ACS Catalysis

Research Article

pubs.acs.org/acscatalysis

Relative Enzymatic Activity Levels from In Silico Mutagenesis.

Caroline Mellot-Draznieks,[†,⊥] Vassili Valayannopoulos,[‡,§] Dominique Chrétien,[§] Arnold Munnich,[‡] Pascale de Lonlay,[‡,§] and Hervé Toulhoat*[,†]

[†]IFP Energies nouvelles, 1&4 Avenue de Bois-Préau, 92852, Rueil-Malmaison Cedex, France

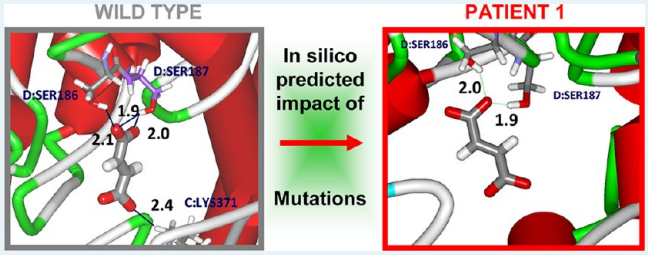
[‡]Centre de Référence des Maladies Héréditaires du Métabolisme, Université de Paris Descartes, Hôpital des Enfants-Malades, Paris, France

[§]Unité INSERM U-781, Hôpital des Enfants-Malades, Paris, France

[⊥]CNRS, Laboratoire de Chimie des Processus Biologiques, Collège de France, Paris, France

Supporting Information

ABSTRACT: We propose a new computational approach for predicting the impact of point mutations on residual enzymatic activities. We build on the general linear trends existing between free energy and enthalpy of transfer of substrates from cytosol to enzyme active sites (protein−ligand binding), therefore linking the docking energies to the binding free energies. In this very first step, we rationalize these trends in terms of a compensation effect decomposed into explicit thermodynamics contributions. In a second step, we combine the latter with the assumption that free energies of transfer, estimated from docking, and free energies of activation are linearly related through a Brönsted−Evans−Polanyi (BEP) relationship, allowing us in fine to predict enzyme activity. As a result, we propose generic Langmuir−Hinshelwood kinetic equations "trained" on the wild type, which provide excellent predictions of rates of catalytic transformations for mutated enzymes from the combination of in silico docking energies to a set of system-specific experimental data. This generalized approach is validated against clinical data on the particular case of human fumarase with major implications for the understanding of hereditary fumarase deficiency.



KEYWORDS: *enthalpy−entropy compensation, protein−ligand binding, impact of mutations on enzyme activity, linear free energy relationship, molecular simulation, BEP relationship, Langmuir−Hinshelwood kinetics, competitive Michaelis−Menten kinetics*

INTRODUCTION

A long-standing goal in biochemistry is the prediction of the thermodynamics of protein−ligand complexes, fuelled by the incentives of structure-based drug design. The availability of ever more high-resolution structural data for functional proteins is currently strongly stimulating efforts to provide insights on enzymatic mechanisms at the molecular scale on the basis of biomolecular simulation studies. There are huge stakes for advances in basic life science and applications in medicine, biotechnologies, or biomimetic chemistry. From the standpoint of physical chemistry, the common challenge is our ability to describe accurately the most likely reaction pathway connecting reactants and products via an enzymatic catalytic "living" site, that is, the pathway with the minimal Gibbs free energy profile. Metabolites undergoing enzymatic transformations in vivo experience complex molecular environments along such a pathway, being transferred typically from an aqueous cytosol to a "site", or specific functional pocket of a protein, evolutionarily determined and encoded in both a relevant sequence of amino acids and the constraints of primary to quaternary spatial structures. Currently, the protein folding problem cannot be solved computationally; however, combining the knowledge of relevant aminoacid sequences, experimental high-resolution XRD and NMR 3D structural data for the relevant proteins, and a reasonable representation of the cytoplasmic solution, it is possible to build in silico atomistic models trying to catch the essential features of inter- and intramolecular interactions at play in the dynamical processes of enzymatic transfers and reactions.

It is important to underline here that one key issue in enzymatic catalysis is the ability of current computational approaches to provide quantitative routes for calculating rate constants of a given reaction using the knowledge of the enzyme crystal structure.[1] An ongoing challenge lies in the development of validated methods for evaluating activation free energies. To that end, the most sophisticated and computationally intensive approach to date involves transition path sampling techniques[2−7] based on DFT-based first-principles

Received: August 9, 2012
Revised: September 27, 2012
Published: October 26, 2012

molecular dynamics simulations.[8] Both to allow calculations with larger atomistic models and in view of the known limitations of the usual exchange-correlation functionals to account for dispersion energy expected of great significance in this context, hybrid QM/MM schemes are among the best compromises between accuracy and tractability.[9] A number of recent studies including empirical valence bond[10−12] or QM/MM methods[13] have demonstrated the capability of quantitative approaches to predict the effect of mutations on enzyme catalysis.

Another extensively used approach consists in computational (or in silico) docking techniques. These have been developed mostly for the purpose of medicinal chemistry, following the empirical concept that drugs generally act as inhibitors or competitors of natural metabolites for specific enzymatic sites. Since the transfer of substrates between cytosols and enzyme sites is driven mostly by electrostatic and nonbinding interactions, empirical forcefields have been expected to achieve successful prediction of relative Michaelis constants for homologous series[14−17] and, therefore, to be useful in the search for new "leads" and optimal analogues. Docking energies are, however, at best "proxies" for enthalpies of transfer, and Michaelis constants actually reflect free energies of transfer. To correlate docking results and experimental Michaelis constants into useful quantitative structure property relationships, a large number of "scoring functions" have been proposed in the literature,[18,19] with a more or less rational background and diverse success. In that tremendously active context, we propose a novel computational approach for predicting the impact of point mutations on residual enzymatic catalytic activities.

In a first step, we build on the general linear trends existing between free energy and enthalpy of complexation upon transfer of substrates from cytosol to enzyme active sites and rationalize these trends in terms of the so-called entropy−enthalpy compensation effect.[20−26] This effect is clearly to be distinguished from the so-called "constable effect" in heterogeneous catalysis, also called "compensation effect" by some authors in the latter field, which denotes the observation of a linear relationship between the apparent energy of activation and the apparent prefactor of the Arrhenian rate equation for a series of related catalysts, tested under the same operating conditions (for recent discussions and explanations, see refs 27 and 28). We build on recent isothermal titration calorimetric measurements that report such a linear correlation between the free energy of transfer of a series of substrates to protein hosts and their enthalpy of transfer. The isothermal titration calorimetry (ITC) technique[29] was, indeed, instrumental in the systematic collection of such data in vitro for numerous protein−ligand systems with a couple of public databases, such as BindingDB and SCORPIO available online.[30,31] Importantly, we propose here to detail the enthalpic and entropic contributions to the transfer into explicit terms, taking into account the desolvation effect and including terms accessible from in silico docking protocols. By separating generic and protein specific contributions, we thus develop an interpretation of the entropy−enthalpy compensation effect.

In a second step, we combine the above rationale with the assumption that free energies of transfer (as estimated from docking) and free energies of activation are linearly related through a Brönsted−Evans−Polanyi (BEP) relationship,[32,33] equivalent in heterogeneous catalysis to the linear free energy relationship (LFER) widely used in protein−ligand complexation. They both offer empirical routes to estimate kinetic parameters from thermodynamic values. BEP relationships have found broad application in heterogeneous catalysis[34] and were successfully documented for more and more diverse systems, thanks to the combination of experimental and DFT studies;[35] however, their use in enzymatic catalysis has been only recently considered.[36] We demonstrate here that the Brönsted−Evans−Polanyi relationship may be transposed from the field of heterogeneous catalysis to the study of mutated protein−ligand systems, that is, that free energies of activation may be correlated for a given reaction catalyzed by mutated variants of a given protein, with the free energy difference between reactant and product in complexes themselves correlated to docking energies in mutated enzymes in a systematic and consistent way.

Finally, relying on the transition state theory together with the well accepted Michaelis−Menten kinetic equations (rather than the Michaelis−Menten formalism, we use their more general expressions known as the Langmuir−Hinshelwood equations in the context of heterogeneous catalysis[37]), we are then in a position to predict enzymatic activities on the sole basis of in silico docking studies. As a result, we propose generic kinetic equations "trained" on the wild type enzyme, which provide further excellent predictions of rates of catalytic transformations for mutated enzymes from the sole input of in silico docking energies.

We test our model in the complex case of point mutations affecting the function of human fumarase in the stereospecific, reversible hydration of fumarate into L-malate, participating in the Krebs cycle within mitochondria, and in the regulation of cytosolic fumarate levels. Fumarate hydratase deficiency is a rare inborn error in the metabolism inherited in an autosomal recessive way, resulting in the accumulation of fumarate and a severe metabolic disorder.[38]

More specifically, we use here in silico docking calculations of fumarate and L-malate substrates in wild type and mutated fumarases to estimate the corresponding protein−ligand complexation enthalpies. The excellent correlation we obtain between in silico enthalpies of complexation for a series of fumarate derivatives in wild type fumarase and experimental Gibbs free energies supports the above framework. The latter is applied to extrapolate Gibbs free energies of complexation of substrates in mutated enzymes. On the basis of these correlations, we further show how our in silico estimates of Gibbs free energies of complexation may be incorporated into classical kinetic equations with competitive inhibition, allowing us to predict the residual catalytic activities of mutated fumarase in reference to the wild type, in remarkable consistency with experimental findings.

The paper is organized as follows: our model and its theoretical background are presented in the Theory section; the Methods section presents both the details of our in silico calculations and the essential experimental characteristics of the test case; in the Results and Discussion section, the computational and experimental results are compared, and the domain of validity of the approach is discussed, allowing one to draw the final conclusions.

