Computational method for relative binding energies of enzyme–substrate complexes

TAO ZHANG AND DANIEL E. KOSHLAND, JR.

Department of Molecular and Cell Biology, Division of Biochemistry, University of California,
and Materials Science Division, U.S. Department of Energy, Lawrence Berkeley Lab, Berkeley, California 94720

(RECEIVED September 12, 1995; ACCEPTED October 28, 1995)

Abstract

A computational method for estimating the relative binding free energies of enzyme–substrate complexes is described that combines electrostatic and solvation models and X-ray crystallographic data. The polar contribution is evaluated by the Poisson–Boltzman equation. The nonpolar contribution is evaluated by solvent transfer data and surface area calculations. This algorithm was used to calculate the relative binding energies of 63 pairs of nine different mutant proteins with seven different substituted R-malate substrates of *Escherichia coli* isocitrate dehydrogenase. Comparison of calculated values with the experimentally observed values shows a high degree of correlation.

Keywords: binding free energy; hydration free energy; isocitrate dehydrogenase; Poisson equation; protein engineering

A key goal of protein engineering is to design binding complexes of high affinity and specificity. Rational designs require a quantitative model that correlates binding free energies to interatomic interactions. The *Escherichia coli* isocitrate dehydrogenase (IDH) provides an experimental system for developing and evaluating such quantitative models. The catalytic mechanism and substrate specificities of IDH have been characterized (Dean & Koshland, 1990, 1993). The protein structure and substrate binding mode were revealed by X-ray crystal structures of the enzyme and the enzyme–substrate complex (Hurley et al., 1990, 1991; Stoddard et al., 1993). The activity of IDH is regulated by phosphorylation of an active site residue, Ser 113, which blocks enzyme action by a combination of electrostatic and steric effects (LaPorte & Koshland, 1982; Thorsness & Koshland, 1987; Dean & Koshland, 1990).

Recently, the Ser 113 residue was mutated in eight different ways, and the catalytic properties of these mutant proteins were evaluated using eight different substrates (Dean et al., 1996, accompanying paper). The binding properties of these substrate compounds and mutant enzymes provide an extensive data set for evaluating computational methods for relative binding free energies of enzyme–substrate complexes.

In this study, we propose an algorithm for calculating the energetic effect of structural perturbations on enzyme–substrate

complexes. This method dissects the contributions of polar and nonpolar interactions to binding free energies. The polar part is calculated by using a Poisson equation that defines electrostatic interactions of solutes (Kirkwood, 1934; Kirkwood & Westheimer, 1938; Hill, 1958; Sharp & Honig, 1990). The nonpolar part is calculated using the empirical solvation parameters for hydrocarbons (Nozaki & Tanford, 1971; Chothia, 1976). We describe this algorithm and apply it to calculate relative binding free energies of 63 pairs of IDH proteins and substrates. For simplicity, we call this approach a molecular statics approach, because it assumes an X-ray structure, in contrast to a molecular dynamics approach, which assumes movement from the structure to new positions by Newton equations.

Results

Structural model

The crystal structure of the enzyme–substrate complex of *E. coli* IDH shows how isocitrate interacts with the catalytic residues in the enzyme active site (Fig. 1). The α -carboxyl and the β -hydroxyl groups of the substrates are close to a bound Mg^{2+} ion. The γ -carboxyl group, on the other hand, is hydrogen bonded to the Ser 113 residue. Substituting this γ -carboxyl group of the ligand with other functional groups (Fig. 2) usually results in reduced substrate binding. Substituting residues at Ser 113 creates a similar effect on substrate binding (Table 1).

Despite the reduced binding constants and catalytic efficiencies, all the IDH mutants are still active on the substrate ana-

Reprint requests to: Daniel E. Koshland, Jr., Department of Molecular and Cell Biology, Division of Biochemistry, 229 Stanley Hall, University of California, Berkeley, California 94720; e-mail: daniel_koshland@maillink.berkeley.edu.

Fig. 1. Plot of the active site of wild-type *E. coli* IDH bound with isocitrate and a metal ion (Ca^{2+}). The residue Ser 113, the γ -carboxyl group of isocitrate, and other active site residues are labeled. The substrate isocitrate and the Mg^{2+} ion are shaded. In the accompanying paper (Dean et al., 1996), the Ser 113 was substituted with eight different amino acids and the γ -moiety was substituted with six different chemical groups. The effects of these structural perturbations on binding of substrates were determined (Dean et al., 1996).

logues, indicating that they bind to IDH in a manner with great similarity to isocitrate. It thus appears that the modified substrates and modified proteins are likely to maintain binding orientations similar to the wild-type protein. Therefore, we did a limited rotamer search to derive the structures of mutant proteins bound with substrate analogues, in which only the side chain of the substituted residue and the γ -moiety of the substrate analogue were adjusted to relieve any overlapping. In terms of

Fig. 2. Modified substrates used in the kinetic studies of the wild-type and mutant *E. coli* IDH enzymes (Dean et al., 1996). The abbreviated names for these substrate analogues are indicated in parentheses. The atomic partial charges used in the calculation are listed next to the corresponding atoms.

the number of side-chain torsional angles, the largest change in the substrate is malate $\rightarrow$ isocitrate or ethynalmalate, and the largest change in the enzyme is Ala $\rightarrow$ Glu.

