Accounting for the instantaneous disorder in the enzyme-substrate Michaelis complex to calculate the Gibbs free energy barrier of an enzyme reaction

Sonia Romero-Téllez, [1,2] Alejandro Cruz, [1] Laura Masgrau, [1,2,3] Àngels González-Lafont [1,2] and José M. Lluch [1,2,*]

[1] Departament de Química and [2] Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona, 08193, Bellaterra, Barcelona, Spain.
[3] Zymvol Biomodeling, Carrer Roc Boronat, 117, 08018 Barcelona, Spain

Email: JoseMaria.Lluch@uab.cat
These authors contributed equally: Sonia Romero-Téllez and Alejandro Cruz

Abstract

Many enzyme reactions present instantaneous disorder. These dynamic fluctuations in the enzyme-substrate Michaelis complexes generate a wide spread of energy barriers that cannot be experimentally observed, but that determines the measured kinetics of the reaction. These individual energy barriers can be calculated using QM/MM methods, but then the problem is how to deal with this dispersion of energy barriers to provide kinetic information. So far, the most usual procedure implies the so-called exponential average of the energy barriers. In this paper we discuss the foundations of this method, and we use the free energy perturbation theory to derive an alternative equation to get the Gibbs free energy barrier of the enzyme reaction. In addition, we propose a practical way to implement it. We have chosen four enzyme reactions as examples. In particular, we have studied the hydrolysis of a glycosidic bond catalyzed by the enzyme *Thermus thermophilus* β-glycosidase, and the mutant Y284P Ttb-gly, and the hydrogen abstraction reactions from C_{13} and C_7 of arachidonic acid catalyzed by the enzyme rabbit 15-lipoxygenase-1.

1. Introduction

Enzymes are very large biological catalysts that speed up the chemical reactions up to many orders of magnitude (nearly 20 in some cases).[1-3] This huge acceleration makes possible that the vital processes for living beings take place in a time scale compatible with life.

Enzymes are very far from being rigid structures. They are highly flexible molecules. Although the origin of the catalytic power of enzymes (electrostatic based catalysis or dynamical contributions) is still under discussion,[4] enzyme dynamics is relevant for the rate constants, no matter by what force enzyme catalysis is driven. Once the enzyme-substrate Michaelis complex is formed, the set of nuclei of the system samples a huge number of configurations on a highly dimensional potential energy surface which can include one or many reactant valleys.[5,6] Different reactant valleys might perform the same function but contribute to the total reaction with different reaction rate constants.[7-9] Thus, catalytic rates of enzymes often display very large dynamic fluctuations over timescales much longer than k_{cat}, mostly associated with slow conformational changes that connect the different reactant valleys. The existence of these relatively stable long-lived conformational reaction valleys in comparison with the time scale of the enzymatic reactions explains the *dynamic disorder* observed in single-molecule studies of some enzyme reactions.[7,8,10-16] In these cases the interconversion rates among the corresponding reactant valleys might be much slower than the catalytic rate constants, and very long molecular dynamics simulations would be required to produce transitions among them and, especially, to reach statistical equilibrium. Thus, ergodicity (that is, that all the states of the ensemble will be populated) does not apply in practice on the full configurational ensemble of the potential energy surface corresponding to the enzyme-substrate Michaelis complex. The population of each reactant valley is not governed by thermodynamic control, but by kinetic control and each reactant valley acts as an independent enzyme with a different rate constant (the overall rate of product formation is determined by an experimentally detected distribution of rate constants).

Dynamic disorder is an experimentally observed phenomenon. When only a reactant valley exists (or in the case of fast interconversion among the valleys), dynamic disorder does not longer appear, and a single rate constant can describe the enzyme reaction. In this case, another dynamic, but entirely different concept, *instantaneous disorder,* might still take place.[17] *Instantaneous disorder* can be defined as instantaneous dynamic fluctuations in the enzyme-substrate Michaelis complex with timescales much smaller

2

than k_{cat}, leading to different energy barriers that cannot be individually detected by the experiment.

A reactant valley includes an ensemble of nuclear configurations of the enzyme-substrate Michaelis complex that can be extensively sampled by means of a long-time molecular dynamics simulation. The interconversions among all the low-energy configurations within the same reactant valley usually consist in relatively easy local conformational changes and they are quite fast at physiological temperatures. These interconversions occur with timescales much smaller than k_{cat}. In these conditions those nuclear configurations of the reactant valley are in local equilibrium (they are populated according to a Boltzmann distribution) with each other, and the whole valley contributes to the reaction flux with a single k_{cat}. The landscape within this reaction valley is highly ridged, with a huge number of minimum-energy structures. Starting from different initial nuclear configurations (snapshots) generated by equilibrium molecular dynamics simulations, quantum mechanics/molecular mechanics (QM/MM) methods can build multiple minimum-energy paths (MEPs), each of them connecting a potential energy minimum inside the reactant valley with a potential energy minimum in the associated product valley, crossing through the corresponding transition state structure. Then, a wide range of different energy barriers might be found.[6,17-32] Thus, for instance, free energy barriers ranging from 14.5 kcal/mol to 51.3 kcal/mol have been calculated [17] for the first step of the reaction catalyzed by HIV-1 protease. This dispersion was attributed to different mechanisms, different conformations of the active center, and to variations in the electrostatic environment of the active site.

The point now is how to take advantage of the calculation of the distribution of the potential energy barriers to obtain information about the experimental rate constant k_{cat}, or, at least, the Gibbs free energy barrier $\Delta G^{o\ddagger}$ of the catalytic reaction when only a reactant valley exists. In other words, how does *instantaneous disorder* determine the kinetics of the reaction? A discussion about this matter is the goal of this paper. We derive, based on free energy perturbation theory, an equation to calculate this free energy barrier that presents some important conceptual differences with respect to the common approaches, and propose a practical way to implement it. We apply all that to four enzyme reactions.