THEORY

Theoretical Frame for the Enthalpy−Entropy Compensation Effect. It is widely acknowledged how favorable thermodynamic changes in binding enthalpies or entropies are attenuated by the so-called enthalpy−entropy compensation, a

ubiquitous, although still not fully understood, phenomenon in biomolecular interactions.[20−26] The enthalpy−entropy compensation effect describes the occurrence of a linear relationship between enthalpy and entropy changes associated with the same change in free energy (Figure 1). This compensation

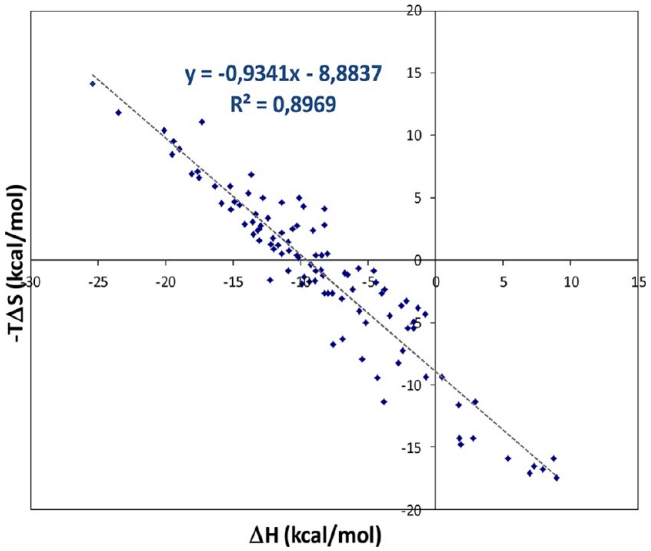


Figure 1. Plot of the entropic versus the enthalpic contribution of protein−ligand interactions measured by ITC for a set of 100 protein−ligand complexes extracted from the BindingDB database.[31]

effect is typically understood in physical adsorption phenomena by a reduction of the degrees of freedom of the "guest" molecule (i.e., a loss of entropy, negative ΔS) in its "host" adsorption site associated with a gain in binding energy (gain of enthalpy; i.e., ΔH negative), the variation in entropy and enthalpy, ΔS and ΔH sharing the same sign. A correlation was recently proposed among molecular confinement, compensation effect, and catalytic activity in the context of adsorption of molecular substrates inside porous crystalline catalysts.[39]

In the field of protein−ligand complexes, no established explicit expression between enthalpic and entropic contributions has been proposed from fundamental thermodynamics so far, although the enthalpy−entropy compensation effect has been universally observed in protein−ligand complexation or other host−guest systems,[20−26,40−44] leaving its physical origin still under controversy. It is worth noting that the utility of high-quality ITC structure−thermodynamic databases on the thermodynamic contributions of enthalpy and entropy changes to free energies of complexation has been highlighted in the context of ligand design, considered as essential in the clarification of compensation effects and structure−thermodynamics correlations.[45]

The empirical linear correlation between entropy and enthalpy of complexation in protein−ligand complexes (Figure 1) may be expressed as follows:

$$-T\Delta S = a\Delta H + b \tag{1}$$

where ΔS represents the total entropy of transfer of the substrate from the aqueous cytosol to the protein site, ΔH represents the total enthalpy of transfer for the same process, T is the temperature, and a and b are constants characteristic of the protein but independent of the substrate.[30,31,45] From these studies, it appears that a is quasiuniversal and approximated to −0.93 (Figure 1) and b varies significantly with the considered protein in the approximate range from −14 to −3 kcal mol^{-1}. In the following, we develop a formal analysis of enthalpy−entropy compensation and develop a strategy to make a proper use of the enthalpic information derived from docking protocols applied to atomistic models of protein−ligands interactions.

We introduce here our central assumptions regarding the compensation effect. We rationalize such a correlation by expressing the respective contributions of enthalpy and entropy in the substrate−enzyme complex and by taking solvation/desolvation effects into account. Primarily, the *enthalpy of transfer* of the substrate (S) from the aqueous cytosol (W) to the protein (P) is defined as the difference of affinity for the substrate between the protein and the solution:

$$\Delta H^{S} = \langle A_P^S \rangle - \langle A_W^S \rangle \quad \text{or} \quad \Delta H^{W} = \langle A_P^W \rangle - \langle A_W^W \rangle \tag{2}$$

where $\langle A_P^{S/W} \rangle$ and $\langle A_W^{S/W} \rangle$ may be calculated from in silico docking calculations and represent the averaged affinities of the substrate or water (S, W as uppercase) for the protein active site (P as an index) and for the cytosol (W as an index), respectively, over the docked poses. Affinities are here defined for a single effect as minus the docking energies of the substrate in complex, which themselves are taken as the sum of intermolecular contributions to the total energy given by the considered forcefield (CHARMm).

In this perspective, we further assume that the transfer of one substrate molecule from solution to the protein active site involves the retrotransfer of n molecules of water back into the solution (desolvation), n corresponding to the number of molecules required to fill the space occupied by the substrate in solution.[46] The *total enthalpy change* involved in the formation of the protein−ligand complex (enthalpy of complexation) is then decomposed as:

$$\Delta H = \Delta H^{S} - n\Delta H^{W} + \Delta H_{conf}^{P} \tag{3}$$

where ΔH_{conf}^{P} represents the conformational enthalpic change of the protein upon complexation. We consistently assumed that this term is substrate-independent, considering that the substrates' sizes and chemical functions studied here are very similar. Similarly, the total entropy is expressed in its three components that represent the entropic change related to the substrate, the water, and the protein, respectively:

$$-T\Delta S = -T\Delta S^{S} + nT\Delta S^{W} - T\Delta S_{conf}^{P} \tag{4}$$

We consider that the larger the substrate, the larger the number of water molecules displaced (noted n). That is precisely the effect we attempt to model with eqs 3 and 4 by splitting the total enthalpy and entropy of transfer into n dependent and n independent contributions: ΔH^{S} and $-T\Delta S^{S}$ are thermodynamic contributions to enthalpic and entropic changes assigned to the substrate for a given solvent, which mostly reflect the changes of specific interactions (e.g., hydrogen bonds, local electrostatic interactions...) between the substrate surface and its surroundings. Because these interactions are of a short-range nature, it is legitimate to consider that the free energy of transfer of a given substrate from the cytosol to the protein site can be split into this surface energetic contribution and a volume-related energetic contribution that is dominated by the replacement of this volume by water molecules (interacting mostly through the network of

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

hydrogen bonds inside the cytosol) initially filling the protein site (retrotransfer). We thus justify our other assumption that ΔH^{P}_{conf} and $-T\Delta S^{P}_{conf}$, the contributions of the protein conformational change upon transfer, are substrate-independent, provided the substrates are not too different in size and chemical functions. These assumptions are, indeed, the basis of our model, which was motivated by the need to interpret the phenomenon illustrated in Figure 1; namely, that in eq 1, a is almost universal (substrate and protein independent) and b is substrate-independent and (weakly) protein-dependent. In the Supporting Information, we give more support to these observations.

Substituting eqs 3 and 4 into 1, we find an n-dependent linear equation:

$$(a\Delta H^{W} + T\Delta S^{W})n + (-T\Delta S^{S} - T\Delta S^{P}_{conf} - b - a(\Delta H^{S} + \Delta H^{P}_{conf})) = 0 \quad (5)$$

that allows determination of both coefficients, a and b, observed in the empirical entropy−enthalpy correlation. The generality of eq 5 for any value of n imposes that both coefficients must be equal to zero.

The slope a may be expressed in general terms as follows:

$$a = \frac{-T\Delta S^{W}}{\Delta H^{W}} \quad (6)$$

and the intercept, b, of the entropy−enthalpy plot with the y axis for a given protein as:

$$b = -T\Delta S^{S} - T\Delta S^{P}_{conf} + \frac{T\Delta S^{W}}{\Delta H^{W}}(\Delta H^{S} + \Delta H^{P}_{conf}) \quad (7)$$

It is then apparent that both constants, a and b, may be estimated by considering the specific case of a single water molecule as a substrate (S = W). Importantly, it results that the constant b is related *exclusively* to the conformational energy changes of the protein upon complexation,

$$b = -T\Delta S^{P}_{conf} - a\Delta H^{P}_{conf} = \Delta G^{P}_{conf} - (1 + a)\Delta H^{P}_{conf} \quad (8)$$

allowing a further simplification of 5 into:

$$(a\Delta H^{W} + T\Delta S^{W})n - T\Delta S^{S} - a\Delta H^{S} = 0 \quad (9)$$

The slope of the entropy−enthalpy correlation, a, for any substrate other than water ($n \neq 1$) may be then written using the above eq 9 as follows:

$$a = \frac{-T\Delta S^{S} + nT\Delta S^{W}}{\Delta H^{S} - n\Delta H^{W}} \quad (10)$$

The elimination of a between eqs 6 and 10 gives:

$$\frac{-T\Delta S^{S}}{\Delta H^{S}} = \frac{-T\Delta S^{W}}{\Delta H^{W}} = a \quad (11)$$

Equations 6 and 11 may be also expressed respectively as:

$$\Delta G^{W} = (1 + a)\Delta H^{W} \quad (12)$$

$$\Delta G^{S} = (1 + a)\Delta H^{S} \quad (13)$$

Given that a is close to −1, ($a \cong -0.934$ according to data available in ITC databases; see Figure 1) eqs 6/11 and 12/13 show that, in general, the free energy change associated with the transfer (of water as well as any other substrate) from cytosol to a protein site is very low compared with the enthalpy change and, therefore, compensated by the entropy change. If a were exactly equal to −1, any protein would lose its selectivity (or substrate specificity) for competing substrates. The total free energy of transfer would be, according to eq 8, equal to the protein-dependent b constant and determined by the conformational change of the protein upon exchange of water with cytosol. One can therefore view b as a measure of the protein hydrophilicity and a as a measure of the protein ability to select substrates. Proteins have evolved as hydrophilic macromolecules ($b < 0$) capable of selectivity bind substrates and generally resistant to substrate overbinding, thanks to the enthalpy−entropy compensation effect. The latter is determined by the properties both of water as a hydrogen-binding solvent and of proteins as hydrogen-binding receptors. This can be viewed as proteins as "living" molecular catalysts "exploiting" the compensation effect to comply to the principle of Sabatier.

Incidentally, assuming that entropy and enthalpy changes upon transfer are almost independent of temperature around ambient conditions, eq 6 implies that a is strictly proportional to T. This allows determining the compensation temperature for which the free energy change associated with the transfer of water would vanish for any protein ($a(T_{comp}) = -1$), i.e. $T_{comp} \cong 319.2$ K ($\cong 46$ °C). At this temperature, similarly to the discussion above, the intercept, b, of the $-T\Delta S = f(\Delta H)$ plot of any substrate−protein complex expresses the free energy of transfer of the substrate from solution to the protein as reduced to the conformational contribution, as shown by eq 8. One then expects serious physiological perturbations for most organisms because any given protein would lose its selectivity (or substrate specificity) for competing substrates. It is likely (however, still to be checked) that proteins from thermophilic organisms may present higher compensation temperatures (i.e., would lie on other correlations with slopes significantly different from above).