In all cases, there are no more than three torsion angles in need of adjustment, and one unique combination of acceptable rotamers was always found to relieve steric overlapping and repulsive charge-charge interactions. Our algorithm described in the Methods section mostly uses the static, experimentally determined crystal structure of the enzyme-substrate complex. This is in contrast to other algorithms, where the crystal structures are treated in a dynamic, flexible manner (Bash et al., 1987; Warshel et al., 1988; Wilson et al., 1991).

In all cases, the relative free binding energy is defined as the difference between the enzyme-substrate complex being considered and the malate enzyme complex, as shown in Figure 3. The two complexes are identical except for the modified substrate in the Q_1 , Q_2 area, where Q_1 refers to the R-substitution and Q_2 is the H of malate.

Nonpolar contribution to free energy

The nonpolar contribution was calculated using the empirical solvation model for transferring straight-chain alkanes from water to organic solvents such as ethanol (Nozaki & Tanford, 1971), octanol (Fauchere & Pliska, 1983), and *N*-methylacetamide (Damodaran & Song, 1986). It was shown in previous studies that these transfer free energies are approximately proportional to the surface area (Nozaki & Tanford, 1971; Chothia, 1976; Eisenberg & McLachlan, 1986). This dependence leads to a relationship between the nonpolar contribution to the binding free energy and the buried surface area of substrate in the binding complex. Substituting the γ -moiety of malate by alkyl groups, such as ethyl and propyl, creates an increasing length of alkane chain. The added alkane chain is less exposed in the enzyme-substrate complex, and change in the accessible surface area for each substrate analogue depends on its binding conformation. For these ethyl and propyl substitutions, the changes of surface area were calculated to be in the range of 60.0–80.0 $\text{\AA}^2$, which corresponds to 1.2–1.6 kcal/mol of hydration free energy, assuming a 20.0-cal $\text{mol}^{-1} \text{\AA}^{-2}$ coefficient. Similarly, the nonpolar free energies due to the γ -substitutions by carboxyl, hydroxyl,

Fig. 3. Schematic diagram of the wild-type and mutant protein-substrate binding complexes. The structural perturbation of the substrate, $L_1 \rightarrow L_2$, affects the free energies of the solvated binding complex ($P \cdot L_1 \rightarrow P \cdot L_2$) and the solvated free substrate ($L_1 \rightarrow L_2$). The net difference between the effect on these two states determines the effect on the binding free energy.

Table 1. Calculated versus observed relative binding energies of the enzyme-substrate complexes of *E. coli* IDH^a

Protein	Substrate	$\Delta G_{\text{PL1F-PL2F}}^{\text{np}}$	$\Delta G_{\text{PL1F-PL2F}}^{\text{p}}$	ΔG^{calc}	ΔG^{exp}	Difference
Ser 113	ISOC	-1.6	-1.1	-2.7	-2.8	0.1
	MMAL	-1.2	0.0	-1.2	-1.5	0.3
	EMAL	-1.6	0.0	-1.6	-1.4	0.2
	PMAL	-1.2	0.0	-1.2	-1.2	0.0
	HMAL	-1.2	-0.2	-1.4	-1.6	0.2
	EHMA	-1.6	-0.3	-1.9	-1.0	0.9
	TART	-1.2	-0.8	-2.0	-1.8	0.2
Asp 113	ISOC	-1.2	0.8	-0.4	0.4	0.8
	MMAL	-1.2	0.0	-1.2	-1.7	0.5
	EMAL	-1.6	0.0	-1.6	-1.9	0.3
	PMAL	-1.2	0.0	-1.2	-1.9	0.7
	HMAL	-1.2	-1.3	-2.5	-1.3	1.2
	EHMA	-1.6	0.9	-0.7	-1.0	0.3
	TART	-1.2	-1.0	-2.2	-2.1	0.1
Leu 113	ISOC	-1.6	-1.0	-2.6	-1.6	1.0
	MMAL	-1.2	0.0	-1.2	-1.6	0.4
	EMAL	-1.6	0.0	-1.6	-1.6	0.0
	PMAL	-1.2	0.0	-1.2	-1.6	0.4
	HMAL	-1.2	-1.1	-2.3	-1.4	0.9
	EHMA	-1.6	-0.5	-1.1	-0.8	0.3
	TART	-1.2	-0.9	-2.1	-1.7	0.4
Tyr 113	ISOC	-1.6	-0.7	-2.3	-1.0	1.3
	MMAL	-1.2	0.0	-1.2	-1.3	0.1
	EMAL	-1.6	0.0	-1.6	-1.3	0.3
	PMAL	-1.2	0.0	-1.2	-1.5	0.3
	HMAL	-1.6	1.2	-0.4	-0.7	0.3
	EHMA	-1.6	0.5	-1.1	-1.2	0.1
	TART	-1.2	0.0	-1.2	-1.4	0.2
Thr 113	ISOC	-1.6	-1.4	-2.8	-2.0	0.8
	MMAL	-1.2	0.0	-1.2	-1.6	0.4
	EMAL	-1.6	0.0	-1.6	-1.6	0.0
	PMAL	-1.2	0.0	-1.2	-1.5	0.3
	HMAL	-1.2	0.3	-0.9	-1.5	0.6
	EHMA	-1.6	-0.2	-1.8	-1.0	0.8
	TART	-1.2	-0.9	-2.1	-1.8	0.3
Ala 113	ISOC	-1.6	-1.1	-2.7	-2.2	0.5
	MMAL	-1.2	0.0	-1.2	-1.7	0.5
	EMAL	-1.6	0.0	-1.6	-1.7	0.1
	PMAL	-1.2	0.0	-1.2	-1.7	0.5
	HMAL	-1.2	-1.1	-2.3	-1.0	0.3
	EHMA	-1.6	0.5	-1.1	-0.6	0.5
	TART	-1.2	-0.9	-2.1	-1.7	0.4
Asn 113	ISOC	-1.6	-1.7	-3.3	-2.9	0.4
	MMAL	-1.2	0.0	-1.2	-1.6	0.4
	EMAL	-1.6	0.0	-1.6	-1.7	0.1
	PMAL	-1.2	0.0	-1.2	-1.6	0.4
	HMAL	-1.2	0.2	-1.0	-1.5	0.5
	EHMA	-1.6	-0.2	-1.8	-1.0	0.8
	TART	-1.2	-0.7	-1.9	-2.0	0.1
Glu 113	ISOC*	-1.6	-1.0	-2.6	0.1	2.7
	MMAL	-1.2	0.0	-1.2	-1.6	0.4
	EMAL	-1.6	0.0	-1.6	-1.6	0.0
	PMAL	-1.2	0.0	-1.2	-1.4	0.2
	HMAL	-1.2	0.4	-0.8	-0.7	0.1
	EHMA	-1.6	0.3	-1.3	0.0	1.3
	TART	-1.2	0.0	-1.2	-1.6	0.4
Gln 113	ISOC	-1.6	-0.5	-2.1	-2.0	0.1
	MMAL	-1.2	0.0	-1.2	-2.0	0.1
	EMAL	-1.6	0.0	-1.6	-1.6	0.0
	PMAL	-1.2	0.0	-1.2	-1.5	0.3
	HMAL	-1.2	-0.1	-1.3	-1.5	0.2
	EHMA	-1.6	-0.7	-2.3	-1.0	1.3
	TART	-1.2	-0.4	-1.6	-1.8	0.2

