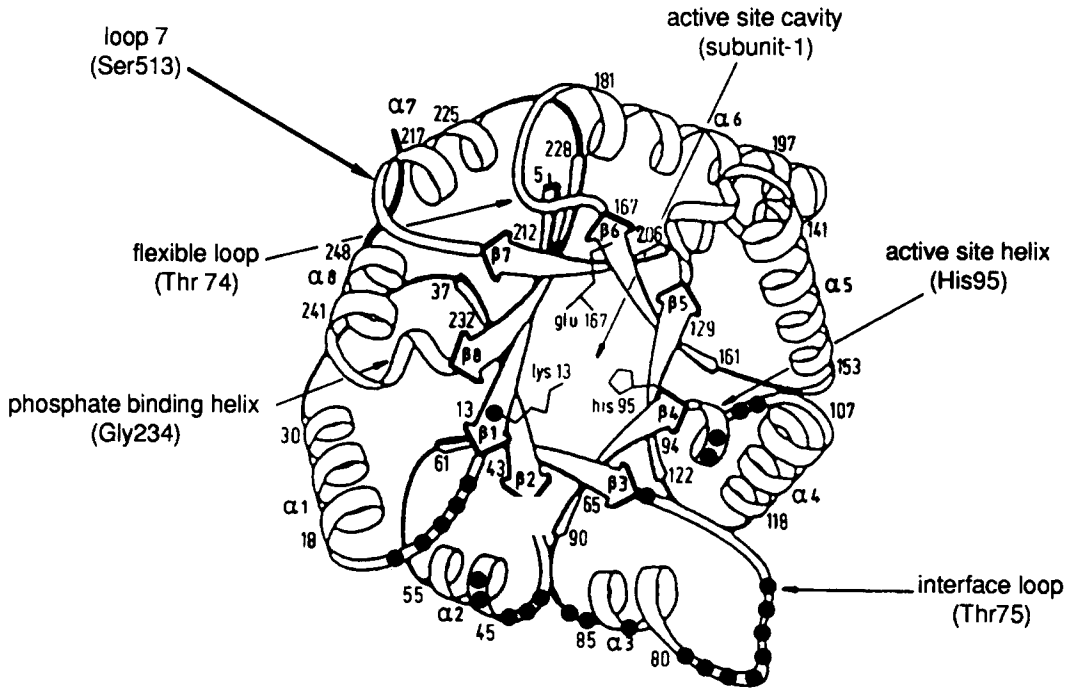

Fig. 2. Schematic fold of one subunit of trypanosomal TIM. The $(\beta\alpha)_8$ fold of trypanosomal TIM is shown along with the location of the flexible loop and the catalytic residues (Lys-13, His-95, and Glu-167). Also indicated is the loop connecting helix α7 with strand β7. Residues involved in the subunit–subunit interface are marked by filled circles. The tip of the interface loop of each subunit (around residue 75) participates in the active site of the other subunit.

tle. Drugs currently used in the treatment of trypanosomiasis are unsatisfactory, giving rise to an urgent search for suitable alternatives. One approach in pursuit of this goal is the rational drug design cycle, which uses biochemical and structural information to exploit differences between host and parasite metabolism.[15] In the case of trypanosomiasis there are many such differences, involving in particular the energy transduction mechanism in host and parasite. The enzymes of the trypanosomal glycolytic pathway have a different cellular location from those of the host, being sequestered in small microbody like organelles called glycosomes.[16] Moreover, the bloodstream form of the trypanosome is entirely dependent on glycolysis for energy production, since it lacks a functional tricarboxylic acid cycle. These unique features, combined with the observation that inhibitors of trypanosomal glycolysis are able to kill trypanosomes,[17] make this pathway an attractive target for drug design. TIM is the first trypanosomal glycolytic enzyme to be studied for this purpose.

Three main aspects of trypanosomal glycolytic enzymes are being considered as points of inhibition which might serve to decrease the rate of trypanosomal glycolysis. These are the import of glycolytic enzymes into their subcellular compartment, the assembly of these enzymes into functionally active oligomers, and the individual mechanisms of the enzymes involved. The binding modes determined here for inhibitors of the reaction mechanism of TIM can serve as entry points into the rational drug design cycle, providing information for the modelling, design and synthesis of the next generation of TIM inhibitors. In this fashion, inhibitors can be iteratively designed toward the ultimate goal of strong and selective inhibition of trypanosomal triosephosphate isomerase.

MATERIALS AND METHODS

Purification, Crystallization, Crystal Soaking, and Data Collection

Trypanosomal triosephosphate isomerase was purified as previously described.[18] Crystals were grown from 2.4 M ammonium sulfate in a 0.2 M 3-(*N*-morpholino)-propanesulfonic acid (MOPS)

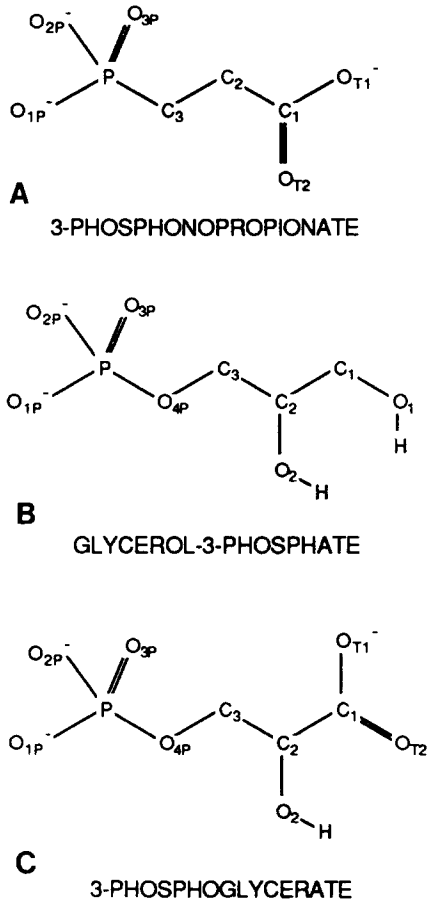

Fig. 3. The inhibitors analysed in complex with TIM. The inhibitors studied here are shown in their prevalent protonation states at pH 7: **(A)** 3-phosphonopropionate (3-PP), **(B)** glycerol-3-phosphate (G3P), and **(C)** 3-phosphoglycerate (3-PGA).

buffer pH 7.0, containing also 1 mM EDTA, 1 mM dithiothreitol, and 1 mM sodium azide as previously described.[19] Under these conditions, crystals grow in the space group $P2_12_12_1$ with cell dimensions a = 113.14 Å, b = 97.56 Å, c = 46.56 Å and good internal order. Crystals obtained in this fashion have been used to produce a 1.83 Å "native" structure, now refined to an R-factor of 18.3% for data between 6.0 and 1.83 Å, with good geometry [e.g., root mean square (RMS) deviation from ideality = 0.020 Å for covalent bond lengths] (Wierenga et al., manuscript in preparation).

In order to allow diffusion of substrate analogues into the active site of TIM, it is necessary to soak crystals in a solution of the substrate analogue which has a low concentration of sulfate ions. This frees the phosphate position of the active site of TIM from sulfate ions which would otherwise compete with the inhibitors for binding to TIM, since sulfate ions are competitive inhibitors of TIM with a K_i of 4.5 mM.[20] To achieve this without disruption of the crystals, trypanosomal TIM crystals were transferred to a solution of the alternative precipitant polyethyleneglycol 6000 (PEG 6000)(44 g in 100 g of final solution) containing 1 mM EDTA, 1 mM dithiothreitol, and 1 mM sodium azide, as well as a reduced concentration of ammonium sulfate (0.1 M) which was found to be necessary to stabilize the crystals in PEG 6000 prior to transfer to a solution of the substrate analogue. The concentration of PEG 6000 used for this protocol was chosen to provide a similar precipitant power to 2.4 M ammonium sulfate.[21] From this solution, crystals were transferred to a solution of PEG 6000 containing 1 mM EDTA, 1 mM dithiothreitol, and 1 mM sodium azide, as well as one of the three compounds shown in Table I at the concentrations shown also in Table I. The chemicals used were of the highest possible quality: PEG 6000 was obtained from Merck, DL-3-phosphoglycerate and DL-glycerol-3-phosphate from Boehringer, and 3-phosphonopropionate from Aldrich. Crystals were soaked under these conditions with occasional transfers into fresh soaking solution for at least a week.

After this period, the crystals were mounted in a small volume of mother liquor in thin glass capillaries which were subsequently sealed at either end. These capillaries were then mounted on the Groningen FAST area detector with an Eliot GX 21 rotating anode generator. A single crystal was used for each substrate analogue to collect a three-dimensional dataset. Data collection and initial data reduction were done using the MADNES package,[22] with further merging and scaling of sets of data being accomplished through programs of the Groningen BIOMOL crystallographic software package. Information on the crystals and the quality of the data is given in Table II.

Starting Model Coordinates, Rigid Body Refinement, and Conventional Least-Squares Refinement

The starting coordinate set used for refinement of each complex was a partially refined model of the native structure of trypanosomal TIM. The 3778 protein atoms of this model, but not the sulphate or water atoms, were used in rigid body refinement by the TNT package[23] against data between 6.0 and 4.5 Å of the G3P dataset. For this refinement, the individual monomers were defined as rigid bodies, to allow a change in relative orientation of the two subunits as well as a repositioning of the dimer in the unit cell. The necessity of rigid body refinement was suggested by the differences in cell dimensions between the native and soaked crystals (Table II), involving particularly the a axis. The rms atomic shifts during this refinement was 0.40 Å for subunit 1 (systematic shift in *xyz:* 0.39, −0.03, −0.02), and 0.36 Å for subunit 2 (systematic shifts in *xyz:* 0.32, 0.03, 0.10). The resulting structure was used as the initial protein structure for all three ligands. After

TABLE I. Inhibition Constants and Soaking Concentrations of Trypanosomal TIM Inhibitors

Compound	Abbreviation	K_i (mM)*	Soaking concentration (mM)
DL-Glycerol-3-phosphate	G3P	0.61	6
DL-3-Phosphoglycerate	3-PGA	1.3	7
3-Phosphonopropionate	3-PP	27†	100

*Binding constants from ref. 20.
†A.M.L., unpublished result.