Still, it is apparent from eq 8 that the intercept, b, of the $-T\Delta S = f(\Delta H)$ plot depends on the protein and might therefore show specific variations upon mutations. In a further attempt to capture the impact of mutations on b, we consider one mutation (mut) and the reference wild type (wt) noted as exponents, and express the universal slope, a, from eq 6:

$$a = \frac{-T\Delta S^{W}_{wt}}{\Delta H^{W}_{wt}} = \frac{-T\Delta S^{W}_{mut}}{\Delta H^{W}_{mut}} \quad \text{or}$$

$$-T\Delta S^{W}_{mut} + T\Delta S^{W}_{wt} = -T\Delta\Delta S^{W} = a(\Delta H^{W}_{mut} - \Delta H^{W}_{wt}) \quad (14)$$

Our second strong hypothesis is that upon a mutation, the change of conformational enthalpy is negligible, whereas the change in conformational entropy is equal to (or exactly absorbs) the difference in variation of entropy experienced by a single water molecule upon transfer from cytosol to the catalytic site. Using eqs 8 and 14, we can now express the impact of a mutation on the intercept, Δb, as follows:

$$\Delta b = b^{mut} - b^{wt} = -T\Delta\Delta S^{P}_{conf} = -T\Delta\Delta S^{W} = a(\Delta H^{W}_{mut} - \Delta H^{W}_{wt}) \quad (15)$$

Kinetic Models for Fumarase Enzymatic Activity. The hydration of fumarate into L-malate is a stereospecific bimolecular reaction, equilibrated by its reverse reaction, the dehydration of L-malate:

fumarate + H_2O ⇔ L-malate

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

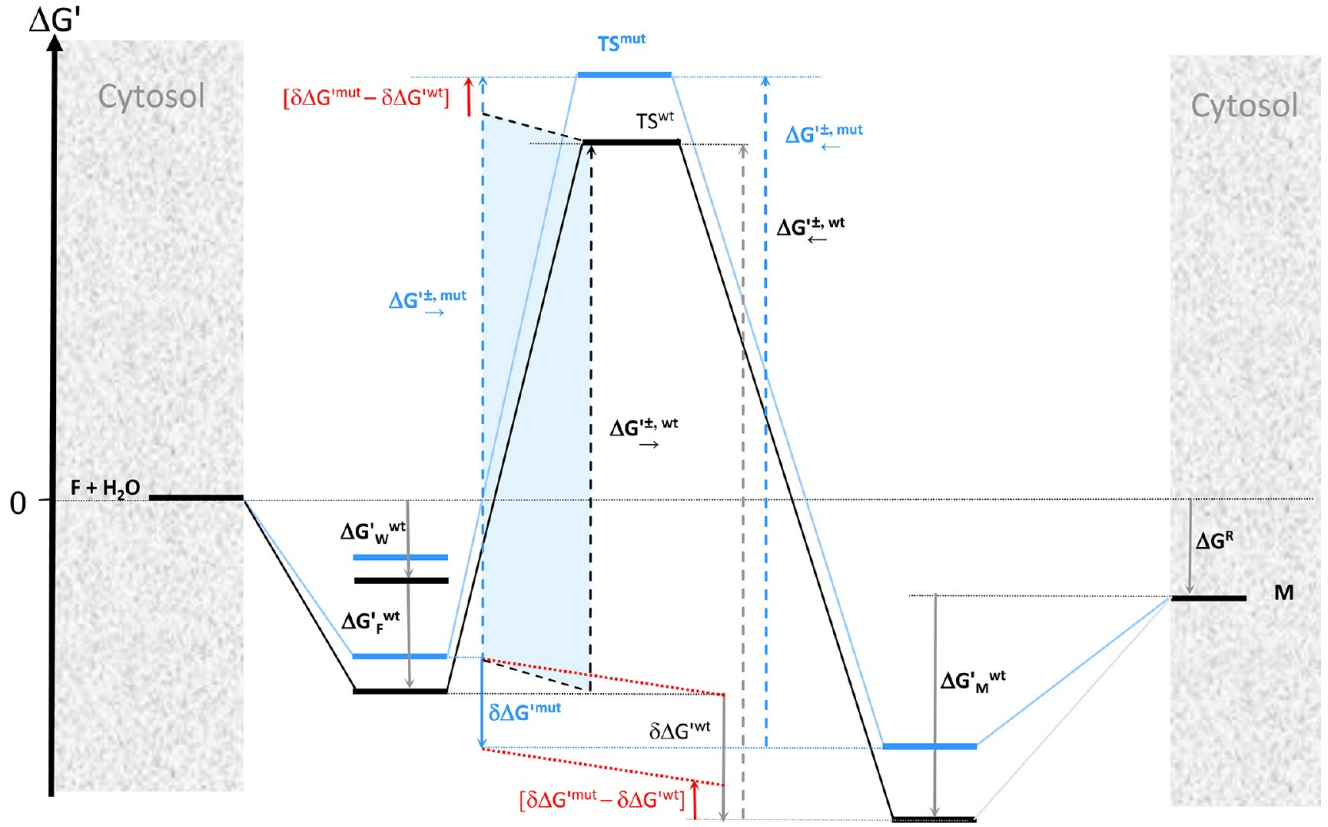


Figure 2. Schematic representation of the free energy diagram of fumarate hydrolysis into L-malate. Energy levels in the wild type and in the mutated proteins are represented in black and blue, respectively.The free energy terms are defined as follows, with respect to the reference state: $\Delta G'^{mut} = \Delta G^{mut} + [\Delta G^{mut} - \Delta G^{wt}]$, expressed for fumarate, malate, and water. From the BEP relationship: $\Delta G'^{\pm,wt}_{\rightarrow} = \lambda\delta\Delta G'^{wt} + \mu$ and $\Delta G'^{\pm,mut}_{\rightarrow} = \lambda\delta\Delta G'^{mut} + \mu$ so that by subtraction, we obtain $\Delta G'^{\pm,mut}_{\rightarrow} = \Delta G'^{\pm,wt}_{\rightarrow} + \lambda[\delta\Delta G'^{mut} - \delta\Delta G'^{wt}]$ (in the present case, $\lambda = 1$), with $\Delta G'^{\pm,mut}_{\leftarrow} = \Delta G'^{\pm,mut}_{\rightarrow} + \delta\Delta G'^{mut}$.

It is an important step in the Krebs cycle, catalyzed in particular by human mitochondrial fumarase (EC 4.2.1.2). According to previous studies,[47−50] the kinetics of these reactions can be described by a classical competitive Michaelis−Menten model in which fumarate and L-malate compete for the same site, which catalyzes the direct and reverse reactions. Water supply is implicitly assumed not limiting. Fumarase activity is strongly affected by pH, with a maximum around pH = 6.5 for the direct reaction and around 7.5 for the reverse reaction at 25 °C.[47] This effect has been successfully accounted for by the Haldane model.[51] At a fixed pH, the competitive Michaelis−Menten interpretation provides apparent (pH-dependent) K_M and V_{max}. A catalytic mechanism and a sophisticated 11-state model estimating kinetic parameters were recently proposed.[52]

In what follows, we propose to describe the kinetics by the more general Langmuir−Hinshelwood model, the correspondence of which with the Michaelis−Menten approach can be easily established (see the Supporting Information). For the time being, we do not account for the effect of pH. We make the assumption of the existence of two distinct "sites", 1 and 2, in the fumarate hydration process: site 1 specifically binds fumarate, the transition state, and L-malate, and site 2, in close proximity to site 1, specifically binds water. We justify this assumption below on the basis of our molecular simulations.

Transition State Theory. As mentioned above, fumarate and L-malate are assumed to compete for a single active site (site 1) at the protein, the amount of which is fixed. According to the Eyring transition state theory, the rate of fumarate hydration and L-malate dehydration, expressed in molecules consumed per site and per second, may be written generically as:

$$\vec{r} = \frac{k_B T}{h}\exp\left(\frac{-\Delta G^{\pm}_{\rightarrow}}{RT}\right)\theta_{2,W}\theta_{1,F} \quad (16)$$

$$\overleftarrow{r} = \frac{k_B T}{h}\exp\left(\frac{-\Delta G^{\pm}_{\leftarrow}}{RT}\right)\theta_{1,M} \quad (17)$$

where k_B and h stand for the Boltzmann and Planck's constants, respectively; R, for the ideal gas constant; and $\Delta G^{\pm}_{\rightarrow}$ and $\Delta G^{\pm}_{\leftarrow}$ are the temperature- and protein-dependent intrinsic free energies of activation for the hydration and dehydration reactions referred to reactants and product in the complex, respectively.

The fractional coverages $\theta_{1,F}$, $\theta_{1,M}$, and $\theta_{2,W}$ in fumarate, L-malate, and water, respectively, are proportional to the corresponding fraction of free sites ($\theta_{1,*}$, $\theta_{2,*}$), according to the law of mass action, expressed here for the wild type enzyme:

$$\theta_{1,F} = \theta_{1,*}\frac{C_F}{C^{wt}_{F,1/2}}\exp\left(\frac{-\Delta G^{wt}_F}{RT}\right) \quad (18)$$

where C_F and $C^{wt}_{F,1/2}$ represent the concentration of fumarate in the cytosol at equilibrium in the wild type fumarase and the concentration of fumarate in cytosol for which half of sites 1 are occupied by fumarate (protein dependent), respectively, and

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

ΔG_F^{wt} is now the free energy of transfer of fumarate from the cytosol to site 1 in the wild type. Similarly, we may write:

$$\theta_{1,M} = \theta_{1,*}\frac{C_M}{C_{M,1/2}^{wt}}\exp\left(\frac{-\Delta G_M^{wt}}{RT}\right) \quad (19)$$

$$\theta_{2,W} = \theta_{2,*}\frac{C_W}{C_{W,1/2}^{wt}}\exp\left(\frac{-\Delta G_W^{wt}}{RT}\right) \quad (20)$$

where $\theta_{1,F}$ and $\theta_{1,M}$ are the fractional coverages of sites 1 by fumarate and L-malate respectively; $\theta_{2,W}$, the fractional coverages of sites 2 by water; and $\theta_{1,*}$ and $\theta_{2,*}$ are the respective fractions of free sites. These relationships are valid for the wild type as well as the mutated enzymes, provided the free energies are appropriately referred, as discussed below. In what follows, we adopt the convention that $C_{F,1/2}^{wt} = C_{M,1/2}^{wt} = 1$ mol L^{-1} for the wild type protein. In other terms, we refer the free energies of transfer of solutes from cytosol to protein sites to the states of half coverage of sites. As a consequence:

$$C_{F,1/2}^{mut} = C_{F,1/2}^{wt}\exp\left(\frac{+\Delta G_F^{wt} - \Delta G_F^{mut}}{RT}\right)\text{mol L}^{-1} \quad (21)$$

where mut refers to a mutant, and wt to the wild type protein. Similar equations stand for water and L-malate (see the Supporting Information (SI)). The balance equations for sites 1 and 2 give:

$$\theta_{1,F} + \theta_{1,M} + \theta_{1,*} = 1 \quad \text{and} \quad \theta_{2,W} + \theta_{2,*} = 1 \quad (22)$$

Eliminating $\theta_{1,*}$ and $\theta_{2,*}$, we finally obtain, for any mutant, master rate equations (see the full expressions of rates in the SI) dependent on a set of parameters and variables listed as:

$$\vec{r}^{mut}, \overleftarrow{r}^{mut} = f(k_B, T, h, \Delta G_F^{wt}, \Delta G_M^{wt}, \Delta G_W^{wt}, \Delta G_F^{mut}, \Delta G_M^{mut}, \Delta G_W^{mut}\Delta G_{\rightarrow}^{\pm,mut}, C_F, C_M, C_W, C_{F,1/2}^{wt}, C_{M,1/2}^{wt}, C_{W,1/2}^{wt}) \quad (23)$$

The corresponding expressions for the wild type are the same, deduced by replacing mut by wt and simplifying.

