Crystallographic and Molecular Modeling Studies on Trypanosomal Triosephosphate Isomerase: A Critical Assessment of the Predicted and Observed Structures of the Complex with 2-Phosphoglycerate

Martin E. M. Noble,† Christophe L. M. J. Verlinde,‡ Hillie Groendijk,‡ Kor H. Kalk,‡ Rik K. Wierenga,*,† and Wim G. J. Hol‡

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany, and Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, NL-9747 AG Groningen, The Netherlands. *Received December 26, 1990*

In the continuation of a project aimed at the rational design of drugs against diseases caused by trypanosomes, the crystal structure of trypanosomal triosephosphate isomerase in complex with the active site inhibitor 2-phosphoglycerate has been determined. Two alternative modeling protocols have been attempted to predict the mode of binding of this ligand. In the first protocol, certain key interactions were restrained in the modeling procedure. In the second protocol, a full search of ligand conformational space was performed. In both cases the protein scaffold was kept static. Both protocols produced models which were reasonably close to the observed structure (rms difference <2.0 Å). Nevertheless, some essential features were missed by each of the protocols. The crystallographic structure of the 2-PGA TIM complex shows that the ligand binds fully within the active site of TIM, with partners for all but one of the ligand's strongly hydrogen bonding groups. Several of the interactions between the ligand and the active site of TIM are seen to be common to all of the complexes so far structurally characterized between trypanosomal triosephosphate isomerase and competitive inhibitors. Such key interactions appear to be the best guide in the prediction of the binding mode of a new inhibitor.

Introduction

The organism *Trypanosoma brucei* is the causative agent of a range of diseases in tropical Africa.[1] Infections of this blood parasite affect both human beings and livestock, causing great human and economic problems. Few treatments exist to aid sufferers, and those which do are harmful to the patient, so that a real need exists to develop new drugs effective against trypanosomiasis.

Trypanosomes have a number of unique biochemical features which may be useful in the development of rationally designed drugs.[2] These differences include in particular the metabolism of the organism which, in its blood-stream form, relies completely upon glycolysis for production of ATP, since it does not carry out a tricarboxylic acid cycle. The enzymes responsible for trypanosomal glycolysis are sequestered in microbodylike organelles called glycosomes, which represent a further unique feature of the energy-transduction system of the parasites. Trypanosomal glycolysis has therefore been selected as the target for a rational drug-design project.[3] The aim of this project is to combine structural information, biochemical testing, and chemical synthesis to produce strong and selective inhibitors of trypanosomal enzymes. It is hoped that one or a combination of such inhibitors could have a successful effect in the treatment of trypanosomiasis.

The first trypanosomal glycolytic enzyme for which a three-dimensional structure has become available is the enzyme triosephosphate isomerase (TIM).[4] This dimeric enzyme acts to interconvert dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) (Figure 1). The kinetics of this reaction have been extensively studied,[5,6] and clues about the mechanism involved have been gained from the X-ray structures of chicken TIM[7] and yeast TIM[8,9] with and without inhibitors. The overall fold of each subunit of the enzyme consists of eight β sheet-α helix units arranged such that the β strands form a parallel central barrel, surrounded by a coat of α helices. Important catalytic roles have been assigned to a lysine residue (equivalent to Lys13 of *T. brucei* TIM), a histidine residue (equivalent to His95 of *T. brucei* TIM), and a glutamic acid residue (equivalent to Glu167 of *T. brucei* TIM) of the active site.[10]

The structure of "native" trypanosomal TIM, from crystals grown in 2.4 M ammonium sulfate, has been solved crystallographically by using the technique of molecular replacement starting with the coordinates of chicken TIM.[4] The asymmetric unit of the native trypanosomal structure contains an α_2 TIM dimer, with 249 amino acids per subunit. In the crystal, the two subunits are distinguished from each other since subunit-1 (residues 2–250) has no sulfate bound at the active site, whereas subunit-2 (residues 302–550) does have a sulfate ion bound at the active site. This difference is reflected by an altered conformation of the "flexible loop" close to the active site of TIM in the two subunits: subunit-1 has an open flexible loop (residues 167–180), whereas subunit-2 has a closed flexible loop (residues 467–480). This loop is thought to play a role in stabilizing intermediates along the reaction pathway and minimizing the phosphate elimination side reaction.[11] The structure has been refined to an *R* factor of 18.3% against data extending to 1.83-Å resolution. The agreement with standard geometric parameters is also good; for example the rms deviation from ideal covalent bond lengths is 0.02 Å, and the deviation from ideal bond angles is 2.9°.

The coordinates of the native structure have also been used in the determination of the structures of cocomplexes

* Author to whom correspondence should be addressed.
† European Molecular Biology Laboratory.
‡ University of Groningen.

(1) Molyneux, D. H.; Ashford, R. W. *The Biology of Trypanosomes and Leishmania, Parasites of Man and Domestic Animals*; Taylor and Francis: London, 1983.
(2) Opperdoes, F. R. *Br. Med. Bull.* **1985**, *41*, 130.
(3) Hol, W. G. J. *Angew. Chem.* **1986**, *25*, 767.
(4) Wierenga, R. K.; Kalk, K. H.; Hol, W. G. J. *J. Mol. Biol.* **1987**, *198*, 109.
(5) Knowles, J. R.; Albery, W. J. *Acc. Chem. Res.* **1977**, *10*, 105.
(6) Albery, W. J.; Knowles, J. R. *Biochemistry* **1976**, *15*, 5627.
(7) 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. *Nature* **1976**, *255*, 609.
(8) Lolis, E.; Alber, T.; Davenport, R. C.; Rose, D.; Hartman, F. C.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6609.
(9) Lolis, E.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6619.
(10) Alber, T.; Banner, D. W.; Bloomer, A. C.; Petsko, G. A.; Phillips, D. C.; Rivers, P. S.; Wilson, I. A. *Philos. Trans. R. Soc. London* **1981**, *B293*, 159.
(11) Pompliano, D. L.; Peyman, A.; Knowles, J. R. *Biochemistry* **1990**, *29*, 3186.

0022-2623/91/1834-2709$02.50/0

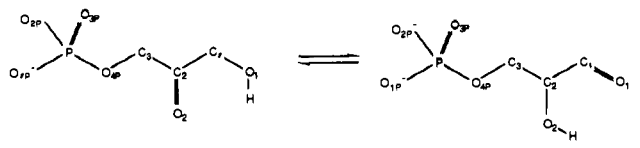

Figure 1. The reaction catalyzed by triosephosphate isomerase. The interconversion of dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate occurs after the aldolase reaction of the glycolytic pathway.

of the enzyme with three inhibitors: 3-phosphonopropionate (3-PP), glycerol 3-phosphate (G3P), or 3-phosphoglycerate (3-PGA) (Figure 2). These inhibitors all bind in the active site of trypanosomal TIM,[12] inducing a fully closed conformation in the flexible loop. As is the case with sulfate in the native structure, ligand binding occurs only in the active site of subunit-2. The equivalent site in subunit-1 is prevented from being occupied by negatively charged ligands, by the proximity of the side chain of Glu323 from a crystallographically related molecule.

In the ligand free, sulfate bound, and substrate analogue bound enzymes, most of the active site residues display a single position and conformation. These include the catalytically important residues Lys13 and His95. By contrast, other residues of the active site occupy extremely different positions dependent upon the presence and nature of an active site bound ligand. The largest difference is the conformation of the flexible loop which can be open (with no ligand bound), "almost closed" (when sulfate is bound), or "fully closed" (in the cases of the phosphate or phosphonate containing compounds).