^a For each protein (mutant or wild type), the corresponding protein-malate binding complex is used as the reference state to calculate the relative binding free energy. For example, all values listed for IDH Ser 113 are binding free energies of Ser 113 mutant-substrate complexes *relative* to that of Ser 113-malate complex. The absolute difference between the calculated and observed binding free energies (Dean et al., 1996) is also listed in the last column. The average difference between the calculated and the observed values for all binding complexes is 0.5 kcal/mol. The calculated free energy change for one binding complex, Glu 113-ISOC, significantly deviates from the observed values (labeled *). ISOC, isocitrate; MMAL, methylmalate; EMAL, ethylmalate; PMAL, propylmalate; HMAL, hydroxymethylmalate; EHMA, ethynalmalate; TART, tartrate.

and ethynal groups (isocitrate, hydroxylmalate, and ethynal-malate) were calculated to be 1.2–1.6 kcal/mol (Table 1).

The calculated surface areas depend on the van der Waals radii for the protein and the solvent atoms. In the CHARMM extended atom type model, only polar hydrogens were included. In our calculation, the van der Waals radii of the heavy atoms were enlarged to account for the attached hydrogens. Recently, a set of van der Waals radii was derived by optimizing the fit between the calculated and experimental hydration free energies of a group of small organic molecules (Sitkoff et al., 1994). Use of this optimized set of van der Waals radii in our calculation altered the calculated nonpolar contributions usually by less than 10%.

Polar contribution to free energy

In our algorithm, the calculated polar contribution to the relative binding free energy is zero for nonpolar hydrocarbons such as the substrate analogues with the γ -alkyl groups, methyl-, ethyl-, and propylmalate (Table 1). This zero polar contribution can be interpreted in two ways. First, the hydrocarbons are assigned neutral partial charges in the extended atom model (Brooks et al., 1983). Second, the overall energetic effect is directly taken into account by calculating the empirical solvation energy in terms of changes of accessible surface area. The carboxyl, hydroxyl, and ethynal groups are polar in nature, and all have a net calculated electrostatic contribution to the change in binding free energy. Their polar contributions range from -1.7 to 1.2 kcal/mol, similar in magnitude to the nonpolar contributions of other groups (Table 1). Similar observations have been made in other enzyme–substrate complexes.

To investigate the effect of partial charge assignment on the calculation of polar contributions, the CHARMM atomic partial charges (Brooks et al., 1983) were adjusted by a random value that varied from 0.0 to 20% of the originally assigned partial charges. The average of these adjustments is 10.0% of the originally assigned CHARMM partial charges. The relative binding free energies of wild-type IDH versus substrate analogues were calculated with these altered partial charges (Table 2). The calculated values are only moderately sensitive to the assigned partial charges. For example, the relative binding energy of isocitrate versus malate changes from -2.7 kcal/mol to -5.0 kcal/mol, a net change of about 2.3 kcal/mol, even though the partial charges were changed randomly by an average of 10% from the standard assignments. The effect is even smaller

for the calculated relative binding free energies of other substrate analogues (Table 2).

Overall model performance

The average difference for the calculated and the observed relative binding free energies is 0.5 kcal/mol for all the 63 pairs of enzyme–substrate complexes (Table 1). For one enzyme–substrate complex (labeled with an asterisk in Table 1), the calculated free energy change deviates significantly from the observed value. Judged by these differences, the performance of our method is comparable to or better than many reported empirical approaches (Novotny et al., 1989; Wilson et al., 1991). The correlation of the calculated and the observed relative binding energies is plotted as a scatter plot in Figure 4A. A linear regression of the calculated and observed binding energies for all but three of the enzyme–substrate complexes has a goodness of fit of 0.32. The goodness of fit was evaluated by the χ^2 test, assuming an average standard deviation of 0.4 kcal/mol for the observed binding energies (corresponding to a twofold variation of K_m at room temperature). When the same plot was made with either the polar terms or the nonpolar terms alone, the goodness of fit is not nearly as good. Therefore, both terms are needed for the theoretical calculation for the substrates of the polar or the polar and nonpolar mixed types.

