

# Inclusion of ionization states of ligands in affinity calculations

Serena Donnini,<sup>1</sup> Alessandra Villa,<sup>2</sup> Gerrit Groenhof,<sup>3</sup> Alan E. Mark,<sup>4</sup>  
Rik K. Wierenga,<sup>1</sup> and André H. Juffer<sup>1\*</sup>

<sup>1</sup>The Biocenter Oulu and the Department of Biochemistry, University of Oulu, Oulu, Finland

<sup>2</sup>Max-Planck-Institute for Polymer Research, Mainz, Germany

<sup>3</sup>Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

<sup>4</sup>School of Molecular and Microbial Sciences, University of Queensland, Australia

## ABSTRACT

When estimating binding affinities of a ligand, which can exist in multiple forms, for a target molecule, one must consider all possible competing equilibria. Here, a method is presented that estimates the contribution of the protonation equilibria of a ligand in solution to the measured or calculated binding affinity. The method yields a correction to binding constants that are based on the total concentration of inhibitor (the sum of all ionized forms of the inhibitor in solution) to account for the complexed form of the inhibitor only. The method is applied to the calculation of the difference in binding affinity of two inhibitors, 2-phosphoglycolate (PGA) and its phosphonate analog 3-phosphonopropionate (3PP), for the glycolytic enzyme triosephosphate isomerase. Both inhibitors have three titrating sites and exist in solution as a mixture of different forms. In this case the form that actually binds to the enzyme is present at relative low concentrations. The contributions of the alternative forms to the difference in binding energies is estimated by means of molecular dynamics simulations and corrections. The inhibitors undergo a  $pK_a$  shift upon binding that is estimated by *ab initio* calculations. An interesting finding is that the affinity difference of the two inhibitors is not due to different interactions in the active site of the enzyme, but rather due to the difference in the solvation properties of the inhibitors.

Proteins 2009; 76:138–150.  
© 2008 Wiley-Liss, Inc.

**Key words:** binding free energy; effective affinity; molecular dynamics;  $pK$  calculations; protonation equilibria; thermodynamic integration; triosephosphate isomerase.

## INTRODUCTION

In free energy calculations of a binding reaction, a thermodynamic cycle (e.g., as shown in Fig. 1) is typically used to relate the calculated properties of the free ligand, the free target, and the ligand-target complex in solution to account for the observed binding affinity. Generally, only the binding reaction itself is considered. Other equilibria, such as acid-base equilibria where for example the ligand and/or enzyme may exist in various ionized forms or protonation states at a given pH are ignored. In such cases the influence of acid dissociation constants ( $pK_a$ ) and also ionic strength, which may affect the apparent  $pK_a$ s, must also be considered.<sup>1,2</sup>

In this work we present free energy calculations of ligand-target binding reactions where the ionization equilibria of the ligands in solution are taken into account. These apply to the general case of various species of a ligand in solution related by  $pK_a$  equilibria where only one such species binds to the target molecule. Quantification of the contribution to the free energy of binding from the ionization equilibria of the ligands is particularly important in view of the fact that experimental estimates of binding affinities—based on the total ligand concentration—involve all equilibria of the ligands (and target) in solution. Whereas calculated affinities are generally based on the form of the inhibitor that binds the enzyme and consider only the binding equilibria. As a consequence one cannot directly compare experimental and calculated binding affinities unless one accounts for the contribution from the other equilibria in solution.

As an example we consider the binding of 2-phosphoglycolate (PGA) and 3-phosphonopropionate (3PP) to the glycolytic enzyme triosephosphate isomerase (TIM).<sup>3,4</sup> These two inhibitors bind to TIM in a specific protonation state (species C in Fig. 2) that differs from the main protonation state observed in water solution at pH 7. Moreover, the protonation state of the inhibitors change upon binding: they are fully deprotonated once they are in the active site of the

Serena Donnini's current address is Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany.  
\*Correspondence to: André H. Juffer, The Biocenter Oulu and the Department of Biochemistry, University of Oulu, P.O. Box 3000, FIN-90014 University of Oulu, Finland. E-mail: Andre.juffer@oulu.fi.

Received 3 September 2008; Revised 14 October 2008; Accepted 4 November 2008  
Published online 18 November 2008 in Wiley InterScience (www.interscience.wiley.com).  
DOI: 10.1002/prot.22326



## METHODS

The form of the inhibitors PGA and 3PP that binds to TIM (species C in Fig. 2) is different from the bound form (species F in Fig. 2), and is also not the dominant state of the inhibitors in water solution, as mentioned in the Introduction. Thus, while the relative free energy of binding can be computed for the fully deprotonated form F, the experimentally determined affinity contains contributions from all accessible protonation states of the ligand and the protein. Consequently, to be able to compare an experimental and computed relative affinity, two additional corrections are required to the values obtained by standard MD free energy calculations:

1. The experimentally measured affinity must be corrected to account for the fact that only species C binds. This step results in the effective free energy of binding of species C (see Effective affinity section).
2. The computed affinity for species F must be corrected for the free energy that is required to transfer a proton to species F. This step results in the predicted free energy of binding of species C (see Predicted relative affinity section).

Note that the different protonation states of the enzyme are not included. We expect that these are not contributing to the affinity difference of the two inhibitors for the enzyme TIM. In particular, as we consider the binding of two inhibitors to the same enzyme, any contribution from the free enzyme protonation equilibria to the affinity would be very similar for the binding of PGA and 3PP, and would therefore cancel when considering the difference in affinities. There is no evidence that the protonation state of the enzyme is different in the two inhibitor-enzyme complexes and we assume that differences in these complexes are all accounted for in the free energy calculations. See the Discussion for more details.

### Effective affinities

The experimentally determined equilibrium association constants  $K_a(\text{exp})$  for the binding of the inhibitors PGA and 3PP to TIM (Table I) can be expressed as:

$$K_a(\text{exp}) = \frac{[\text{TIM} \cdot \text{C}]}{[\text{TIM}][\text{I}]} \quad (1)$$

where [I] represents the total concentration of all forms of inhibitor in solution and C denotes the specific form of the inhibitor that binds to the enzyme TIM (Fig. 2). The effective affinity is defined as the affinity of the form C of the inhibitor for TIM. The effective association constant,  $K_a(\text{eff})$ , is related to  $K_a(\text{exp})$  by the following relation:

**Table I**  
Experimental<sup>7,8</sup> (exp) and Theoretical Results for the Affinity of PGA and 3PP to TIM

	exp		exp	eff
$K_{i,\text{PGA}}$	$0.027 \pm 0.005$	$\Delta G_{\text{PGA}}$	$-26.1 \pm 0.4$	$-40.7$
$K_{i,\text{3PP}}$	27	$\Delta G_{\text{3PP}}$	-9.0	-15.5
		$\Delta G_{\text{PGA}} - \Delta G_{\text{3PP}}$	-17.1	-25.2
$K_{i,\text{DHAP}}$	$0.66 \pm 0.06$	$\Delta G_{\text{DHAP}}$	$-18.2 \pm 0.2$	
$K_{i,\text{DGAP}}$	$0.008 \pm 0.002$	$\Delta G_{\text{DGAP}}$	$-29.1 \pm 0.7$	
	$2.303RT (\text{p}K_{1a,2a,(1a-2a)})$			MD + TI
$\Delta G_{1a}$	29.1	$\Delta G_3$		$28.3 \pm 4.3$
$\Delta G_{2a}$	39.4	$\Delta G_4$		$24.1 \pm 7.2$
$\Delta G_{1a} - \Delta G_{2a}$	-10.3	$\Delta G_3 - \Delta G_4 = \Delta \Delta G (\text{calc})$		$4.2 \pm 8.4$
	$\Delta \Delta G (\text{pred}) = \Delta G_{1a} - \Delta G_{2a} + \Delta \Delta G (\text{calc}) = -6.1$			

Binding affinities of dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (DGAP) are also reported.<sup>7</sup>  $\Delta G(\text{exp})$  is obtained as  $\Delta G = RT \ln K_i$ , with  $K_i$  the experimental inhibition constant,  $R$  the gas constant, and  $T$  the temperature (298 K).  $\Delta G(\text{eff})$  (effective) of PGA and 3PP are derived from the experimental ones as described in the Methods. Free energy differences  $\Delta G$  in the lower panel are labeled according to Figure 1.  $\text{p}K_{1a}$  and  $\text{p}K_{2a}$  are listed in Figure 2 and are calculated as described in the Methods.  $K_i$  in mM,  $\Delta G$  in  $\text{kJ mol}^{-1}$ . MD, molecular dynamics; TI, thermodynamic integration.