2. Theoretical basis

Let us assume that an enough lengthy molecular dynamics simulation that covers a single reaction valley has been carried out for a given enzyme-substrate Michaelis complex. Then, a wide set of snapshots (that is, instantaneous nuclear configurations) are randomly chosen from the generated molecular dynamics trajectory. The energy optimization of these snapshots leads to a set of minimum energy structures that can be used as starting structures to build minimum potential energy paths that lead to the corresponding products. As mentioned above, a wide dispersion of QM/MM potential energy barriers $\Delta V^{\ddagger}$ (or energy barriers $\Delta E^{\ddagger}$ if the thermal and zero-point energy corrections are included) are expected to be obtained because of the *instantaneous disorder*. From here on we assume that $\Delta V^{\ddagger}$ and $\Delta E^{\ddagger}$ do not significantly differ. Several different procedures to average these barriers have been proposed.[33] The so-called exponential average has been the most used:[24,27-35]

$$\Delta E_{AV}^{\ddagger} = -RT ln \left\{ \frac{1}{N} \sum_{i=1}^{N} exp \left(\frac{-\Delta E_i^{\ddagger}}{RT} \right) \right\} \tag{1}$$

where R is the gas constant, T is the temperature, $\Delta E_i^{\ddagger}$ is the energy barrier for each snapshot, and N is the number of snapshots selected.

Equation 1 is obtained if several assumptions are fulfilled: a) The reaction valley can be decomposed into a huge number of strictly non-overlapping reactant sub-valleys. b) The ensemble of structures inside each reactant sub-valley α evolves to the corresponding product sub-valley with a rate constant k_{α} that can be calculated through the conventional transition state theory as a function of the Gibbs free energy barrier associated to that sub-valley α, $\Delta G_{\alpha}^{\circ\ddagger}$:

$$k_{\alpha} = \frac{k_B T}{h} exp \left(\frac{-\Delta G_{\alpha}^{\circ\ddagger}}{RT} \right) \tag{2}$$

where k_B is the Boltzmann constant and h is the Planck constant;

c) Because of all the reactant sub-valleys composing the whole reactant valley are also in local equilibrium among them, the overall rate constant k_{cat} can be obtained by arithmetic average of a representative number of k_{α} weighed according to the population P_{α} of their corresponding reactant sub-valleys α (that is, according to their relative free energies):

$$k_{cat} = \frac{k_B T}{h} exp \left(\frac{-\Delta G^{\circ\ddagger}}{RT} \right) = \sum_{\alpha} P_{\alpha} k_{\alpha} \tag{3}$$

d) Each snapshot i chosen along the molecular dynamics trajectory can be classified as belonging to a particular reactant sub-valley α. After energy optimization of this snapshot, the corresponding minimum potential energy path is calculated and the value of the

energy barrier $\Delta E_i^{\ddagger}$ is obtained. e) All the snapshots belonging to the same reactant sub-valley α have roughly the same value of $\Delta E_i^{\ddagger}$, that can be taken as a good estimate of $\Delta G_\alpha^{\circ\ddagger}$. f) The population P_α is given by the number of times a snapshot i with $\Delta E_i^{\ddagger}$ belonging to the reactant sub-valley α appears along the molecular dynamics simulation, which holds a Boltzmann distribution. Under these conditions, equation 1 is reached by combining equations 2 and 3, estimating each $\Delta G_\alpha^{\circ\ddagger}$ by the common value of $\Delta E_i^{\ddagger}$ that is supposed to be obtained for whatever snapshot belonging to the reactant sub-valley α, and using $\Delta E_{AV}^{\ddagger}$ as a reasonable approximation to the Gibbs free energy barrier $\Delta G^{\circ\ddagger}$ of the catalytic reaction when only a reactant valley exists.

Nevertheless, there is no practical way to classify a given snapshot into a particular reactant sub-valley α, as assumed in d). Snapshots that should be associated to the same value of $\Delta G_\alpha^{\circ\ddagger}$ and, according to assumption e), to the same value of $\Delta E_i^{\ddagger}$ (so just contributing to the weight of this value), are actually associated, as a result of the calculation, to different (or even very different) values of $\Delta E_i^{\ddagger}$. Then, we should be aware that an important dispersion of values of $\Delta E_i^{\ddagger}$ can actually appear for snapshots corresponding to the same reactant sub-valley α, in disagreement with assumption e). Thus, let us suppose, only for the sake of a very simple example, that snapshots 1 and 2 are chosen, with calculated energy values $\Delta E_1^{\ddagger} = 15$ kcal/mol and $\Delta E_2^{\ddagger} = 22$ kcal/mol. It is impossible to know if both belong to the same reactant sub-valley α, or if snapshot 1 is in the reactant sub-valley α and snapshot 2 in the reactant sub-valley β. In the second case, both values should be introduced once in equation 1, but we are not sure if $\Delta E_1^{\ddagger}$ and $\Delta E_2^{\ddagger}$ are a good estimation of their corresponding $\Delta G_\alpha^{\circ\ddagger}$ and $\Delta G_\beta^{\circ\ddagger}$, respectively. In the first case, either $\Delta E_1^{\ddagger}$ or $\Delta E_2^{\ddagger}$ (but just one of them) should be introduced twice in equation 1 as an estimate of $\Delta G_\alpha^{\circ\ddagger}$ (twice due to the population factor P_α), but we are not sure about which of them is a good estimate of $\Delta G_\alpha^{\circ\ddagger}$, or even if none of them is. In a nutshell, the application of equation 1 as a result of the assumptions on which it is based turns out to be doubtful.

Anyway, the main problem when equation 1 is applied lies in the way how the considered snapshots are chosen. Because in the sum in equation 1 the smaller the value of $\Delta E_i^{\ddagger}$ the bigger the contribution, it has been suggested that the best result should be obtained by selecting or filtering in some way, e.g. based on catalytic parameters, a few snapshots with expected low energy barriers.[24] The limit case would consist of just picking up the

snapshot with the smallest $\Delta E_i^{\ddagger}$. Both approaches are commonly used in QM/MM studies of enzyme reactions. However, it has been accepted [33] that the high-energy barrier snapshots do not significantly contribute to that sum, but to the number N of snapshots used. So, if those high-energy barrier snapshots are neglected, $\Delta E_{AV}^{\ddagger}$ will be underestimated.