TABLE II. Some Data Concerning Trypanosomal TIM Native and Complexed Crystals

Dataset	Native	G3P	3 PGA	3 PP
Maximum resolution (Å)	1.83	2.2	2.15	2.6
Reflections	38819	18064	18169	10879
Completeness (%)	90.0 (to 1.9 Å)	72.8 (to 2.3 Å)	72.8 (to 2.3 Å)	75.7 (to 2.8 Å)
Completeness at higher resolution (%)	55 (1.9–1.85 Å)	41 (2.4–2.2 Å)	40 (2.4–2.2 Å)	22 (2.8–2.6 Å)
R-merge (%)*	9.0	4.3	5.1	2.3
A (Å)	113.14	112.36	112.40	112.36
B (Å)	97.69	97.59	97.55	97.54
C (Å)	46.56	46.65	46.63	46.71

$$*R\text{-merge} = \frac{\sum_h \sum_i |(\bar{I}_h - I_{h,i})|}{\sum_i \sum_h |\bar{I}_h|} \times 100\%.$$

minor rebuilding of the protein model, models of the three inhibitors were built into difference electron density. Initial inhibitor models were built using the CHEM-X molecular modeling and energy minimization package. For the chiral ligand 3-PGA, the much better inhibition of TIM by a racemic mixture compared to pure (L) stereomer[20] suggested clearly that the occupant of the active site in the complex would be the (D) stereomer. This was confirmed by the failure of attempts to fit the (L) stereomer of 3-PGA into difference electron density. Also for G3P, the second chiral ligand, the difference electron density in the active site of subunit 2 clearly indicated that the (D) stereomer, and not the (L) stereomer was bound.

At this point in refinement, work was suspended until a more highly refined native structure became available. This was superimposed, allowing independent monomer movement, upon the protein structures of the complexes derived from the preceding work, and used for further TNT refinement without rigid body constraints. Initial values for individual temperature factors were taken from the native structure. TNT refinement was carried out initially at low resolution (50.0–3.0 Å), with the upper resolution limit being extended stepwise to the upper limit of the dataset. During refinement, a bulk solvent scattering model was used.[24] Modeling bulk solvent makes the calculated phases of low-resolution reflections more accurate, allowing the use of these reflections in refinement. The result is an increase of about 1500 observations by using a low-resolution cut-off of 50 Å rather than 6 Å. For TNT refinement, the structures of the inhibitors were defined with freely varying torsion angles, but with bond lengths, angles, and trigonal atom planarity defined from crystal structures of the three compounds (3-PP[25], G3P[26], 3-PGA[27]). At the end of this first stage of TNT refinement, the R-factors of the three models were as follows: G3P = 19.7% (50.0–2.2 Å), 3-PGA = 18.8% (50.0–2.2 Å), 3-PP = 17.6% (50.0–2.6 Å).

Molecular dynamics refinement using the GROMOS MDX/EMX package[28,29] was initiated at this point. MDX refinement was chosen to allow the protein and substrate analogues to find a refinement minimum away from the local minima which limit the radius of convergence of conventional least squares methods, and free from the subjective bias associated with manual fitting. Again refinement was started at low resolution and extended stepwise to the limit of the available data. At each resolution step, refinement started with energy minimization including a crystallographic pseudo-energy term (EMX), continued by a 1 psec molecular dynamics refinement run (MDX), and finished with further EMX.

The weight applied to the crystallographic pseudo-potential was chosen during EMX runs, by repeatedly setting the parameter SIGF to equal the RMS

TABLE III. Mean Atomic *B*-Factors of Subsets of Atoms

Units of Å^2	Native	G3P	3 PGA	3 PP
Subunit 1	23.2	20.8	24.5	20.4
Subunit 2	28.8	23.7	28.1	23.9
Ligand (average)	50.8	36.8	46.8	25.9
Ligand (individual atoms)	S:47.5	P:35.8	P:38.7	P:21.8
	O1:48.7	O1P:41.8	O1P:58.7	O1P:31.5
	O2:65.5	O2P:68.6	O2P:49.3	O2P:27.6
	O3:37.3	O3P:41.3	O3P:48.7	O3P:22.2
	O4:54.8	O4P:42.2	O4P:38.0	C3:28.7
		C3:34.8	C3:58.3	C2:19.7
		C2:22.2	C2:50.4	C1:25.9
		O2:26.0	O2:32.2	OT1:33.7
		C1:20.9	C1:35.0	OT2:22.3
		O1:44.4	OT1:54.8	
			OT2:50.7	

difference coefficient for consecutive EMX runs until an approximately constant value was achieved. A value of SIGF equal to the rms difference coefficient has been found appropriate for GROMOS refinement in work by other authors.[29] When this convergence occurred, MDX was begun with this value of SIGF, and a simulated constant temperature of 300 K to which the protein was linked with a relaxation time of 0.01 psec.[30] It was found necessary to use time steps of only 1 fsec in some of the MDX runs as opposed to the default value of 2 fsec. This adjustment prevented crashing of the SHAKE routine which was being used to ensure conformity to ideal bond length and angle values.[31] At the end of the final MDX run, an averaging procedure was used to produce the structure for energy minimization by EMX. This procedure involved the averaging of the last few structures of the MDX trajectory, at which point the molecules being simulated are assumed to be oscillating about their position of minimum energy. Typically, the last 0.1 psec of a 1psec run was sampled at 20 fsec intervals, yielding five intermediate structures to be averaged.

It was found that the structures produced by GROMOS MDX/EMX could be further refined by carrying out cycles of TNT refinement. For the MDX/EMX produced structure of the G3P complex, this further TNT refinement produced a 2% drop in *R*-factor, while causing an RMS movement of atomic positions of only 0.11 Å. Presumably this movement represents adaptation of the protein to the different stereochemical potential of the second refinement package. In this case, the figure of 0.11 Å for the RMS atomic shifts represents an estimate of the error caused by using a different set of geometric restraints. After 10 cycles of TNT positional refinement performed on the final structures produced from GROMOS EMX, the *R*-factors of the various complexes were as follows: G3P = 19.4% (50.0–2.2 Å), 3-PGA = 17.2% (50.0–2.2 Å), 3-PP = 16.4% (50.0–2.6 Å).

Each model was then subject to five cycles of TNT *B*-factor refinement with imposed correlation of *B*-factors for covalently bonded atoms. This restraint was given twice as much weight in refinement of the 3-PP model as in refinement of the other two, due to the low resolution of the 3-PP dataset. Electron density maps were calculated using all available reflections, bulk solvent scattering modeled as for the refinement, and weights provided by the SIGMAA program.[32] Details of the protein were rebuilt in a $2mF_o-DF_c$ exp $i\alpha c$ electron density map, using the program FRODO[33] running on an Evans and Sutherland PS390. At the same time, the substrate analogues were slightly remodelled in an mF_o-F_c exp $i\alpha c$ map calculated with the substrate analogue absent from the phasing model. In addition, some well-defined water molecules were selected from a list of the peaks greater than six times the RMS electron density of these difference density maps. Possible waters were screened on the basis of stereochemical feasibility, as well as a subjective appraisal of their apparent electron density and position.

For all structures the manual alterations were generally minor, except in the region of the protein loop around residue 512. In this region, clear electron density suggested the flip of the 512–513 peptide plane. Further refinement of the complexes consisted of a combination of manual inspection of the structures in $2mF_o-DF_c$ electron density, and automatic refinement with both TNT and low temperature (100 K) molecular dynamics. Waters were allowed individual B-factors in refinement of the G3P and 3-PGA structures, but due to the limited resolution of the 3-PP data, waters in this structure were constrained to all have the same B-factor, which was allowed to vary during the refinement. The final mean atomic *B*-factors of the protein subunits, and the atomic B-factors of all ligand atoms are given in Table III. The individual *B*-factors of ligand atoms are relatively high, a phenomenon which could be a

TABLE IV. Refinement Statistics of TIM Substrate Analogue Complexes

	3-PGA	G3P	3-PP
R-factor*	14.0	13.7	12.5
Resolution range (Å)	50.0–2.2	50.0–2.2	50–2.6
Reflections	16813	16881	9691
Model content			
Protein atoms	3778	3778	3778
Water atoms	150	91	38
Ligand atoms	11	10	9
rms deviation from ideality:			
Covalent bond lengths (Å)	0.016	0.014	0.014
Bond angles (degrees)	2.5	2.5	2.5

$$*R\text{-factor} = \frac{\sum_h |(F_{\text{obs},h} - F_{\text{calc},h})|}{\sum_h |F_{\text{obs},h}|} \times 100\%.$$

result of incomplete occupancy, although the resolution of the data was judged insufficient to allow inclusion of ligand occupancy parameters in refinement.

At an intermediate point in the refinement of the 3-PGA complex, a slight ambiguity arose about the positioning of the carboxyl and hydroxyl groups of the ligand. The apparent fit into electron density of the ligand was not significantly altered by exchanging the positions of these two groups. The two resulting conformations were each refined to convergence with low temperature molecular dynamics and TNT refinement (data not shown). The correct option was finally chosen by an analysis of the protein–ligand interactions resulting from the ligand in the two different conformations. GROMOS calculation of the nonbonded interactions in the two conformations showed a Lennard–Jones potential term of -34.2 kJ mol^{-1} for the rejected solution, compared to -68.2 kJ mol^{-1} for the chosen conformation, whilst the electrostatic potential term of the rejected conformation was 128.0 kJ mol^{-1} compared to 115.2 kJ mol^{-1} for the chosen conformation. In support of this choice, the atomic *B*-factors of the ligand refined in the incorrect conformation were seen to be inconsistent, with the hydroxyl oxygen having a low relative *B*-factor (O2:19 Å^2), and the oxygens of the carboxyl group having much higher *B*-factors (OT1:77 Å^2 and OT2:54 Å^2). These figures can be compared with the more consistent atomic *B*-factors of the correct conformation of 3-PGA given in Table III. Final confirmation of the correctness of the chosen ligand confirmation came from the excellent fit of the ligand into omit electron density calculated at the end of the refinement (Fig. 5). The final refinement statistics of the complexes are given in Table IV.

For the purposes of comparison, the refined structures were all superimposed upon the native structure on the basis of the core of subunit 2, i.e., the 105 Cα atoms corresponding to the 8 β-sheets and 8 α-helices of subunit 2. Solvent accessibility calculations were carried out using the program WHAT IF,[34] which allows selection of target residues for which solvent accessibility is to be calculated, and the protein molecules to be used as an environment. Individual atomic radii were used in the calculations, and a probe radius of 1.4 Å was chosen.

RESULTS

Determination of the structures of the three complexes here described leads to refinement statistics given in Table IV. Substrate analogues were observed to bind only in the active site of subunit 2, the same subunit which contains a sulphate ion in the "native" structure. In all of the complexes, the flexible loop of subunit 2, consisting of residues 467–480, is in its characteristic "fully closed" conformation. The structure of subunit 1 was seen to be the same in these structures as in the native structure.

Coordinate Accuracy

Before analysis of the structures is undertaken, it is important to consider their accuracy and achieve an estimate of the possible errors in the coordinates determined. The final refinement statistics of the three complexes, each with *R*-factors less than 15% and RMS deviation from covalent bond length ideality of less than 0.017 Å (Table IV), suggest generally good structures, but cannot give a quantitative estimate of coordinate errors. One estimate of this value is given by the Luzzati plot (Fig. 4)[35] which indicates an RMS coordinate error of less than 0.2 Å.