$$K_a(\text{eff}) = \frac{[\text{TIM} \cdot \text{C}]}{[\text{TIM}][\text{C}]}$$

$$K_a(\text{eff}) = \frac{[\text{TIM} \cdot \text{C}][\text{I}]}{[\text{TIM}][\text{I}][\text{C}]} \quad (2)$$

$$K_a(\text{eff}) = K_a(\text{exp}) \frac{[\text{I}]}{[\text{C}]}$$

where [C] is the equilibrium concentration of C free in solution. Equation (2) can be expressed in terms of effective binding free energy  $\Delta G(\text{eff})$ :

$$\Delta G(\text{eff}) = -RT \ln K_a(\text{eff})$$

$$\Delta G(\text{eff}) = \Delta G(\text{exp}) - RT \ln \frac{[\text{I}]}{[\text{C}]} \quad (3)$$

The difference in the effective binding free energy  $\Delta \Delta G(\text{eff})$  of the two ligands PGA and 3PP can therefore be written as:

$$\Delta \Delta G(\text{eff}) = \Delta G(\text{eff})_{\text{PGA}} - \Delta G(\text{eff})_{\text{3PP}}$$

$$\Delta \Delta G(\text{eff}) = \Delta G(\text{exp})_{\text{PGA}} - \Delta G(\text{exp})_{\text{3PP}}$$

$$- RT \ln \left( \frac{[\text{C}_{3\text{PP}}][\text{I}_{\text{PGA}}]}{[\text{C}_{\text{PGA}}][\text{I}_{3\text{PP}}]} \right) \quad (4)$$

$$\Delta \Delta G(\text{eff}) = \Delta \Delta G(\text{exp}) - RT \ln \left( \frac{[\text{C}_{3\text{PP}}][\text{I}_{\text{PGA}}]}{[\text{C}_{\text{PGA}}][\text{I}_{3\text{PP}}]} \right)$$

where  $\text{C}_{\text{PGA}}$  and  $\text{C}_{3\text{PP}}$  are the C species of the inhibitors PGA and 3PP, respectively, and  $[\text{I}_{\text{PGA}}]$  and  $[\text{I}_{3\text{PP}}]$  are the total equilibrium concentrations of inhibitors PGA and 3PP, respectively, in solution.

$\Delta \Delta G(\text{eff})$  in Eq. (4) is the difference in the free energy of binding of species C of PGA and 3PP for TIM, which

can be compared with the predicted relative binding free energy  $\Delta\Delta G(\text{pred})$  [Eqs. (14) and (17)].  $\Delta\Delta G(\text{eff})$  corresponds to the first correction mentioned at the beginning of the Methods section. It remains now to estimate the equilibrium concentrations of C and I.

### Determination of equilibrium concentrations

The equilibrium concentrations of the different species in solution can be calculated using the following equations:

$$[I]_0 = [I] + [\text{TIM} \cdot \text{C}] \quad (5)$$

$$[\text{TIM}]_0 = [\text{TIM}] + [\text{TIM} \cdot \text{C}] \quad (6)$$

Here,  $[I]_0$  and  $[\text{TIM}]_0$  represent the initial concentrations of inhibitor (I) and enzyme (TIM), respectively, and TIM·C is the complex between the inhibitor and TIM.

Combining Eqs. (1) and (6) gives:

$$[\text{TIM}]_0 = \frac{[\text{TIM} \cdot \text{C}]}{K_a(\text{exp})[I]} + [\text{TIM} \cdot \text{C}] \quad (7)$$

The inhibitor I exists in solution in different ionization species (Fig. 2) and [I] can be written as the sum of the concentrations of all different species in solution:

$$[I] = [A] + [B] + [C] + [D] + [E] + [F] \quad (8)$$

On the basis of the experimentally determined acid dissociation constants ( $pK_a$ s) of PGA and 3PP it is estimated that at pH 7.6 species A and D are present in negligible concentration (as is explained in the Results), and for simplicity in the following we consider  $[A] = [D] = 0$ . If Eq. (8) is now expressed as a function of [C], the species which binds to the enzyme, we obtain:

$$[I] = \frac{[C][\text{H}^+]}{K_3} + [C] + \frac{K_2[C]}{K_4} + \frac{K_2[C]}{[\text{H}^+]} \quad (9)$$

where  $K_2$ ,  $K_3$ , and  $K_4$  are associated with the proton dissociation reactions  $C \rightarrow F$ ,  $B \rightarrow C$ , and  $E \rightarrow F$  (Fig. 2), respectively, and are obtained from the corresponding calculated  $pK$  or experimental  $pK_a$  values. The proton concentration  $[\text{H}^+]$  is calculated from the pH of the solution at which the experiment was carried out (7.6).<sup>7</sup>

When [I] in Eqs. (5) and (7) is expressed as in Eq. (9), a system of two equations with [C] and [TIM·C] as unknowns is obtained:

$$[I]_0 = \frac{[C][\text{H}^+]}{K_3} + [C] + \frac{K_2[C]}{K_4} + \frac{K_2[C]}{[\text{H}^+]} + [\text{TIM} \cdot \text{C}] \quad (10)$$

$$[\text{TIM}]_0 = \frac{K_i(\text{exp})[\text{TIM} \cdot \text{C}]}{\frac{[C][\text{H}^+]}{K_3} + [C] + \frac{K_2[C]}{K_4} + \frac{K_2[C]}{[\text{H}^+]}} + [\text{TIM} \cdot \text{C}] \quad (11)$$

The program Mathematica<sup>9</sup> was used to solve such a system, with  $[I]_0 = [\text{TIM}]_0 = 1$  M and  $K_i(\text{exp})$  the experimental inhibition constant ( $1/K_a(\text{exp})$ ) in Table I. The equilibrium concentrations of the other species can then also be calculated with Eq. (9). [C] and [I] ( $[I]_0 - [\text{TIM} \cdot \text{C}]$ ) are now determined and are used to estimate the difference in the effective binding free energy  $\Delta\Delta G(\text{eff})$  in Eq. (4).

### Predicted relative affinity

The form of the inhibitors that bind to TIM (species C in Fig. 2) carries a negative charge of  $-2e$  on the phosphate and phosphonmethyl moiety of PGA and 3PP, respectively, just as the natural substrates of the enzyme: dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (DGAP).<sup>10,11</sup> In the first step of the isomerization catalyzed by TIM, a proton is abstracted by the catalytic glutamic acid (Glu) in the active site of the enzyme from the substrate's hydroxylic carbon.<sup>3,4,12</sup> Accordingly, upon binding, the carboxylate proton of PGA and 3PP is donated to the catalytic Glu of TIM (this form of TIM is denoted as TIM-H). Consequently, the inhibitors are fully deprotonated in the active site of the enzyme (carrying an overall charge of  $-3e$ , species F in Fig. 2).<sup>5,6</sup>

Figure 1 details the thermodynamic cycle that is employed to estimate the difference in the free energy of binding  $\Delta\Delta G(\text{pred})$  (predicted relative affinity) between species C of PGA and 3PP for TIM. As shown in Figure 1, the binding reaction is divided in two steps: (a) species C of the inhibitor donates a proton to the solution, resulting in species F, and the catalytic Glu of TIM takes up a proton from the solution, resulting in TIM-H; (b) species F binds TIM-H. The deprotonation of the inhibitor/protonation of TIM and the binding reaction are therefore described by two different reactions.  $\Delta G_1$  and  $\Delta G_2$ , the affinities of species C of PGA and 3PP for TIM, respectively, are consequently obtained as:

$$\Delta G_1 = \Delta G_{1a} + \Delta G_{1b} \quad (12)$$

$$\Delta G_2 = \Delta G_{2a} + \Delta G_{2b} \quad (13)$$

$\Delta\Delta G(\text{pred})$  can now be calculated as:

$$\begin{aligned} \Delta\Delta G(\text{pred}) &= \Delta G_1 - \Delta G_2 \\ \Delta\Delta G(\text{pred}) &= \Delta G_{1a} - \Delta G_{2a} + \Delta G_{1b} - \Delta G_{2b} \end{aligned} \quad (14)$$

The difference in (de)protonation free energies  $\Delta G_{1a} - \Delta G_{2a}$  in Eq. (14) can be obtained from:

$$\Delta G_{1a} - \Delta G_{2a} = 2.303RT(\text{p}K_{1a} - \text{p}K_{2a}) \quad (15)$$

where  $\text{p}K_{1a}$  and  $\text{p}K_{2a}$  are the  $\text{p}K$ s relative to the deprotonation of the carboxylic groups of the binding form of PGA and 3PP, respectively, in solution to yield the fully charged  $-3e$  inhibitors. The method used to calculate these  $\text{p}K$ s is explained in the “ $\text{p}K$  calculations” section below.  $\Delta G_{1a} - \Delta G_{2a}$  is the correction to the computed relative affinity to account for the transfer of the proton to species F. TIM is protonated (TIM  $\rightarrow$  TIM-H in Fig. 1) in both processes  $\Delta G_{1a}$  and  $\Delta G_{2a}$ . Because the enzyme is in this step of the thermodynamic cycle free in solution, the protonation reactions occur independently of the inhibitor. Therefore, they contribute equally to  $\Delta G_{1a}$  and  $\Delta G_{2a}$  and do not make a net contribution to the difference in these two free energies.