Also observed in the first three trypanosomal TIM plus inhibitor complexes was the adaptability of certain residues of the active site in response to ligand binding. In particular the catalytic residue Glu467 was seen to occupy one of two discreet conformations dependent upon the nature of the ligand bound at the active site. In complex with G3P, Glu467 occupies a "swung in" conformation, where it hydrogen bonds with the ligand hydroxyl groups. In complex with the carboxylic acid 3-PP, Glu467 again occupies a "swung in" conformation, with carboxyl oxygens so close to the carboxyl group of 3-PP that there is probably a hydrogen bond between the two. This suggests that the pK_a of either the ligand or the catalytic base is altered in the environment of the complex, so as to allow the protonation of one or the other. By contrast, the complex between trypanosomal TIM and 3-PGA has the catalytic glutamate occupying a "swung out" conformation, well away from the carboxyl group of the ligand. It is unclear why 3-PGA is unable to form a similar interaction with the catalytic glutamate to that observed in the complex with 3-PP.

It was noted that drugs based upon these active site inhibitors would have a problem in achieving specificity of inhibition, since the active site of TIM is highly conserved in terms of primary and tertiary structure across evolution. A sequence alignment of the closely related human and chicken TIMs (see e.g., ref 9) allows the prediction of the spatial location of residues of human TIM which are different in nature from the occupants of the equivalent sequence loci of trypanosomal TIM. This analysis shows that within 10 Å of the active site, there is 85% sequence identity between trypanosomal and human enzymes. The additional constraint that residues which may be exploited for drug selectivity must be solvent accessible leaves few options as targets for drug development. Figure 3 shows the Cα fold of one monomer of trypanosomal TIM with the ligand G3P bound to indicate the site of inhibitor binding. The preferred target sequence, Ala400-Tyr401-Tyr402, is also indicated. The equivalent sequence locus in human TIM contains a His, a Val, and a Phe residue. The residues of the flexible loop of subunit-2 (467–480) are colored light blue.

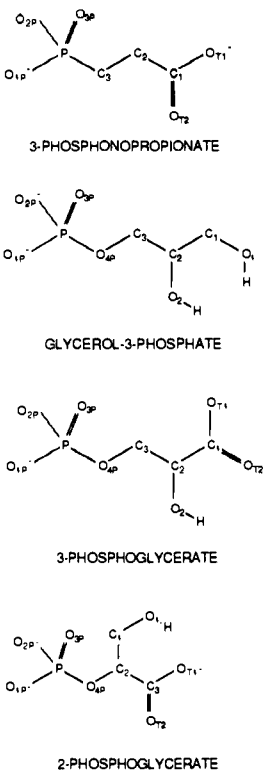

Figure 2. Four active site inhibitors of TIM. The nomenclature for the atoms used in the text derives from this figure. The inhibition constants of the first three compounds for trypanosomal TIM are 27 mM (3-phosphonopropionate) (A. M. Lambeir, unpublished results), 0.6 mM (glycerol 3-phosphate[13]), and 1.3 mM (3-phosphoglycerate[13]). The complexes between trypanosomal TIM and these ligands have been analyzed previously.[12] The ligand 2-phosphoglycerate has an inhibition constant of 6.9 mM[13] and is the inhibitor described in this paper.

The structure of trypanosomal TIM complexed with the three inhibitors mentioned above showed that active site inhibitors were bound in a pocket almost completely closed off from solvent by the closure of the flexible loop. It was apparent that no carbon atoms of these ligands were accessible to solvent in such a way as to allow freedom in inhibitor-development toward regions of selectivity. It was concluded that in order to get selective inhibitors, it was necessary to produce compounds either different in their mode of binding, or branched close to the entrance of the active site.

Among the compounds tested for this purpose are the molecules [2-(*N*-formyl-*N*-hydroxyamino)ethyl]phosphonic acid (Witmans et al., paper in preparation) and 2-phosphoglycerate (2-PGA) (Figure 2). 2-Phosphoglycerate is an inhibitor of trypanosomal TIM with an inhibition constant of 6.9 mM,[13] which resembles a previously analyzed inhibitor, 3-PP, since both have a carboxyl group separated by three covalent bonds from the phosphorus

(12) Noble, M. E. M.; Wierenga, R. K.; Lambeir, A. M.; Opperdoes, F. R.; Thunnissen, A. M. W. H.; Groendijk, H.; Hol, W. G. J. *Proteins* **1991**, *10*, 50.

(13) Lambeir, A. M.; Opperdoes, F. R.; Wierenga, R. K. *Eur. J. Biochem.* **1987**, *168*, 69.

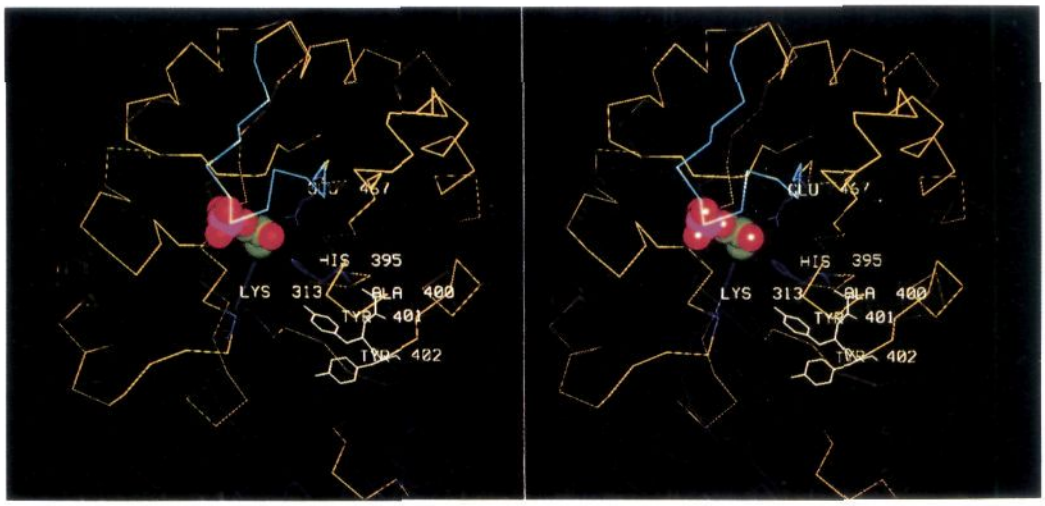

Figure 3. One subunit of triosephosphate isomerase. The Cα tracing of subunit-2 of the complex between triosephosphate isomerase and G3P is shown in yellow, with the flexible loop of subunit 2, residues 467–480, highlighted in pale blue. A CPK model of glycerol 3-phosphate shows the position of the active site, at the C terminal end of the parallel β barrel. The catalytic residues Lys313, His395, and Glu467 are shown in blue. Also shown in white are residues Ala400, Tyr401, and Tyr402. This tripeptide is the nearest sizeable stretch of solvent accessible amino acids to the active site which are specific to trypanosomal TIM and so is the preferred target for inhibitor development. The faint blue tracing indicates Cα positions of the other subunit.

atom of either a phosphate group (2-PGA) or a phosphonate group (3-PP) (see Figure 2). 2-PGA is a larger molecule than 3-PP, due to an extra CH_2OH group. The steric bulk of this extra group was hoped to cause a change in binding mode relative to the binding mode of previously studied inhibitors in such a way that the CH_2OH group would point out of the active site.

Described here is the structure of trypanosomal TIM with bound 2-PGA. In addition, the results of two alternative modeling protocols performed with the BIOGRAF software package[14] are presented. In the first, termed a priori, 2-PGA was forced to adopt a conformation similar to the previously studied inhibitors and then allowed to relax in a static protein environment. The energy expression governing the relaxation of the molecule in this first protocol was simplified to discount electrostatic interactions, but altered to include constraints on the phosphate position and on certain protein–ligand interactions based on previously studied complexes. In the second, termed a posteriori, a full ligand conformational space search was conduced with only the initial position and orientation of the ligand phosphate group presumed from previously determined complexes. Each of the possible conformations available to the ligand was allowed to relax in a static protein environment under the influence of BIOGRAF's Dreiding force field, with a distance-dependent dilectric treatment of the electrostatic protein–ligand interactions. The results of modeling and crystallographic studies have been compared, and the success of the computer-aided modeling has been assessed.