Discussion

We have used here a computational method that combines a Poisson equation for the polar factor and a calculated surface area solvent transfer model for the nonpolar factor of the binding free energy. We identify this method as a molecular statics (MS) method for brevity because it uses the X-ray crystal structure and makes no extended analysis of molecular dynamics or Monte Carlo methods. The steric overlap caused by the structural perturbation is relieved by using the Ponder and Richards (1987) side-chain rotamer library before energy evaluation. The empirical atomic solvation model and the finite difference solution of the Poisson–Boltzmann equation were then used to evaluate the nonpolar and polar contributions of the relative binding free energy, respectively. This algorithm calculated the values of the relative binding free energies of 63 pairs of IDH mutant protein–substrate analogue complexes with an average deviation of ± 0.5 kcal/mol. Because no parameter optimization was used, the agreement between the calculated and observed values is significant. Our method demonstrates that macroscopic models can be used to calculate changes of free energy with small structure perturbations of both substrate and enzyme.

Despite recent advances (Bash et al., 1987; Warshel et al., 1988; Novotny et al., 1989; Wilson et al., 1991), the computational methods for evaluating binding free energies are still limited by accuracy and speed. The MS method described here uses far fewer computational resources than the free energy perturbation (FEP) method (Bash et al., 1987; Warshel et al., 1988). In the FEP method, a mutational effect on free energy is calculated by a simulation of the dynamic transition from the native to the mutated state, in which the Hamiltonian of the native state is gradually changed to that of the mutated state. Although FEP is a “first-principle” method and has been successful in calculating relative binding free energies in several reports, it has some practical limitations. First, the FEP method generally re-

Table 2. Calculated relative binding energy of wild-type IDH with altered partial charges^a

Substrate	$\Delta G_{\text{PL1F-PL2F}}^{\text{np}}$	$\Delta G_{\text{PL1F-PL2F}}^{\text{p}}$	ΔG^{calc}	ΔG^{exp}	Difference
ISOC	-1.6	-3.4	-5.0	-2.8	2.2
MMAL	-1.2	0.0	-1.2	-1.6	0.4
EHMA	-1.6	-0.4	-2.0	-1.0	1.0
TART	-1.2	-1.0	-2.2	-1.8	0.4

^a See text for additional information. ISOC, isocitrate; HMAL, hydroxylmethylmalate; EHMA, ethynal-malate; TART, tartrate.

quires a large computational resource, which limits its use in screening energetic effects of a large number of structural perturbations. For example, simulating a 103-ps dynamic trajectory of a moderate-sized protein requires days of CPU time on most of today's computers (Debolt et al., 1994). Some critical molecular motions such as a side-chain rotation in a packed protein interior require simulation of a long dynamic trajectory (McCammon & Harvey, 1987; Tidor & Karplus, 1991). Second, the accuracy of the simulated trajectory depends on the molecular mechanics force field. In some studies, a long dynamic trajectory, generated by simulation without constrained atoms (such as the protein backbone), usually leads to conformations that significantly deviated from the crystal structures (Ferguson et al., 1991; Simonson & Brunger, 1992). In a 60-ps simulated trajectory of the solvated active site of the enzyme-substrate complex of *E. coli* IDH, the atoms of the bound substrate moved by an average 2.1 Å from their crystallographically determined positions.

In contrast, empirical approaches (Novotny et al., 1989; Wilson et al., 1991) for calculating binding energies are based on direct energy evaluation, sometimes combined with conformational searches, sampling of rotamer libraries, and empirical solvation models (Wilson et al., 1991). These algorithms are usually fast and can screen the energetic effects of a number of structural perturbations. Because solvent molecules are not explicitly included in these empirical models, solvent screening and polarization effects are either not included or are approximated by a distance-dependent dielectric constant (Wilson et al., 1991). A better treatment of electrostatic interactions will improve model accuracy.

In the MS method, the binding free energy is dissected into two parts: nonpolar and polar. It has been shown that the free energy of transferring hydrocarbons from aqueous solutions into organic solvents (Nozaki & Tanford, 1971; Chothia, 1976) and the hydrophobic interior of proteins (Eriksson et al., 1992) is proportional to the change of solvent-accessible surface area. Therefore, the nonpolar contribution to the binding free energy can be directly estimated by surface area calculations. However, the contributions from specific salt bridges or hydrogen bonds in the protein-ligand binding complexes are not counted by the water-ethanol or water-octanol transfer models. The energetic effects of these specific charge-charge interactions must be calculated by electrostatic models of the solvated protein-ligand binding complexes. A Poisson equation describes the electrostatic interactions of polar solutes. The finite difference solution of the Poisson equation gives results with accuracy comparable to the explicit solvent simulation but at less computational cost (Sharp & Honig, 1990). Combining a Poisson equation (polar contribution) and the solvation model (nonpolar contribution), we are able to predict the relative binding energies of the substrate analogues for the IDH mutants (Table 1; Fig. 4). We also show that either calculated nonpolar or polar contributions alone correlate poorly with observed values.