BEP and Rationale for the Impact of Mutations. Linear free energy relationships imply that changes in activation free energies are linearly correlated with concomitant changes in free energy differences between reactant and product, that is, free energy of reaction (see for instance eq 1 in ref 53).[53,54] Theoretical studies have lent them strong support, and their validity is now widely accepted because many reactions follow such relationships.[55] LFERs are thus extensively used in enzyme catalysis to correlate reaction rates with the corresponding equilibrium constants and develop quantitative approaches.[53,54] However, for a given reaction catalyzed by related but different enzymes with variable activity, for example, wild type and mutated, because the free energy of reaction does not change but reaction barriers may change, the only way a LFER may hold is if the barrier correlates with the free energy difference between reactant and product in complexes. This has been recognized for a long time in the context of heterogeneous catalysis,[32−35] in which LFERs are called Brönsted−Evans−Polanyi relationships.

We assume the general validity of a BEP linear relationship between the free energy of activation in the forward direction and the free energy difference $\delta\Delta G$ between the reactants in the complex and the product in the complex (Figure 2):

$$\Delta G_{\rightarrow}^{\pm} = \lambda(\delta\Delta G) + \mu \quad (24)$$

where λ and μ do not depend on the mutations of the protein.

Using the above BEP concept, Figure 2 gives a schematic representation of the free energy diagram of fumarate hydrolysis into L-malate, summarizing the various thermodynamic and kinetic components of both the wild type and mutated enzymes. For a given protein (wild type or mutated), introducing ΔG_R as the free energy difference associated with the hydration reaction of fumarate into L-malate in aqueous solution, we can write, properly taking into account the various free energy references:

$$\delta\Delta G^{mut} = (\Delta G_R - \Delta G_M^{wt} + 2\Delta G_M^{mut}) - (-\Delta G_W^{wt} + 2\Delta G_W^{mut} - \Delta G_F^{wt} + 2\Delta G_F^{mut}) \quad (25)$$

Finally, as illustrated in Figure 2, the consistency requirement for the free energy profile connecting fumarate in solution to L-malate in solution along the reaction pathway involving the reactants and common transition state in complex is expressed as:

$$\Delta G_{\leftarrow}^{\pm,mut} = \Delta G_{\rightarrow}^{\pm,mut} + \delta\Delta G^{mut} \quad (26)$$

Combining eq 24 for mutated and wild type proteins, we deduce:

$$\Delta G_{\rightarrow}^{\pm,mut} = \Delta G_{\rightarrow}^{\pm,wt} + \lambda[\delta\Delta G^{mut} - \delta\Delta G^{wt}] \quad (27)$$

Combining now eqs 26 and 27 for mutated and wild type proteins, we obtain:

$$\Delta G_{\leftarrow}^{\pm,mut} = \Delta G_{\leftarrow}^{\pm,wt} + (\lambda - 1)[\delta\Delta G^{mut} - \delta\Delta G^{wt}] \quad (28)$$

Importantly, the BEP relationship expresses the fact that the conformation of the transition state and of the product in complex would be closely related and merely translated by similar amounts on the free energy scale as a consequence of point mutations expressed in the host enzyme.

In summary, eqs 1−15 allow us to estimate ΔG_F^{mut}, ΔG_M^{mut}, and ΔG_W^{mut} directly from docking studies. In addition, ΔG_R is well-known from experiment so that the connection with forward and reverse free energy barriers is provided by the BEP relationship, eqs 24−28. This lets us predict enzymatic rates in silico for any mutated fumarase using eq 23 once λ and μ have been determined from a "training" data set. In the following, we test this method against experimental data available for patients affected by inherited fumarase deficiency.

METHOD

Recently, the high resolution structure of wild type human mitochondrial fumarase (EC 4.2.1.2) has been reported by Kavanagh et al.[56,57] from the Structural Genomics Consortium (PDB file 3E04). This disclosure offers an opportunity for direct in silico assessment of the normal hydratation of fumarate into L-malate and its pathogenic alterations recently available from enzymatic assays.[58] The enzyme occurs as a tetrameric complex, as shown in Figure 3. Previous crystallographic studies of the analog fumarase C expressed in *Escherichia coli* by Weaver et al.[59−63] have established two distinct sites, I and II, that can bind carboxylic acids. Site I, shown in Figure 3, lies at the interface of three monomers, being considered as the putative active site, and site II is believed to be involved in the allosteric regulation of the enzymatic activity. Site II includes His129, Asn131, Asp132,

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

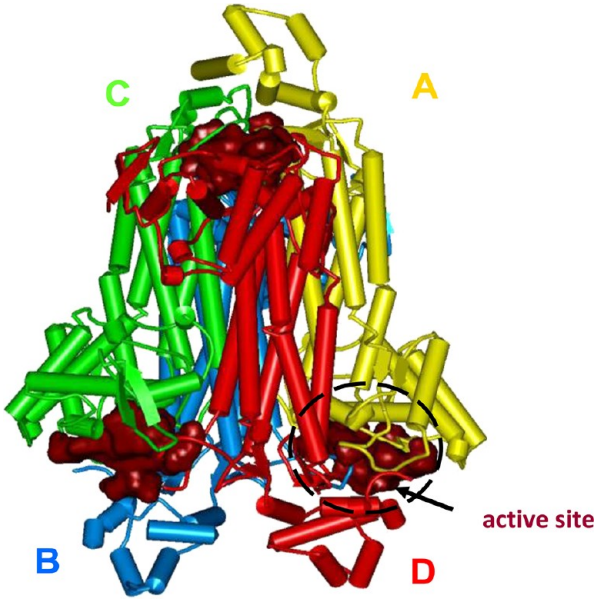


Figure 3. Schematic representation of human fumarase (PDB 3E04) highlighting the tetrameric assembly of the protein. Each monomer (A, B, C, D) is colored distinctively. The four active sites I, as determined from preliminary site searching, are highlighted in red, each of them localized at the interface of three monomers.

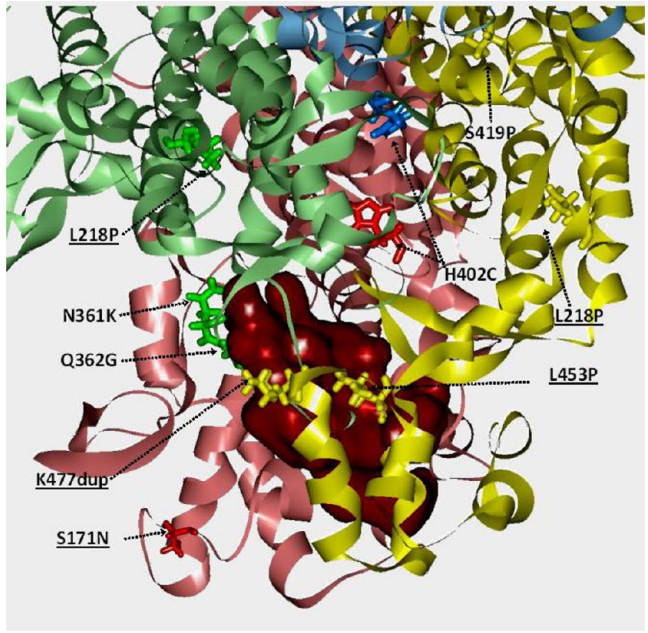


Figure 4. Schematic ribbon representation of the position of the various mutated residues considered in this work, around one of the active sites I (solid red) here at the interface of chains A, C, and D. Mutated residues carried by chains A, B, C, and D are located by their stick representation in yellow (chain A), blue (chain B), green (chain C), and red (chain D). The newly reported FH mutations from ref 58 are underlined in black.

and Asp133, whereas site I involves Thr 100, Ser 137, and Ser 138.

We looked for the binding sites in the wild type human fumarase from its recently disclosed crystal structure. Through sequence alignment, we find the corresponding residues for EC 4.2.1.2; namely, His176-Pro177-Asn178-Asp179 for site II and Thr149-Ser186-Ser187 for site I.

In the present work, we considered four new point mutations that have been recently clinically characterized by some of us in three patients:[58] L453P (referred to as P1), L218P and K477dup (referred to as P2), and S171N (referred to as P3). In a further attempt to test the predictive value of our model, four additional mutants from patients exhibiting acknowledged FH deficiency or renal cancel (RCC) were also considered relying on the recent database of FH mutations by Bayley et al.,[64,65] borne by different exons at the DNA level. We have considered His402Cys[66] (FH, exon 8); Asn361Lys[67] (RCC, exon 7); Ser419Pro[68] (RCC, exon 9); and the first reported FH mutation by Bourgeron et al. in 1994, Glu262Gln[69] (FH, exon 7). Figure 4 reports the spatial localization of the point mutations considered in this work with respect to the putative catalytic site I, represented by a dark red surface. We have selected a representation showing their closest positions around one of the four active sites of the tetrameric enzyme, allowing visual differentiation of mutations directly affecting the active site from those involved, rather, in its peripheral region.

The mutations considered here may affect various regions of the protein. Mutations such as N361K, Q362G, L453G, and K477dup are located in the immediate environment of the active site, whereas other mutations lie rather in more peripheral regions (S171N, H402C, L218P, and S419P). We verified further that the considered mutations do not significantly affect the computed tetramerisation energies of fumarase, which remains stabilizing, within 8 ± 0.5% of the total energy within the CHARMm forcefield. We therefore predict that fumarase tetrameric complexes may remain integrally expressed in cells of the patients bearing those duplication or missense mutations.