Results

(1) The Structure of the Complex between Trypanosomal TIM and 2-PGA. The structure obtained by crystallography for the complex between trypanosomal TIM and 2-PGA has refined to an *R* factor of 14.9% for all data between 15.0 and 2.4 Å. A Ramachandran plot[15] of the structure shows that only Lys13, Lys313, and Ser513 lie in unfavorable regions of ϕ–ψ space (Figure 4). The aberrant torsion angles of these residues are clearly defined in the electron density maps and may be necessary determinants of an optimal active-site architecture. In both cases, the affected residues form contacts with the ligand: Lys 13/313 is crucial to catalysis, and the amide nitrogen of Ser513 forms a direct hydrogen bond to the phosphate

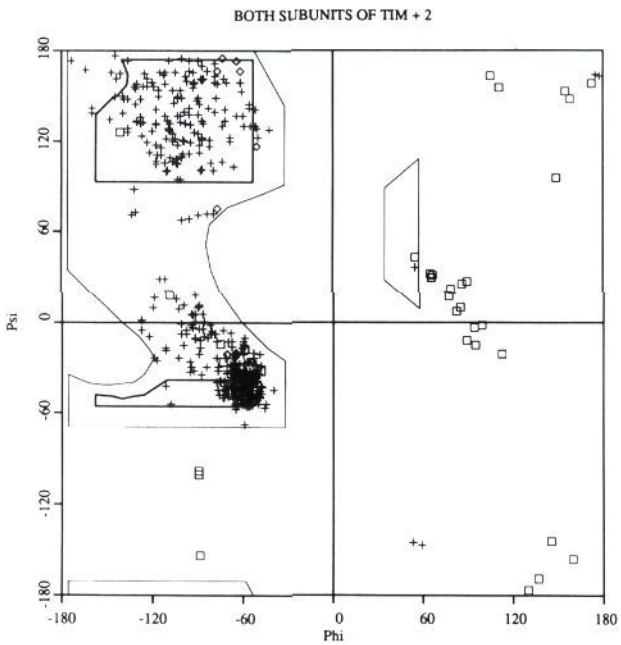

Figure 4. The Ramachandran plot of TIM in complex with 2-PGA. Main-chain torsion angles for all residue classes are shown with plusses, with the exception of prolines (diamonds), and glycines (squares). The thin lines surround allowed regions of ϕ–ψ space for a relaxed van der Waals constraint, and the thicker lines surround fully allowed regions according to Ramachandran and Sasisekharan.[15] The good clustering of main-chain torsion angles within "allowed" regions indicates a well-refined structure. The two plusses clearly outside an allowed region, ($\phi \approx 60$, $\psi \approx -150$) belong to the well-defined active site lysine residues. The other plus near to the right-handed α helix region ($\phi \approx 56$, $\psi \approx 37$) belongs to residue Ser513.

(14) BIOGRAF 2.10; Biodesign Inc, 199 S. Los Robles Ave., Passadena, CA 91101; 1989.

(15) Ramachandran, G. N.; and Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 283.

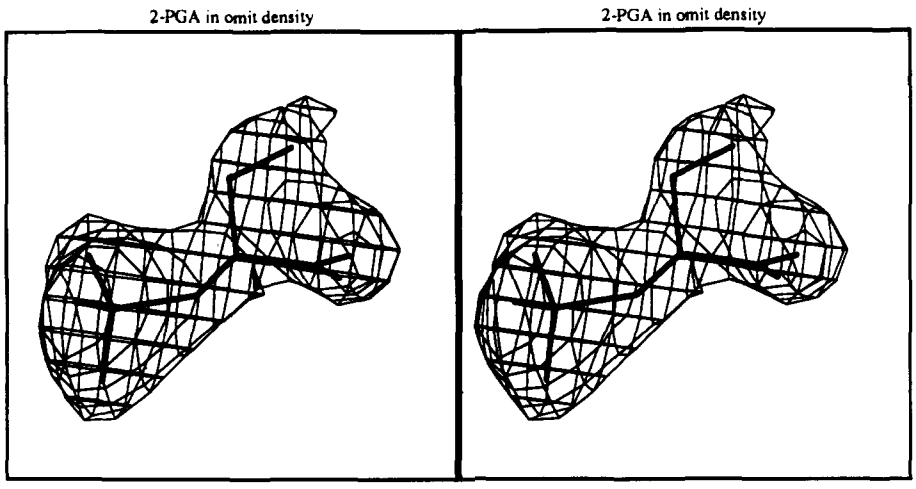

Figure 5. The fit of 2-PGA into omit density. After completion of the refinement of the comdplex, 2-PGA was deleted from the model, which was then subjected to 10 cycles of positional and five cycles of B factor refinement. The resulting model was used to calculate F_c and phases for the calculation of an $mF_o - DF_c$ SIGMAA-weighted map. Shown here is the fit of the final 2-PGA model into the resulting electron density. The map is contoured at 4 times its standard deviation. The atoms at the edge of the electron density (C2 and OT2) have the highest B factors (Table I).

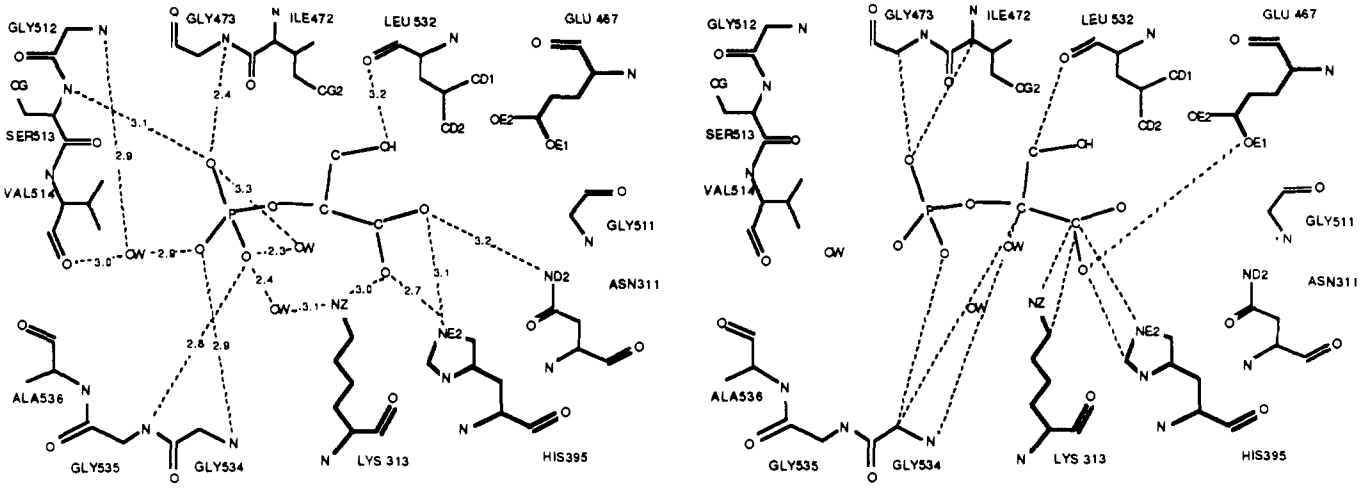

Figure 6. Interactions between 2-PGA and trypanosomal TIM. Hydrogen bonds in part A were chosen as contacts between the protein and ligand shorter than 3.2 Å involving a suitable donor–acceptor pair, regardless of bond angle. The van der Waal's contacts in part B are the other direct protein ligand contacts of less than 3.5 Å. The shortest of the contacts marked as a van der Waal's interaction is 3.0 Å between O3P and the Cα of Gly473.

group of 2-PGA, as well as the other substrate analogues so far studied. A Luzzati plot[16] of the agreement with the crystallographic data suggests that the rms coordinate error is less than 0.2 Å (data not shown). Pairwise comparison of the new set of protein coordinates with all of the previously determined structures of trypanosomal TIM complexes yields rms positional differences of well-defined atoms ($B < 50$ Å^2) ranging from 0.27 Å (compared to the complex with G3P) to 0.33 Å (compared to the complexes with sulfate, 3-PGA, and 3-PP). This figure is made up of random error in each of the pairs, as well as concerted structural differences in atomic position caused by the different ligands bound to TIM. A further assessment of the quality of the final model is the quality of omit density calculated for the ligand after omit refinement. The fit of (*S*)-2-PGA into such an omit map is shown in Figure 5. This figure shows that generally the density around the ligand atoms is good, but that it is weakest for atoms C2 and OT2. This observation correlates with the high individual atomic B factors of these atoms given in Table I. These figures can be compared with an overall mean B for all atoms of the protein molecule of 25 Å^2.