The accuracy of $\Delta E_{AV}^{\ddagger}$ as an estimation of the Gibbs free energy barrier $\Delta G^{\circ\ddagger}$ is usually [33] assessed taking as a reference a calculation of the potential of mean force (PMF) [36, 37] from umbrella sampling simulations.[38] At this point it would be interesting to think about the meaning of the free energy. Free energy is used to account for the contribution of not only a unique structure, but of an ensemble of structures. If only a unique structure is considered, the logarithmic transformation of the unique Boltzmann term leads to an energy. If an ensemble of structures is considered, the sum of the corresponding Boltzmann terms gives the partition function, and its logarithmic transformation leads to the free energy. However, we cannot in practice obtain the global partition function for the whole reactant valley. Then, if only one reactive minimum is chosen, with only one associated MEP and only one PMF, only an estimate of the value $\Delta G_{\alpha}^{\circ\ddagger}$ corresponding to a reactant sub-valley α will be obtained. That is, in building a PMF just a quite limited region of the configuration space around the MEP is actually sampled, very far from the whole reactant valley. Both, the reduced time length (up to tens of picoseconds in QM/MM calculations) per each window and the biasing potential to sample the window associated to each value of the reaction coordinate, confine the sampling just to the neighborhood of the selected MEP. By the way, what MEP? Depending on the chosen MEP, a high or a low free energy barrier will be obtained. Even in the case of a low free energy barrier, the probability of appearance of the corresponding starting reactant structure (and, hence, of its low barrier) remains unknown. The free energy perturbation method as usually applied[38] as an alternative to umbrella sampling presents a similar problem. In a nutshell, in cases where *instantaneous disorder* is relevant, with the appearance of a huge dispersion of nuclear configurations and barriers throughout the reactant valley, none of the particular values of $\Delta G_{\alpha}^{\circ\ddagger}$ might be actually representative of the free energy barrier $\Delta G^{\circ\ddagger}$ of the catalytic reaction. So, the comparison of $\Delta E_{AV}^{\ddagger}$ obtained by means of eq. 1 (where the snapshots i will be obtained from, e.g., a 100 ns long molecular dynamics trajectory) with a $\Delta G_{\alpha}^{\circ\ddagger}$ value calculated from umbrella sampling

simulations (with per window samplings up to tens of picoseconds) makes not much sense.

We propose here a more reliable way to deduce and apply an equation quite similar to equation 1. This deduction does not make any of the assumptions a) to f) detailed above. Free energy perturbation (FEP)[38] theory can be used to calculate the free energy difference between a reference system (0) with a Hamiltonian H_0 and a perturbed system (1) characterized by a Hamiltonian H_1. For a molecular dynamics simulation in an isothermal-isobaric ensemble (NPT) the Gibbs free energy difference between the perturbed 1 and the reference 0 systems can be classically written as

$$\Delta G_{0 \to 1} = G_1 - G_0 = -RT \ln \frac{\int e^{-\frac{H_1(\vec{r}^N, \ \vec{p}^N)+PV_1}{RT}} \, d\vec{r}^N d\vec{p}^N}{\int e^{-\frac{H_0(\vec{r}^N, \ \vec{p}^N)+PV_0}{RT}} \, d\vec{r}^N d\vec{p}^N} =$$

$$-RT \ln \langle e^{-\frac{(H_1 - H_o)+P(V_1 - V_0)}{RT}} \rangle_0 \tag{4}$$

where $\vec{r}^{(N)}$ and $\vec{p}^{(N)}$ denotes the set of 3N atomic Cartesian coordinates and conjugate momenta, respectively, H is the classical nuclear Hamiltonian, V_0 and V_1 are the volume of the systems 0 and 1, respectively, at each configuration (the additive term PV in the exponential of the integral is absent in a canonic (*NVT*) ensemble), R is the gas constant, and $\langle \ \rangle_0$ denotes an ensemble average over the isothermal-isobaric ensemble sampled from the equilibrated reference state. Because the molecular dynamics simulations only sample one reaction valley, there are not significant differences between V_0 and V_1 nor significant changes of the volume V_0 along the simulation. Then the volumes can be ignored in equation 4, and the simulation can be even carried out in the canonical ensemble (NVT) without appreciable error. Thus, to get the free energy difference between state 1 and 0, at each configuration of the equilibrated reference system the corresponding Hamiltonian is switched to the one corresponding to the perturbed system, and the differences $\Delta H = H_1 - H_o$ are collected to calculate the ensemble average. Using Cartesian coordinates, the kinetic term is separable from the potential term and cancels out when the difference is done.

To study the *instantaneous disorder* in an enzyme reaction we propose the following protocol. We define the reactant ensemble as the reference system 0, and the transition state (TS) ensemble as the perturbed system 1. A molecular dynamics simulation of the enzyme-substrate Michaelis complex is run to generate a lot of nuclear configurations. Many snapshots must be selected at random (that is, populating many different regions

of the reaction valley). For each selected snapshot, a potential energy minimum is located, and the corresponding MEP is built using a QM/MM method. Thus, each selected snapshot of the reference system 0 is associated to a transition state structure that belongs to the perturbed system 1. In this case, for each snapshot i $\Delta H = \Delta V_i^{\ddagger}$ (or even $\Delta E_i^{\ddagger}$), the QM/MM potential energy barrier of the corresponding MEP. Taking into account that the set of snapshots along the molecular dynamics simulation appears according to a Boltzmann distribution in the reference ensemble, the average in the second member of equation 1 is just a discretization of the integral in the last member of equation 4, provided that a big enough number of selected snapshots are included in equation 1. On the other hand, the minimization in the orthogonal directions to the MEP permits to assume that the low-energy configurations in the transition state ensemble are sampled enough by means of the simulation in the reactant ensemble, in such a way that the FEP calculation converges reasonably. Thus, the exponential average in the second member of equation 1 directly corresponds to the Gibbs free energy barrier $\Delta G^{o\ddagger}$ of the catalytic reaction when only one reactant valley exists. At this point it is worth mentioning that the free energy barrier depends (see equation 4) on the quotient between the partition functions of the transition state and the reactants. The greater the accessible fluctuations in a given system, the bigger the value of the corresponding partition functions. The adequate generation of those fluctuations in reactants is warranted by a long enough sampling in the reactant ensemble. On the other hand, the fluctuations in the transition state ensemble (usually less pronounced than in the reactants) are a consequence of the dispersion of the reactants that translates to the dispersion of the corresponding MEPs and the multiplicity of maxima of these MEPs.