From the thermodynamic cycle (Fig. 1), the difference  $\Delta G_{1b} - \Delta G_{2b}$  can be obtained as:

$$\Delta G_{1b} - \Delta G_{2b} = \Delta G_3 - \Delta G_4 \quad (16)$$

where  $\Delta G_3$  and  $\Delta G_4$  are the free energies of mutating species F of PGA into 3PP in solution and in the active site of the enzyme, respectively.  $\Delta G_3$  and  $\Delta G_4$  can be computed via thermodynamic integration and molecular dynamics simulation techniques (described in the following two sections). Note that  $\Delta G_3 - \Delta G_4$  is the relative affinity of species F for TIM which we refer to as  $\Delta\Delta G(\text{calc})$

Equation (14) now becomes:

$$\begin{aligned} \Delta\Delta G(\text{pred}) &= \Delta G_{1a} - \Delta G_{2a} + \Delta G_3 - \Delta G_4 \\ \Delta\Delta G(\text{pred}) &= \Delta G_{1a} - \Delta G_{2a} + \Delta\Delta G(\text{calc}) \end{aligned} \quad (17)$$

### Thermodynamic integration

The free energy difference between two states A and B was estimated using the thermodynamic integration formula:

$$\Delta G = \int_0^1 d\lambda \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda \quad (18)$$

Here  $\lambda$  is a coupling parameter and H is the (classical) Hamiltonian of the system.  $\langle \partial H / \partial \lambda \rangle$  was estimated as the average  $\partial H / \partial \lambda$  during an equilibrium molecular dynamics (MD) simulation with  $\lambda$  kept constant. A total of 18 independent simulations were performed with different  $\lambda$  values between 0 and 1 to model the transformation of PGA into 3PP in water and in the active site of the protein. For every  $\lambda$  point, 100 ps of equilibration was followed by 200 ps of data collection. Reverse transformations were also carried out

**Table II**  
Bonded Parameters<sup>a</sup> of the Fully Deprotonated Forms of PGA and 3PP

Bonds	PGA		3PP			
	$b_0$ (nm)	$k^b$ (kJ mol <sup>-1</sup> nm <sup>-4</sup> )	$b_0$ (nm)	$k^b$ (kJ mol <sup>-1</sup> nm <sup>-4</sup> )		
C-Od1/Od2	0.125	1.34e+07				
C-CH2 <sub>a</sub>	0.153	7.15e+06				
C-CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub>	0.143	8.18e+06	0.153	7.15e+06		
Op4/CH2 <sub>b</sub> -P	0.161	4.84e+06	0.1797	4.84e+06		
P-Op1/Op2/Op3	0.148	8.60e+06				
Angles	$\theta_0$ (°)	$k^\theta$ (kJ mol <sup>-1</sup> )	$\theta_0$ (°)	$k^\theta$ (kJ mol <sup>-1</sup> )		
Od1-C-Od2	126.0	770				
Od1/Od2-C-CH2 <sub>a</sub>	117.0	635				
C-CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub>	109.5	520				
CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub> -P	120.0	530	113.0	545		
Op4/CH2 <sub>b</sub> -P-Op1/Op2/Op3	109.6	450				
Op1-P-Op2/Op3	120.0	780				
Op2-P-Op3	120.0	780				
Proper dihedrals	$\Phi_0$ (°)	$k^\Phi$ (kJ mol <sup>-1</sup> )	$n$	$\Phi_0$ (°)	$k^\Phi$ (kJ mol <sup>-1</sup> )	$n$
Od1-C-CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub>	180.0	4.305	2	0.0	3.502	2
C-CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub> -P	0.0	3.77	3	0.0	5.86	3
CH2 <sub>a</sub> -Op4/CH2 <sub>b</sub> -P-Op1	0.0	1.05	3	0.0	2.93	3
Improper dihedrals	$\varepsilon_0$ (°)	$k^\varepsilon$ (kJ mol <sup>-1</sup> degree <sup>-2</sup> )	$\varepsilon_0$ (°)	$k^\varepsilon$ (kJ mol <sup>-1</sup> degree <sup>-2</sup> )		
C-Od1-Od2-CH2 <sub>a</sub>	0.0	0.051				

For 3PP, only the parameters that differ from PGA are reported explicitly. Atoms labeled according to Figure 3.

<sup>a</sup>Bonded potential energy functions:

Bond potential:  $V_b(b) = \frac{1}{4} k^b (b^2 - b_0^2)^2$ ;

Angle potential:  $V_\theta(\theta) = \frac{1}{2} k^\theta (\cos\theta - \cos\theta_0)^2$ ;

Proper dihedral potential:  $V_\Phi(\Phi) = k^\Phi (1 + \cos(n\Phi - \Phi_0))$ ;

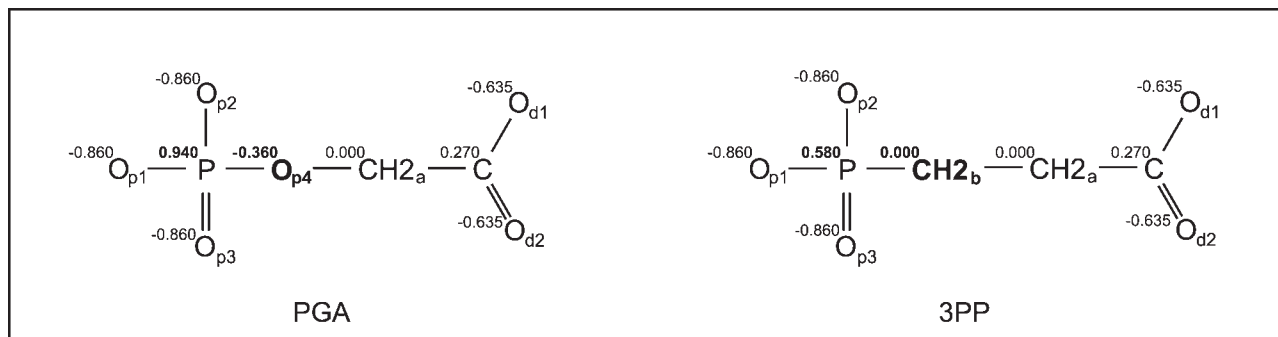
Improper dihedral potential:  $V_{id}(\varepsilon) = \frac{1}{2} k^\varepsilon (\varepsilon - \varepsilon_0)^2$ .

using the resulting structure at  $\lambda = 1$  as the starting configuration. The simulations are described in detail in the following section.

The transformation was achieved by the mutation of an oxygen atom into a carbon atom with all associated bonded and nonbonded interactions being mutated accordingly (the parameters used are reported in Table II and Fig. 3). For this a dual topology approach was applied. A molecule was constructed containing the two COO<sup>-</sup> groups of PGA and 3PP. During the transformation, the interactions of the COO<sup>-</sup> atoms of PGA with the rest of the system were gradually reduced to zero, whereas the interactions of the COO<sup>-</sup> atoms of 3PP were gradually increased. The mass of the atoms was held constant and the COO<sup>-</sup> groups of PGA and 3PP did not interact with each other. Only PGA of subunit A of TIM was mutated into 3PP.

The numerical integration of Eq. (18) was carried out using the trapezoidal method. The error in the  $\langle \partial H / \partial \lambda \rangle$  was calculated using a block averaging procedure.<sup>13,14</sup> Nonbonded interactions between the initial and the final





Mulliken atomic charges (RHF 3-21G\*)

Ligand	Op1,2,3	P	Op4/CH2b	CH2a	C	Od1,2
PGA	-0.965,-0.965,-0.944	1.425	-0.767	0.180	0.668	-0.846,-0.786
3PP	-0.972,-0.963,-0.963	1.352	-0.368	-0.137	0.691	-0.821,-0.818

**Figure 3**

Schematic representation of PGA and 3PP in the fully deprotonated form ( $-3e$ ). Atomic charges ( $e$ ) are indicated in the upper left corner of each atom. The atom and charges that differ between PGA and 3PP are indicated in bold. For comparison Mulliken charges are also listed.

state were interpolated using a soft-core potential<sup>15</sup> as implemented in the GROMACS package.<sup>16</sup> The soft-core parameter alpha was set to 1.51.