Table I. Atomic B Factors (Å^2) of 2-PGA Bound to *T. brucei* TIM

atom	B	atom	B	atom	B
P	39	O4P	35	C3	65
O1P	42	C1	38	OT1	26
O2P	44	O1	52	OT2	66
O3P	34	C2	66		

The active site interactions made by 2-PGA are summarized in Figure 6. The phosphate group is directly coordinated by amide nitrogens from three stretches of the polypeptide chain and is not within hydrogen-bonding distance of any charged protein group. This situation is identical with that observed in the other substrate analogue complexes. As with these other complexes, there are also a number of solvent-mediated hydrogen bonds between the ligand and the protein, one of which connects the phosphate group with NZ of the catalytic lysine residue

(16) Luzzati, P. V. *Acta Crystallogr.* **1952**, *25*, 802.

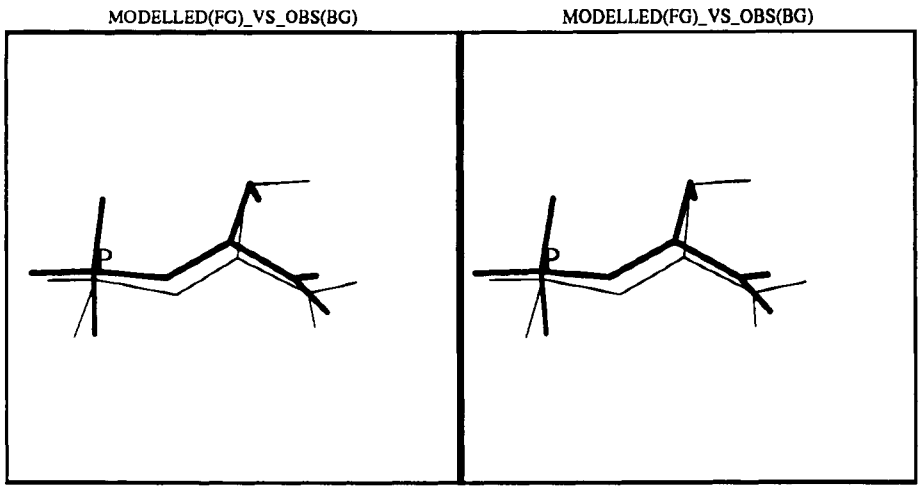

Figure 7. Stereoscopic comparison of a priori modeled 2-PGA with the observed structure. The a priori modeled conformation of 2-PGA is shown in bold lines, and the observed conformation in thin lines. The smallest difference occurs for the P atoms (0.1 Å), and the largest for O1 (1.73 Å); see also Table II.

Table II. Errors (Å) in the a Priori Modeling of 2-PGA

atom	errors	atom	errors	atom	errors
P	0.08	O4P	0.49	C3	0.46
O1P	0.67	C1	0.58	OT1	0.89
O2P	0.17	O1	1.73	OT2	0.54
O3P	0.58	C2	0.47	rms	0.73

313. Of the two carboxyl oxygens of 2-PGA, one forms a hydrogen bond with NZ of Lys313, and the other forms a hydrogen bond with ND2 of Asn311. In addition, both form hydrogen bonds with NE2 of His395. This pattern of hydrogen bonds resembles the coordination of the carboxyl group of 3-PP in the complex between TIM and that ligand.

All potential hydrogen-bonding groups of the ligand have hydrogen-bonding partners within the protein. The longest, and therefore weakest, hydrogen bond involves the O1 atom of 2-PGA, which sits 3.2 Å from the main chain oxygen of Leu532. The strength of this hydrogen bond is also decreased by an unfavorable carbon–donor–acceptor angle of 68°. The environment of this oxygen could, however, also be stabilized by electrostatic interaction with the catalytic glutamate, the OE1 atom of which sits some 3.7 Å away. This distance falls outside the arbitrarily chosen hydrogen-bonding cutoff range, but within the likely range of a significant charge–dipole interaction. The observed conformation of the catalytic glutamate corresponds to the "swung out" conformation.

Figure 6 shows that despite the additional bulk of 2-PGA relative to, for example, the previously studied ligand 3-PP, it is able to sit within the active site without too many close contacts. One reason for this is the additional space in the active site made by the catalytic glutamate occupying a "swung out" position. There may, however, also be other small adaptations in the protein structure which do not appear significant at the level of accuracy of these structures (see also Figure 10).

(2) Comparison of Modeled and Observed Structures. (a) A Priori Modeling. The observed conformation of (*S*)-2-PGA and the structure of 2-PGA derived from the a priori modeling protocol are shown in Figure 7. The differences in atomic position represented in this picture are quantified in Table II. Figure 7 shows that the observed structure agrees generally well with the prediction, except for the torsion angle O4P–C2–C1–O1, which can be seen to have been misassigned. The rms error of 0.73 Å calculated for all 11 non-hydrogen ligand atoms is decreased to 0.54 Å when the hydroxyl oxygen is left out of the calculation. The reason for the misassignment of the O4P–C2–C1–O1 torsion angle was clearly the assumption that the glutamate would occupy a "swung in" conformation in the complex. This location of the glutamate effectively blocked the observed position of the 2-PGA hydroxyl group.

Table III. Results of the a Posteriori Conformational Search[a]

	R,in	*R*,out	*S*,in	*S*,out	relaxed[b]
P	2.06	1.42	2.15	0.71	0.65
O1P	4.59	2.27	3.74	0.60	0.85
O2P	1.57	3.46	3.65	0.74	0.71
O3P	4.23	2.39	3.49	0.85	0.94
O4P	0.84	1.01	0.92	0.43	0.41
C1	4.68	0.62	5.06	2.06	1.56
O1	6.89	1.51	7.12	1.27	1.18
C2	3.01	1.56	2.92	0.87	0.77
C3	4.87	4.39	5.41	0.92	0.35
OT1	6.80	5.76	6.72	1.66	0.27
OT2	4.78	6.16	6.60	1.71	0.34
rms	4.45	3.32	4.75	1.19	0.82
energy (kJ mol^{-1})	−30.3	−36.6	−33.5	−51.2	−62.2

[a] The distances (in Å) quoted are achieved by superimposing the protein scaffold used in the modeling protocol upon the observed protein structure, and then comparing the positions of ligand atoms. The value quoted for the energy is the sum of the covalent energy terms of the inhibitor, together with the nonbonded energy terms of the interactions of ligand atoms with each other, and the nonbonded energy terms of ligand atoms with atoms of the protein scaffold. [b] The figures for the relaxed column represent the movement of the crystallographically determined ligand atoms when allowed to relax in a fixed protein scaffold under the control of a similar energy expression to that used during the a posteriori modeling. The final energy value of −62.2 kJ mol^{-1} can be compared with a calculated energy value for the observed protein ligand complex of +71.1 kJ mol^{-1}.

The interactions enforced during the modeling protocol, i.e between OT1 of 2-PGA and NZ of Lys313, and between OT2 of 2-PGA and ND2 of Asn311, are seen to be present also in the observed structure of the 2-PGA TIM complex (Figure 6). The observed lengths of these hydrogen bonds, 3.0 and 3.2 Å, respectively, are very close to the value of 3.0 Å used in restraining the molecule during modeling. The position in which the phosphate was fixed during modeling was also fairly accurate, with each of the atoms of the phosphate moiety of the a priori modeled ligand lying within 0.7 Å of their observed positions (Table II).