Using the wild type and mutated human fumarases, docking calculations were performed for water, fumarate, fumarate derivatives, and L-malate sequentially at site I and in an explicit solvation box. Energies of docking were used to evaluate enthalpies of transfer as expressed in eq 3, neglecting in a first approximation the conformation enthalpic change upon complexation and estimating that seven water molecules were transferred back into the solution upon the substrate complexation to the protein.[46] All calculations were performed within the Discovery Studio version 2.5.5 suite of programs (DS 255) as licensed by Accelrys Inc.[70] Docking computational details are given in the Supporting Information.

RESULTS AND DISCUSSION

In a first step, docking simulations were started on fumarate related substrates in wild type fumarase. We have used the experimental study by Teipel et al.[47−50] which explores the substrate specificity of wild type fumarase for fumarate. Using our computed energies of docking for the various derivatives together with those of water, in silico enthalpies of transfer of eight fumarate derivatives and related substrates have been derived from eq 3, neglecting for the time being the term ΔH^{P}_{conf}, assumed constant (see the Theory section). It is worth recalling that the *n*-term expresses the enthalpic contribution of the release of water molecules from the protein site to the cytosol, occurring upon substrate complexation.

Figure 5 plots the experimental Gibbs free energies of complexation for the eight considered substrates in wild type fumarase, deduced from Michaelis constants measured by Teipel et al.,[47−50] as function of in silico approximation of enthalpies of transfer, which we note ΔH^*, with $\Delta H^* = \Delta H^S -$

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

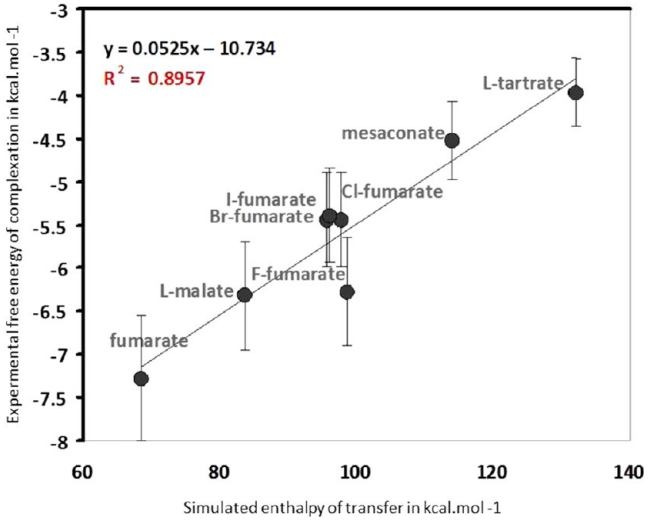


Figure 5. Linear correlation found between experimental free energies of complexation of various substrates by fumarase (adapted from refs 47−50) and in silico average affinity differences between the site and an explicit water solvation box (enthalpy of transfer, see the Methods section). Inset: equation of the regression line and value of the squared coefficient of correlation R^2.

$n\Delta H^{W}$. These in silico enthalpies correlate linearly remarkably well with experimental free energies of complexation, with a regression coefficient of $R^2 = 0.896$, reflecting that our sequential docking calculations correctly capture the interactions at play. This linear correlation is represented by:

$$\Delta G = a^*\Delta H^* + b^* = 0.0525\Delta H^* - 10.734 \quad (29)$$

expressed in kilocalories per mole.

We notice that in combining eqs 3 and 4, one can write:

$$\Delta G = \Delta G^{S} + \Delta G^{P}_{conf} - n\Delta G^{W} \quad (30)$$

and using 8, 12 and 13, we obtain:

$$\Delta G = (a+1)[\Delta H^{S} - n\Delta H^{W}] + [b + (a+1)\Delta H^{P}_{conf})] \quad (31)$$

which gives a theoretical basis to the empirical correlation, eq 29. Considering the statistical errors associated with empirical correlations, the prediction $a^* = (a+1)$ is verified within 20%, with $a^*/(a+1) \approx 0.8$. We also obtain $b^* = b + (a+1)\Delta H^{P}_{conf}$, offering an experimental evaluation of ΔH^{P}_{conf}. Comparing eqs 1 and 29 and keeping the experimental discrepancy observed, it is straightforward to recover eq 3, rewritten:

$$\Delta H = \left(\frac{a^*}{a+1}\right)\Delta H^* + \Delta H^{P}_{conf} \quad (32)$$

(for the wild type fumarase, $\Delta H^{P}_{conf} \approx -28$ kcal mol^{-1})

Therefore, the interval of ΔH spanned by fumarase substrates shown in Figure 5 would be approximately 25−80 kcal mol^{-1}. In a next step, it is straightforward to deduce the in silico estimate of entropies of complexation from eq 32, which turn out to be systematically positive. Because our in silico estimates of the enthalpies of complexation are also (strongly) positive, the conclusion is that the protein−ligand complexation is entropically driven (Table 1).

To evaluate ΔH for the complexation of fumarate, L-malate, and water in mutated enzymes, we rely on eq 15 so that:

$$\Delta H = \Delta H^* + \left(\frac{b^* - (b + \Delta b)}{a+1}\right) \quad (33)$$

Importantly, Figure 6 further reveals the remarkable consistency between our linear $-T\Delta S = f(\Delta H)$ correlation on WT fumarase (red dots) and the correlation obtained from isothermal titration calorimetry (ITC) data (deep blue dots) on a breadth of protein−ligand complexes (i.e., ∼110 protein−ligand complexes from Binding DB database).[31] The renormalization implied by eq 33 also leads to consistent predictions of the free energy of complexation of fumarate, L-malate, and water in mutated fumarases (green dots), relying on the general compensation effect presented as the starting point of our theoretical analysis in the theory section.

Although we are aware that CHARMm-based docking calculations may yield only "effective" enthalpies of transfer (in contrast with "true" enthalpies), our correlative approach based on a training set of experimental K_m for the wild type allows us to predict "true" free energies of transfer.

Figure 6 shows that our computed $-T\Delta S$ components versus ΔH do correlate, suggesting that the quality of the forcefield used is not a major issue.

Here, we should emphasize the general pattern and principle of the so-called "enthalpy−entropy compensation effect" at play in confined host−guest systems. In the context of molecular physisorption of organic molecules in porous zeolites, for example,[71−75] it was experimentally observed that the enthalpic gain upon adsorption (driving force) is systematically compensated by an entropic loss.

Table 1. Comparison of Free Energies of Complexation of Fumarate and Fumarates Derivatives in Wild Type Fumarase from Experiment and Calculations

	K_m mol L^{-1} (exp[a])	ΔG kcal mol^{-1} (exp[b])	ΔH kcal mol^{-1} (sim[c])	ΔS kcal mol^{-1}K^{-1} (sim[c])	$-T\Delta S$ kcal mol^{-1} (sim[c])
fumarate	5.0×10^{-6}	−7.28	26.4	0.11	−33.6
L-malate	2.5×10^{-5}	−6.32	38.5	0.15	−44.8
F-fumarate	2.7×10^{-5}	−6.27	50.4	0.19	−56.6
Br-fumarate	1.1×10^{-4}	−5.44	48.1	0.18	−53.5
I-fumarate	1.2×10^{-4}	−5.38	48.3	0.22	−53.7
Cl-fumarate	1.1×10^{-4}	−5.44	49.8	0.18	−55.2
mesaconate	5.1×10^{-4}	−4.52	62.7	0.22	−67.2
L-tartrate	1.3×10^{-3}	−3.96	77.1	0.27	−81.1
water	1.5×10^{-8}	−10.73	0	0.036	−10.73

[a]Experimental values of K_m are taken from Teipel et al.[47−50] [b]At equilibrium conditions, the Michaelis constant K_m is related to the Gibbs free-energy change of transfer, ΔG, through: $\Delta G = RT \ln K_m$. [c]ΔH sim were estimated using eq 32 taking the number of retrotransfer water molecules as $n = 7$.

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

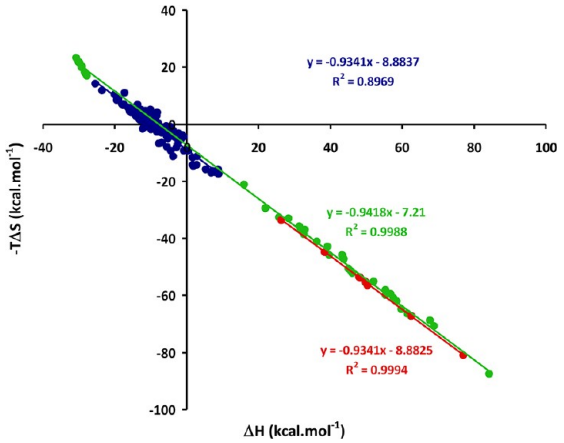


Figure 6. Plot of the calculated enthalpies of complexation, ΔH, of fumarate, L-malate, and water versus the entropy component, $-T\Delta S$, (from eq 32) in mutated human fumarase (green dots, and linear regression for this plot, green line, equation and squared coefficient of correlation in inset with green characters), as compared with experimental ITC data extracted from Binding DB database (deep blue dots; and linear regression for this plot, deep blue line; equation and squared coefficient of correlation in inset with deep blue characters). Also shown is the plot for the complexation of fumarate derivatives in the wild type from eq 32. (red dots; and linear regression for this plot, red line; equation and squared coefficient of correlation in inset with red characters).

This was further computationally exemplified and theoretically analyzed by one of us.[39] According to this latter analysis, above the so-called "confinement compensation temperature", entropic losses overcompensate the enthalpic gain so that physisorption in microporous solids is no longer spontaneous.

In the case of protein−ligand complexations, the above-mentioned experimental correlations show that if variations of enthalpy are frequently negative, they may also be positive. In these latter cases, the corresponding variations of entropy are also systematically positive.

In the case of halogenated fumarates, mesaconate, and L-tartrate vs fumarase, we find in silico highly positive ΔH and deduce also very positive ΔS (Table 1). This can be understood in light of the nice very recent results of Lai et al.[76] From molecular dynamics simulations allowing calculations of absolute binding free energies and entropies of binding of HIV-1 protease inhibitors Nelfinavir and Amprenavir, the latter authors conclude that the favorable entropic contribution to binding is dominated by the ligand desolvation entropy. It can thus be inferred that for highly hydrophilic substrates (i.e. strongly solvated in water (cytosol)), binding may be disfavored from an enthalpic standpoint, but still favored by the entropy gain upon desolvation (water release). In our cases, we dock fumarate derivatives of related dicarboxylic anions, therefore very hydrophilic, and find accordingly higher affinities for water than for the protein sites. Equation 2 leads, therefore, to very positive ΔH values, and from correlation 29, we infer positive ΔS values. We understand from the argument of Lai et al. that the latter is dominated by the strong desolvation entropy of our substrates.