Fig. 4. Scatter plots of calculated relative binding energies versus experimentally observed relative binding energies. **A:** Calculated relative binding energies containing both polar and nonpolar contributions. All data in Table 1 except for Glu 113-ISOC were included. **B:** Calculated relative binding energies containing only polar contributions. **C:** Calculated relative binding energies containing only nonpolar contributions.

Caveats

Several factors affect the evaluation of the electrostatic energy by a Poisson equation. The calculated electrostatic potential is sensitive to the dielectric constants of solvent and solute. The aqueous solvent dielectric constant has a consensus value of 80.0, and the protein dielectric constant is generally believed to be in the range of 2.0–5.0. Because the protein–ligand interface is usually solvent inaccessible, the calculated electrostatic interactions are quite sensitive to the chosen protein dielectric constant. In this study, we chose a protein dielectric constant of 2.0 because this value is commonly used as a default for the protein core. Improvements in the method will allow better estimation of the dielectric constant of a protein core.

The finite difference solution of a Poisson equation is not free of errors when compared to an analytical solution (Gilson et al., 1987). For a single point charge in a low-dielectric cavity surrounded by a high-dielectric solvent, the errors are larger at grid points close to the charge or the dielectric boundary. In our study, the charged ligand atoms are all at least 3.0 Å away from the protein atoms and at least 1.4 Å (water probe radius) below the surface of the dielectric boundary. Therefore, the errors of the calculated electrostatic potentials should be less than 10%, using the error estimate for a single point charge model (Klapper et al., 1986). Entropy also contributes to the binding free energy. The solvation entropy contribution to the relative binding free energy of complexes $P \cdot L_1$ to $P \cdot L_2$ was partly accounted for by the $\Delta G_{F,P \cdot L_1 \rightarrow P \cdot L_2}^{np}$. The contributions to relative binding energy from configurational and translational entropies are currently omitted in our method. Because we focus here on the relative binding free energies of structurally related substrate analogues, these terms will cancel out in the difference calculations at the core of our method. The configurational entropy would need to be estimated to calculate the absolute binding free energy (Vajda et al., 1994).

In this study, we calculated the *relative* binding free energies of structurally related ligands, and only limited rotamer searching for binding conformation was done, in contrast to the exhaustive conformational searching and sampling of other algorithms (Wilson et al., 1991). Predicting the energetic and structural effects of large perturbations may require more conformational searching than is currently employed in our method.

Conclusions

We propose a molecular structure method to estimate the relative binding energies of protein–ligand complexes. The predicted relative binding free energies of 63 pairs of enzyme–substrate complexes of *E. coli* IDH showed a good correlation with the experimentally observed values (Fig. 4A). Both nonpolar and polar contributions must be included in the model to achieve a high degree of correlation (Fig. 4B,C). This approach will be applied to more complex examples in the future.

Methods

The binding free energies of two enzyme–ligand complexes PL_1 and PL_2 are defined as:

$$\Delta G_{F,P \cdot L_1} = \Delta G_{P+L_1 \rightarrow P \cdot L_1} = G_{P \cdot L_1} - (G_P + G_{L_1}) \quad (1)$$

$$\Delta G_{F,P \cdot L_2} = \Delta G_{P+L_2 \rightarrow P \cdot L_2} = G_{P \cdot L_2} - (G_P + G_{L_2}), \quad (2)$$

where $P \cdot L$ and $P + L$ refer to the solvated enzyme–ligand complex and the solvated free enzyme and ligand, respectively. L_1 and L_2 are two structurally related substrate analogues, and ΔG_F is the free energy of formation of a binding complex (Fig. 3). The difference of binding free energies of the two substrate analogues L_1 and L_2 is:

$$\begin{aligned} \Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)} &= \Delta G_{F,P \cdot L_1} - \Delta G_{F,P \cdot L_2} \\ &= (G_{P \cdot L_1} - G_{P \cdot L_2}) - (G_{L_1} - G_{L_2}) \\ &= \Delta G_{P \cdot L_1 \rightarrow P \cdot L_2} - \Delta G_{L_1 \rightarrow L_2}, \end{aligned} \quad (3)$$

where $\Delta G_{P \cdot L_1 \rightarrow P \cdot L_2}$ and $\Delta G_{L_1 \rightarrow L_2}$ are differences of free energy of the enzyme–substrate complexes $P \cdot L_1$, $P \cdot L_2$ and free substrates L_1 , L_2 , respectively.

Free energy G is dissected into a polar part (G^p) and a nonpolar part (G^{np}), so that the total relative binding free energy is calculated as the sum of a polar part and a nonpolar part:

$$\Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)}^{total} = \Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)}^{np} + \Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)}^p, \quad (4)$$

where

$$\begin{aligned} \Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)}^{np} &= \Delta G_{P \cdot L_1 \rightarrow P \cdot L_2}^{np} - \Delta G_{L_1 \rightarrow L_2}^{np} \\ \Delta G_{F,(P \cdot L_1 \rightarrow P \cdot L_2)}^p &= \Delta G_{P \cdot L_1 \rightarrow P \cdot L_2}^p - \Delta G_{L_1 \rightarrow L_2}^p. \end{aligned} \quad (5)$$

Nonpolar contribution

According to the empirical solvation model (Nozaki & Tanford, 1971; Chothia, 1976; Fauchere & Pliska, 1983), the free energy of transferring nonpolar atoms from an aqueous solution to the hydrophobic interior of a protein can be approximated as:

$$\Delta G^{np} = \sigma \sum A_i, \quad (6)$$

where $\sum$ indicates a summation over atoms, ΔG^{np} is the transfer free energy, σ is the empirical solvation parameter, and A_i is the solvent-accessible surface area.