(b) A Posteriori Modeling. The a posteriori modeling of the 2-PGA ligand yielded similarly encouraging results. Table III lists the errors in atomic positions of the lowest energy conformation generated in each of the four con-

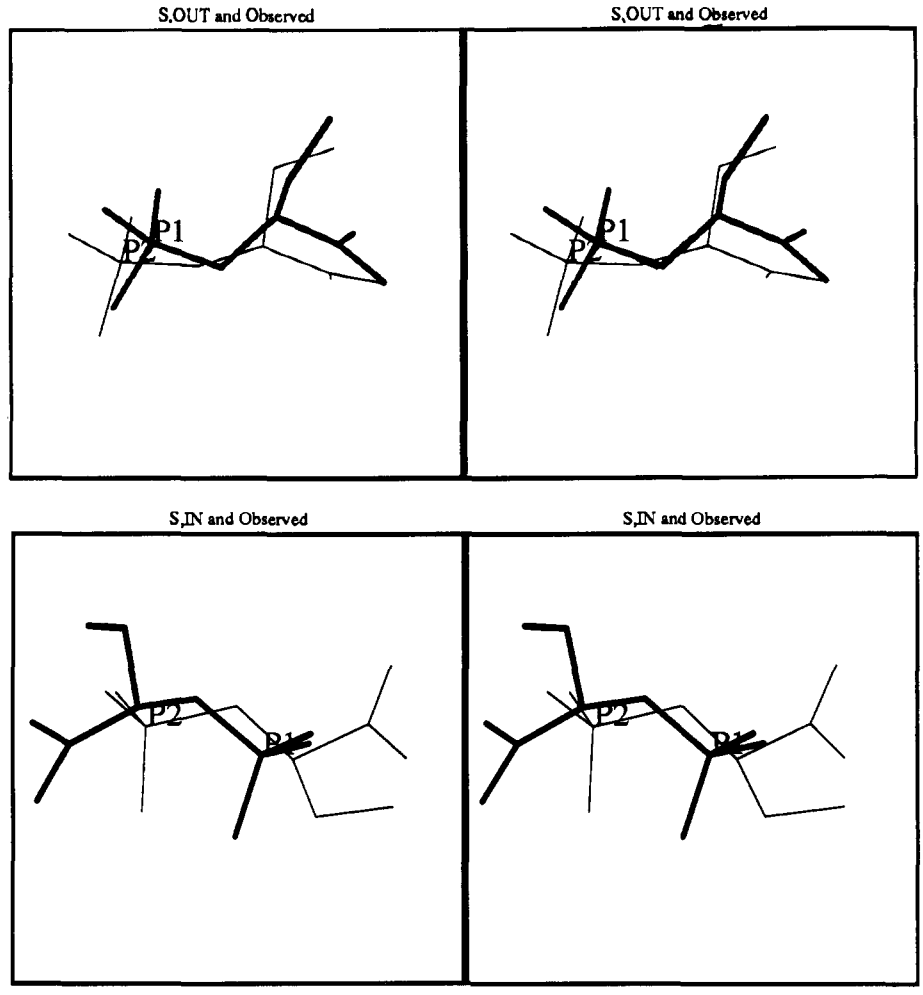

Figure 8. Stereoscopic comparison of a posteriori modeled 2-PGA with the observed structure. Part A features the lowest energy conformation of the (*S*,out) search in bold lines compared with the observed 2-PGA conformation shown in thin lines. Part B has the equivalent comparison for the (*S*,in) conformation. The distances between the individual atoms of the observed and modeled ligand are given in Table III.

formational searches which were conducted. The last row of Table III gives the energy associated with this conformation according to the Dreiding force field used in modeling. It can immediately be seen that the overall lowest energy complex occurred for the combination of ligand chirality and glutamate conformation which was actually observed in the crystal structure, i.e (*S*)-2-PGA in an active site with a "swung out" glutamate. From this table, it can also be seen that this lowest energy conformation agrees reasonably with the observed structure: the rms error in atomic positions is 1.19 Å. The conformations predicted by the (*S*,out) search is compared with the observed structure in Figure 8A. The carboxyl group can be seen to be misplaced due to an accumulation of errors of torsion angles starting from the essentially correct phosphate moiety. For reference, Figure 8B shows the equivalent comparison for the worst search result in terms of rms error (i.e *S*,in). In this latter case, the conformation with the lowest energy has turned around completely within the active site. This turning around is made possible where the start conformation for energy minimization, as generated by setting the ligand torsion angles, has a severe steric clash with the protein. In such a case, considerable gradients exist at the start of energy minimization which can result in large movements of the ligand atoms.

To assess the meaning of these results, the observed conformation of 2-PGA in the environment of the observed protein structure was subjected to 50 steps of energy minimization. The force field and conditions used were the same as those used in the modeling process, and again, only the ligand was allowed to adjust. The results in terms of atomic shifts and change in "energy" are also shown in the last column of Table III. The rms error of 0.82 Å produced by this relaxation is a reasonable indication of the minimum error which could have been produced by the search protocol used. The 1.19 Å actually achieved is therefore quite reasonable. Another test of the quality of the predicted structure is a comparison of the predicted and observed protein–ligand interactions. Figure 9 summarizes the contacts of 2-PGA in the predicted (*S*,out) conformation. These contacts can be compared with the observed contacts shown in Figure 6. A reasonable agreement of interactions is more relevant to a drug-design problem than the actual figure achieved for rms error. Although several of the major interactions have been reproduced by the modeled compound, such as the interactions of the phosphate group and the hydrogen bonds from His395 and Asn311 to one of the carboxyl oxygens, several interactions are also missing. Notably, the hydrogen bond observed between a carboxyl oxygen of 2-PGA and the NZ of Lys313 is missing in the computer-predicted complex. Also, some interactions which are not observed in reality are found in the computer modeled structure. In particular, the complex predicted by the conformational search includes a hydrogen bond between the bridging phosphate oxygen of 2-PGA and the NZ of Lys313. Such an interaction is not observed in any of the crystallographically determined complexes of TIM with inhibitors.