Since three similar structures have been determined here, an estimate of combined coordinate error is available from pairwise comparison of atomic positions in the three models along with that of the native structure. Root mean square coordinate differences for all protein atoms with *B*-factors less than 50 and all possible pairwise structure comparisons are tabulated in Table V. Although the structures of the complexes were all derived from the na-

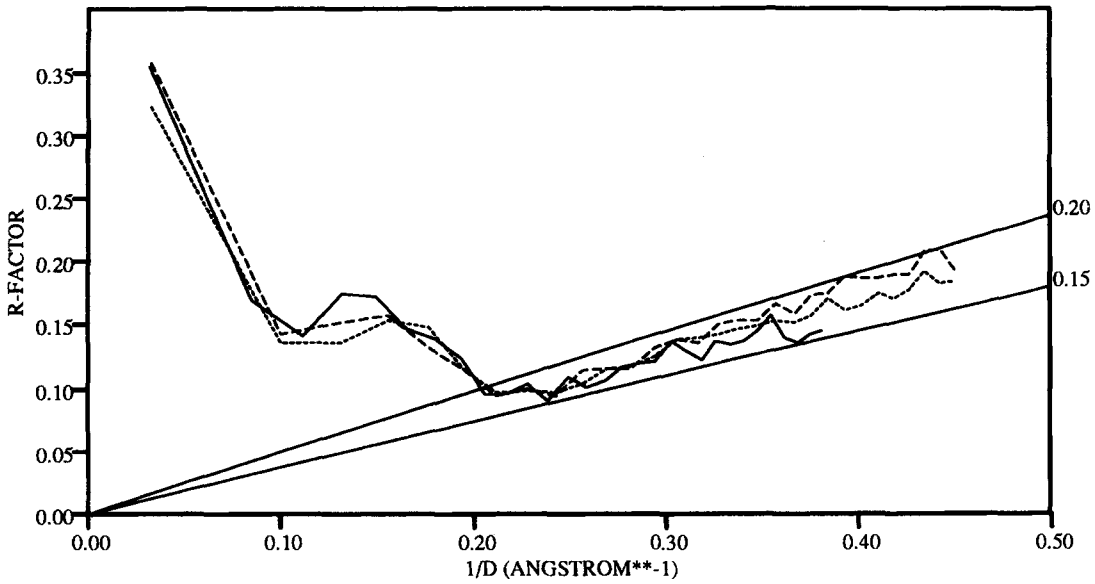

Fig. 4. Luzzati plots of trypanosomal TIM complexes. The variations of *R*-factor as a function of 1/resolution for the complexes with 3-PP (solid line), G3P (short dashed line), and 3-PGA (long dashed line) are compared with theoretical curves corresponding to coordinate errors of 0.15 and 0.20 Å (solid lines labeled at the right-hand side of graph).

tive structure, the extensive use of MDX in the refinement, giving ample opportunity for structure divergence, makes this means of estimation of coordinate error reasonable. The pairwise differences of about 0.4 Å suggests an overall coordinate error of 0.2–0.3 Å in each structure.

Another way of assessing the local accuracy of a part of a crystallographically determined structure is to look at the fit of that region into so called "omit density." Omit density is mF_o-F_c electron density calculated with phases derived from a structure which has been refined for some cycles without the region under examination in the atom list. The fit of each of the three ligands into omit density is shown in Figure 5, from which it may be seen that the ligands, especially G3P and 3-PGA, are well defined. This implies that the accuracy of the coordinates of the ligands is likely to be within the upper limit estimate deduced from Luzzati plots and coordinate comparisons (Fig. 4). The poor electron density around C2 and C3 of 3-PP, despite the high level of structure refinement, leaves some uncertainty about the positioning of these atoms. Clear density for the phosphonate moiety and the carboxyl moiety suggests, however, that only the central portion of the ligand is poorly defined. The poor density around the C2 and C3 atoms of 3-PP is not reflected by high individual *B*-factors for these atoms (Table III). This apparent inconsistency is probably due to the imposition during refinement of correlation between *B*-factors for covalently bonded atoms, which prevents the C2 and C3 atoms of 3-PP from becoming very high. From these considerations, and with the aforementioned reservations about the 3-PP complex, an accuracy of around 0.3 Å can be assumed for further study of the atomic models.

Protein–Ligand Interactions

In considering a complex as a basis for rational drug design it is important to try to understand the interactions responsible for determining the binding mode of the ligand. The protein ligand contacts of each complex are shown in Figure 6, divided into two classes. In Figure 6A, C, and E are shown the protein–ligand hydrogen bonds, both direct and water mediated. An arbitrary cut off distance of 3.2 Å has been used to select potential hydrogen bonds. Figure 6B, D, and F show all direct protein–ligand contacts of less than 3.5 Å in length, excluding those already shown in Figure 6A, C, and E. A three-dimensional view of the disposition of active site residues around each of the three ligands is shown in Figure 7.

The interactions will be considered in three categories; first, direct protein–ligand hydrogen bonds, second, water-mediated protein–ligand hydrogen bonds, and finally protein–ligand van der Waals interactions.

Direct protein–ligand hydrogen bonds

In all of the complexes, the phosphate/phosphonate group receives a similar pattern of hydrogen bonds from the protein (Fig. 6). The coordination of this group is performed by peptide NH groups from three discreet regions of the sequence, namely (1) the flexible loop residue Gly-473, (2) the loop 7 res-

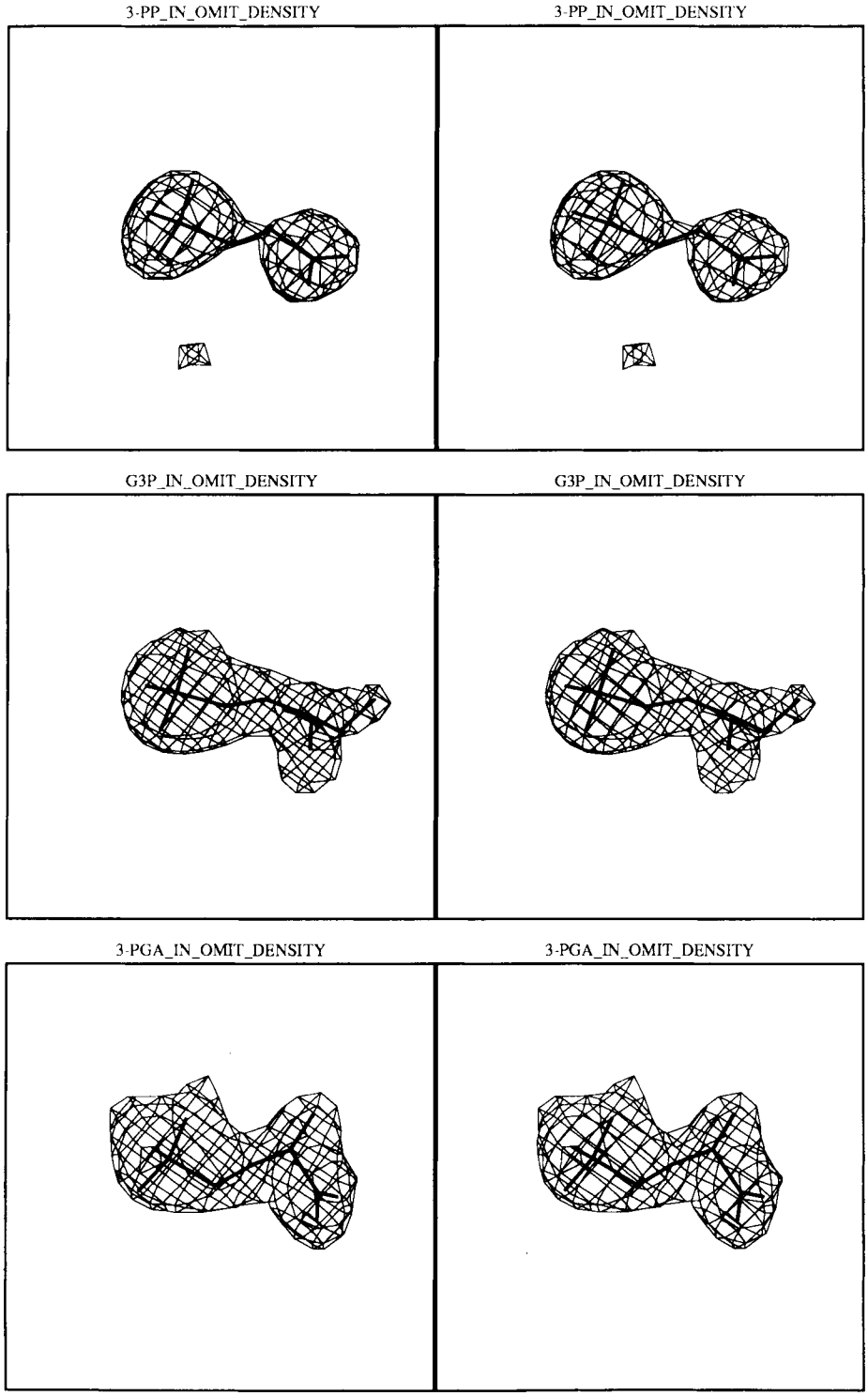

Fig. 5. Stereoscopic picture of the fit of substrate analogues into omit density. The atomic models of ligands (thick lines) are superimposed upon "omit" electron density calculated at the end of the refinement. The maps are contoured at eight standard deviations. Note that only one of the possible stereoisomers of each of the chiral compounds 3-PGA and G3P can be built into the density.

idue Ser-513, and (3) two residues of the phosphate binding loop, Gly-534 and Gly-535. The hydrogen bond from Gly-473 is possible in each of the protein ligand complexes due to the fully closed conformation of the flexible loop, while the hydrogen bond with Ser-513 is allowed by the restructuring of loop