### Molecular dynamics simulations

The GROMOS96 force field 43a1<sup>17</sup> was used to describe the system. Aliphatic hydrogen atoms were treated as united atoms, together with the carbon atom to which they were attached. Force field parameters (charges and Lennard-Jones parameters) of the phosphate and phosphonomethyl group of the inhibitors were derived from the force field parameters of phosphoserine.<sup>18,19</sup> Quantum mechanical calculations, performed with the Gaussian03 program,<sup>20</sup> were used to determine the minimum energy conformations of the inhibitors in vacuum at the RHF/6-31G\* level of theory. The torsion around the CH<sub>2</sub><sub>a</sub>—C bond (Fig. 3) was parametrized using the results from the ab initio calculations and the X-ray structures of the free inhibitors.<sup>21,22</sup> 1-4 Lennard-Jones interactions involving the mutated atom (O→C) and carboxylic oxygens were excluded. The partial atomic charges and the bonded parameters are reported in Table II and Figure 3.

All molecular dynamics simulations were performed with the GROMACS suite of programs.<sup>16,23,24</sup> The starting structure of the inhibitor-enzyme complex was the RCSB-PDB<sup>25</sup> atomic resolution (0.83 Å) X-ray structure of liganded TIM (PDB code 1N55).<sup>26</sup> The free inhibitor was placed in a cubic box and the inhibitor-enzyme complex (and 542 crystallographic water molecules per monomer) in a dodecahedral box which were

subsequently filled with 1509 and 18,718 SPC (Simple Point Charge) water molecules,<sup>27</sup> respectively. A twin range cut-off was used for the Coulomb and Lennard-Jones interactions. Interactions between atoms within 0.9 nm were evaluated every step, while interactions between atoms lying between 0.9 and 1.4 nm were evaluated every 5 steps. To correct for the neglect of electrostatic interactions beyond the long range cutoff (1.4 nm), a reaction field (RF) correction with  $\epsilon_{RF} = 78.0$  was employed. Constant pressure and temperature were maintained by weakly coupling the system to an external bath at 1 bar and 298 K using the Berendsen barostat and thermostat<sup>28</sup> with coupling times of 1.0 and 0.1 ps, respectively. A leap-frog integrator was used.<sup>29</sup> The integration time step was 2 fs. The bond distances and the bond angle of water were constrained using the SETTLE algorithm.<sup>30</sup> All other bond distances were constrained using the LINCS algorithm.<sup>31</sup> Before the simulations, the potential energy of each system was minimized using a steepest descent approach, followed by a 20 ps molecular dynamics (MD) simulation with position restraints (force constant equal to 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup>) on the inhibitor and enzyme to relax the waters. A 200 ps simulation was performed to equilibrate each system before initiating the free energy calculations. In all calculations, position restraints were applied on all heavy atoms of the enzyme using a harmonic force constant of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup>.

### pK calculations

Theoretical estimates of pKs were obtained from the standard free energies of deprotonation as computed

with the Liptak and Shields approach<sup>32</sup> using the program Gaussian03.<sup>20</sup> The free energy of a proton dissociation reaction was estimated in vacuum, and the difference between the solvation free energies of the reactants and products was then added. The B3LYP/6-31G\* level of theory was applied in all computations. The geometry of each species was optimized in vacuum and the energy estimated. To correct for changes in vibrational entropy, a frequency calculation was performed with Gaussian03. To estimate the free energy of solvation of each species, the polarizable continuum model (PCM)<sup>33,34</sup> was applied.

pKs were calculated for the following deprotonation reactions of PGA and 3PP in solution: B→E, B→C, C→F, and E→F (pK<sub>1</sub>, pK<sub>3</sub>, pK<sub>2</sub>, and pK<sub>4</sub> in Fig. 2, respectively) as described in the previous paragraph. The pK<sub>2</sub>s of reactions C→F of PGA and 3PP correspond to pK<sub>1a</sub> and pK<sub>2a</sub>, respectively [Eq. (15)] and are required to calculate the difference  $\Delta G_{1a} - \Delta G_{2a}$  (Fig. 1). A number of experimental pK<sub>a</sub> values for the free inhibitors in solution are available (Table III) and were assigned to the reactions A→B, B→E, and E→F (indicated as pK<sub>a</sub>I, pK<sub>a</sub>II, and pK<sub>a</sub>III, respectively, in Fig. 2). This assignment is discussed in detail in the Results section.

As the calculations of absolute pKs are prone to very large systematic errors, the shift in the pKs for reactions C→F (pK<sub>2</sub>) with respect to B→E (pK<sub>1</sub>),  $\Delta pK_{2-1}$ , and B→C (pK<sub>3</sub>) with respect to E→F (pK<sub>4</sub>),  $\Delta pK_{3-4}$ , were used in preference to the raw values obtained from the calculations.  $\Delta pK_{2-1}$  and  $\Delta pK_{3-4}$  can be interpreted as the shifts in the pK of a certain protonation equilibria when the overall charge on the molecule changes (Fig. 2). pK<sub>2</sub> and pK<sub>3</sub> were then obtained by combining the experimentally measured pK<sub>a</sub> values with the calculated pK shifts:

$$pK_2 = pK_{aII} + \Delta pK_{2-1} \quad (19)$$

$$pK_3 = pK_{aIII} + \Delta pK_{3-4} \quad (20)$$

## RESULTS

The difference in the free energy of binding ( $\Delta\Delta G$ ) or relative binding affinity of the two inhibitors PGA and 3PP for the enzyme TIM was calculated. Both inhibitors contain three titrating sites and in solution they exist as a mixture of alternatively protonated forms of which only species C (Fig. 2) binds to TIM. In a solution of the enzyme and inhibitor, therefore, both the binding equilibrium between inhibitor and enzyme and ionization equilibria of the inhibitor (and enzyme) will occur. The experimental affinity  $\Delta G(\text{exp})$  is obtained from the total inhibitor concentration, which is the sum of all forms of

**Table III**  
Experimentally Determined pK<sub>a</sub> Values of PGA10 and 3PP<sup>35</sup> in Solution

Ligand	pK <sub>a</sub> I	pK <sub>a</sub> II	pK <sub>a</sub> III
PGA	<2	3.6	6.8
3PP	2.26 ± 0.04	4.63 ± 0.02	7.75 ± 0.02

inhibitor in solution. This implies that all ionization equilibria of the inhibitors in solution contribute to  $\Delta G(\text{exp})$ . This work determines the magnitude of this contribution. Note that the ionization equilibria of the enzyme which may also contribute to the binding properties were not included. The computationally obtained estimates of relative binding affinity  $\Delta\Delta G(\text{pred})$  typically consider only the binding species itself, species C in this particular case, such that any contribution to the affinity from other equilibria are neglected. Consequently,  $\Delta\Delta G(\text{pred})$  and  $\Delta\Delta G(\text{exp})$  cannot be compared directly.

To correct for the difference between  $\Delta\Delta G(\text{exp})$  and  $\Delta\Delta G(\text{pred})$  one must determine the contribution of the ionization equilibria to the experimental affinity and subtract this from  $\Delta\Delta G(\text{exp})$ , the result of which is referred to as  $\Delta\Delta G(\text{eff})$  (effective) because it is a measure of the (effective) affinity of the binding form C of the inhibitor for TIM, without contributions from the other equilibria in solution.  $\Delta\Delta G(\text{eff})$  can be estimated if the ratio of the equilibrium concentrations of the binding form (in this case species C) of the inhibitor and the total inhibitor I in solution are known [Eq. (4) in the Methods]. The relative concentrations of the different species can in turn be obtained from the pK<sub>a</sub>s of the ionizable groups [see Eqs. (5) – (9) in the Methods section].

The binding species C of PGA and 3PP correspond to only 0.27 and 7.2% of the total concentration of the inhibitors (Fig. 2). Given that the  $\Delta\Delta G(\text{exp})$  is  $-17.1 \text{ kJ mol}^{-1}$  (Table I), we obtain from Eq. (4) in  $\text{kJ mol}^{-1}$ :

$$\Delta\Delta G(\text{eff}) = \Delta\Delta G(\text{exp}) - 8.1 = -25.2 \quad (21)$$

That is the true affinity difference of species C,  $\Delta\Delta G(\text{eff})$ , is about  $8 \text{ kJ mol}^{-1}$  more negative than the experimental value which include all the nonbinding forms. In other words, PGA has effectively an  $-8.1 \text{ kJ mol}^{-1}$  higher affinity for TIM with respect to 3PP than experimentally measured. This is a consequence of the fact that at pH 7.6, the pH at which the affinity measurements were performed,<sup>7,8</sup> the concentration of the binding form of PGA was much lower than that of 3PP.