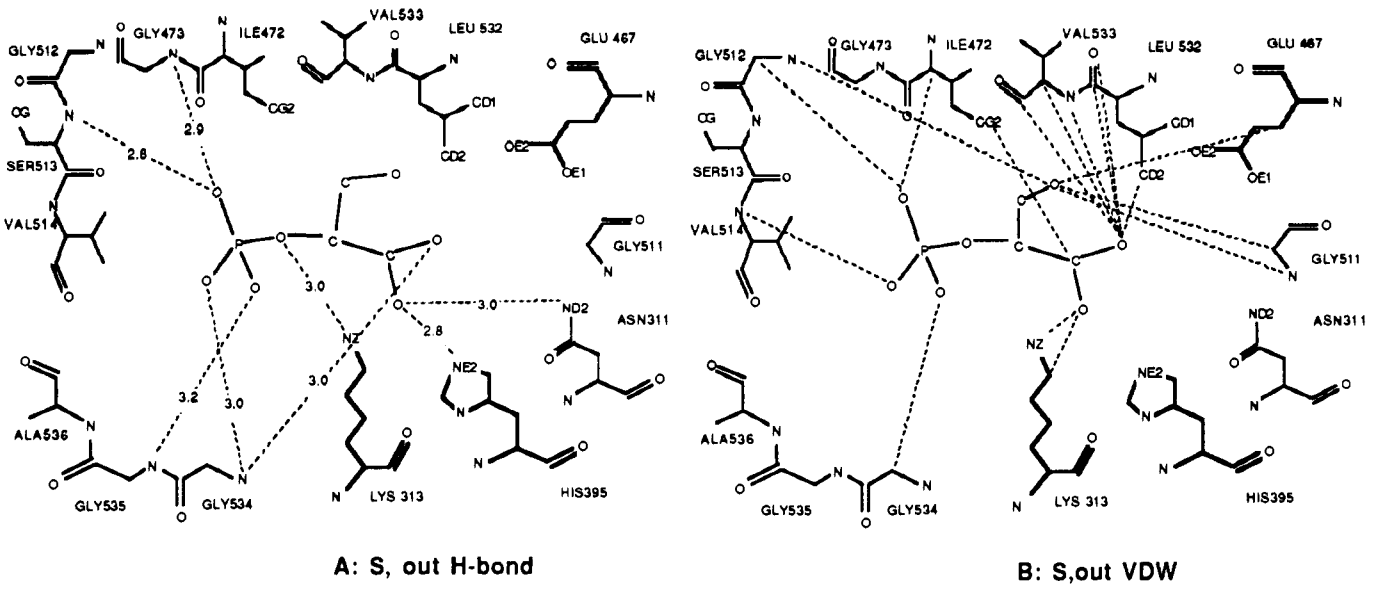

Figure 9. Protein–ligand contacts of a posteriori modeled 2-PGA (*S*,out) search result. Hydrogen bonds in part A were chosen as contacts between the protein and ligand shorter than 3.2 Å involving a suitable donor–acceptor pair, regardless of bond angle. The van der Waal's contacts in part B are the other direct protein ligand contacts of less than 3.5 Å. The shortest of these is 3.0 Å between a ligand carboxyl oxygen and the carbon of Val533.

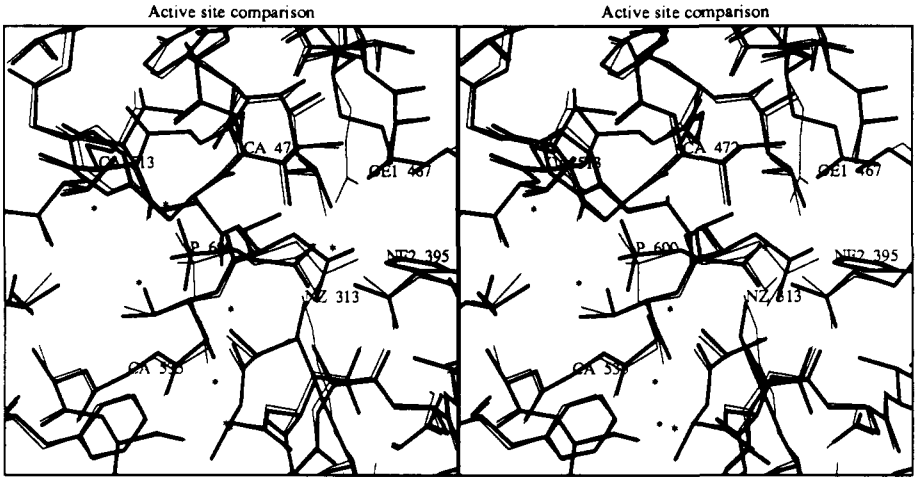

Figure 10. Protein adaptation to ligand binding. The structure of active site residues around the ligand 2-PGA are shown in bold lines, as is the atomic model of the ligand itself. The background model, in thin lines, represents the superimposed atomic model of the protein used during the modeling procedures, i.e. the structure of trypanosomal TIM in complex with G3P. The catalytically active residues Glu467, His395, and Lys313 are labeled, as are residues Ser513, Gly535, and Ile472 and the inhibitors (labeled P600). The asterisks represent water positions determined in the 2-PGA complex.

(3) Protein Adaptation to Ligand Binding. One limitation of the modeling protocol used was the assumption of a fixed protein scaffold against which the idealized ligand molecule was fitted. To assess the validity of this assumption in this case, the protein structure of the observed enzyme–ligand complexes can be compared with the protein structure used in the modeling experiments, i.e. with the protein structure observed in the TIM-G3P complex. This comparison for the 2-PGA protein structure shows only one large difference (Figure 10). This difference involves the catalytic glutamate, which sits in the "swung out" position in the 2-PGA complex, compared with the "swung in" position in the protein model used for the first phase of modeling. There also appears to be a slight shift of Leu532, although this displacement is near the noise level.

Discussion

The crystallographically determined mode of binding of 2-PGA to trypanosomal TIM is reasonably well-defined, as assessed from the quality of omit electron density, contoured at 4σ, surrounding the ligand (Figure 5). However, this is not so true at higher contouring levels, and the relatively high thermal factors of ligand atoms suggest that crystals of the complex contain some static disorder or flexibility in the bound ligand. This may relate to the fact that 2-PGA, which is not a very strong inhibitor of trypanosomal TIM, is quite dissimilar in shape and chemical nature to the substrate, product, or intermediates of the TIM reaction.

2-PGA binds totally within the active site of trypanosomal TIM, with a closed flexible loop shielding all but phosphate atoms from the aqueous environment, as calculated by the program WHATIF.[17] The phosphate moiety occupies the previously determined phosphate binding location, and one oxygen atom of the carboxyl group interacts with the active site lysine and histidine residues. This binding mode resembles closely the binding modes of previously determined trypanosomal TIM inhibitors,

(17) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52.

Table IV. Statistics of the Data Collection and Refinement

data statistics		refinement statistics		model content	
unique reflections (15.0–2.4 Å)	17456	R factor[b] (15.0–2.4 Å)	14.9%	protein atoms	3778
completeness (%)	84.1%	rms deviation for bonds	0.013 Å	water atoms	130
R merge[a]	5.8%	rms deviation for angles	2.3°	ligand atoms	11

[a] R merge = $[\sum_h\sum_i|(\bar{I}_h - I_{h,i})|/\sum_h\sum_i|\bar{I}_h|]100\%$. [b] R factor = $[\sum_h|(F_{o,h} - F_{c,h})|/\sum_h|F_{o,h}|]100\%$.

despite marked differences in ligand shape, emphasizing the anchoring nature of these interactions.

The qualitative nature of these protein–ligand interactions agrees well with the assumptions used in modeling of the compound prior to structure determination. The use of these assumptions, based upon the complexes of TIM with G3P, 3-PP, and 3-PGA, allowed an essentially correct prediction of the complex with 2-PGA.

A much wider conformational search was carried out in the a posteriori modeling protocol. This protocol differed in two further ways from the apriori modeling: firstly, no anchoring interactions were restrained, and secondly, electrostatic interactions were included in the calculations. Also in this protocol a model in reasonable agreement with the observed structure was obtained, although certain key interactions were missing. This shows that the modeling method has to be further optimized. Such optimization would have to include improvement of the modeling force field, a consideration of protein–solvent and ligand–solvent interactions, and an allowance for protein flexibility. The potential for improvement of the force field is evidenced by the movement of atoms of the observed 2-PGA structure by up to 1.6 Å upon relaxation, a distance much further than the errors in the crystallographically determined coordinates. A treatment of protein flexibility would increase the complexity greatly, since changes in protein structure may also give rise to a reordering of tightly bound water molecules. Given these necessary improvements, the success of the automated modeling procedure may be surprising. The accuracy achieved in this case is probably possible because (1) the differences between the ligands of unknown and known binding mode were comparatively small, and (2) the principle determinant of ligand binding to TIM are the well-defined interactions of many main-chain nitrogens with the phosphate group. Such an anchoring interaction is unlikely to be present in all drug-design problems.

It is also interesting to note that a reasonable prediction of the position and orientation of the ligand phosphate group was possible despite the neglect in the modeling of water-mediated hydrogen bonds which are seen in the crystallographically determined complex to sit between the ligand phosphate moiety and residues of the active site. This suggests that the effect of the three water-mediated hydrogen bonds is secondary to the effect of the four direct hydrogen bonds between protein and phosphate atoms in this case.

For the further development of inhibitors of trypanosomal TIM in particular, this study has not provided very hopeful results. The squeezing of 2-PGA into the active site of TIM suggests a strong preference for this orientation of phosphate containing compounds. This preference is undesirable for drug design, since the active site is highly conserved across evolution, and so a poor target for a selective inhibitor. In order to extend a phosphate containing inhibitor away from this region, it therefore seems necessary to use an even larger phosphate compound, or a phosphate compound unable to form interactions with the residues lining the active site, such as the active site lysine and histidine residues. It is to be hoped that one of these approaches will succeed in starting the development of trypanosomal TIM inhibitors away from the active site, and toward regions of sequence specificity.