At this point it is crucial to understand that any attempt to select only a few "suitable" snapshots with "convenient" configurations that lead to low energy barriers to be included in equation 1 is an inadequate procedure, which only fortuitously will provide a good estimation of the experimental free energy barrier. Conversely, as many selected snapshots as possible, belonging to as many different regions of the reaction valley as possible, should be included in the exponential average. Evidently, this fact could imply the calculation of a too huge number of QM/MM MEPs to be feasible with a reasonable computational cost. To circumvent this problem, a tested criterion (depending on each particular enzyme reaction) should be used to identify a priori those snapshots corresponding to precatalytic structures, that is, belonging to the limited part of the configurational space sampled by molecular dynamics simulations that might be adequate

for catalysis (with a not foreseen excessively high energy barrier). As explained above, the high-energy barrier snapshots do not significantly contribute to the sum in the second member of equation 1 (they just add roughly zeros), but they do increase the value of N in the denominator and do have to be accounted.

Taking into account all that, if N is the high number of randomly selected snapshots along the molecular dynamics simulation, and n is the number of precatalytic structures identified after the corresponding filter according to the adopted criterion, we can reach the following equation

$$\Delta G^{\circ\ddagger} = -RT ln\frac{n}{N} - RT ln\left\{\frac{1}{n}\sum_{i=1}^{n} exp\left(\frac{-\Delta E_i^{\ddagger}}{RT}\right)\right\} \tag{5}$$

This equation reflects the fact that the Gibbs free energy barrier of the enzyme reaction is due to two different contributions. The first term of the second member of the equation indicates how difficult it is for a precatalytic structure to appear throughout the sampling. The second term comes from the exponentially averaged contribution of the precatalytic structures (note that this average can be calculated using just a representative set of the precatalytic structures, in such a way that n in this second term becomes the number of structures included in that representative set). As a consequence, both the non-precatalytic and the precatalytic structures contribute to the Gibbs free energy barrier of the enzyme reaction, and the weight of each part is unknown a priori. Even more, in general, the precatalytic structures will present a considerable dispersion of energy barriers, and the existence of a number of high energy barriers will also provoke an increase of the Gibbs free energy. Then, it must be emphasized that the free energy barrier according to equation 5 is the result of a long sampling over the reactant valley. Conversely, the free energy barrier resulting from a PMF calculation just includes a quite short sampling in the neighborhood of only one MEP, and therefore it just explores a quite narrow region of the reactant valley.

To illustrate all those concepts, we will focus on four enzyme reactions. Firstly, the hydrolysis of a glycosidic bond catalyzed by the enzyme *Thermus thermophilus* β-glycosidase (Ttb-gly), a retaining glycosyl hydrolase, [39] and the mutant Y284P Ttb-gly. Next, the hydrogen abstraction reactions from C_{13} and C_7 of arachidonic acid (AA) catalyzed by the enzyme rabbit 15-lipoxygenase-1 (15-rLOX-1).[28]

3. Hydrolysis of a glycosidic bond catalyzed by the enzyme *Thermus thermophilus* β-glycosidase

We started with an enzyme reaction where both the molecular dynamics simulations of the enzyme-substrate Michaelis complex and the potential energy profiles corresponding to each selected snapshot were calculated at a QM/MM level, specifically, at the QM(SCC-DFTB)/AMBER level.

Ttb-gly catalyzes the hydrolysis of terminal, non-reducing β-D-glucosyl residues, for example from 4-nitrophenyl β-*D*-fucopyranoside (*p*NP-Fuc). The first step of the reaction (glycosylation) consists of the nucleophilic attack of the residue Glu338 to the anomeric carbon C_1 of the fucose (Fuc) ring and the cleavage of the glycosidic bond of the pNP-Fuc substrate. A covalent glycosyl-enzyme intermediate with Glu338 is formed. A protonated residue Glu164 makes this step easier by protonation of the glycosidic oxygen of the leaving group *p*NP. In the second step (deglycosylation) Glu164 now deprotonates the incoming water molecule, that attacks as a nucleophile on the carbon C_1 of the fucose ring in a hydrolysis reaction that breaks the glycosidic bond between fucose and Glu338 and produces fucose (see Figure 1). In the presence of an appropriate acceptor substrate, like *N*-methyl-*O*-benzyl-*N*-(β-*D*-glucopyranosyl)-hydroxylamine (BnON(Me)-Glc), transglycosylation can compete with hydrolysis in the deglycosylation step and lead to the formation of the *N*-methyl-*O*-benzyl-*N*-(β-*D*-fucopyranosyl(1-4)β-*D*-glucopyranosyl)-hydroxylamine product.
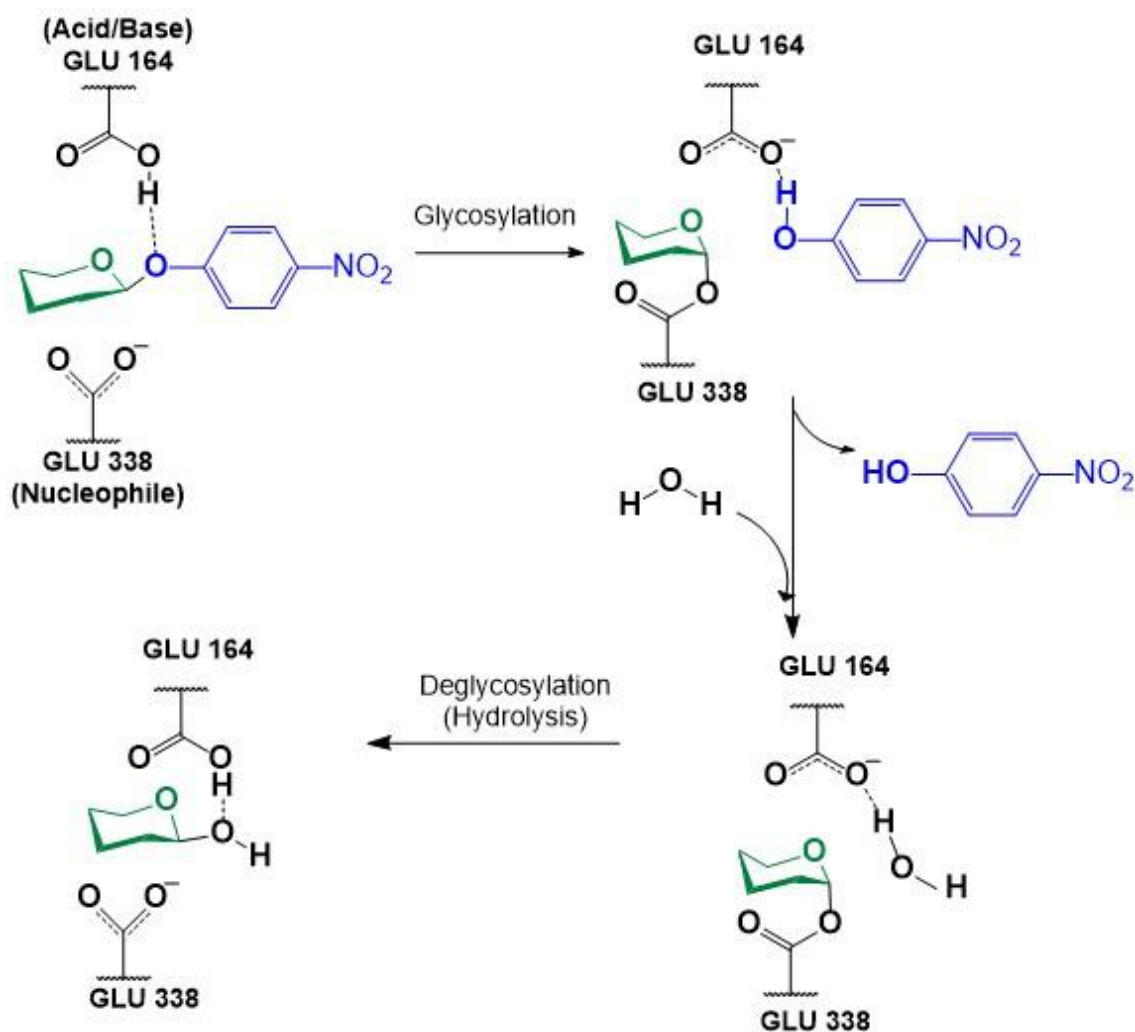
Figure 1. Catalytic mechanism of the glycosylation and deglycosylation (hydrolysis) of 4-nitrophenyl β-*D*-fucopyranoside catalyzed by the enzyme Ttb-gly. The fucose ring and the *p*NP leaving group are pictured in green and blue, respectively. A water molecule needs to be placed in the right place for the hydrolysis reaction to be possible.