Having accounted for the contribution of the ionization equilibria of the inhibitors in solution to the binding affinity, it is now possible to relate the calculated binding affinity,  $\Delta\Delta G(\text{pred})$ , to the experimental value, using the thermodynamic cycle in Figure 1 where only the binding of the C form of the inhibitors is considered.

As can be seen in Figure 1, the binding reaction has been divided into two steps. The inhibitors are first deprotonated while in solution and then allowed to bind to the enzyme. In Figure 1, 3PP-H/PGA-H indicates the binding form of the inhibitors (species C in Fig. 2) and PGA/3PP the bound form (species F in Fig. 2). The relative free energy of binding of the bound form F of the inhibitors ( $\Delta G_{1b} - \Delta G_{2b}$ ) can be computed using molecular dynamics (MD) simulations, and is indicated  $\Delta\Delta G(\text{calc})$  (calculated). This difference in free energy must then be corrected for the free energy that is required to transfer a proton from TIM to species F, to yield the relative affinity for species C,  $\Delta\Delta G(\text{pred})$ . This correction is given by the difference  $\Delta G_{1a} - \Delta G_{2a}$  in Figure 1 and was estimated from the pKs of the deprotonation reactions of the inhibitors in solution [ $pK_{1a}$  and  $pK_{2a}$  for PGA and 3PP, respectively; Eqs. (14), (15), (19), and (20) in the Methods section]. Values of 5.1 and 6.9 were obtained for  $pK_{1a}$  and  $pK_{2a}$ , respectively ( $pK_2$  in Fig. 2), corresponding to a relative free energy difference of  $-10.3 \text{ kJ mol}^{-1}$  for  $\Delta G_{1a} - \Delta G_{2a}$  (Table I). The free energy difference for the deprotonation of species C of PGA in solution was therefore more favorable than that of 3PP. The protonation of the catalytic glutamic acid of the free enzyme TIM also occurs during processes  $\Delta G_{1a}$  and  $\Delta G_{2a}$  (TIM→TIM-H in Fig. 1). However, this contribution is the same in  $\Delta G_{1a}$  and  $\Delta G_{2a}$  and therefore effectively cancels when the difference is considered.

To estimate  $pK_{1a}$  and  $pK_{2a}$ , experimental  $pK_a$  values (Table III) and ab initio pK calculations (Table IV) were combined. Note that we use the more generic connotation pK to indicate computational estimates of  $pK_a$ s. In Figure 2, the final pK values can be read next to the corresponding deprotonation reactions. Also the experimental  $pK_a$ s are shown in Figure 2. If on a molecule there are multiple protonatable sites, as in the case of PGA and 3PP, the experimental estimate of a  $pK_a$  cannot be assigned to a particular (de)protonation reaction in a straightforward manner. In the following it is discussed how the experimental  $pK_a$ s were interpreted in view of the deprotonation reactions in Figure 2. PGA and 3PP have both three titratable sites and the experimental  $pK_a$  values are indicated in Table III as  $pK_{aI}$ ,  $pK_{aII}$ , and  $pK_{aIII}$ . If one compares these values to the standard  $pK_a$  values of phosphoric acid (2.1, 7.2, and 12.7), phosphonic acid (2.0 and 6.7) and carboxylic group of aspartic acid (3.9) in aqueous solution,<sup>36,37</sup> to a first approximation  $pK_{aI}$  and  $pK_{aIII}$  can be assigned to the deprotonation reaction of the first and second proton of the phosphate and phosphonomethyl moiety of PGA and 3PP, respectively, and  $pK_{aII}$  to the deprotonation of the carboxylic group of the inhibitor. This last assignment is consistent with the observation that  $pK_{aIII}$  of a phosphonate is about 1  $pK_a$  unit higher than that of the corresponding phosphate ester.<sup>38,39</sup> However, because in solution different ionized forms of the inhibitor coexist,

**Table IV**

Ab Initio pKs for the Deprotonation Reactions of PGA and 3PP in Solution at pH = 7.6

Ligand	$pK_1$ B → E	$pK_2$ C → F	$pK_3$ B → C	$pK_4$ E → F	$\Delta pK_{2-1}$	$\Delta pK_{3-4}$
PGA	23.3	24.8	30.5	32.1	1.5	-1.6
3PP	25.2	27.5	35.0	37.4	2.3	-2.4

$\Delta pK_{2-1}$  and  $\Delta pK_{3-4}$  together with the experimental  $pK_a$  values (Table III) are used to estimate  $pK_2$  and  $pK_3$  as described in the Methods [Eqs. (19) and (20)]. The differently protonated forms of PGA and 3PP are indicated with letters according to Figure 2.

$pK_{aI}$ , for example, can be associated with two reactions: deprotonation of the first proton of the phosphate/phosphonate group when the carboxylic group at the opposite end of the molecule is (a) protonated and (b) deprotonated, that is, reactions A→B and D→E in Figure 2. As a consequence  $pK_{aI}$  cannot be assigned to a specific reaction, either A→B or D→E. Nevertheless one can consider that the  $pK_a$  of a carboxylic group is 3.9 and at a pH corresponding to  $pK_{aI}$  (about 2), the concentration of a protonated carboxylic group (as in species A) is almost 80 times larger than that of the deprotonated one (as in species D). As the measured  $pK_a$  will be dominated by the most abundant species, the pK of reaction A→B, during which the carboxylic group remains protonated, was approximated by the measured  $pK_{aI}$ . Based on similar arguments, the pKs of reactions B→E ( $pK_1$  in Fig. 2) and E→F ( $pK_4$  in Fig. 2) were assigned to the experimental values of  $pK_{aII}$  and  $pK_{aIII}$ , respectively. One can also now see that  $pK_2$ , the pK of interest for the deprotonation reaction of the binding form C to F of the inhibitors, cannot be estimated from such considerations. For this reason ab initio calculations were used to estimate its value.

From the estimated pKs in Figure 2, one can observe the increase of the pK value of a certain deprotonation reaction when the molecule carries one additional negative charge ( $\approx 1.5$  and  $\approx 2.3$   $pK_a$  units for PGA and 3PP, respectively): compare in this respect reactions 1 with 2 and 3 with 4 in Figure 2. As expected, the higher the negative charge on the molecule, the lower the tendency to lose a proton.

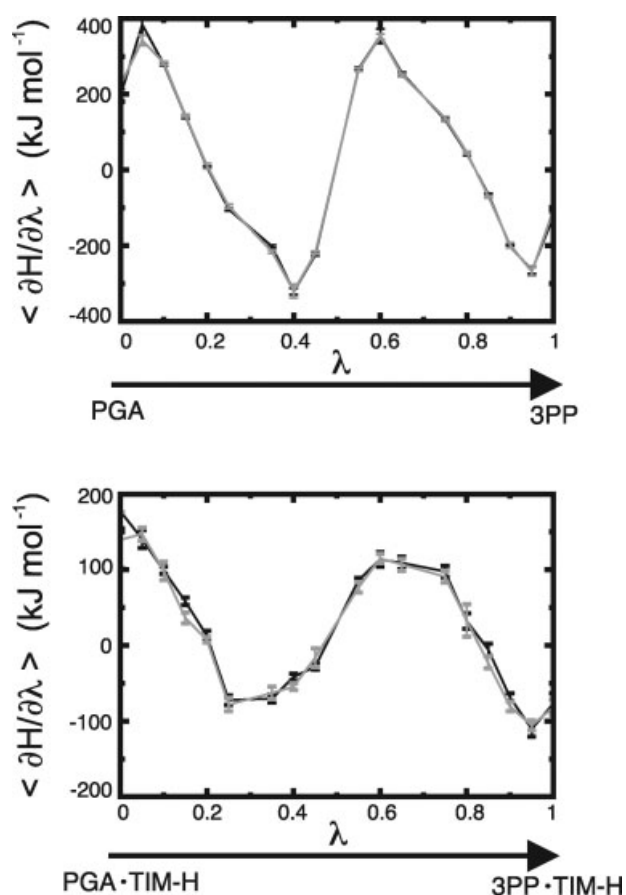
It remains now to estimate  $\Delta\Delta G(\text{calc})$ , the relative affinity of the bound form F of the inhibitors for TIM.  $\Delta\Delta G(\text{calc})$  can be obtained from the difference  $\Delta G_3 - \Delta G_4$  (Fig. 1 and Eq. (16) in the Methods section). Recall that  $\Delta\Delta G(\text{calc}) = \Delta G_{1b} - \Delta G_{2b} = \Delta G_3 - \Delta G_4$  in Figure 1.  $\Delta G_3$  and  $\Delta G_4$  are the free energy differences between the F species of the inhibitors in solution and in the active site of the enzyme, respectively, and were computed using molecular dynamics (MD) simulations and the thermodynamic integration formula. The calculated  $\Delta G_3$  and  $\Delta G_4$  were both positive and of comparable magnitude,  $28.3 \pm 4.3 \text{ kJ mol}^{-1}$  and  $24.1 \pm 7.2 \text{ kJ mol}^{-1}$ ,



respectively.  $\Delta\Delta G(\text{calc})$  was, therefore, not significant ( $4.2 \pm 8.4 \text{ kJ mol}^{-1}$ ). This indicates that the fully deprotonated F species of PGA and 3PP have a comparable affinity for TIM.