Experimental Section

(1) Compounds Used and Soaking Procedure. The enzyme triosephosphate isomerase from the organism *T. brucei* brucei was obtained in a copurification procedure with the other glycolytic enzymes of *T. brucei* as previously described.[18] Crystals in spacegroup $P2_12_12_1$ were grown as before[19] from 2.4 M ammonium sulfate in 3-(*N*-morpholino)propanesulfonic acid (MOPS), pH 7.0, from a solution also containing 1 mM dithiothreitol, 1 mM sodium azide, and 1 mM ethylenediaminetetraacetic acid (EDTA). Binding of 2-PGA to the active site of the protein in one such crystal was achieved by transferring it to a solution of the alternative precipitant polyethylene glycol 6000 (PEG 6000) in MOPS, pH 7.0, containing 100 mM ammonium sulfate, 1 mM EDTA, 1 mM dithiothreitol, and 1 mM sodium azide. From this solution, the crystal was transferred to a similar solution of PEG6000 containing 2-PGA in place of the ammonium sulfate, at a concentration of 30 mM. The protocol used in the transfer has been described previously.[20] Racemic 2-PGA was obtained from Boehringer.

(2) Data Collection and Processing. The crystal was soaked for 7 days in the solution described above, during which time the soaking solution was frequently refreshed. After the soaking period, the crystal was mounted in a glass capillary and a three-dimensional data set to 2.4-Å resolution was collected. The cell dimensions were a = 112.6 Å, b = 97.3 Å, c = 46.7 Å, $\alpha = \beta = \gamma = 90°$. For data collection, the Groningen FAST system was used in conjunction with an Elliot GX21 rotating-anode generator. The protocol used during data collection was chosen with the help of the STRATEGY program, implemented by Dr. I. Viekovic, University of Groningen, as a sublevel of the MADNES package.[21] This program uses prediction of data set completeness for a given crystal orientation and set of detector parameters in order to help choose the optimal data collection strategy. Initial data reduction was done with the program MADNES, with profile fitting being carried out by the program PROCOR.[22] Scaling and merging of reflections was completed by use of programs from the Groningen BIOMOL suite. For standardization, the reflection data derived from the 2-PGA complex were linearly scaled to the data of the native crystals which had themselves been placed on an approximately absolute scale. Further statistics of the data are given in Table IV.

(3) Starting Model and Structure Refinement. The starting model used in refinement was the trypanosomal TIM structure observed in the G3P-TIM complex, which has an R factor of 14.7% for data between 15.0 and 2.2 Å and an rms deviation from expected covalent bond lengths of only 0.015 Å. This structure was chosen in preference to the higher resolution (1.83 Å) native structure since the protein structure of TIM in substrate analogue complexes has been observed to be different from the protein structure in complex with sulfate.[12] The coordinates of the two

(18) Misset, O.; Bos, O. J. M.; Opperdoes, F. R. *Eur. J. Biochem.* **1986**, *157*, 441.
(19) Wierenga, R. K.; Hol, W. G. J.; Misset, O.; Opperdoes, F. R. *J. Mol. Biol.* **1984**, *178*, 487.
(20) Schreuder, H. A.; Van der Laan, J. M.; Groendijk, H.; Wierenga, R. K. *J. Appl. Crystallogr.* **1988**, *21*, 426.
(21) Pflugrath, J. W.; Messerschmidt, A. *MADNES: Users Guide*; Max-Planck-Institut für Biochemie: Martinsried, Germany, 1986.
(22) Kabsch, W. *J. Appl. Crystallogr.* **1988**, *21*, 916–924.

protein subunits of this structure were refined as rigid bodies against data between 6.0 and 4.0 Å, by using the TNT package.[23] This refinement lead to a drop in the R factor for this resolution range of 22.5% to 20.2%. The resulting model was used to phase SIGMAA-weighted $2mF_o - DF_c$ and $mF_o - F_c$ electron density maps.[24] These were examined on an Evans and Sutherland picture system by using the interactive graphics package FRODO. The difference electron density maps suggested that 2-PGA bound only at the active site of subunit-2 of the enzyme. Positive and negative features of the difference electron density in this active site suggested that the conformation of Glu467 in the 2-PGA complex was different from its conformation in the phasing model.

In order not to bias the active site electron density toward any particular conformation of the ligand, only protein atoms were included in the first part of structure refinement. This refinement began with least-squares refinement using the TNT package at gradually increasing levels of resolution. This was followed by a brief protocol of molecular dynamics crystallographic (MDX) and energy refinement (EMX) using the GROMOS package.[25] GROMOS refinement was implemented by using a similar protocol to that previously described for the refinement of the first three complexes of trypanosomal TIM. After this refinement, the structure was subjected to a further 10 cycles of TNT positional refinement, and then five cycles of TNT B factor refinement before being used for the calculation of further $mF_o - F_c$ and $2mF_o - DF_c$ maps. At this stage in the refinement, the R factor of the model was 17.7% (15.0–2.4 Å). Water molecules were selected from peaks in the difference electron density above 4 times the standard deviation of the map, provided that their environment was chemically sensible. Also, an atomic model of the inhibitor was manually built into the difference electron density by using the program FRODO.[26] The stereochemistry of the ligand had originally been derived from the molecular modeling and energy minimization module of the BIOGRAF[14] program. Only the S enantiomer of 2-PGA was seen to be consistent with the shape of the difference electron density within the active site of subunit-2.

The model was then subjected to further cycles of TNT-restrained least-squares refinement of positional and thermal parameters. The atomic positions of 2-PGA were included in this refinement, with ideal geometries defined in the TNT package from BIOGRAF-deduced values. Electron density surrounding the model for 2-PGA was, however, still unsatisfactory, so that a further round of GROMOS MDX/EMX was initiated. The tightly bound waters which had been determined from the difference electron density maps were included in the refinement, but the ligand itself was omitted. After 1 ps of MDX refinement, 2-PGA was added back to the mode, and 10 cycles of TNT positional refinement and five cycles of TNT B factor refinement were carried out. The final structure was used to calculate electron density maps which showed an essentially complete tracing of the ligand molecule and a protein molecule well-defined by electron density. Final statistics of the refinement are given in Table IV.

(4) A Priori Modeling of 2-PGA. Computer modeling was carried out by using the program BIOGRAF.[14] For the modeling of 2-PGA, the protein structure used was that observed in the complex between trypanosomal TIM and glycerol 3-phosphate (G3P), with all water molecules and the G3P ligand deleted. To produce an environment of manageable size, protein atoms were selected within 9 Å of the ligand G3P. The modeling of 2-PGA into this active site environment was accomplished in two phases. Firstly an atomic model was produced of the molecule 2-phosphoglycolate ($HOOCCH_2OPO_3^{2-}$), which can be considered as a substructure of the ligand 2-PGA, and which is a potent inhibitor of TIM.[27,13] This model was moved manually into the active site of TIM and positioned roughly by using the following three assumptions. (1) The phosphate atoms would superimpose upon the phosphate atoms of G3P as observed in the G3P–TIM complex. (2) One of the two ligand carboxyl atoms would sit between the NZ atom of the positively charged catalytic residue Lys313 and the NE2 atom of the catalytic histidine His395. (3) The other carboxyl oxygen would occupy a position suitable for accepting a hydrogen bond from the ND2 atom of residue Asn311. These assumptions were based on observations from the first series of compounds analyzed complexed with trypanosomal TIM. These compounds bound in the active site of TIM, with their phosphorus atoms superimposing to within 0.1 Å of each other and a hydrogen-bonding pattern consistent with the constraints imposed upon phosphoglycolate.