In a second stage, we made use of the correlation found above to predict reaction rate constants of hydrolysis of fumarate into L-malate with respect to the mutated proteins as built in silico. We have demonstrated above in the Theory section how this can be achieved by a combination of a Langmuir−Hinshelwood microkinetic model and the BEP linear free energy relationship.

Figure 7 compares the experimental residual enzymatic activities of mutated fumarases relative to the wild type from

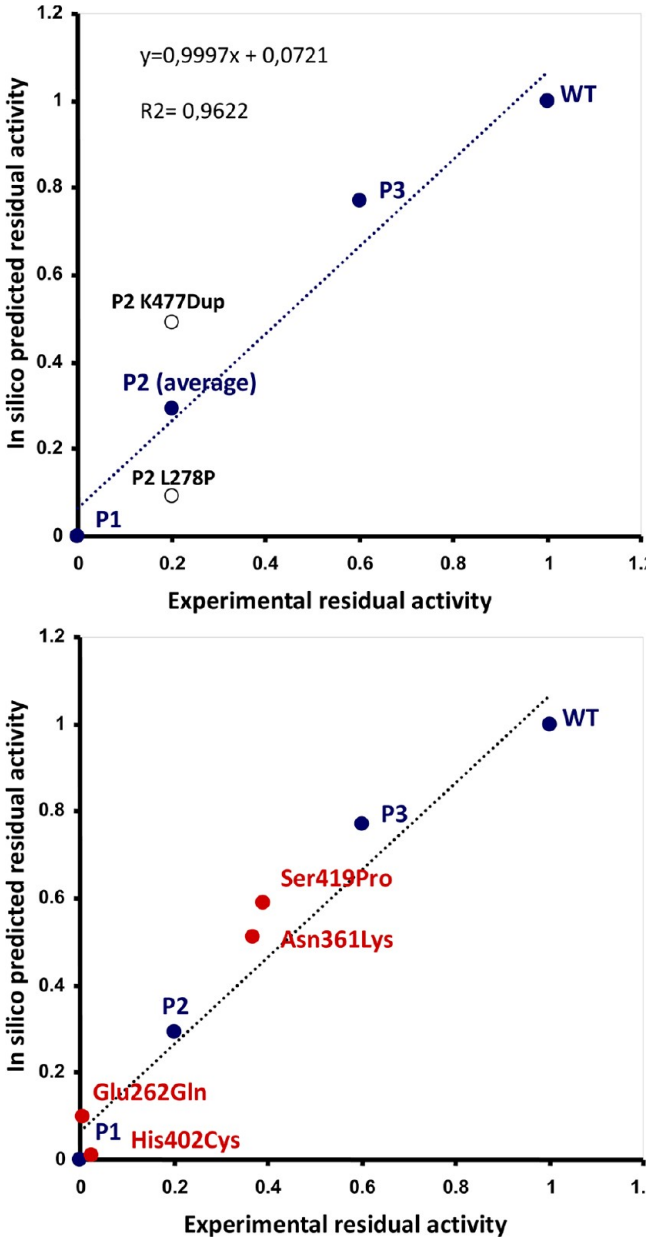


Figure 7. (a) Correlation between experimental relative residual activities taking the wild type fumarase activity as the reference and in silico relative rates using Langmuir−Hinshelwood and BEP models. Four mutations were used as a training set (Leu453Pro, Lys477Dup, Leu278Pro, Ser171Asp). (b) Prediction of relative rates of four mutations from in silico calculations.

three different patients[58] and our predicted residual activities using the Langmuir−Hinshelwood model as processed from the in silico docking calculations. We used as a training set these three most recent clinical cases of hereditary fumarase deficiency, that is, L453P (P1), L278P and K477dup (P2), and S171N (P3), in which it was highlighted that the residual enzymatic activities measured in fibroblasts correlates well with the syndrome severity of patients. In the later work, the test

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

Table 2. Free Energies of Complexation of Fumarate, L-Malate in Mutated Fumarase, Derived from Docking Calculations[a]

	K_m mol L^{-1}	ΔGkcal mol^{-1}	ΔH kcal mol^{-1}	ΔS kcal mol^{-1} K^{-1}	$-T\Delta S$ kcal mol^{-1}
wild type					
fumarate	$5. \times 10^{-6}$	−7.28	26.4	0.11	−33.64
L-malate	2.5×10^{-5}	−6.32	38.5	0.15	−44.81
patient 1					
fumarate	0.00287	−3.49	43.86	0.16	−47.35
L-malate	0.04615	−1.83	68.99	0.24	−70.83
patient 2					
fum. (L218P)	0.00030	−4.83	36.35	0.14	−41.19
L-mal. (L218P)	0.00314	−3.43	57.54	0.20	−60.98
fum. (K477dup)	3.1×10^{-6}	−7.56	21.96	0.10	−29.52
L-mal. (K477dup)	4.5×10^{-5}	−5.96	46.16	0.17	−52.12
patient 3					
fumarate	9.6×10^{-6}	−6.89	25.76	0.11	−32.65
L-malate	8.1×10^{-5}	−5.62	45.03	0.17	−50.64
Glu362Gln					
fumarate	0.00014	−5.28	31.75	0.12	−37.04
L-malate	0.00275	−3.51	52.62	0.21	−62.14
His402Cys					
fumarate	0.00060	−4.42	31.43	0.12	−35.85
L-malate	0.00990	−2.75	56.76	0.19	−59.52
Asn361Cys					
fumarate	3.9×10^{-5}	−6.05	32.61	0.13	−38.67
L-malate	0.00048	−4.54	55.45	0.20	−60.0
Ser419Pro					
fumarate	0.00016	−5.22	16.05	0.07	−21.27
L-malate	0.00205	−3.69	39.33	0.14	−43.02

[a]The respective contribution on the enthalpic change and entropic change upon transfer from the solution to the protein active site are given, as obtained from docking calculations and free energy/enthalpy correlations (eq 33). Numerical values for water are given in the SI.

involves the measurement of the rate of absorbance at 250 nm (detection of fumarate) in the initial linear regime upon addition of 100 μL of a 500 mM solution of L-malate to 0.9 mL of a phosphate-buffered 20% Triton suspension of fibroblasts. For control assays (wild type fumarase), a mean activity of 370 nmol min^{-1} mg $protein^{-1}$ (4.38×10^6 s^{-1}) was recorded, with ~±30% repeatability. For the three patients, P1 (L453P), P2 (L278P/K477dup), and P3(S171N), residual activities were reported to be 0% (undetectable), 20%, and 60%, respectively. Note that patient 2 bears two different mutations, one on each allele, so that the experimental assay is expected to reflect an average of the residual activities, which could be virtually assigned to cells from two patients, each being homozygous for one of these mutations.

Considering the experimental conditions, we estimate that the prevailing L-malate and fumarate concentrations at the recording point are 49×10^{-3} mol L^{-1} and 1×10^{-3} mol L^{-1}, respectively. Using eqs 16−23, we compute for each mutant the predicted relative rate rr_{prod}^{mut} of production of fumarate in the assay conditions as:

$$rr_{pred}^{mut} = \frac{(\vec{r}^{mut} - \vec{r}^{mut})}{(\vec{r}^{wt} - \vec{r}^{wt})} \quad (34)$$

We have determined the coefficients λ and μ of the BEP relationship 24 as those providing the best coefficient of correlation in the parity diagram of predicted vs experimental relative residual rates of fumarate production, involving each mutation of patients 1 and 3 and the average for the two mutations reported for patient 2.

Because this procedure was consistent with a value of λ close to 1, it was fixed at 1, which is expected in general for such free energy linear relationships when the transition state configuration is very close to that of the product (here, L-malate).

Left with a single free parameter to refine, we converge for μ = 14.29 kcal mol^{-1}. The corresponding parity diagram shown in Figure 7 exhibits an almost perfect linear correlation, with an excellent regression coefficient, R^2 = 0.967 (slope of 1.004; intercept of 0.064). Forcing the regression line to intercept zero degrades only slightly this correlation, with R^2 = 0.956 (slope of 1.086).

We made no attempts to account for the error margin involved in the experimental measurements. Notice that when λ = 1, the reverse intrinsic activation energy $\Delta G_{\leftarrow}^{\pm,mut}$ becomes constant (mutation-independent) by virtue of eq 28. It ensues from eq 23 that the relative rates of the inverse reaction are strictly correlated to the coverage of sites by L-malate. In other terms, the experimental assay of residual enzymatic activities consisting of measuring the initial rate of consumption of L-malate measures this coverage.

Overall, the estimates of residual activity obtained within the Langmuir−Hinshelwood model provides an excellent correlation with the results obtained from in vitro enzymatic assays of the isolated proteins. Typically, patient 1 (L453P) has the most severe fumarase deficiency, and its fumarase appears experimentally practically inactive. This is very well reflected in the in silico predictions by a reduction of its residual enzymatic activity by 100%. In the special case of patient 2, Figure 7 shows the individual predicted relative rates for each mutation taken individually, L278P and K477dup (open circles). The impacts of both mutations were further averaged in the estimation of rate constants to account for the fact that the fumarase

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673−2686

tetrameric complexes may randomly express one or the other of the two mutations over the four tetramers.

It is noteworthy that although each mutation has distinct impacts on fumarase activity (~10% and ~50% residual activities), the averaged rate falls remarkably well on the regression line, with a residual activity of ~30%, supporting the correctness of the individual rates.

It should be noticed that at this point, to obtain such a result, it is important to effectively include the mutation-dependent correction on the intercept b of the correlation $\Delta G = f(\Delta H)$, according to eq 15 proposed above. If this correction is omitted, the predicted free energy of complexation of water in the protein site becomes invariant to mutations. Keeping in that case $\lambda = 1$ and $\mu = 14.29$ kcal mol^{-1}, our Langmuir–Hinshelwood model still yields an excellent correlation between in silico predictions and experiments ($R^2 = 0.98$) but loses parity because a relative residual activity of 0.36 instead of ~0 is now predicted for P1. In view of the considerations above on the consequence of setting $\lambda = 1$, this reflects actually a systematic overestimation of the coverage in L-malate for all mutations.

To further investigate the extent to which our approach might be used to predict fumarate deficiencies and the related residual activities, we considered other cases of FH mutations reported in the literature, including three biallelic and one monoallelic mutations, that is, on one hand Glu362Gln, His402Cys, Asn361Lys; and on the other hand, Ser419Pro, acknowledged to be mainly responsible for FH deficiency (Glu362Gln and His402Cys) or renal cancer (Asn361Lys and Ser419Pro).