To test for convergence, the reverse transformations associated with  $\Delta G_3$  and  $\Delta G_4$  were also calculated. The results for the reverse transformations were comparable (within error) to the forward transformations ( $\Delta G_3 = 27.3 \pm 4.6 \text{ kJ mol}^{-1}$  and  $\Delta G_4 = 19.8 \pm 9.5 \text{ kJ mol}^{-1}$ ), suggesting that the computed free energies have converged.  $\partial H/\partial\lambda$  profiles as a function of the coupling parameter  $\lambda$  are shown in Figure 4 for both the reverse and forward transformations.

$\Delta G_3$  and  $\Delta G_4$  were decomposed in contributions corresponding to specific bonded and nonbonded interactions (Table V). Although this decomposition has little physical meaning because the individual components are not state functions and are therefore dependent on the path along which the transformation occurs, the analysis



**Figure 4**  
 $\langle \partial H/\partial\lambda \rangle$  (in  $\text{kJ mol}^{-1}$ ) as a function of  $\lambda$  for the transformation of the fully deprotonated PGA into 3PP in water (upper panel) and in the solvated active site of TIM (lower panel). Forward and reverse transformation are reported in black and gray, respectively.

**Table V**

Decomposition of  $\Delta G_3$ ,  $\Delta G_4$ , and  $\Delta\Delta G(\text{calc})$  ( $\Delta G_3 - \Delta G_4$ ) in Contributions ( $\text{kJ mol}^{-1}$ ) from Inhibitor (I), Water ( $\text{H}_2\text{O}$ ), and Enzyme (TIM) Interactions

	PGA $\rightarrow$ 3PP in $\text{H}_2\text{O}$ $\Delta G_3$	PGA $\rightarrow$ 3PP in TIM + $\text{H}_2\text{O}$ $\Delta G_4$	$\Delta\Delta G$
I-I	52.4	54.4	-2.0
I- $\text{H}_2\text{O}$	-24.0	-34.1	10.1
I-TIM	—	3.9	-3.9
( $\text{COO}^-$ )- $\text{H}_2\text{O}$	11.1	0.9	10.2
( $\text{COO}^-$ )-TIM	—	-2.3	2.3
(O/CH <sub>2</sub> -PO <sub>3</sub> <sup>2-</sup> )- $\text{H}_2\text{O}$	-34.8	-34.7	-0.1
(O/CH <sub>2</sub> -PO <sub>3</sub> <sup>2-</sup> )-TIM	—	6.2	-6.2

The inhibitor contributions I- $\text{H}_2\text{O}$  and I-TIM are furthermore decomposed between the carboxylic ( $\text{COO}^-$ ) and phosphate/phosphonomethyl (O/CH<sub>2</sub>-PO<sub>3</sub><sup>2-</sup>) moieties.

provides an indication of the relative importance of different terms within the force field. As can be seen in Table V, the contributions to the relative free energy difference from the inhibitor-inhibitor (I-I), inhibitor-solvent (I- $\text{H}_2\text{O}$ ), and inhibitor-enzyme (I-TIM) interactions were  $-2.0$ ,  $10.1$ , and  $-3.9 \text{ kJ mol}^{-1}$ , respectively. Intramolecular interactions of the inhibitors (I-I) and interactions of the inhibitors in the active site with TIM (I-TIM) contributed relatively little to  $\Delta\Delta G(\text{calc})$ . In particular, the contribution of the phosphate/phosphonomethyl of the inhibitor ( $-6.2 \text{ kJ mol}^{-1}$ ) indicates that the phosphate group is more stable in the active site of TIM with respect to the phosphonomethyl group.

We have now obtained a  $\Delta\Delta G(\text{calc})$  (difference in affinity of species F of PGA and 3PP for TIM) of  $4.2 \text{ kJ mol}^{-1}$  and a difference  $\Delta G_{1a} - \Delta G_{2a}$  of  $-10.3 \text{ kJ mol}^{-1}$ . This yields a  $\Delta\Delta G(\text{pred})$  (difference in affinity of species C of PGA and 3PP for TIM) of  $-6.1 \text{ kJ mol}^{-1}$ . This quantity can directly be compared to the  $\Delta\Delta G(\text{eff})$  of  $-25.2 \text{ kJ mol}^{-1}$  estimated above [Eq. (21)].  $\Delta\Delta G(\text{pred})$  is almost  $20 \text{ kJ mol}^{-1}$  smaller than the effective relative affinity. Table I summarizes the experimentally measured and calculated affinities, as well as the measured affinities for the natural substrates of TIM. The calculated affinities for species F ( $\Delta\Delta G(\text{calc})$ ) — derived from molecular dynamics and thermodynamic integration) and the predicted affinities as determined for species C of PGA and 3PP are given separately.

## DISCUSSION

### Effective affinity

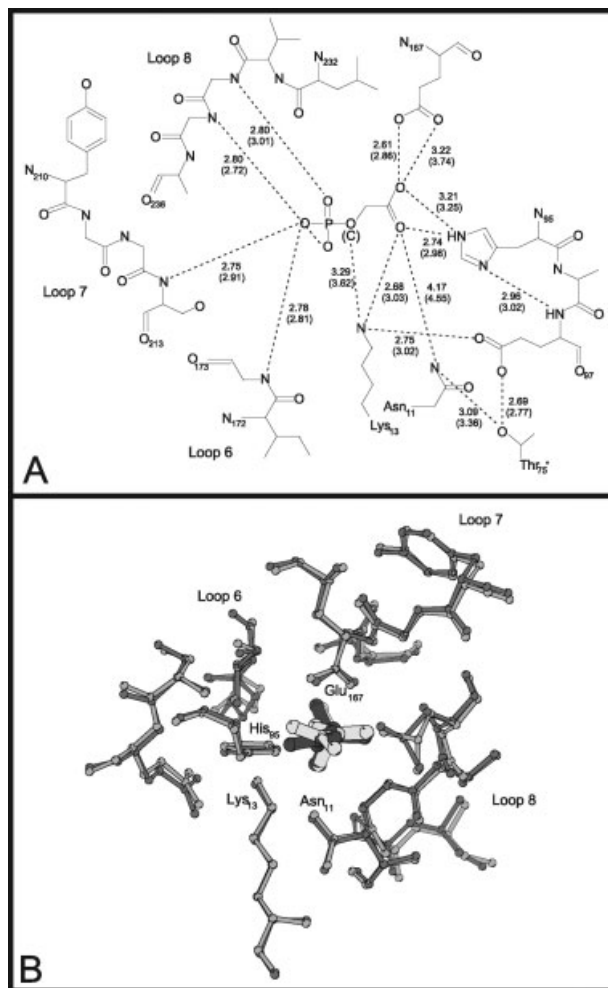
In this work we have introduced the concept of effective affinity. The effective affinity must be considered in cases where the inhibitor can exist in alternative protonation states when free in solution and where the dominant species in solution differs from the species bound to the

enzyme. Because experimental affinities are generally based on the total concentration of inhibitor, the protonation equilibria of the inhibitors in solution make a contribution to the measured affinity. This implies that experimental affinities cannot be directly compared with calculated values, unless one corrects for nonbinding species. The difference between the effective and measured affinities depends on the ratio of the concentration of the binding form of the ligand with respect to the total concentration [Eq. (2)]. If this ratio approaches one, effective affinities coincide with the measured ones. In the case of the TIM inhibitors PGA and 3PP, the binding species is just a small fraction of the total inhibitor concentration in solution. The effective affinities of the binding species of PGA and 3PP for TIM increased by  $-14.6$  and  $-6.5$   $\text{kJ mol}^{-1}$ , respectively, with respect to the measured values (Table I). Ignoring the various protonation equilibria and focusing on just the binding species would lead to an overestimation of the predicted binding affinity to TIM. An effective affinity therefore allows for a proper interpretation of experimental affinities and a direct comparison of calculated and measured affinities.