In a previous paper[39] we studied the reaction mechanism of those glycosylation and deglycosylation steps (both for hydrolysis and transglycosylation). Now we take advantage of those simulations to build the setup for the study of the deglycosylation step corresponding to the hydrolysis of *p*NP-Fuc. We will present here the results corresponding to both wild-type Ttb-gly and the mutant Y284P Ttb-gly.

3.1. Methodological details for the enzyme *Thermus thermophilus* β-glycosidase

The starting structure to study the hydrolysis reaction catalyzed by wild-type Ttb-gly in this paper was the covalent glycosyl-enzyme intermediate obtained in our previous

work[39] as a result of the glycosylation process. The corresponding intermediate located in the case of the mutant Y284P Ttb-gly (result not published) was used for the hydrolysis reaction catalyzed by that mutant. The intermediate was modified by deleting all the pNP coordinates except those of the just formed hydroxyl group (due to the proton transfer from Glu164) that was converted into a water molecule using the PyMol program.[40] The system was solvated in a cubic box of preequilibrated TIP3P[41] water molecules and three sodium ions were added to neutralize it. More than 70,200 atoms were included in all. The coordinates of the system were relaxed at the MM level by restraining all protein atoms and the just built water molecule (sugar acceptor). Then, a short-restrained MD simulation (210 ps) at the MM level to allow the rearrangement of water molecules in the active site was performed under periodic boundary conditions (PBC) in the *NVT* canonical ensemble at 300 K. The temperature was controlled by Langevin dynamics.[42] All the MM calculations were done using the ff14SB[43] Amber force field for the protein, and the GLYCAM06j[44] atom types and parameters were employed for the sugar moiety. Along the MD simulations the covalent bonds containing hydrogen were constrained by means of the SHAKE algorithm[45], and the long-range electrostatic interactions were treated using the particle-mesh Ewald method.[46] A 1fs time step was used. in all the MD trajectories. All simulations were carried out employing the AMBER16 software (GPU (CUDA) version of the PMEMD [47, 48] package.

As will be explained the section 3.2, the final snapshot of the MD simulation at the MM level was chosen to initiate a QM(SCC-DFTB)/AMBER MD simulation, from which several snapshots were chosen to generate a lot of potential energy profiles corresponding to the hydrolysis reaction (see below). The potential energy profiles were calculated at the QM(SCC-DFTB)/AMBER level with the modular program package ChemShell [49, 50] employing TURBOMOLE[51] to obtain the QM energies and gradients at the SCC-DFTB[52] (Self-Consistent Charge Density Functional Tight-Binding) level. MM energies and gradients were evaluated using DL_POLY[53], implemented in the ChemShell package, using the AMBER force field. The electrostatic embedding scheme[54] was employed to let the MM point charges to polarize the electronic density of the QM region. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. The cubic box of water molecules was reduced to a 30 Å sphere surrounding the full protein for all the QM/MM calculations. All residues and water molecules within 15 Å from the anomeric carbon C_1 were allowed to move during the optimization process (active region, around 2100 atoms), while the remaining atoms were kept fixed. The QM region includes the fucose

ring, eight nearby waters, and the sidechains of Glu164, Glu338, and Tyr284 (67 atoms in all). Three link atoms were used to define the QM/MM boundary with the charge-shift approach.[36] The total charge of the QM regions is -1 a.u.

The QM/MM optimizations were performed using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS)[55] algorithm combined with the Hybrid Delocalized Internal Coordinate Scheme.[56] The reaction paths were scanned by performing harmonically restrained optimizations along a suitable reaction coordinate (see below) in steps of 0.1 Å. For each potential energy profile, the highest potential energy structure was taken as an approximation to the corresponding transition state structure.