Despite the quite large number of mutations known to date,[64,65] the number of mutations studied here was restricted to those for which measurements of FH enzyme activity were available and performed on fibroblasts. This was to ensure a comparison of data with the patients P1, P2 ,and P3 collected in similar conditions. In the case of Ser419Pro, the experimental residual activity measured on fibroblasts of the heterozygous patient is predicted by the average of predictions for WT (relative activity 1) and for the in silico mutation. Indeed, as in the case of patient 2, the residual enzymatic activity is assumed to result from the average expression of the genes borne by the two alleles.

The corresponding predictions of residual FH activity for these four additional mutations are reported in red on the parity diagram of Figure 7. The regression line involving both the initial training set and the additional mutations set is not significantly modified, with $R^2 = 0.951$ (slope = 1.036, intercept = 0.074).

It appears that the considered mutations always increase the Michaelis constants, K_M, (Table 2) in a rather systematic fashion for both fumarate and L-malate substrates with respect to the wild type, as could be expected from weaker enzyme–substrate interactions in mutated enzymes.

However, the predicted rates do not correlate with the observed residual fumarase activities if we use a noncompetitive Michaelis–Menten model. Taking the competitive inhibition by L-malate into account, the Langmuir–Hinshelwood model, which we have shown identical to the competitive Michaelis–Menten model, yields an excellent correlation between the observed and the predicted reaction rates. From the above, it is apparent that the impact of mutations on the residual enzymatic activity is extremely complex, consisting of independent effects on individual Michaelis constants K_M, that is, individual free energies of complexation of fumarate, malate, and water (ΔG_F^{mut}, ΔG_M^{mut}, ΔG_W^{mtt}) on one hand, and on activation barriers that result from the linear combination of free energies (ΔG_F^{mut}, ΔG_M^{mut}, ΔG_W^{mtt}, ΔG_R) on the other hand. It obviously follows that a simple comparison of docking calculations on substrate-mutated enzymes with respect to the wild type is not sufficient to accurately predict relative enzymatic activities. For instance, the free energies of complexation of fumarate correlate badly with experimental residual activities ($R^2 \sim 0.49$), and the free energies of complexation of L-malate are only slightly better ($R^2 \sim 0.62$).

Still, in light of the successful prediction of residual activities, we have analyzed the host–guest interactions of substrates in the wild type and mutated enzymes, in an attempt to get molecular insights into FH deficiency and identify interactions that might be responsible for the deterioration of enzymatic activities. We have localized the catalytic site as the cavity within the wild type fumarase complex not only large enough to accept substrate docking attempts but also on a physicochemical consistency criterion.

As shown in Figure 8a for the WT protein, S186 and S187, together with L371, were identified as the key amino acids. The docked fumarate develops multiple hydrogen bonds between one of its carboxylate group and –OH groups of two adjacent serines, Ser186 and Ser187, from 1.9 to 2.1 Å, while the other carboxylate group interacts with the proximal $-CH_2-$ group in the α position with respect to the positively charged terminal amino of Lys371, at a larger distance of 2.4 Å. We believe this predicted mode of anchoring of fumarate in the wild type active site prevents the rotation of the substrate around the axis of its double bond, a favorable prerequisite for the stereoselective nucleophilic attack of water in the remaining free space of the active site pocket, leading exclusively to the L-malate product.

The comparison of fumarate locations in mutated enzymes reveals from mild to marked modifications in substrate–ligand interactions (Figure 8b and SI Figure S1). We observe that point mutations in the (either immediate or peripheral) region of the active site induce variously severe alterations of the above substrate–ligand interactions, especially when precluding the formation of the multiple hydrogen bonds between both oxygen atoms of one of the carboxylate groups and the two adjacent serines Ser 186 and Ser 187 (a common feature to all mutations considered).

Moreover, for patient 1 (L453P), we note that the combination of the above effect with the suppression of interactions between the other carboxylate group and the L371 amino acid in the mutated enzyme results in a severe deterioration of fumarate anchorage in the active site, reflected in the particularly low absolute value of the computed free energy of binding (high K_M), in correlation with the nearly total loss of fumarase activity of patient 1. Conversely, patient 3, in whom both types of interactions are maintained although altered, exhibits a mild FH deficiency, suggesting that binding interactions at both carboxylate ends may be required to maintain some enzymatic activity.

CONCLUSION

The present work successfully predicts enzyme activity changes upon mutation on the basis of two generally observed linear relationships, that is, the linear correlation between enthalpy and entropy changes resulting from protein–ligand binding, on one hand, and the BEP relationship between the free energy change of a reaction (between reactants and products in a

dx.doi.org/10.1021/cs300538z | ACS Catal. 2012, 2, 2673–2686

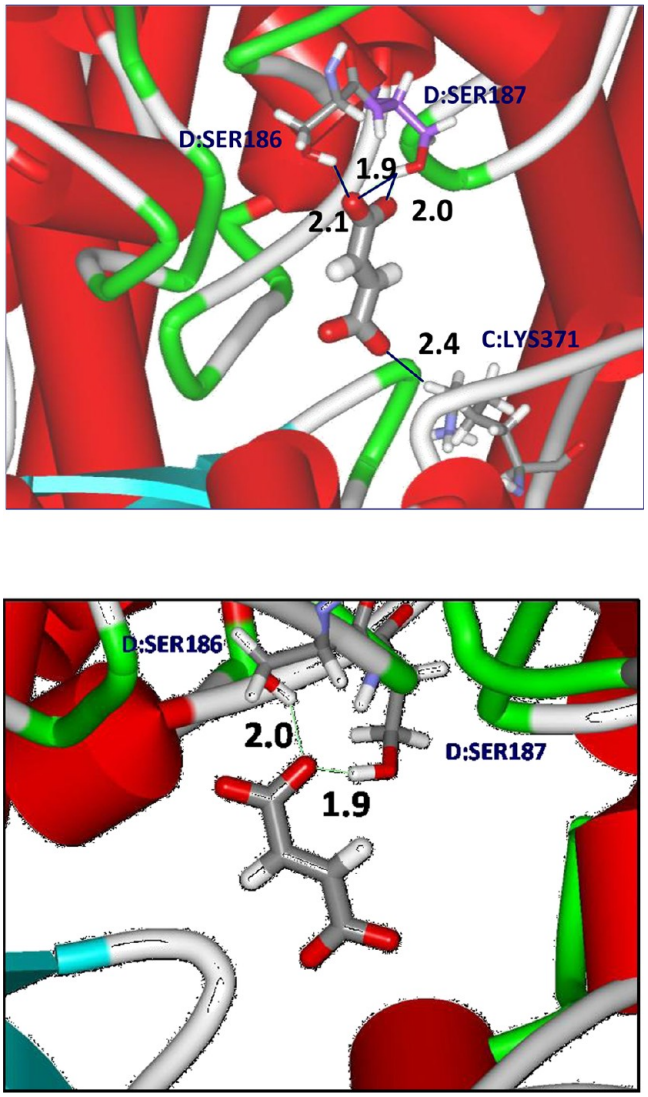


Figure 8. (a) Schematic view of fumarate in the active site in wild type fumarase (3E04), as obtained from the higher-energy docking pose, in interaction with Ser186 and Ser187 of chain D and with Lys371 of chain C. Hydrogen bonds are represented as black lines; (b) Schematic view of fumarate in the active site of the mutated enzyme (patient 1, L453P) from the docking calculations.

complex) to the free energy of activation or reaction rate, a special case of linear free energy relationship (LFER), on the other hand. An important implication of the present work is to provide a universal theoretical framework allowing interpretation of the enthalpy−entropy compensation effect empirically observed in protein−ligand complexes. It includes a generalized approach of the impact of point mutations: (i) In the $-T\Delta S = f(\Delta H)$ entropy−enthalpy plot, the slope, *a*, is temperature-dependent but independent of the protein, and may be expressed as the ratio of the entropy and enthalpy components of the Gibbs free energy of transfer of water or any other substrate to a protein. (ii) The intercept, *b*, of the $-T\Delta S = f(\Delta H)$ plot depends on the protein, while being related *exclusively* to the conformational energy changes of the protein upon complexation. (iii) Taking advantage of the compensation effect, free energies of complexation in mutated enzymes may be derived from docking studies in silico using a "training" set provided by diverse homologous substrates in complex with the wild type enzyme. (iv) Intrinsic activation energies for the direct and reverse enzymatic reaction in complex are connected to the free energies of complexation of reactants and products through a simple BEP relationship. (v) The enzymatic kinetics are explicitly described by the Langmuir−Hinshelwood model, in a form that we show to be equivalent to the competitive Michaelis−Menten model.

We have tested this general theoretical framework in the in silico prediction of the residual activity of mutated human fumarases with implications for the medical prognostic of hereditary fumarase deficiency. Fumarase, or fumarate hydratase (FH), is an important member of the Krebs cycle catalytic cascade occurring inside the intramitochondrial space of chemotrophic eukaryotes. Thanks to the recently released high-resolution structure of the human wild type fumarase, we have been able to compare this structure to the derived 3D atomistic models involving the point mutations causing patients' FH deficiency. Using specific biomolecular simulation tools and the CHARMm forcefield, we have localized active sites and evaluated free energies of transfer of the reactants and product, water, fumarate, and L-malate by performing docking studies. The latter were the input in our predictions of residual activities, which we could finally compare with experimental data from assays on patients cells (fibroblastes).

Our in silico results are in parity with the measured residual fumarase activity, which in turn correlates well with the severity of the fumarase hereditary deficiency syndrome. We conclude that our calculations are, indeed, predictive. We anticipate that the combination of the proposed computational approach, which we propose to designate by the acronym R.E.A.-L.I.S.M.,[77] with a calibration against experimental Michaelis constants for a set of homologous substrates and the wild type, opens the way to a successful general strategy for understanding predictively the impact of mutations on biological functions.

■ ASSOCIATED CONTENT

Supporting Information

Additional computational methods for docking calculations, equivalence of the Langmuir−Hinshelwood model with the classical Michaelis−Menten model, free energies of complexation of water with wild type and mutated fumarase, complexation geometries of fumarate in FH of P2 (K477dup, L218P) and P3 (S171N) patients, experimental support for *b* independent of substrate in eq 1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: herve.toulhoat@ifpen.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Chris Ottolenghi and Dr. Yves de Keyser for fruitful discussions.

■ REFERENCES

(1) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.

(2) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.

(3) Dellago, C.; Bolhuis, P.; Geisselr, P. *Adv. Chem. Phys.* **2002**, *123*, 1.