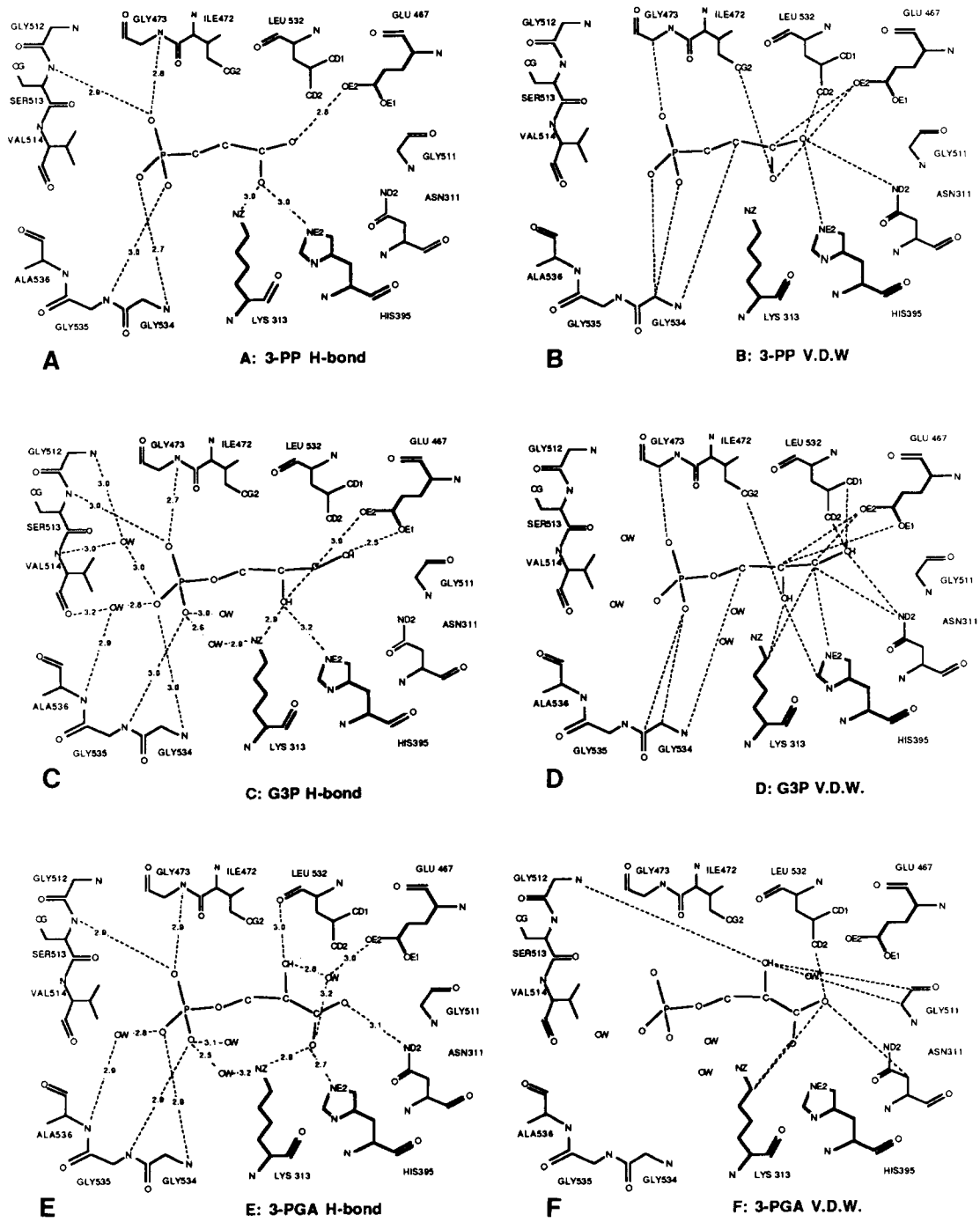

Fig. 6. Protein ligand interactions. **A, C,** and **E:** Direct and water-mediated hydrogen bonds of 3-PP, G3P, and 3-PGA, respectively. The distances between donor and acceptor atoms are indicated (Å). Hydrogen bonds were decided on the basis of a donor acceptor distance of less than 3.2 Å. **B, D,** and **F:** Protein–ligand contacts of 3-PP, G3P, and 3-PGA, respectively. All direct protein ligand contacts of less than 3.5 Å are shown in dotted lines, with the exception of those contacts shown in **(A), (C),** and **(E).**

7 relative to the ligand free or sulfate bound structure. In none of the complexes are any direct contacts made with the "bridging" group of the phosphorous (O4P of 3-PGA and G3P, and C3 of 3-PP).

All three ligands bind in such a way that an oxygen atom sits in a location suitable for hydrogen bonding with NZ of Lys-313 and NE2 of His-395. This oxygen is provided by the hydroxyl group of G3P, or by one oxygen atom of the carboxyl groups of 3-PGA or 3-PP. The environment occupied by the

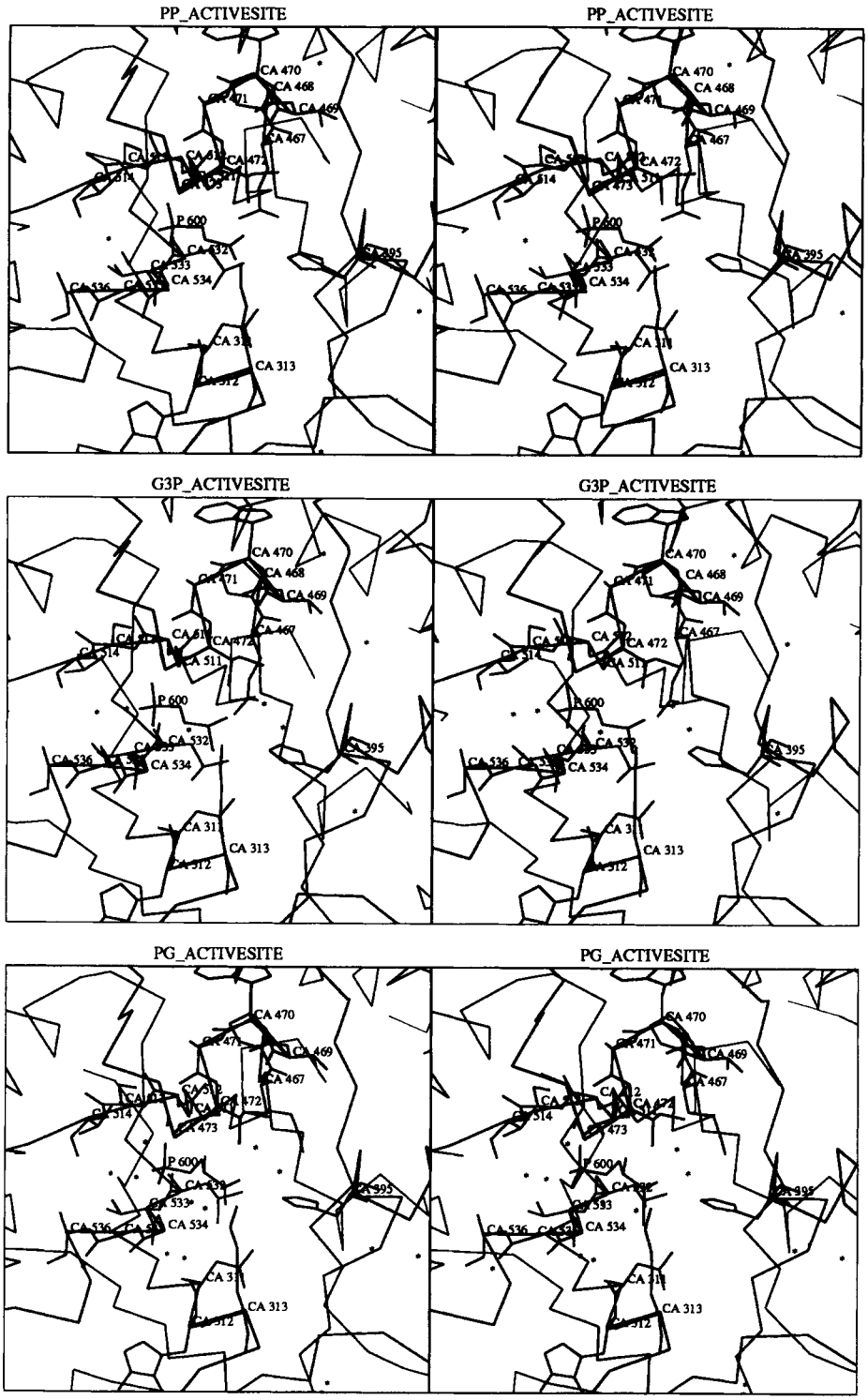

Fig. 7. Stereoscopic pictures of the active site of the complexes of 3-PP, G3P, and 3-PGA with trypanosomal TIM. The residues lining the active sites of the three complexes are drawn in full, upon a Cα trace of the molecule. The active site residues are labeled at their Cα positions, and the ligands are labeled at the phosphorous positions (P 600). Residues 11–13 are at the end of β-strand 1; residue 395 is at the end β-strand 4 and at the beginning of the active site helix; residues 467–473 form part of loop 6, the flexible loop; residues 511–514 form loop 7; and residues 532–536 form the phosphate-binding helix at the end of β-strand 8 (see also Fig. 2).

TABLE V. RMS Differences in Atomic Coordinates Between Trypanosomal TIM Structures*

	Native	G3P	3 PGA	3 PP
Native	—	0.35 Å	0.33 Å	0.41 Å
G3P	0.39 Å	—	0.37 Å	0.39 Å
3-PGA	0.43 Å	0.36 Å	—	0.40 Å
3-PP	0.34 Å	0.31 Å	0.38 Å	—

*RMS coordinate differences of all atoms with *B*-factors less than 50 Å^2 for models after superposition as in Materials and Methods. The top right block contains subunit 1 comparisons, and the bottom left contains subunit 2 comparisons.

second carboxyl oxygen of 3-PP also is similar to that occupied by the second carboxyl oxygen of 3-PGA, involving a hydrogen bond with ND2 of Asn-311. In no case is a direct hydrogen bond between the NZ of Lys-313 and the phosphate or phosphonate group of the ligand observed.

With the exception of the phosphoester oxygen mentioned earlier, all ligand oxygen atoms have protein hydrogen bonding partners. Even 3-PGA, which has a bulk and functional group disposition rather different from the reaction substrate, is able to adopt a conformation such that all oxygens make hydrogen bonding interactions. This conformation is somewhat different from that of the other two ligands, as shown in Figure 8 where the conformations of ligand and catalytic residues are shown for the three complexes described here and subunit 2 of the native structure.

Water-mediated protein–ligand hydrogen bonds

Due to the low resolution of the 3-PP structure few clearly defined water molecules have been included in the model, so that an incomplete picture of water mediated protein–ligand interactions is observed in this complex. In both the 3-PGA and the G3P complexes, however, a number of water-mediated hydrogen bonds are observed. One common interaction is the water mediation of a hydrogen bond from the catalytic residue Lys-313 to the ligand phosphate group. This indirect interaction is the only one connecting the ligand phosphate with a charged protein group. There are several other water mediated hydrogen bonds from loop 7 and the phosphate binding loop to the ligand phosphate group, although due to slight differences in length more are included in the hydrogen bond list of G3P than of 3-PGA.