3.2. Results for wild-type Ttb-gly

As explained in section 3.1., the final snapshot of the MD simulation at the MM level was chosen as starting point for a QM(SCC-DFTB)/AMBER DM simulation without any restraints. Firstly, the system was heated in the *NVT* canonical ensemble increasing the temperature along 60 steps, 100 fs each, up to 300 K. Then, an equilibration period of 20 ps was performed. A final production run of 100 ps was done.

In order to apply equation 5 we selected N = 2000 snapshots uniformly distributed at equal intervals along that MD simulation. It is not possible in practice to calculate the hydrolysis potential energy paths corresponding to these *N* snapshots. For this reason, we looked for the precatalytic structures. It is evident that only those snapshots that involve a water molecule placed in the right place for the hydrolysis reaction to be possible are precatalytic, in the sense that only those will imply a reasonable potential energy barrier for the hydrolysis. Thus, we searched precatalytic snapshots where (see Figure 1) there is a water molecule with one of its hydrogen atoms closer than 2 Å to the proton acceptor oxygen atom ($OE2_{GLU164}$) of Glu164, and with its oxygen atom closer to the anomeric carbon atom C_1 of the fucose ring than any of its hydrogen atoms. According to this filter, only *n* = 146 precatalytic snapshots were identified. The energy of each one was minimized at the QM(SCC-DFTB)/AMBER level to obtain the associated reactant minimum. From these structures the corresponding QM(SCC-DFTB)/AMBER hydrolysis potential energy paths were calculated using the difference between the distance H_{WAT}-O_{WAT} and the distance C_1-O_{WAT} (O_{WAT} is the oxygen atom of the water molecule that is in the right place to produce the hydrolysis reaction, and H_{WAT} is its hydrogen atom that is going to protonate Glu164) to define the reaction coordinate R2. As expected, a wide dispersion of potential energy barriers was obtained (they range from

13

20.9 kcal/mol up to 31.8 kcal/mol), as seen in Figure 2. The arithmetic mean is 26.2 kcal/mol with a standard deviation of 2.4 kcal/mol. Although these values can be indicative of the size of the barriers and their dispersion, they cannot be used to estimate the free energy.
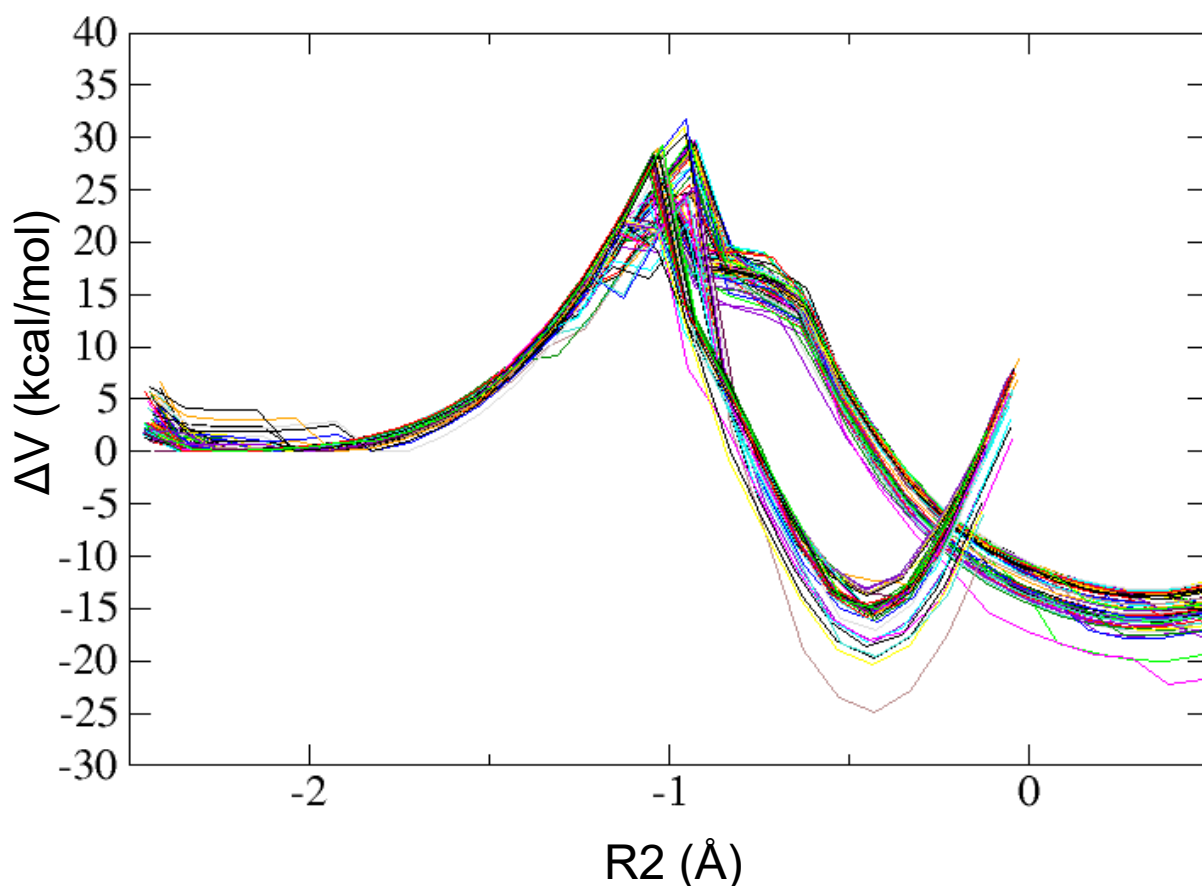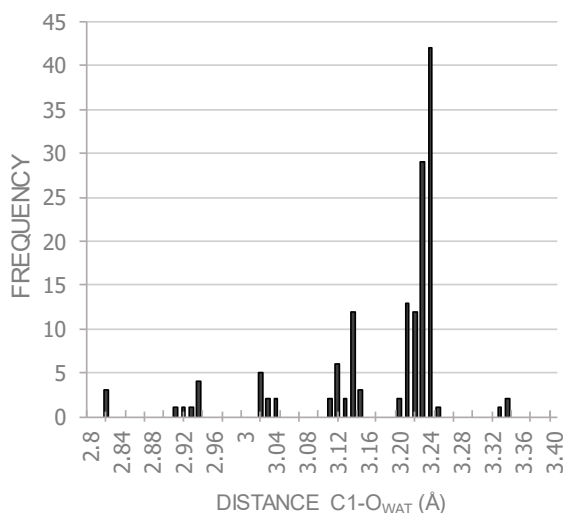

Figure 2. Potential energy paths calculated starting from each one of the 146 precatalytic snapshots identified for the hydrolysis reaction catalyzed by wild-type Ttb-gly. The energies are given in kcal/mol and the reaction coordinate R2 in Å.