(4) Carter, E.; Ciccotti, G.; Hynes, J.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
(5) Den Ottern, W.; Briels, W. *Mol. Phys.* **2000**, *98*, 773.
(6) Darve, E.; Wilson, M.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113.
(7) Fleurat-Lessard, P.; Ziegler, T. *J. Chem. Phys.* **2005**, *123* (8), 084101.
(8) Car, R.; Parrinello, M. *P.R.L.* **1985**, *55* (22), 2471.
(9) Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93* (7), 2523.
(10) Liu, H.; Warshel, A. *Biochemistry* **2007**, *46*, 6011−6025.
(11) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (39), 16869−16874.
(12) Frushicheva, M. P.; Cao, J.; Warshel, A. *Biochemistry* **2011**, *50*, 3849−3858.
(13) Marti, S.; Andres, J.; Silla, E.; Moliner, V.; Tunon, I.; Bertran, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 286−290.
(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(15) Aqvist, J; Medina, C.; Samuelsson, J. *Protein Eng.* **1994**, *7* (3), 385.
(16) Wu, G.; Robertson, D. H.; Brooks, C. L.; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549.
(17) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. *J. Comput. Chem.* **2003**, *24*, 1549.
(18) Jain, A. N. *Curr. Protein Pept. Sci.* **2006**, *7*, 407.
(19) Moitessier, N.; Englebienne, P; Lee, D; Lawandi, J; Corbeil, C. R. *Br. J. Pharm.* **2008**, *153*, S7−S26.
(20) Gilli, P.; Ferretti, V.; Gillli, G.; Borea, P. A. *J. Phys. Chem.* **1994**, *98*, 1515−1518.
(21) Dunitz, J. D. *Chem. Biol.* **1995**, *2*, 709−712.
(22) Calderone, C. T.; Williams, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 6262−6267.
(23) Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann. *M. Biophys. Chem.* **2001**, *93*, 215.
(24) Talhout, R.; Villa, A.; Mark, A. E.; Engberts, J. B. F. N. *J. Am. Chem. Soc.* **2003**, *125*, 10570.
(25) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596.
(26) Krishnamurthy, V. M.; Bohall, B. R.; Semetey, V.; Whitesides, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 5802.
(27) Bond, G. C.; Keane, M. A.; Kral, H.; Lercher, J. H. *Catal. Rev. Sci. Eng* **2000**, *42*, 323.
(28) Guernalec, N.; Geantet, C; Cseri, T; Vrinat, M; Toulhoat, H; Raybaud, P. *Dalton Trans.* **2010**, *39*, 8420.
(29) O'Brien, R.; Ladbury, J. E; Chowdry, B. Z. In *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*; Harding, S. E., Chowdry, B. Z., Eds.; Oxford University Press: New York, 2000; Chapter 10.
(30) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007**, *35*, D198−D201.
(31) See also: http://www.bindingdb.org/bind/index.jsp and http://www.biochem.ucl.ac.uk/scorpio/scorpio.html.
(32) Brønsted, N. *Chem. Rev.* **1928**, *5*, 231.
(33) Evans, M. G.; Polanyi, N. P. *Trans. Faraday Soc.* **1938**, *34*, 11.
(34) Bligaard, T.; Norskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. *J. Catal.* **2004**, *224*, 206−217.
(35) Nørskov, J. K.; Scheffler, M.; Toulhoat, H. *MRS Bull.* **2006**, *31*, 669.
(36) Assary, R. S.; Broadbelt, L. J.; Curtiss, L. A. *J. Mod. Model.* **2012**, *18*, 145−150 (This DFT study for a decarboxylation reaction by a thiamine-containing enzyme demonstrates that the BEP relationship is valid for correlating free energy of formation of the transition state and the free energy of the reaction, allowing prediction of the activation barriers of 17 chemically similar novel reactions).
(37) See, for instance: Van Santen, R. A.; Niemantsverdriet, J. W. *Chemical Kinetics and Catalysis*; Plenum Press: New York and London, 1995.
(38) Eng, C.; Kiuru, M.; Fernandez, M. J.; Aaltonen, L. A. *Nat. Rev. Cancer* **2003**, *3*, 193.
(39) Toulhoat, H.; Lontsi Fomena, M.; de Bruin, T. *J. Am. Chem. Soc.* **2011**, *133*, 2481.
(40) Rekharsky, M. V.; Inoue, Y. *Chem. Rev.* **1998**, *98*, 1875.
(41) Liu, L.; Guo, Q.-X. *Chem. Rev.* **2001**, *101*, 673.
(42) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872.
(43) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007**, *35*, D198−D201.
(44) Winzor, D. J.; Jackson, C. M. *J. Mol. Recognit.* **2006**, *19*, 389.
(45) Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. *J. Mol. Biol.* **2008**, *384*, 1002−1017.
(46) To evaluate *n*, the number of water molecules desolvated from the substrate upon complexation with the protein, two water boxes were built and relaxed, one around a single water molecule, the other around a single substrate molecule using the same size parameters for the resulting water box. The difference between the total number of water molecules between the two systems yielded *n*.
(47) Alberty, R. A.; Massey, V.; Frieden, C.; Fuhlbrigge, A. R. *J. Am. Chem. Soc.* **1954**, *76*, 2485.
(48) Frieden, C.; Alberty, R. A. *J. Biol. Chem.* **1955**, *212*, 859.
(49) Teipel, J. W.; Hill, R. L. *J. Biol. Chem.* **1968**, *243*, 5679.
(50) Teipel, J. W.; Hass, G. M.; Hill, R. L. *J. Biol. Chem.* **1968**, *243*, 6684.
(51) Haldane, J. B. S. *Enzymes, Monographs on Biochemistry*, London and New York, 1930.
(52) Mescam, M.; Vinnakota, K. C.; Beard, D. A. *J. Biol. Chem.* **2011**, *286*, 21100.
(53) Warshel, A.; Schweins, T.; Fothergill, T. *J. Am. Chem. Soc.* **1994**, *116*, 8437−8442.
(54) Alkherraz, A.; Kamerlin, S. C. L.; Feng, G.; Sheikh, Q. I.; Warshel, A.; Williams, N. H. *Faraday Discuss.* **2010**, *145*, 281−299.
(55) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley: New York, 1991.
(56) Picaud, S.; Kavanagh, K. L.; Yue, W. W.; Lee, W. H.; Muller-Knapp, S.; Gileadi, O.; Sachettini, J.; Oppermann, U. *J. Inherited Metab. Dis.* **2011**, *34*, 671.
(57) http://www.rcsb.org/pdb/explore/explore.do?structureId=3E04.
(58) Ottolenghi, C.; Hubert, L; Allanore, Y.; Brassier, A.; Altuzarra, C.; Mellot-Draznieks, C.; Bekri, S.; Goldenberg, A.; Veyrieres, S.; Boddaert, N.; Barbier, V.; Valayannopoulos, V.; Slama, A.; Chrétien, D.; Ricquier, D.; Marret, S.; Frebourg, T.; Rabier, D.; Munnich, A.; de Keyzer, Y.; Toulhoat, H.; de Lonlay, P. *Hum. Mutat.* **2011**, *32*, 1046.
(59) Weaver, T. M.; Banaszak, L. *Biochemistry* **1996**, *35*, 13955.
(60) Weaver, T.; Lees, M.; Banaszak, L. *Protein Sci.* **1997**, *6*, 834.
(61) Estévez, M.; Skarda, J.; Spencer, J.; Banaszak, L.; Weaver, T. M. *Protein Sci.* **2002**, *11*, 1552.
(62) Rose, I. A.; Weaver, T. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3393.
(63) Weaver, T. M. *Acta Crystallogr.* **2005**, *D61*, 1395.
(64) Bayley, J. P.; Launonen, V.; Tomlinson, I. P. M. *BMC Med. Genet.* **2008**, *9*, 20.
(65) See also: http://chromium.liacs.nl/lovd_sdh/home.php.
(66) Philips, T. M.; Gibson, J. B.; Ellison, D. A. *Pediatr. Neurol.* **2006**, *35*, 150.
(67) Alam, N. A.; Rowan, A. J.; Wortham, N. C.; Pollard, P. J.; Mitchell, M.; Tyrer, J. P.; Barclay, E.; Calonje, E.; Manek, S.; Adams, S. J.; Bowers, P. W.; Burrows, N. P.; Charles-Holmes, R.; Cook, L. J.; Daly, B. M.; Ford, G. P.; Fuller, L. C.; Hadfield-Jones, S. E.; Hardwick, N.; Highet, A. S.; Keefe, M.; MacDonald-Hull, S. P.; Potts, E. D. A.; Crone, M.; Wilkinson, S.; Camacho-Martinez, F.; Jablonska, S.; Ratnavel, R.; MacDonald, A.; Mann, R. J.; Grice, K.; Guillet, G.; Lewis-Jones, M. S.; McGrath, H.; Seukeran, D. C.; Morrison, P. J.; Fleming, S.; Rahman, S.; Kelsell, D.; Leigh, I.; Olpin, S.; Tomlinson, I. P. M. *Hum. Mol. Genet.* **2003**, *12*, 1241.
(68) Pithukpakorn, M.; Wei, M.-H.; Toure, O.; Steinbach, P. J.; Glenn, G. M.; Zbar, B.; Linehan, W. M.; Toro, J. R. *J. Med. Genet.* **2006**, *43*, 755.
(69) Bourgeron, T.; Chrétien, D.; Poggi-Bach, J.; Doonan, S.; Rabier, D.; Letouzé, P.; Munnich, A.; Rötig, A.; Landrieu, P.; Rustin, P. *J. Clin. Invest.* **1994**, *93*, 2514.

(70) http://accelrys.com/products/discovery-studio/
(71) Eder, F.; Stockenhuber, M.; Lercher, J. A. *J. Phys. Chem. B* **1997**, *101*, 5414.
(72) Denayer, J. F.; Baron, G. V.; Martens, J. A.; Jacobs, P. A. *J. Phys. Chem. B* **1998**, *102*, 3077.
(73) Ruthven, D. M.; Kaul, B. K. *Adsorption* **1998**, *4*, 269.
(74) Tielens, F.; Denayer, J. F. M.; Daems, I.; Baron, G. V.; Mortier, W. J.; Geerlings, P. *J. Phys. Chem. B* **2003**, *107*, 11065.
(75) Atkinson, D.; Curthoys, G. *J. Chem. Soc. Faraday Trans. 1* **1981**, *77*, 897.
(76) Deng, N.-J.; Zhang, P.; Cieplack, P.; Lai, L. *J. Phys. Chem. B* **2011**, *115*, 11902.
(77) Relative Enzymatic Activity Levels from In Silico Mutagenesis.