An additional water-mediated hydrogen bond is observed in the 3-PGA complex, where a water sits between the ligand carboxyl group and the carboxyl group of the catalytic glutamate. This water is also hydrogen bonded to the hydroxyl group of 3-PGA which otherwise forms only one weak hydrogen bond (3.0 Å to the peptide oxygen of Leu-532).

van der Waals Contacts Between Ligand and Protein

Figure 6B, D, and F shows considerable number of contacts between protein and ligand which due to the length or nature of participating atoms are not hydrogen bonds. In the cases of G3P and 3-PP, some of these contacts involve the catalytic residue Glu-467 which occupies a position near ligand carbon atoms. The existence of many more protein ligand van der Waals contacts shows the tightness of fit of the ligands in the active site. Part of the protein surface forming the binding site around the inner end of the ligand binding pocket is made up of the hydrophobic sidechains Leu-532 and Ile-472. These residues are seen to contact ligand oxygen atoms in the complexes of 3-PP and G3P with uncomfortably short carbon oxygen distances around 3.1 Å.

Structural Adaptation of the Enzyme to Ligand Binding

Determining the adaptation of the protein to the binding of ligands may be done by comparing the protein structure observed in enzyme–ligand complexes, with the protein structure observed without the ligands of interest bound. In the case of trypanosomal TIM, there are two structures which might be used as a point of comparison, i.e. the structure of native subunit 1, in which no sulfate is bound to the active site, and the structure of native subunit 2, which has a bound sulfate ion and an "almost closed" flexible loop. The differences observed in the former of these comparisons are dominated by the conformational changes associated with closure of the flexible loop. These gross changes have been closely analysed (Wierenga et al., in press, 1991).

The latter comparison, between ligand and sulfate bound protein structures, shows more subtle changes in protein structure which depend upon the nature of the occupant of the active site. A plot corresponding to this comparison is shown in Figure 9, in which the RMS backbone coordinate differences between each of the three sets of subunit 2 coordinates determined here, and subunit 2 of the native structure are plotted against residue number. It is interesting to note that all of the structures seem to have diverged from the native subunit 2 structure in similar ways, with conserved regions of divergence around residues 330, 401, 473, and 512, as well as the conformationally mobile N- and C-termini.

The shift near residue 330 occurs in a region poorly defined by electron density, which is probably disordered within the crystal. The shift around residue 401 is of interest, since it involves two well-defined residues, Tyr-401 and Tyr-402. These tyrosines are located at the C-terminus of the "active site" helix, which has it's N-terminus pointing into the active site. The catalytic residue His-395, and

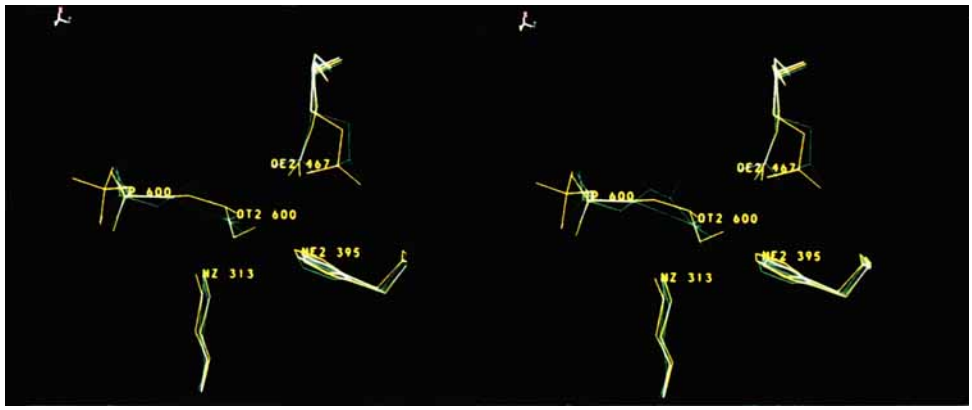

Fig. 8. Stereoscopic picture of the conformations of ligands and catalytic residues. The superimposed atomic models of ligand and catalytic residues (Lys-313, His-395, and Glu-467) are shown for the native (orange), 3-PP (blue), G3P (yellow), and 3-PGA (green) structures. Note the common position in all three ligands of OT2 600. The oxygen at this position interacts with the protein through good hydrogen bonds with NZ of Lys-313, and NE2 of His-395. The distance from the sulfur position to the phosphorous position is in all cases about 1 Å.

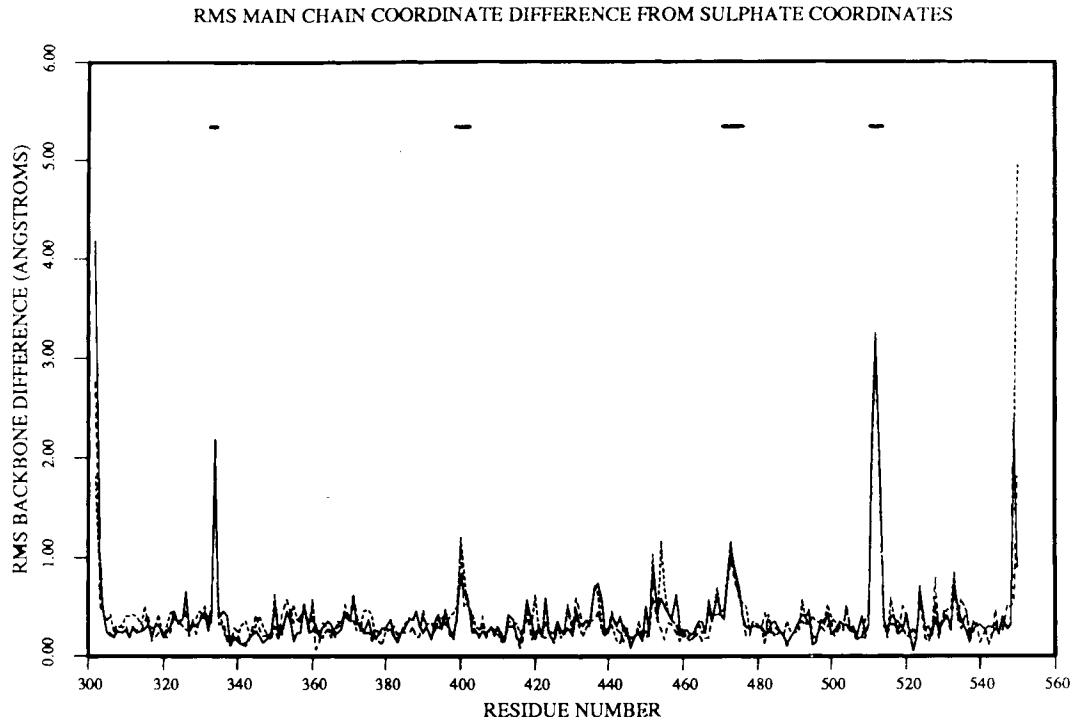

Fig. 9. Coordinate comparison of subunit 2 of the TIM–sulfate complex with each of the TIM–substrate analogue complexes. Rms coordinate differences between the subunit 2 backbone atoms of either (1) native and G3P (solid line), (2) native and 3-PGA (long dashed line), or (3) native and 3-PP (short dashed line) are plotted as a function of residue number. The short solid bars at the top of the figure mark the regions discussed in the text: 333–334, 399–402, 472–476, and 511–514.

residues Ser-396 and Glu-397, which are involved in the architecture of the active site, are situated on the N-terminus of this helix. It seems therefore that Tyr-401 and Tyr-402 might alter conformation on ligand binding due to either the propagation of small conformational changes along the active site helix, or a rigid body movement of this helix. It is less likely that the structural difference results from changes in crystal contacts of these surface residues, since in none of the structures are there any interactions of less than 4 Å made by these residues to crystallographically related molecules.

The shift around residue 473 is best described as a further closure of the flexible loop relative to the flexible loop conformation in a sulfate-containing active site. This shift can be correlated with the po-

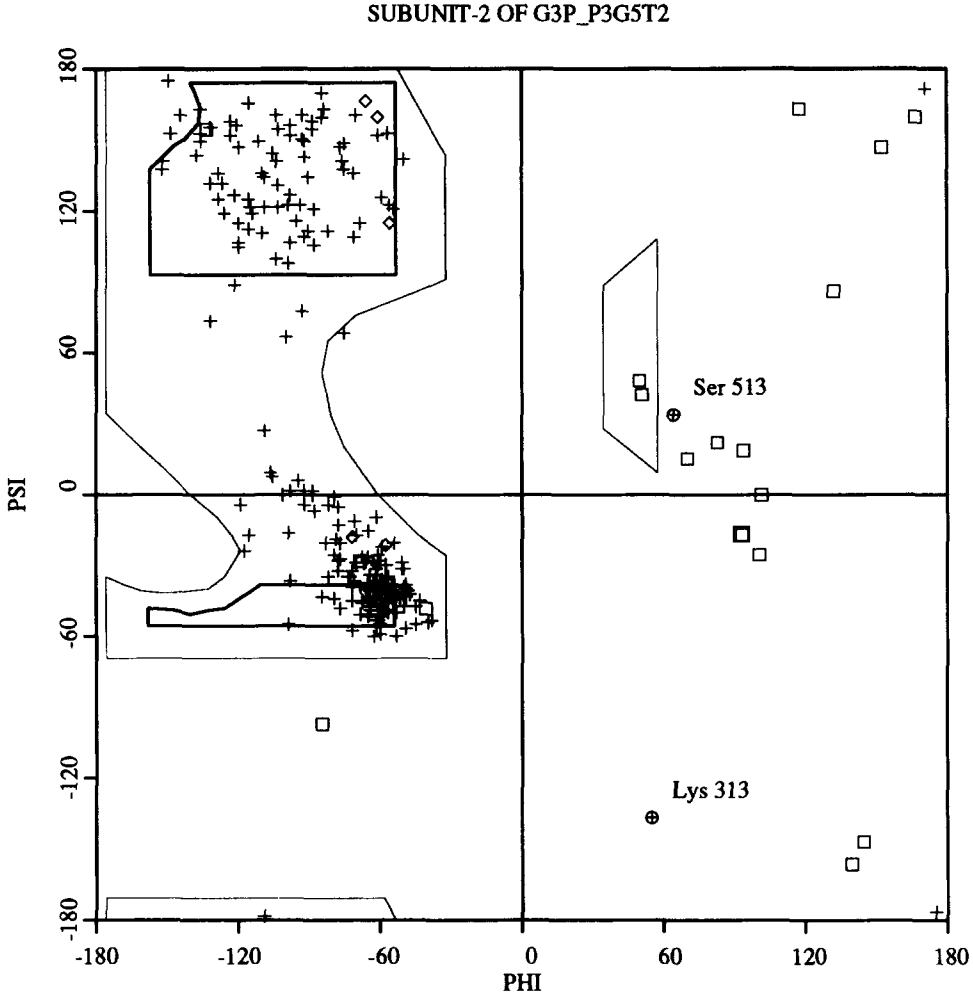

Fig. 10. Ramachandran plot of subunit 2 of the G3P complex. The scattergram of main chain torsion angles of glycine residues (boxes), proline residues (diamonds), and all other residues (plusses) is shown. Encircled are the torsion angles of Lys-313 and Ser-513. The bounded area is the preferred region of ϕ–ψ space described by Ramachandran et al.[42]

sition of the phosphate/phosphonate moieties of the substrate analogues relative to the position of sulfate in the native structure. The phosphorous atoms of the ligands studied here sit further into the active site than does the equivalent sulfur atom in the native structure; the coordinate difference between sulfur and phosphorous atoms after structure superposition is 0.9 Å for 3-PGA, 1.1 Å for G3P, and 1.2 Å for 3-PP (see also Fig. 8). This positioning of the ligand phosphate groups "pulls" further closed the flexible loop through the conserved hydrogen bond between the ligands and the peptide nitrogen of residue 473.