Although it is not truly relevant four our purpose in this paper, we can mention that two kinds of potential energy paths are obtained. Some of them reach the minima corresponding to the products at a value of R2 about -0.5 Å, and the others already at the positive region of R2. This delay is caused by the fact that the proton transfer to $OE2_{GLU164}$ is not included in R2. Consequently, some hysteresis appears in the region of products of some energy paths, delaying the hydrogen transfer and reaching the product minima at

more advanced values of R2 (that is, allowing that Glu338 gets further away from the anomeric carbon atom C_1 of the fucose ring).

It is interesting to observe that the filter we have used leads to 146 precatalytic snapshots where the distribution of distances C_1-O_{WAT} (that is not included as a criterium for the filter) is relatively concentrated near 3.2 Å (see Figure 2a), with some structures with distances C_1-O_{WAT} slightly shorter (by 0.2 Å – 0.3 Å). In addition, we can see in Figure 3b that there is no correlation at all between these distances C_1-O_{WAT} and the potential energy barriers. Likewise, there is no correlation either between the distances corresponding to the proton to be transferred - oxygen atom ($OE2_{GLU164}$) of Glu164 (also very grouped) and the potential energy barriers, although in this case a very slight increasing trend of the barriers as the distance augments can be glimpsed (see Figure 3c). In a nutshell, the distribution of distances C_1-O_{WAT} and H_{WAT} - $OE2_{GLU164}$ are quite narrow within the set of precatalytic snapshots and cannot explain the considerable dispersion of potential energy barriers. Many other geometrical parameters of the Michaelis complex play here a role determining those barriers, and we cannot focus just on the more relevant geometrical distances from the point of view of the chemical reaction.

a)
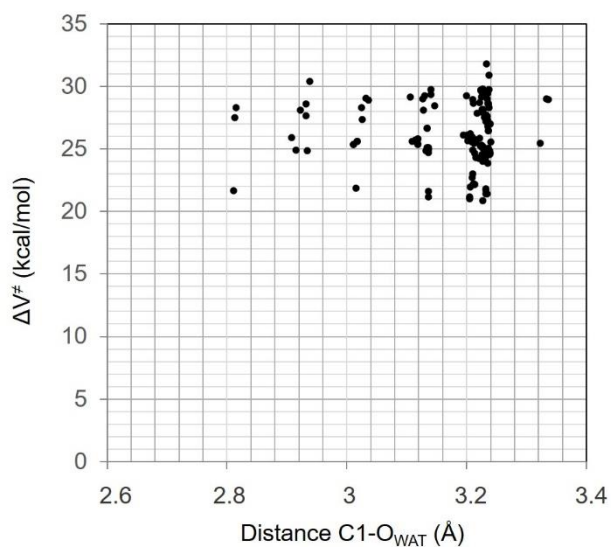
b)
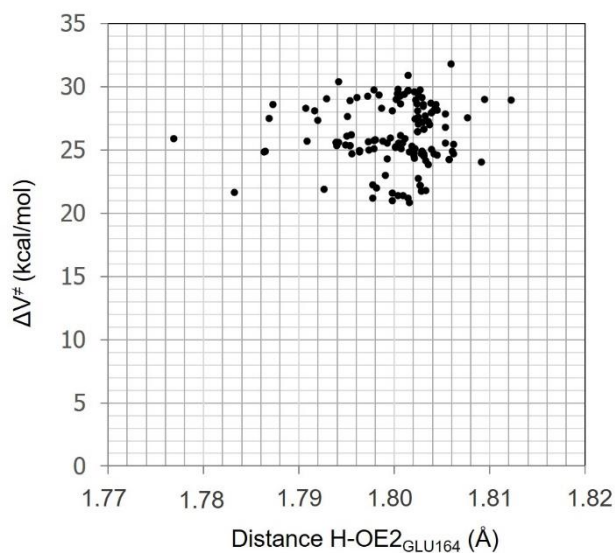
c)

Figure 3. Histogram of the distances C_1-O_{WAT} for the 146 precatalytic snapshots identified for the hydrolysis reaction catalyzed by wild-type Ttb-gly (a); potential energy barriers as a function of the distances C_1-O_{WAT} (b) or the distances proton to be transferred (H_{WAT}) - oxygen atom ($OE2_{GLU164}$) of Glu164 (c) in these precatalytic snapshots. Energies are given in kcal/mol and distances in Å.

Going now to equation 5, with $n = 146$ and $N = 2000$ the contribution of the non-precatalytic structures (first term of the second member of the equation) to the Gibbs free energy barrier is 1.6 kcal/mol. The contribution of the 146 precatalytic snapshots is the

exponential average (second term of the second member of the equation) leading to 22.8 kcal/mol. In all, the Gibbs free energy barrier amounts to 24.4 kcal/mol. It can be clearly seen that the contribution of the non-precatalytic structures is not negligible at all, and that it appears even in the case of short molecular dynamics simulations (100 ps).

3.3. Results for the mutant Y284P Ttb-gly.

Then, we studied the same reaction enzyme but with quite smaller energy barriers. To this aim we choose the mutant Y284P Ttb-gly. We proceeded with this mutant in the same way as for wild-type Ttb-gly. Again, we selected N = 2000 snapshots uniformly distributed at equal intervals along the QM(SCC-DFTB)/AMBER 100 ps molecular dynamics simulation without any restraints. Using the same filter, we found this time $n = 202$ precatalytic snapshots. The energy of each one was minimized at the QM(SCC-DFTB)/AMBER level to obtain the associated reactant minimum. From these structures the corresponding QM(SCC-DFTB)/AMBER hydrolysis potential energy paths were calculated using the difference between the distance H_{WAT}-O_{WAT} and the distance C_1-O_{WAT} to define the reaction coordinate R2. Now, even a wider dispersion of potential energy barriers was obtained (they range from 12.4 kcal/mol up to 33.6 kcal/mol), as seen in Figure 4. The arithmetic mean is 23.1 kcal/mol with a standard deviation of 4.7 kcal/mol. Some delay of the hydrogen transfer also appears in some cases, so causing again the existence of two kinds of potential energy paths differing in the product's region.