### Predicted relative affinity

Although the preceding discussion shows that the calculations overestimate the experimental binding affinity, from Table I it is observed that the predicted relative affinity  $\Delta\Delta G(\text{pred})$  in fact significantly underestimates the experimental one. There are a number of possible reasons for this difference.

The first is related to the calculation of  $\Delta G_4$ , the difference in the free energy of the bound form of the inhibitors in the active site of the enzyme TIM (Fig. 1).  $\Delta G_4$  is used to calculate  $\Delta\Delta G(\text{calc})$ . Upon ligand binding, one of the active site loops of TIM (loop 6) undergoes a large conformational change as the tip of this loop moves by about  $8 \text{ \AA}$  to close off the active site from bulk solvent.<sup>40–42</sup> To avoid sampling changes in the conformation of loop 6, all heavy atoms of TIM were restrained to the positions of the PGA-TIM complex during the calculations. This implies that during the calculation of  $\Delta G_4$ , only the closed bound conformation (of the PGA-TIM complex) is sampled. The free energies associated with the restraining of the active site containing PGA and 3PP were not computed as it was assumed that they were of comparable magnitude. It is reasonable to assume that the closed conformation is dominant in the bound state and the two complexes PGA-TIM and 3PP-TIM show a practically identical mode of binding (Fig. 5). In particular, the root mean square deviation (RMSD) between the structures of the two complexes was  $0.04 \text{ nm}$  for backbone atoms and  $0.08 \text{ nm}$  for active site atoms in a sphere of  $0.4 \text{ nm}$  around the inhibitor (Fig. 5). Nevertheless, this is obviously a source of error.



**Figure 5**

(A): Active site scheme of PGA-TIM and 3PP-TIM as observed in the X-ray structures 1N55<sup>26</sup> and 1IIG<sup>8</sup>, respectively. The bridging oxygen of PGA is replaced by a CH<sub>2</sub> in 3PP (indicated as (C)). All hydrogen bonds between inhibitor and enzyme atoms and other important hydrogen bonds are highlighted. Distances in 1N55 and 1IIG (in parenthesis) are in Å. TIM is a dimer consisting of two identical subunits of 250 residues. Thr75\* is in loop 3 of the adjacent subunit. Figure adapted from Kursula and Wiererga.<sup>26</sup> (B): Active site X-ray structures of PGA-TIM (1N55)<sup>26</sup>, in dark gray, and 3PP-TIM (1IIG)<sup>8</sup>, in light gray, superimposed on the backbone atoms of subunits B. Inhibitors are drawn with thicker bonds (in the center). Catalytic residues are indicated. Figure generated with Molscript.<sup>43</sup>

The second potential source of error is related to the accuracy of the  $pK$  calculations. We applied the B3LYP/6-31G\* level of theory in the ab initio calculations—a compromise between accuracy and computational effort—and found that the calculated absolute  $pK$  values were far off the experimental values (Table IV). The use of relative differences of the calculated  $pK$ s together with the experimentally measured  $pK$ s will reduce systematic errors in the computations. The residual error might still be as large as  $1\text{--}2$   $pK$  unit per calculated  $pK$  ( $\sim 6\text{--}12$   $\text{kJ mol}^{-1}$ ).

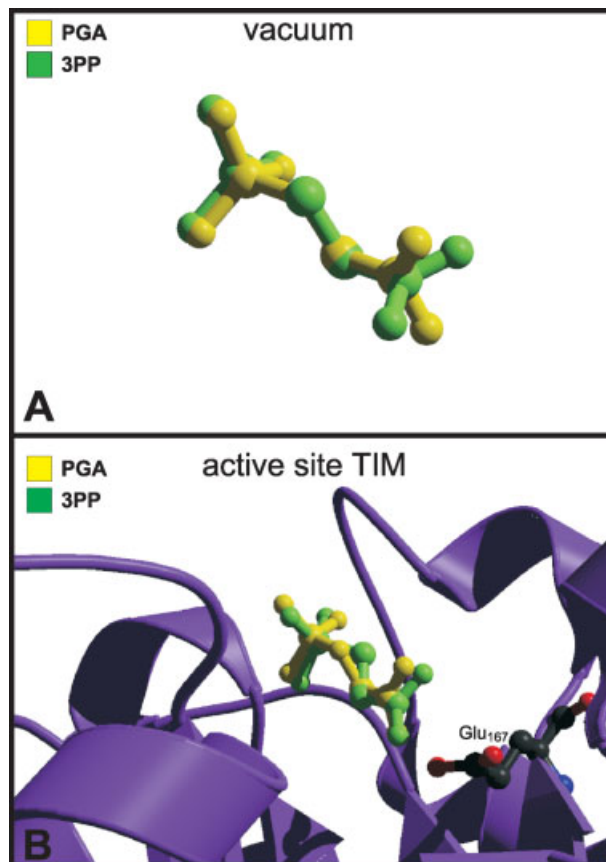
The relative affinity of the two inhibitors for TIM was obtained by adding together free energy differences ( $\Delta G$ ) from ab initio calculations and molecular dynamics simulations. The first relies upon a continuum solvation model for the solvent while for the second a discrete solvation model was employed. Because we are interested in relative free energy differences, the quantities that were summed were  $\Delta\Delta G$ s ( $\Delta G_{1a} - \Delta G_{2a}$  and  $\Delta G_3 - \Delta G_4$ ), rather than  $\Delta G$ s. Consequently, systematic errors introduced by each model are expected to cancel.

Another potential source of error is concerned with the importance of the low barrier hydrogen bond (LBHB)<sup>44</sup> for the binding free energy of TIM. In the atomic resolution structure of the PGA-TIM complex one such very strong hydrogen bond between the inhibitor and the enzyme was identified.<sup>26</sup> While this work assumes that this interaction is equally important for PGA and 3PP, its actual contribution to the affinity difference of PGA and 3PP for TIM is not clear. Even though not trivial, a mixed quantum mechanical/molecular mechanical approach to model this interaction might be helpful in elucidating this particular aspect.

Despite the above mentioned uncertainties, some qualitative conclusions can be drawn from the results. The reactions considered involve the binding of two inhibitors, PGA and its phosphonate analog 3PP to the glycolytic enzyme TIM. While it is known that phosphonate analogs of TIM ligands bind more weakly to the enzyme, the physical basis of this difference is not fully understood.<sup>11,38,45</sup> As can be seen in Figure 6, upon binding the torsion involving the carboxylic group of 3PP changes from staggered to eclipsed. It has been speculated that this deformation may influence the strength of ligand binding.<sup>8,38,45</sup> This work shows that the energetic cost of this deformation is in fact rather small (about  $2 \text{ kJ mol}^{-1}$ , Table V) and therefore cannot account for the observed difference in affinity. Also the contribution of the inhibitor-enzyme interactions to  $\Delta\Delta G(\text{pred})$  was small ( $-3.9 \text{ kJ mol}^{-1}$ , Table V), in agreement with the similarity of the bound PGA-TIM and 3PP-TIM structures. Instead, the main contribution to the predicted binding free energy difference was due to the difference in the pKs of the carboxylic group of the inhibitors in solution ( $-10.3 \text{ kJ mol}^{-1}$ ,  $\Delta G_{1a} - \Delta G_{2a}$ , Table I and Fig. 1). These results underline the importance of considering the appropriate thermodynamic cycle for free energy calculations, especially in cases where properties of the ligand in the bound state differ from that free in solution. Such considerations are frequently ignored in computational drug design.<sup>47,48</sup>

## CONCLUSIONS

The contribution of the protonation equilibria of PGA and 3PP to the binding affinity has been estimated,



**Figure 6**

Superimposition of (A) the unbound structures of PGA (yellow) and 3PP (green) (from vacuum calculations) and (B) the bound structures in the active site of TIM (from the X-ray structures PGA-TIM (PDB code 1N55)<sup>26</sup> and 3PP-TIM (PDB code 1IIG)<sup>8</sup>). In (B), the active sites atoms of TIM (defined as a sphere of radius 0.4 nm centered at the inhibitor) were superimposed. Only the backbone atoms of PGA-TIM (purple) and the catalytic glutamic acid (Glu 167; oxygen, carbon and nitrogen atoms colored in red, black, and blue, respectively) are shown. The inhibitors are drawn with the phosphate/phosphonomethyl moiety on the left side and the carboxylic group on the right side of the reader. Figures generated with Molscript<sup>43</sup> and Raster3-D.<sup>46</sup>

which can be used to obtain what we refer to as the effective affinity of the inhibitor for TIM by subtracting this correction from the experimental value. The binding species of PGA effectively binds about  $8 \text{ kJ mol}^{-1}$  stronger to TIM than the binding species of 3PP, with respect to the experimental affinity difference. This correction is required because of the relatively low concentration of the binding form of PGA in solution as compared to the binding form of 3PP at the pH used experimentally. By incorporating this correction it is possible to directly compare the predicted affinity difference of PGA and 3PP for TIM with the effective affinity difference, as both refer to the form of the inhibitor that is actually recognized by the enzyme. In this particular case, the theoretical prediction of the affinity difference of  $-6.1 \text{ kJ mol}^{-1}$

between PGA and 3PP for TIM underestimates the effective affinity difference of  $-25.2 \text{ kJ mol}^{-1}$  by about  $20 \text{ kJ mol}^{-1}$ . A failure to properly account for the effect of ionization could lead to theoretical estimates of free energies closer to experimental values than in reality. This underpins the importance to include ligand (solvation) properties in predictions of affinities for proteins.