The last region of difference from the sulfate structure is loop 7 comprising residues 510 to 514. In the sulfate structure, the conformation of this loop is poorly defined, as can be seen from its *B*-factors (Fig. 11). In both the ligand-free structure and the substrate analogue complexes, the loop is well defined but shows one of two conformations, differing from each other significantly. These differences include a 180° flip of the peptide plane of Gly-512. The restructuring of this loop relative to its ligand-free conformation allows the formation of a hydrogen bond between the peptide nitrogen of Ser-513 and a terminal phosphate/phosphonate oxygen. The loop conformation observed in the substrate analogue complexes also allows for the OG of Ser-513 to form hydrogen bonds with residues of the flexible loop (2.6 Å to the peptide oxygen of Gly-471, and 3.3 Å to the peptide nitrogen of Gly-475). The strength of these interactions is witnessed by the somewhat unfavorable main chain torsion angles of Ser-513 which fall close to the less frequently occupied left handed α-helical region, as seen in the Ramachandran plot of trypanosomal TIM-G3P complex subunit 2 (Fig. 10). The only other residue far from a preferred region of ϕ–ψ space is the catalytic residue

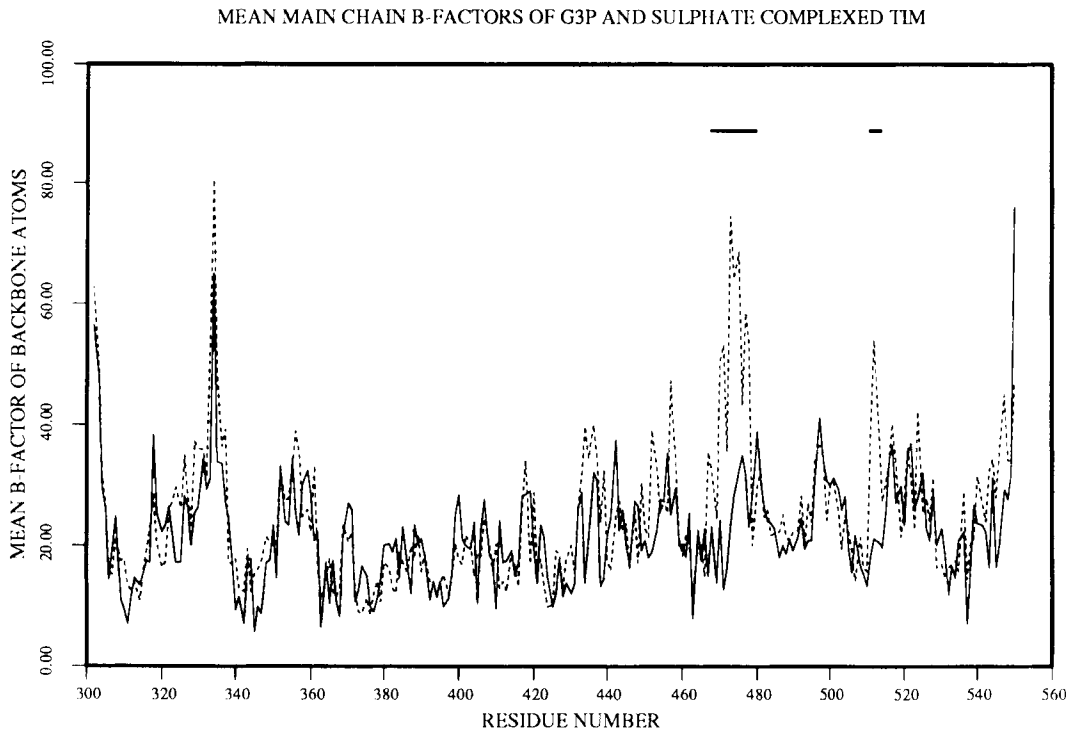

Fig. 11. Mean main chain *B*-factors of native subunit 2 and G3P subunit 2. The mean main chain *B*-factors (N, Cα, C, and O atoms) of G3P subunit 2 (solid), and native subunit 2 (dashed) are plotted as a function of residue number. The short solid bars drawn at the top of the figure mark the regions discussed in the text: 468–580, and 511–514.

Lys-313. The Ramachandran plots of the other complexes shown similar results (data not shown).

The aforementioned differences in backbone conformation relative to the native subunit 2 structure are reflected in changes in atomic *B*-factors. The variation of mean main chain B-factor per residue with subunit 2 sequence number is plotted in Figure 11 for the G3P complex and the native structure. The high *B*-factors of the flexible loop atoms observed in subunit 2 native have been decreased by the exchange of sulfate for substrate analogue at the active site. The residues near Gly-512 also have lower atomic *B*-factors in the G3P complex than in the sulphate complex. Similar results are observed for main chain *B*-factor plots of the subunit 2 residues of the 3-PP and 3-PGA complexes (data not shown).

Another way in which the active site changes in response to ligand binding is in the loss of some tightly bound water molecules, and the repositioning of others. Analysis of the changes in active site water structure brought about by ligand binding are best done by comparing the water structure of the ligand free subunit (subunit 1), with the structure of the ligand bound subunits. The protein structural changes which occur when the ligands bind make it hard to say how many well-defined waters of the ligand free complex are displaced by ligand binding, but a qualitative estimate is possible. The combination of protein structural changes and ligand binding leads to the loss of three waters from the active site cleft of trypanosomal TIM + G3P or 3-PP compared to the uncomplexed enzyme. In the 3-PGA complex, however, one less water is lost from the active site. The active site water which is observed in the 3-PGA complex but not in the G3P complex is the water which contacts the carboxyl and hydroxyl groups of 3-PGA and the carboxyl group of Glu-467 (Fig. 6). Although this water has no equivalent in ligand free TIM, there is a water in an equivalent position in the sulfate complex. Around the phosphate groups of G3P and 3-PGA it appears that most waters present in the complexes have equivalents in the ligand free structure. An exception to this is the water which hydrogen bonds only to a phosphate oxygen atom of G3P or 3-PGA (Fig. 6). As might be expected, there is no apparent equivalent to this water in the ligand-free structure. The water molecules displaced by ligand binding are generally not replaced by oxygen atoms of the ligands. The only exception is a water of the ligand-free structure which is replaced by a carboxyl atom of either 3-PP or 3-PGA in those complexes. The equivalent location is occupied by the C1 methylene group in the G3P complex.

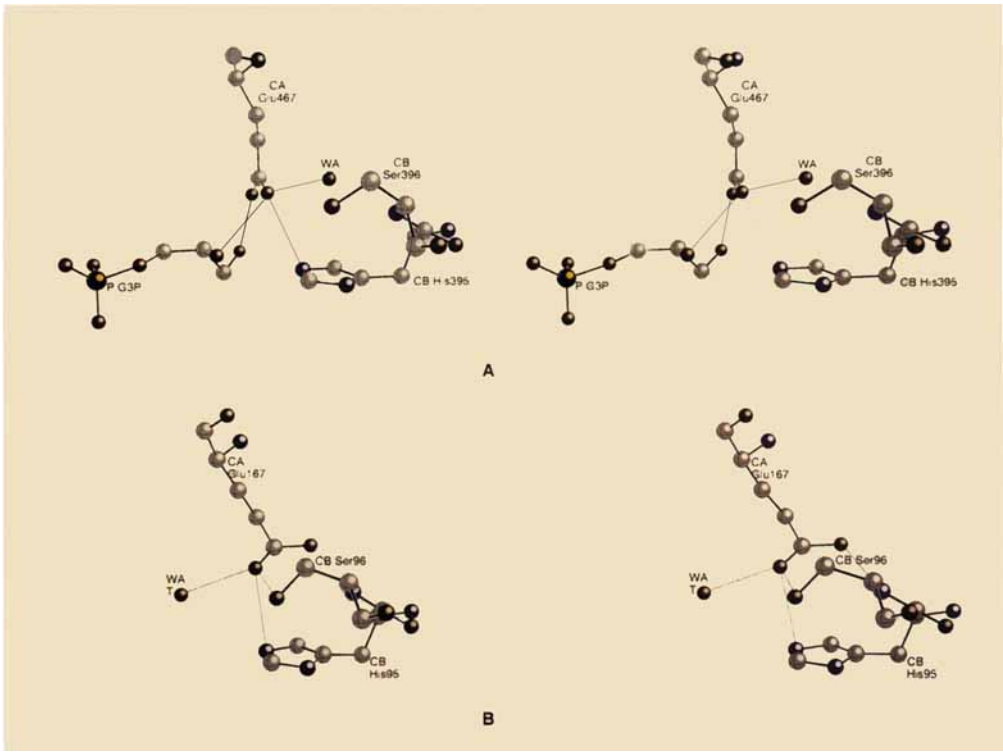

Fig. 12. Stereoscopic picture of the interactions of Glu-467 in two conformations. The hydrogen bonding interactions of Glu-467 in the "swung out" subunit 1 conformation **(A)** and the "swung in" conformation observed in complex with G3P **(B)** are shown (thin dark lines). Potential hydrogen bonds were decided on the basis of a cut off distance between donor and acceptor atoms of 3.2 Å. Notice that the position vacated by the movement of the glutamate is filled by the binding of a water molecule.

Conformational Variation of Glu-467

The structures of the TIM active sites in the complexes with 3-PP, 3-PGA, or G3P are all very similar, despite the marked differences between these three ligands. For residues interacting with the ligands, the RMS differences in atomic coordinates for different complexes are in the region of 0.2–0.4 Å. The only side chain which has significantly different conformations in the three complexes is the catalytic residue Glu-467. This residue adopts a range of conformations to locate its carboxyl moiety at a position dependent on the nature of the bound species (Fig. 8). The carboxyl group is found in a position close to the substrate analogues in complex with G3P and 3-PP, while in the 3-PGA complex the carboxylate is nearly 4 Å away from any ligand atom, occupying a position similar to that observed in subunit 1 and subunit 2 of the native structure.