A consensus σ value of 20.0–25.0 cal mol^{−1} Å^{−2} for nonpolar hydrocarbon molecules was derived in several studies (Nozaki & Tanford, 1971; Fauchere & Pliska, 1983; Damodaran & Song, 1986). In a recent report, Eriksson et al. (1992) studied the structural response of T4 lysozyme to cavity-forming mutants and its relation to the hydrophobic effect. When the energetic effect of cavity formation was taken into account, the average atomic solvation parameter of hydrocarbons was calculated to be 20.0 cal mol^{−1} Å^{−2}. We used this value in our calculation of the nonpolar contribution to binding free energy. The solvent-accessible surface area A_i is calculated by the Lee and Richards (1971) algorithm using a 1.4-Å-radius water probe. We also use the extended atom type model in which the van der Waals radii of the heavy atoms take into account the attached hydrogens (Brooks et al., 1983).

The nonpolar contribution to the relative binding free energy of complexes $P \cdot L_1$ versus $P \cdot L_2$ is proportional to the difference in the change of solvent-accessible surface area (upon binding) of the complexes:

$$\Delta G_{P \cdot L_1 \rightarrow P \cdot L_2}^{np} = \sigma \left(\sum A_i^{P \cdot L_1} - \sum A_i^{P \cdot L_2} \right)$$

$$\Delta G_{L_1 \rightarrow L_2}^{np} = \sigma \left(\sum A_i^{L_1} - \sum A_i^{L_2} \right), \quad (7)$$

where $\sum$ indicates a summation over the corresponding groups of atoms, $A_i^{P \cdot L}$ and A_i^L are the calculated solvent-accessible surface areas of atom i in the solvated enzyme-substrate complexes $P \cdot L$ and the free ligand L , respectively (Fig. 3).

Polar contribution

The electrostatic energy (G^p) of a group of atoms is defined as:

$$G^p = \sum \phi_i q_i,$$

where q_i is the partial charge, ϕ_i is the electrostatic potential, and $\sum$ indicates a summation over the polar atoms. The difference of electrostatic energy for two enzyme-substrate complexes $P \cdot L_1$ and $P \cdot L_2$ is:

$$\Delta G_{P \cdot L_1 \rightarrow P \cdot L_2}^p = \sum_{P \cdot L_1} \phi_i^{P \cdot L_1} q_i - \sum_{P \cdot L_2} \phi_i^{P \cdot L_2} q_i \quad (8)$$

$$\approx \sum_{Q_1} \phi_i^{P \cdot L_1} q_i - \sum_{Q_2} \phi_i^{P \cdot L_2} q_i, \quad (9)$$

where $\phi_i^{P \cdot L_1}$ and $\phi_i^{P \cdot L_2}$ are the calculated electrostatic potentials for the solvated enzyme-substrate complex $P \cdot L_1$ and $P \cdot L_2$, respectively (Fig. 3).

Similarly, the difference of electrostatic energy for two free substrates L_1 and L_2 is:

$$\Delta G_{L_1 \rightarrow L_2}^p = \sum_{L_1} \phi_i^{L_1} q_i - \sum_{L_2} \phi_i^{L_2} q_i \quad (10)$$

$$\approx \sum_{Q_1} \phi_i^{L_1} q_i - \sum_{Q_2} \phi_i^{L_2} q_i, \quad (11)$$

where $\phi_i^{L_1}$ and $\phi_i^{L_2}$ are the calculated electrostatic potentials for the solvated free substrate L_1 and L_2 , respectively (Fig. 3).

The electrostatic potentials of solvated atoms can be calculated as the sum of a Coulomb potential and a "polarization" potential:

$$\phi = \phi^{\text{coulombic}} + \phi^{\text{polarization}}. \quad (12)$$

The Coulomb potential at an atom site i due to other charges is evaluated by Coulomb's law:

$$\phi_i^{\text{coulombic}} = \sum_{j \neq i} \frac{q_j}{r_{i,j}}, \quad (13)$$

where $r_{i,j}$ is interatomic distance between atom i and atom j .

The polarization potential involves electrostatic interactions between the polar atoms and the polarization induced by the polar atoms in the solvent, and can be calculated by solving the Poisson equation:

$$\nabla[\epsilon(r)\nabla\phi(r)] + 4\pi\rho(r) = 0, \quad (14)$$

where $\phi(r)$, $\rho(r)$, and $\epsilon(r)$ are the electrostatic potential, charge density, and dielectric function, respectively. The Poisson equations of a desolvated and solvated molecule were solved numerically, and the differences of the two sets of calculated potentials are the polarization potential (Sharp & Honig, 1990).