## ACKNOWLEDGMENTS

The authors thank Cesar Araujo and Päivi Pirilä from the University of Oulu, Finland, for precious contributions, and Anne-Marie Lambeir from the University of Antwerp, Belgium, Markus Alahuhta and Marco Casteleijn from the University of Oulu, Finland, for useful discussions.

## REFERENCES

- Juffer AH, Vogel H.  $pK_a$  calculations of calbindin  $D_{9k}$ : effects of  $Ca^{2+}$  binding, protein dielectric constant, and ionic strength. *Proteins* 2000;41:554–567.
- Donnini S, Mark AE, Juffer AH, Villa A. Incorporating the effect of ionic strength in free energy calculations using explicit ions. *J Comput Chem* 2005;26:115–122.
- Rose IA. Mechanism of C-H bond cleavage in aldolase and isomerase reactions. *Brookhaven Symp Biol* 1962;15:293–309.
- Knowles JR. Enzyme catalysis: not different, just better. *Nature* 1991;350:121–124.
- Campbell ID, Jones RB, Kiener PA, Richards E, Waley SG, Wolfenden R. The form of 2-phosphoglycolic acid bound by triosephosphate isomerase. *Biochem Biophys Res Commun* 1978;83:347–352.
- Campbell ID, Jones RB, Kiener PA, Waley SG. Enzyme-substrate and enzyme-inhibitor complexes of triose phosphate isomerase studied by  $^{31}\text{P}$  nuclear magnetic resonance. *Biochem J* 1979;179:607–621.
- Lambeir AM, Opperdoes FR, Wierenga RK. Kinetic-properties of triose-phosphate isomerase from *trypanosoma-brucei-brucei*: a comparison with the rabbit muscle and yeast enzyme. *Biochemistry* 1987;168:69–74.
- Noble MEM, Wierenga RK, Lambeir AM, Opperdoes FR, Thunnissen AMWH, Kalk KH, Groendijk H, Hol WGJ. The adaptability of the active site of trypanosomal triosephosphate isomerase as observed in the crystal structures of three different complexes. *Proteins* 1991;10:50–69.
- Wolfram Research. Mathematica. Champaign, IL: Wolfram Research, Inc., 2003.
- Hartman FC, LaMuraglia GM, Tomozawa Y, Wolfenden R. The influence of pH on the interaction of inhibitors with triosephosphate isomerase and determination of the  $pK_a$  of the active-site carboxyl group. *Biochemistry* 1975;14:5274–5279.
- Belasco JG, Herlihy JM, Knowles JR. Critical ionization states in the reaction catalyzed by triosephosphate isomerase. *Biochemistry* 1978;17:2971–2978.
- Albery WJ, Knowles JR. Free energy profile of the reaction catalyzed by triosephosphate isomerase. *Biochemistry* 1976;15:5627–5631.
- Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford: Oxford Science Publication; 1987.
- Hess B. Determining the shear viscosity of model liquids from molecular dynamics simulations. *J Chem Phys* 2002;116:209–217.
- Beutler TC, Mark AE, vanSchaik RC, Geber PR, vanGunsteren WF. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett* 1994; 222:529–539.
- Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7: 306–317.
- vanGunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: GROMOS96 manual and user guide. Zürich, Groningen: BIOMOS b.v.; 1996.
- Hansson T, Nordlund P, Åqvist JA. Energetics and nucleophile activation in a protein tyrosine phosphatase. *J Mol Biol* 1997;265:118–127.
- [http://www.gromacs.org/contributions\\_ffg43a1p](http://www.gromacs.org/contributions_ffg43a1p). [Accessed 15 May 2002].
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 03, Revision C.02. Wallingford, CT: Gaussian, Inc.; 2004.
- van Bolhuis F. 3-phosphonopropionic acid-ion. In: Structural Report MSD015, Farmacie. Groningen, The Netherlands: Rijksuniversiteit Groningen; 1988.
- Lis T. Structure of phosphoglycolate (PG) in different ionization states. Structure of phosphoglycolic acid ( $\text{H}_3\text{PG}$ ) and its salts:  $\text{KH}_2\text{PG}$ ,  $\text{NaH}_2\text{PG}$ ,  $\text{K}_3\text{H}_3(\text{PG})_2$ ,  $\text{Na}_2\text{HPG} \cdot 0.5\text{H}_2\text{O}$ ,  $\text{Na}_2\text{H}(\text{PG})_2 \cdot 4\text{H}_2\text{O}$  and  $\text{Na}_3\text{PG} \cdot 4\text{H}_2\text{O}$ . *Acta Crystallogr C* 1993;49:696–705.
- Berendsen HJC, van der Spoel D, vanDrunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp Phys Commun* 1995;91:43–56.
- van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A, Berendsen H. Gromacs: fast, flexible and free. *J Comput Chem* 2005;26: 1701–1718.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P. The protein data bank. *Nucleic Acids Res* 2000;28:235–242.
- Kursula I, Wierenga RK. Crystal structure of triosephosphate isomerase complexed with 2-phosphoglycolate at  $0.83\text{-\AA}$  resolution. *J Biol Chem* 2003;278:9544–9551.
- Berendsen HJC, Postma JPM, vanGunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Reidel, Dordrecht; 1981. pp 331–342.
- Berendsen HJC, Postma JPM, vanGunsteren WF, Nola AD. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
- Hockney RW, Goel SP, Eastwood J. Quiet high-resolution computer models of a plasma. *J Comp Phys* 1974;14:148–158.
- Miyamoto S, Kollman PA. SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J Comput Chem* 1992;13:952–962.
- Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997; 18:1463–1472.
- Liptak MD, Shields GC. Accurate  $pK_a$  calculations for carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *J Am Chem Soc* 2001; 123:7314–7319.
- Miertus S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for a prevision of solvent effects. *Chem Phys* 1981;55:117–129.



34. Mennucci B, Tomasi J. Continuum solvation models. A new approach to the problem of solute's distribution and cavity boundaries. *J Chem Phys* 1997;106:5151–5158.
35. Heubel P-HC, Popov AI. Acid properties of some phosphonocarboxylic acids. *J Solution Chem* 1979;8:615–625.
36. Perrin DD. Dissociation constants of organic bases in aqueous solution. London: Butterworths; 1965.
37. Perrin DD. Dissociation constants of organic bases in aqueous solution. Supplement, 1972. London: Butterworths; 1972.
38. Dixon HBF, Sparkes MJ. Phosphonomethyl analogues of phosphate ester glycolytic intermediates. *Biochem J* 1974;141:715–719.
39. Orr GA, Knowles JR. The interaction of the phosphonate analogue of 3-phospho-d-glycerate with phosphoglycerate kinase. *Biochem J* 1974;141:721–723.
40. Joseph D, Petsko GA, Karplus M. Anatomy of a conformational change: hinged “lid” motion of the triosephosphate loop. *Science* 1990;249:1425–1428.
41. Davenport RC, Bash PA, Seaton BA, Karplus M, Petsko GA, Ringe D. Structure of triosephosphate isomerase-phosphoglycolohydroxamate complex: an analogue of the intermediate on the reaction pathway. *Biochemistry* 1991;30:5821–5826.
42. Wierenga RK, Noble ME, Davenport RC. Comparison of the refined crystal structures of liganded and unliganded chicken, yeast and trypanosomal triosephosphate isomerase. *Biochemistry* 1992;38:4389–4397.
43. Kraulis P. Molscript—a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991;24:946–950.
44. Cleland WW, Frey PA, Gerlt JA. The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 1998;273:25529–25532.
45. Lolis E, Petsko GA. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5 Å resolution: implication for catalysis. *Biochemistry* 1990;29:6619–6625.
46. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. *Methods Enzymol* 1997;277:505–524.
47. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;3:935–949.
48. Jain AN. Virtual screening in lead discovery and optimization. *Curr Opin Drug Discov Dev* 2004;7:396–403.