The conformation of Glu-467 observed in complex with G3P allows hydrogen bonds from ligand hydroxyl groups to each of the carboxyl oxygens. In this position, the OE2 atom of Glu-467 is situated 3.0 Å from the C2 of the ligand. In the complex with 3-PP the glutamate occupies a similar position, able to hydrogen bond to the carboxyl group of 3-PP. Figure 12 compares the environment of the glutamate in complex with G3P (Fig. 12A), with the environment in the ligand free subunit 1 (Fig. 12B).

In the complex with 3-PGA, no direct interaction is observed between the ligand and Glu-467, although there is a water molecule mediating a contact between the carboxyl groups of ligand and catalytic base (Fig. 6). Instead, the carboxyl group of 3-PGA interacts with NE2 of His-395, NZ of Lys-313, and the ND2 of Asn-311, while the carboxyl group of Glu-467 forms an anchoring hydrogen bond in the "swung out" position to the main chain nitrogen of Ser-396. This difference between the protein interactions of the two carboxylic acids 3-PP and 3-PGA is surprising in view of the fact that the carboxyl group of 3-PGA has a location and orientation similar to that of the carboxyl group of 3-PP.

Conformation of Bound Ligands

Since crystal structures exist for each of the three substrate analogues not complexed with enzyme,[25–27] there is also the opportunity to assess the effect of the protein on the ligand conformation. The resolution of the protein–ligand complexes is of course too poor to examine changes in bond lengths or angles, but the torsion angles describing the conformation of the molecules can be compared. It is

TABLE VI. Atomic Solvent Accessibility of G3P in Different Flexible Loop Contexts*

Accessible surface area (Å^2)	G3P subunit 2 (fully closed)	Native subunit 2 (almost closed)	Native subunit 1 (Open)	No protein
P	3.7	5.6	9.4	21.2
O1P	7.6	8.5	8.9	39.7
O2P	0.8	3.8	11.0	41.4
O3P	0.8	1.3	0.8	34.2
O4P	—	—	—	8.0
C3	—	—	—	19.9
C2	—	—	—	13.3
O2	—	—	3.0	38.5
C1	—	—	—	43.6
O1	—	—	—	41.4

*Atomic van der Waals radii used in the calculations were C = 1.8 Å, O = 1.4 Å, N = 1.7 Å, S = 2.0 Å, P = 2.0 Å. Probe radius was 1.4 Å.

seen that in the crystal structure of uncomplexed 3-PP, the molecule is mostly extended, with O1P, P, C3, C2, and C1 trans-coplanar, but with the carboxyl group almost perpendicular to this plane. This conformation is nearly ideal for the minimization of eclipsing interactions. In the enzyme–ligand complex, however, the conformation is somewhat different, with the carboxyl group strained into the plane of the rest of the molecule. G3P shows a similar difference between free and complexed conformation. When G3P is not bound to TIM, O2 lies 76° out of the plane defined by O1P, P, O4P, C3, and C2, whereas the corresponding torsion angle in the protein ligand complex is only 12°.

The most notable difference between the conformation of complexed and uncomplexed 3-PGA is the change in the orientation of the carboxyl group. In uncomplexed 3-PGA, this group is coplanar with C2 and O2, allowing the formation of an internal hydrogen bond from the hydroxyl to the carboxyl group which stabilizes the resulting eclipsing interaction. In the enzyme–ligand complex, however, this group lies far from coplanar, and no internal hydrogen bond is possible.

A further analysis of the ligand complexes which has been carried out is calculation of solvent accessible surfaces. The figures for solvent accessibility of G3P atoms in the context observed here is given in Table VI, where it is compared with theoretical accessibilities calculated for this molecule in the environment of the active sites observed in native subunit 1 and subunit 2. For comparison, the surface accessibility of the ligand in its bound conformation but without protein environment is also given. It can be seen that the effect of the loop closure is to isolate the molecule from water, increasing not only the concealment of phosphate atoms, but also of O2, the analogue of the substrate ketone oxygen.

DISCUSSION

This work, in conjunction with earlier work, identifies at least three conformations available to the flexible loop of TIM, that is open (subunit 1), almost closed (native subunit 2), and fully closed (subunit 2 of the substrate analogue complexes). It also demonstrates concerted changes elsewhere in the protein associated with the almost closed to fully closed transition, involving particularly the restructuring of residues around Gly-512 in loop 7. The restructuring of this loop affects the nature of the ligand binding site, and must be caused by a mechanism sensitive to the difference between a sulfate molecule and a substrate analogue.

The other most notable conformational difference is the variability in the conformation of Glu-467. This residue has to be flexible to allow it to adopt its active "swung in" conformation. In both sulfate-bound and sulfate-free subunits of the native structure, Glu-467 is seen to be stably hydrogen bonded to Ser-396 in a conformation unsuitable for catalysis (Fig. 12). To adopt its active conformation therefore, the glutamate has to change its hydrogen bonding pattern and a corresponding change in active site water structure has to occur. Such a reordering could be understood in terms of a switching mechanism, with two stable hydrogen bonding patterns defining the two discrete states of this side chain.

In general, the binding of a ligand to an enzyme can be described by the following chemical equation:

$$\text{Enzyme·water} + \text{Ligand·water} \rightarrow \text{Enzyme·Ligand·water} + \text{Bulk water}$$

The free energy associated with this reaction can be divided into three major parts: first, the free energy of desolvation of the enzyme and ligand; second, the free energy changes in vacuum of protein and ligand adopting the conformations they have when complexed in solution; and third, the energy of the protein ligand and water interactions in the enzyme.ligand.water complex. Each of these parts will be made up of an enthalpic and entropic contribution, and the net free energy change of ligand binding will be a small difference between the resulting balance of large energy changes. Although it is there-

fore currently impossible to predict binding constants from the structures of the enzyme–ligand complexes, it is possible to identify structural features which will contribute to the balance. Since G3P is the compound studied here most closely related to substrate, it shall be used as the example to be discussed.

The first observation is that no previously unsatisfied protein hydrogen bond is satisfied by ligand binding. Also, no ligand or protein group is deprived of a hydrogen bonding partner by the process of complex formation. The net binding free energy due to hydrogen bonding will therefore depend upon subtle difference in hydrogen bonding strengths rather than the full bond strength of a lost or gained hydrogen bond. The delicacy of the balance between the free energy gained from a protein ligand hydrogen bond, and the corresponding free energy loss due to desolvation energy has been demonstrated recently in several thermodynamic perturbation studies.[36,37] By analogy with G3P, it is possible to conclude that the substrate also will have fully satisfied hydrogen bonding potential. Although, as mentioned above, each hydrogen bond may make a limited contribution to the net free energy change of complex formation, the prerequisite of satisfying all potential hydrogen bonds of a ligand will determine the mode of binding of that ligand to the enzyme. This can be exploited by nature in orienting substrates relative to catalytic residues, as well as by drug designers aiming to exploit differences in the patterns of hydrogen bonding groups on a protein surface.

The entropic contribution to the binding energy will depend on the entropic gain of a net decrease in constrained water molecules from the active site and the ligand, opposed by the loss of entropy as the ligand becomes bound to the enzyme. Another contributing factor is the rigidification of the flexible loop, which is seen to have high *B*-factors in the ligand free enzyme, but low B-factors in the enzyme-substrate analogue complexes. The implications of this change in flexibility have been discussed previously.[38]

The catalytic glutamate is seen to be involved in a protein–protein hydrogen bond in the free enzyme, but in a protein–ligand hydrogen bond in complex with G3P (Fig. 12). It would seem that this residue could make a greater contribution to net binding energy if it had no hydrogen bonding partner in the free state. Such a change would increase the binding energies of all species along the catalytic path, and so alter the "uniform binding criterion" discussed by Albery and Knowles.[39] Since TIM is an energetically optimized enzyme, this change would be expected to lead to lower catalytic efficiency if applied to the wild type enzyme, but might improve the efficiency of a sluggish mutant enzyme related to the wild type by a point mutation. This observation could explain recent results of genetic pseudo-reversion analysis of sluggish chicken TIM mutants altered in either their catalytic base (Glu-165-Asp)[40] or histidine residue (His-95-Gln). The rescue to these chicken TIM mutants was best effected by mutation of Ser-396, the "ground state" hydrogen bonding partner of Glu-165, to proline, a residue unable to form the same hydrogen bonds.

Interestingly, the complexed ligands themselves are seen to exist in conformations different from their expected energy minima, or their uncomplexed crystal structures. The energy differences involved in inducing an eclipsed ligand conformation may be as large as a few kilocalories per mole, a figure comparable to the net energy gains reported for an enzyme-substrate hydrogen bond.[41] These energies therefore play a part in deciding the strength and mode of ligand binding. In this case, a similar deformation is seen to occur to both G3P and 3-PP, i.e., the straining of the oxygen corresponding to O2 of G3P toward an eclipsing interaction with the phosphate/phosphonate bridging group. That the enzyme should apply such a strain to substrate, and so by analogy to substrate analogues, can be understood from a stereoelectronic consideration of the isomerisation reaction. It has been pointed out[9] that an extended conformation of the enediolate intermediate reaction is important to minimize W-elimination, a side reaction which can proceed from the enediolate reaction intermediate leading to inorganic phosphate and toxic methyl glyoxal.

A similar conformation is observed for phosphoglycollate in complex with yeast TIM[45] as is observed for phosphonopropionate here in complex with trypanosomal TIM. In both cases there is an eclipsing interaction between a carboxyl oxygen and the bridging group to the phosphate/phosphonate moiety. Since in phosphonopropionate this bridging group is a methyl group and in phosphoglycollate an oxygen atom, the energetic price paid is likely to be different in two cases.

It is possible to speculate that the protein–ligand interactions providing the strain which forces the extended conformation upon substrate or intermediate analogues complexed with TIM are the coordination of the phosphate/phosphonate group at one end, involving a hydrogen bond from the flexible loop, and the coordination of an oxygen corresponding to the ketone oxygen of the substrate (O2 in G3P, OT2 in 3-PP and 3-PGA) at the other. These interactions are seen to be conserved in all of the enzyme inhibitor structures. It is also interesting to note that the ketone oxygen of the substrate has been shown to be tightly bound by TIM by means of fourier transform infrared spectroscopy.[4]

The closeness of the carboxyl group of Glu-467 to the carboxyl group of 3-PP implies that one or the other of these carboxyl groups is protonated, so that a hydrogen bond can exist between the two. This is

supported by studies on the binding of 2-phosphoglycollate, the phosphate analogue of 3-PP, to chicken TIM.[43] The complex between chicken TIM and phosphoglycollate is found to involve the trianionic species of the ligand, and an additionally protonated protein group, probably the active site glutamate. The pK_a of the carboxyl group in the environment of the complex is presumably raised by the strength of the interaction possible between two carboxyl groups sharing a proton. That Glu-467 remains "swung out" in the 3-PGA complex suggests that proton uptake does not occur when 3-PGA binds to TIM, so that both ligand and catalytic base carboxyl groups remain charged, and therefore repulsive to each other.