Assuming the solution ions follow a Boltzmann distribution around charged solutes, we have:

$$\rho(r) = \rho(r)_{\text{solute}} + \rho_{\text{ion}} \quad (15)$$

$$4\pi\rho_{\text{ion}} = -\bar{\kappa}^2\phi(r) [1 + \phi(r)^2/6 + \phi(r)^4/120 + \dots] \quad (16)$$

$$\bar{\kappa}^2 = \epsilon_0\kappa^2 = (8\pi e^2 N_A I)/(1,000kT) \quad (17)$$

where ϵ_0 , e , N_A , I , k , and T are the vacuum dielectric constant, electron charge, Avogadro's constant, ionic strength, Boltzmann constant, and absolute temperature, respectively. The inverse of κ is the Debye-Huckel screening length. Omitting the higher-order terms in Equation 19, the linear Poisson-Boltzmann equation is:

$$\nabla[\epsilon(r)\nabla\phi(r)] - \bar{\kappa}^2\phi(r) + 4\pi\rho_{\text{solute}}(r) = 0. \quad (18)$$

The above equation can be solved numerically by the finite difference method (Press et al., 1988; Sharp & Honig, 1990). The charge distribution, dielectric constant, and ionic strength functions are all represented by their values at discrete grid points. These discrete functions are obtained by mapping a protein molecule onto a finite box of equally spaced grid points (Sharp & Honig, 1990). A solvent-accessible surface envelope was calculated by rolling a solvent probe of 1.4 Å radius to determine the solute/solvent boundary. The dielectric constants inside and outside this protein envelope were set to 2.0 and 80.0, respectively. The ionic strength of the solvent was set to 0.1 M, corresponding to a Debye-Huckel screening length of 9.98 Å. The electrostatic potential was evaluated in two steps. The Poisson-Boltzmann equation (Equation 19) was first solved on a large, coarse grid box (grid space 1.2 Å). The coarse grid box contains the solute molecule and at least a 15-Å shell of solvent. The potentials at the grid box surface (boundary condition) were set according to the Debye-Huckel formula:

$$\phi = \frac{q}{\epsilon r} \exp(-\kappa r) \quad (19)$$

where q is a point charge and r is the distance from the grid point to a point charge. The coarse grid solution can give an estimate of boundary conditions for a small, fine grid box (Klapper et al., 1986). By successively *focusing* this large, coarse grid onto a small, fine grid, the grid space was reduced to 0.3 Å. The electrostatic potential at the position of atom i is linearly extrapolated from the electrostatic potentials at the surrounding grid points.

The atomic partial charges for heavy atoms and polar hydrogens used in this study were set according to the CHARMM force field (Brooks et al., 1983). The partial charges of substrate analogues (Fig. 2) were derived according to the similar carboxyl (aspartates or glutamates) and hydroxyl amino acids (serines or threonines).

Structural model of enzyme-substrate complex

We used the X-ray crystal structure of the wild-type *E. coli* IDH bound with the isocitrate and the metal ion (Ca^{2+}) (Hurley et al., 1991; Stoddard et al., 1993) as the starting model for all calculations. It was assumed that other substrates bind to IDH in the same way as isocitrate. The missing polar hydrogens in the crystal structure were built using the H-build routine of CHARMM (Brooks et al., 1983). To derive the structural model of a mutant enzyme-substrate complex, we did a limited conformational search starting with the structure of the wild-type enzyme-substrate complex (Fig. 5). Steric overlap and repulsive charge-charge interactions (such as two closely placed positively charged hydrogens, etc.) caused by mutations were relieved by first adjusting the side chain of the mutated residue or the modified substrate. For some enzyme-substrate complexes, no structural adjustments were needed. For example, the γ -carboxyl group of isocitrate was simply removed from the wild-type enzyme-substrate model to generate the model of the wild-type IDH-methylmalate complex. The substitution of the γ -carboxyl group by H creates the structural model for the wild-type IDH-malate complex with an extended atom type (Brooks et al., 1983). Because all the substrate analogues and the site-specific mutations involved structural modifications of no more than four heavy atoms (malate to isocitrate or ethynalmalate were among the biggest structural perturbations considered) or no more than two side-chain torsional angles, this limited conformational search approach succeeds in relieving the steric overlaps.

Fig. 5. Algorithm for calculating the relative binding free energies of enzyme-substrate complexes of *E. coli* IDH. The flowchart shows the stepwise derivation of the enzyme-substrate binding complexes, and the evaluation of the polar and nonpolar contributions to the change of the binding free energies. Details are described in the Methods section.

Observed binding free energy

The binding properties of the IDH Ser 113 mutants versus the eight substrates (Fig. 2) have been determined by kinetic studies (Dean et al., 1996). These structural perturbations include nonpolar, polar, and charged functional groups and therefore allow evaluation of different energetic aspects of the computational method. The Michaelis constant of isocitrate has been shown to be equal to its equilibrium dissociation constant (Dean & Koshland, 1993) and therefore was used as the equilibrium dissociation constant. The experimentally determined relative binding energies of two substrate analogues for a protein were calculated as:

$$\Delta G_{F,PL_1} - \Delta G_{F,PL_2} = -RT \ln \left(\frac{K_1}{K_2} \right),$$

where K_1 and K_2 are the dissociation constants of the two ligands.

An example

The binding of isocitrate and hydroxymethylmalate to wild-type IDH provides an example of how the relative binding free energies were calculated.

Protein = P = wild-type IDH

Ligand 1 = L_1 = isocitrate

Ligand 2 = L_2 = hydroxymethylmalate.

We define Q_c as the common part of the two substrates, which is the malate moiety of isocitrate in this case (Fig. 3). Therefore,

According to Equation 10,

$$\begin{aligned} \Delta G_{F,P \cdot L_1 - P \cdot L_2}^{np} &= 20.0 \text{ cal}/\text{\AA}^2 (-82.4 \text{ \AA}^2) \\ &\quad - 20.0 \text{ cal}/\text{\AA}^2 (-61.2 \text{ \AA}^2) \\ &\approx -0.4 \text{ kcal.} \end{aligned}$$

According to Equations 12 and 14,

$$\Delta G_{F,P \cdot L_1 - P \cdot L_2}^p = -1.1 - (-0.2) = -0.9 \text{ kcal/mol.}$$

Therefore,

$$\Delta G_{F,P \cdot L_1 - P \cdot L_2} = -0.4 + (-0.9) = -1.3 \text{ kcal/mol.}$$

Acknowledgments

This work was supported by the National Science Foundation and the Materials Science Division, U.S. Department of Energy, under contract DE-AC-76SF00098. In addition, T.Z. was supported by an NIH post-doctoral fellowship in the General Medical Sciences.