When the position and torsion angles of the ligand had been manually adjusted to satisfy these constraints, the ligand conformation was allowed to relax by applying energy minimization using the Dreiding forcefield, as incorporated in the BIOGRAF program, without consideration of the electrostatic term of the potential function. Convergence was assessed on the basis of BIOGRAF's graphical output of the energy of the system as a function of step number. During this minimization, the phosphorus atom was held fixed and an harmonic restraint was applied to the distances between the ligand carboxyl atoms and residues Lys313 and Asn311. No change in protein structure was allowed. The minimized structure was used as the basis for the addition of the CH_2OH group which relates 2-phosphoglycolate to 2-phosphoglycerate. Addition of this group introduces chirality into the molecule, so that a decision had to be made as to which enantiomer would bind. It was clear from inspection that only the S enantiomer would fit into the active site of trypanosomal TIM if the above constraints were to be obeyed. The appropriate positioning of the CH_2OH group for this enantiomer was chosen and the ligand conformation was minimized with constraints as before. Although the selected enantiomer of 2-PGA had reasonable interactions with surrounding protein groups, there was still a clash observed between the catalytic residue Glu467 and the ligand hydroxyl group. Although this clash could have been removed by allowing movement of the glutamate side chain in the further relaxation of the model, it was decided to suspend further modeling studies at this point, since the feasibility of a complex with an inward oriented 2-PGA molecule had already been satisfactorily established.

(5) A Posteriori Modeling of 2-PGA. After the solution of the structure, it was clear that in the complex with 2-PGA, the catalytic glutamate residue actually occupied the "swung out" conformation, explaining a poor correspondence between the a priori modeled and observed ligand conformation in the position of the 2-PGA hydroxyl group. In order to see whether the BIOGRAF software was able to suitably determine the mode of binding of 2-PGA with the aid of fewer subjective assumptions, a further series of modeling experiments was initiated. These experiments involved a full search of ligand conformational space, carried out with each of the four possible combinations of ligand chirality and glutamate conformation, i.e. (i) (S)-2-PGA with a "swung out" glutamate (S,out), (ii) (S)-2-PGA with a "swung in" glutamate (S,in), (iii) (R)-2-PGA with a "swung out" glutamate (R,out), and (iv) (R)-2-PGA with a "swung in" glutamate (R,in). In this series of experiments, no positional or interatomic distance restraints were applied, and the electrostatic term in the energy expression was switched on. The movement of the glutamate to the "swung out" conformation was achieved in practice by allowing it to relax in a static protein environment with a strongly applied harmonic potential pulling it toward its hydrogen-bonding partner in previously determined "swung out" complexes, i.e. the main-chain nitrogen of Ser396. Calculations for this phase of the modeling were run on a Silicon Graphics personal Iris machine.

The ligand model used in these experiments was the same as that used for a priori modeling, and again the protein environment used was a 9-Å substructure of TIM from the TIM-G3P complex. This environment was chosen for the a posteriori modeling so as not to unduly bias the conformational search toward the observed result. The only assumption in the initial positioning of the search model was the superposition of phosphate atoms upon the phosphate position observed in the G3P structure. The electrostatic term of the energy expression was used, with a dielectric constant equal to the interatomic separation. Ligand charges were assigned by using the BIOGRAF Gasteiger routine. The charges generated by this routine are shown in Table V, where they are compared with the Mulliken population deduced charges calcu-

(23) Tronrud, D. E.; Ten Eyck, L. F.; Matthews, B. W. *Acta. Crystallogr.* **1987**, *A43*, 489–501.
(24) Read, R. J. *Acta Crystallogr.* **1986**, *A42*, 140.
(25) Gros, P.; Fujinaga, M.; Dijkstra, B. W.; Kalk, K. H.; Hol, W. G. J. *Acta Crystallogr.* **1989**, *B45*, 488.
(26) Jones, T. A. *Methods Enzymol.* **1985**, *115*, 157.
(27) Wolfenden, R. *Nature* **1969**, *233*, 704.

Table V. 2-PGA Charges Used in the Search Protocol[a]

atom	BIOGRAF	MNDO	atom	BIOGRAF	MNDO
P	0.321	1.273	O1	−0.552	−0.289
O1P	−0.545	−0.921	H	0.000	0.098
O2P	−0.558	−0.935	C2	0.391	0.222
O3P	−0.543	−0.969	C3	0.466	0.319
O4P	−0.430	−0.526	OT1	−0.754	−0.642
C1	−0.020	0.05	OT2	−0.776	−0.679

[a] The 2-PGA model used during energy calculations featured all non-hydrogen atoms and all hydrogen atoms attached to non-carbon atoms. The charges given in this table for the MNDO calculation represent the result of collapsing hydrogen atom charges onto the corresponding heavier atom.

Table VI. Scope of the Conformational Search

dihedral angle	initial value	final value	increment
O1P–P–O4P–C2	30.0	360.0	30.0
P–O4P–C2–C1	30.0	360.0	30.0
O4P–C2–C1–O1	30.0	360.0	30.0
O4P–C2–C3–OT1	30.0	180.0	30.0

lated after a single point MNDO calculation. The BIOGRAF-derived charges were chosen over MNDO charges in order that the ligand charges should be as consistent as possible with the protein charges, which were themselves derived from a BIOGRAF charge library. The lack of charges on protons attached to hetero atoms in the Gasteiger charges is consistent with the explicit hydrogen bonding potential used by the Dreiding forcefield. Heteroatom protons of the ligand and protein were generated by the standard BIOGRAF routines. In all of the modeling studies, the catalytic glutamate was treated as protonated in the assignment of atomic partial charges.

A thorough search of ligand conformational space was performed, such that all combinations of the four torsion angles defining the ligand conformation were examined. This procedure was implemented by using the BIOGRAF search option. The four torsion angles involved and the ranges and search grid used are shown in Table VI. For each of the 10368 conformations generated, 10 steps of unconstrained conjugate gradient energy minimization were performed to allow the ligand conformation to find its nearest local energy minimum. The value of the calculated energy for each conformation was the sum of the ligand intramolecular energy and the energies of the protein–ligand nonbonded interactions. As before, the atoms of the protein were kept in fixed positions. This strategy searches ligand conformational space, but is prejudiced by the position and orientation of the phosphate group from which the torsion angles were defined. This prejudice is partly corrected for by the application of energy minimization without any positional restraints which was carried out on each of the search conformations. A full search including translational and orientational parameters of the ligand would be both computationally prohibitive and also more thorough than the accuracy of the in vacuo energy expression could justify. For each combination of ligand chirality and glutamate conformation, the lowest energy conformation was extracted and analyzed.

(6) Structure Comparisons. For the purposes of comparing structures, all protein molecules were superimposed upon the native structure. Superposition was done after the method of Kabsch,[28] using the 105 Cα atoms of the eight β strands and eight α helices of the core of subunit-2. This set of atoms was chosen because they had been observed to be insensitive to the nature of any bound ligand. Analysis of the structures was carried out on an Evans and Sutherland picture system using either the FRODO package or the BIOGRAF package.

Acknowledgment. It is a great pleasure to thank Drs. Fred Opperdoes, Annemie Lambeir, and Klaus Mueller for many stimulating discussions. This research received financial support from the UNDP/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases (through a grant to W.G. J.H.), from Hoffmann-La Roche, Basle (through a fellowship for C.L.M.J.V.), and by the Dutch Foundation for Chemical Research (SON). The coordinates of the TIM-2-PGA complex have been submitted to the Brookhaven Protein Data Bank (4TIM).

(28) Kabsch, W. *Acta Crystallogr.* **1978**, *A34*, 827.

(29) **Abbreviations**: TIM, triosephosphate isomerase (EC 5.3.1.1); DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; G3P, glycerol 3-phosphate; 3-PGA, 3-phosphoglycerate; 3-PP, 3-phosphonopropionate; 2-PGA, 2-phosphoglycerate; rms, root mean square; MOPS, 3-(*N*-morpholino)-propanesulfonic acid; PEG, polyethylene glycol; EDTA, ethylenediaminetetraacetic acid; MNDO, modified neglect of differential overlap.