The solvent accessibility calculation shows that in an open loop environment O2 of G3P would be exposed to solvent, whereas with the loop closed this is not so. The contact missing from this oxygen when the molecule is placed in an open loop context is the contact to the hydrophobic residue Ile-472. Together with Leu-532, this residue provides a hydrophobic environment around the O1 and O2 atoms of G3P and thus by analogy around the substrate ketone and hydroxyl groups. Such an environment will have different electrostatic properties to one either exposed to solvent, as in the open-loop structure, or surrounded by polar residues, a fact which might in turn be expected to influence catalysis.

For drug design, it is necessary that the ligands should be increased in size to allow them to bridge the 10 or so Å separating the active site from regions of sequence specificity. Any potential drug requires interactions with such regions in order to provide selectivity of inhibition, so that clinical usage should not be harmful to the patient. The solvent accesibility calculations show that with the loop fully closed around a ligand, there is very little free space within which the inhibitors may be allowed to grow. Indeed, the only exposed ligand atoms are the phosphorous atom and the terminal oxygen atoms attached to it. For this reason, a limited number of means of development are suggested. Using a phosphate or phosphonate as an anchor, compounds worth examining should be either (1) too bulky to sit within the active site cleft of TIM, so that they are reoriented toward the outside of the protein, (2) designed to sit inside the substrate binding pocket, but branched in such a way that part of the molecule leads out of the active site, or (3) phosphodiester compounds with one branch designed to lead from the outside of the active site in the direction of species specific sequence loci.

Work is currently underway to explore some of these possibilities. It is to be hoped that in subsequent turns of the rational drug design cycle, further understanding of ligand–enzyme interactions can be achieved. The number of interactions contributing to the overall binding energy of a ligand to a protein molecule makes such understanding essential to the task of producing accurate knowledge-based predictions of molecules able to selectively inhibit trypanosomal enzymes.

ACKNOWLEDGMENTS

The experiments with 3-phosphonopropionate were suggested by the later professor Dr. Alan Horn, whom we will remember for his active interest in our project.

We thank Drs. Jeremy Knowles, Greg Petsko, and Elias Lolis for stimulating discussions and for sending us manuscripts prior to publication. This research had financial support from the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases through grants to WGJH and FRO.

REFERENCES

1. Knowles, J.R., Albery, W.J. Perfection in enzyme catalysis: The energetics of triosephosphate isomerase. Acc. Chem. Res. 10:105–111, 1977.
2. Albery, W.J., Knowles, J.R. Free-energy profile for the reaction catalysed by triosephosphate isomerase. Biochemistry 15:5627–5631, 1976.
3. Reider, S.V., Rose, I.A. The mechanism of the triosephosphate isomerase reaction. J. Biol. Chem. 234:1007–1010, 1959.
4. Belasco, J.G., Knowles, J.R. Direct observation of substrate distortion by triosephosphate isomerase using Fourier transform infrared spectroscopy. Biochemistry 19: 472–477, 1980.
5. Banner, D.W., Bloomer, A.C., Petsko, G.A., Philips, 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. Structure of chicken muscle triosephosphate isomerase determined crystallographically at 2.5 Å resolution, using the amino acid sequence data. Nature (London) 255:609–614, 1975.
6. Alber, T., Hartman, F.C., Johnson, R.M., Petsko, G.A., Tsernoglou, D. Crystallization of yeast triosephosphate isomerase from polyethylene glycol. J. Biol. Chem. 256: 1356–1361, 1981.
7. Wierenga, R.K., Kalk, K.H., Hol, W.G.J. Structure determination of the glycosomal triosephosphate isomerase from *Trypanosoma brucei brucei* at 2.4 Å resolution. J. Mol. Biol. 198:109–121, 1987.
8. Chothia, C. The 14th barrel rolls out. Nature (London) 333:598–599, 1990.
9. Alber, T., Banner, D.W., Bloomer, A.C. Petsko, G.A., Phillips, D.C., Rivers, P.S., Wilson, I.A. On the three-dimensional structure and catalytic mechanism of triosephosphate isomerase. Phil. Trans. R. Soc. London B293:159–171, 1981.
10. Pompliano, D.L., Peyman, A., Knowles, J.R. Stabilization of a reaction intermediate as a catalytic device: Definition of the functional role of the flexible loop of triosephosphate isomerase. Biochemistry 29:3186–3194, 1990.
11. Davenport, R.C. Jr., Ph.D. Thesis (Massachusetts Institute of Technology, Cambridge, MA), 1986.
12. Straus, D., Raines, R.T., Kawashima, E., Knowles, J.R., Gilbert, W. Active site of triosephosphate isomerase: *In vitro* mutagenesis and characterization of an altered enzyme. Proc. Natl. Acad. Sci. U.S.A. 82:2272–2276, 1985.
13. Nickbarg, E.B., Davenport, R.C., Petsko, G.A., Knowles, J.R. Triosephosphate isomerase: Removal of a putatively electrophilic histidine residue results in a subtle change in catalytic mechanism. Biochemistry 27:5948–5960, 1980.
14. Molyneux, D.H., Ashford, R.W. "The Biology of Trypanosomes and Leishmania, Parasites of Man and Domestic Animals." London: Taylor and Francis, 1983.
15. Hol, W.G.J. Protein crystallography and computer graph-

ics: Towards rational drug design. Angew. Chem. 25:767–778, 1986.
16. Opperdoes, F.R. Biochemical peculiarities of trypanosomes, African and South American. Br. Med. Bull. 41: 130–136, 1985.
17. Fairlamb, A.H., Opperdoes, F.R., Borst, P. New approach to screening drugs for activity against African trypanosomes. Nature (London) 265:270–271, 1985.
18. Misset, O., Bos, O.J.M., Opperdoes, F.R. Glycolytic enzymes of *Trypanosoma brucei;* simultaneous purification, intraglycosomal concentrations and physical properties. Eur. J. Biochem. 157:441–453, 1986.
19. Wierenga, R.K., Hol, W.G.J., Misset, O., Opperdoes, F.R. Preliminary crystallographic studies of triosephosphate isomerase from the blood parasite *Trypanosoma brucei brucei.* J. Mol. Biol. 178:487–490, 1984.
20. Lambeir, A.-M., Opperdoes, F.R., Wierenga, R.K. Kinetic properties of triosephosphate isomerase from *Trypanosoma brucei brucei.* Eur. J. Biochem. 168:69–74, 1987.
21. Schreuder, H.A., Van der Laan, J.M., Wierenga, R.K. The transfer of protein crystals from their original mother liquor to a solution with a completely different precipitant. J. Appl. Crystallogr. 21:426–429, 1988.
22. Pflugrath, J.W., Messerschmidt, A. MADNES: Users Guide. Max-Planck-Institut für Biochemie, Martinsried, Federal Republic of Germany, 1986.
23. Tronrud, D.E. Ten Eyck, L.F., Matthews, B.W. An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Crystallogr. A43: 489–501, 1987.
24. Fermi, G., Perutz, M.F., Shaanen, B., Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74Å resolution. J. Mol. Biol. 175:159–174, 1984.
25. van Bolhuis, F. 3-phosphonopropionic acid-ion. Structural Report MSD015, Farmacie, Rijksuniversiteit Groningen, The Netherlands, 1988.
26. McAlister, J., Sundralingam, M. Redetermination of disodium DL-α-glycerophosphate hexahydrate. Acta Crystallogr. B36:1652–1654, 1980.
27. Fewster, P.F., Fenn, R.H. Disodium D-3-phosphoglycerate, a substrate to phosphoglycerate kinase. Acta Crystallogr. B38:282–284, 1982.
28. Fujinaga, M., Gros, P., van Gunsteren, W.F. Testing the method of crystallographic refinement using molecular dynamics. J. Appl. Crystallogr. 22:1–8, 1989.
29. Gros, P., Fujinaga, M., Dijkstra, B.W., Kalk, K.H., Hol, W.G.J. Crystallographic refinement by incorporation of molecular dynamics: Thermostable serine protease thermithase complexed with Eglin C. Acta Crystallogr. B45:488–499, 1989.
30. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Dinola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
31. Van Gunsteren, W.F., Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
32. Read, R.J. Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. A42:140–149, 1986.
33. Jones, T.A. Interactive computer graphics: FRODO. Methods Enzymol. 115:157–171, 1985.
34. Vriend, G. WHAT IF: A molecular modeling and drug design program. J. Mol. Graphics 8:52–56, 1990.
35. Luzzati, V. Traitement Statistique des Erreurs dans la Determination des Structures Cristallines. Acta Crystallogr. 5:802–810, 1952.
36. Fleischman, H., Brooks, C.L. III. Protein-drug interactions: Characterisation of inhibitor binding in complexes of DHFR with trimethoprim and related derivatives. Proteins: Structure, Function and Genet. 7:52–61, 1990.
37. Hirono, S., Kollman, P.A. Calculation of the relative free energy of 2′GMP and 2′AMP to ribonuclease T1 using molecular dynamics/free energy perturbation approaches. J. Mol. Biol. 212:197–209, 1990.
38. Alber, T., Gilbert, W.A., Ponzi, D.R., Petsko, G.A. The role of mobility in the substrate binding and catalytic machinery of enzymes. In: "Mobility and Function in Proteins and Nucleic Acids." Ciba Foundation Symposium 93. London: Pitman, 1982: 4–24.
39. Albery, W.J., Knowles, J.R. Evolution of enzyme function and the development of catalytic efficiency. Biochemistry 15:5631–5640, 1976.
40. Hermes, J.D., Blacklow, S.C., Knowles, J.R. Searching sequence space by defineably random mutagenesis: Improving the catalytic potency of an enzyme. Proc. Natl. Acad. Sci. U.S.A. 87:696–700, 1990.
41. Fersht, A.R. Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. Biochemistry 26:8031–8037, 1987.
42. Ramachandran, G.N., Sasisekharan, V. Conformation of polypeptides and proteins. Adv. Prot. Chem. 23:283–437, 1968.
43. Campbell, I.D., Jones, R.B., Kiener, P.A., Richards, E., Waley, S.G., Wolfenden, R. The form of 2-phosphoglycollic acid bound by triosephosphate isomerase. Biochem. Biophys. Res. Commun. 83:347–352, 1978.
44. Lolis, E., Alber, T., Davenport, R.C., Rose, D., Hartman, F.C. Petsko, G.A. Structure of yeast triosephosphate isomerase at 1.9 Å resolution. Biochemistry 29:6609–6618, 1990.
45. Lolis, E., Petsko, G.A. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5 Å resolution. Biochemistry 29:6619–6625, 1990.