References

- Bash P, Singh U, Langridge R, Kollman P. 1987. Free energy calculations by computer simulation. *Science* 236:564.
- Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. 1983. Charmm: A program for macromolecular energy minimization and molecular dynamics calculations. *J Comput Chem* 4:187-217.
- Chothia C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105:1.
- Damodaran S, Song K. 1986. The role of solvent polarity in the free energy of transfer of amino acid side chains from water to organic solvents. *J Biol Chem* 261:7220.
- Dean A, Koshland DE Jr. 1990. Electrostatic and steric contributions to regulation at the active site of isocitrate dehydrogenase. *Science* 249:1044-1046.
- Dean A, Koshland DE Jr. 1993. Kinetic mechanism of *Escherichia coli* isocitrate dehydrogenase. *Biochemistry* 32:9302-9309.
- Dean AM, Shiau AK, Koshland DE Jr. 1996. Determinants of performance in the isocitrate dehydrogenase of *Escherichia coli*. *Protein Sci* 5:341-347.
- Debolt S, Pearlman D, Kollman P. 1994. Free energy perturbation calculations on parallel computers—Demonstrations of scalable linear speedup. *J Comput Chem* 15:351.
- Eisenberg D, McLachlan A. 1986. Solvation energy in protein folding and binding. *Nature* 319:199.
- Eriksson A, Baase W, Zhang X, Heinz D, Blaber B, Baldwin B, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255:178.
- Fauchere J, Pliska V. 1983. *Eur J Med Chem Chim Ther* 18:369.
- Ferguson D, Radmer R, Kollman P. 1991. Determination of the relative binding free energies of peptide inhibitors to the HIV-1 protease. *J Medicinal Chem* 34:2654.
- Gilson M, Sharp K, Honig B. 1987. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J Comput Chem* 9:327.
- Hill T. 1958. Influence of electrolyte on effective dielectric constants in enzymes, proteins and other molecules. *J Phys Chem* 60:253.
- Hurley JH, Dean AM, Koshland DE Jr, Stroud RM. 1991. Catalytic mechanism of NADP⁺-dependent isocitrate dehydrogenase: Implications from the structures of magnesium-isocitrate and NADP⁺ complexes. *Biochemistry* 30:8671-8678.
- Hurley JH, Dean AM, Stohl JL, Koshland DE Jr, Stroud RM. 1990. Regulation of an enzyme by phosphorylation at the active site. *Science* 249:1012-1016.
- Kirkwood J. 1934. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J Chem Phys* 2:351-361.
- Kirkwood J, Westheimer F. 1938. The electrostatic influence of substituents on the dissociation constants of organic acids. I. *J Chem Phys* 6:506-512.
- Klapper I, Hagstrom R, Fine R, Sharp K, Honig B. 1986. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins Struct Funct Genet* 1:47.
- LaPorte D, Koshland DE Jr. 1982. A protein with kinase and phosphatase activities involved in regulation of tricarboxylic acid cycle. *Nature* 300:458.
- Lee B, Richards F. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 55:379.
- McCammon J, Harvey S. 1987. *Dynamics of proteins and nucleic acids*. New York: Cambridge University Press.
- Novotny J, Bruccoleri R, Saul F. 1989. On the attribution of binding energy in antigen-antibody complexes MCPC 603, D1.3, and HYHEL-5. *Biochemistry* 28:4735.
- Nozaki Y, Tanford C. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J Biol Chem* 246:2211-2217.
- Ponder J, Richards F. 1987. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193:775.
- Press W, Teukolsky S, Vetterling W, Flannery B. 1988. *Numerical recipes in C*. New York: Cambridge University Press.
- Sharp K, Honig B. 1990. Electrostatic interactions in macromolecules—Theory and applications. *Annu Rev Biophys Biophys Chem* 19:301.
- Simonson T, Brunger A. 1992. Thermodynamics of protein-peptide interactions in the ribonuclease-S system studied by molecular dynamics and free energy calculations. *Biochemistry* 31:8661.
- Sitkoff D, Sharp K, Honig B. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 98:1978.
- Stoddard BL, Dean AM, Koshland DE Jr. 1993. Structure of isocitrate dehydrogenase with isocitrate, nicotinamide adenine dinucleotide phosphate, and calcium at 2.5-Å resolution: A pseudo-Michaelis ternary complex. *Biochemistry* 32:9310-9316.
- Thorsness PE, Koshland DE Jr. 1987. Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate. *J Biol Chem* 262:10422-10425.
- Tidor B, Karplus M. 1991. Simulation analysis of the stability mutant R96H of T4 lysozyme. *Biochemistry* 30:3217.
- Vajda S, Weng Z, Rosenfeld R, DeLisi C. 1994. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry* 33:13977.
- Warshel A, Sussman F, Hwang J. 1988. Evaluation of catalytic free energies in genetically modified proteins. *J Mol Biol* 201:139.
- Wilson C, Mace J, Agard D. 1991. Computational method for the design of enzymes with altered substrate specificity. *J Mol Biol* 220:495.