According to equation 5, with $n = 202$ and $N = 2000$ the contribution of the non-precatalytic structures (first term of the second member of the equation) to the Gibbs free energy barrier is 1.4 kcal/mol. The contribution of the 202 precatalytic snapshots is the exponential average (second term of the second member of the equation) leading to 15 kcal/mol. In all, the Gibbs free energy barrier amounts to 16.4 kcal/mol.
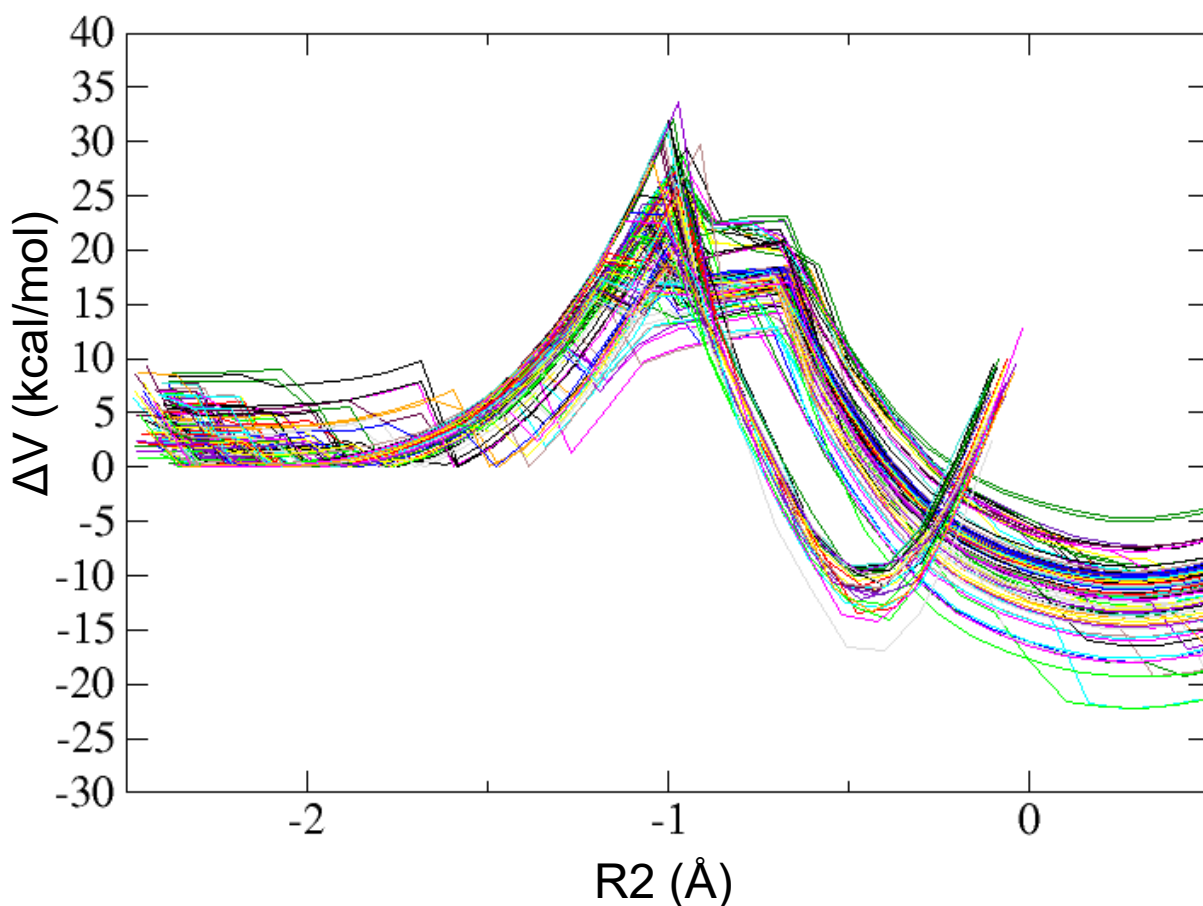
Figure 4. Potential energy paths calculated starting from each one of the 202 precatalytic snapshots identified for the hydrolysis reaction catalyzed by mutant Y284P Ttb-gly. The energies are given in kcal/mol and the reaction coordinate R2 in Å.

4. Hydrogen abstraction reactions from arachidonic acid catalyzed by the enzyme rabbit 15-lipoxygenase-1.

Finally, we studied two enzyme reactions where the molecular dynamics simulations of the enzyme-substrate Michaelis complex were run at MM level and the potential energy profiles corresponding to the selected snapshots were calculated at QM(DFT)/MM level. The MM level for the molecular dynamics simulations makes it possible to run trajectories a thousand times longer than in the case of Ttb-gly.

Lipoxygenases (LOXs) are non-heme iron dioxygenases that catalyze the oxidation of polyunsaturated fatty acids, producing conjugated hydroperoxide fatty acids. These hydroperoxidations turns out to be very highly regio- and stereospecific. The rate determining step of the overall hydroperoxidation process is the initial hydrogen atom

abstraction from a bisallylic methylene of one of the (Z,Z)-1,4-pentadiene units of the substrate fatty acid by the Fe(III)-OH$^-$ cofactor, yielding Fe(II)-OH$_2$ and a pentadienyl π radical.[57-61] In mammals, the main substrate of LOXs is arachidonic acid, whose hydroperoxidation initiates the biosynthesis of signaling lipid mediators that are crucial for human health and are implicated in inflammatory processes.[62]

The crystallographic structure for rabbit reticulocyte-type 15-lipoxygenase-1 (15-rLOX-1) dimer was the first one resolved for a mammalian LOX (PDB code 2P0M).[63] 15-rLOX-1 converts AA into hydroperoxyeicosatetraenoic acids (HPETEs) with the following distribution of products: 97 % of 15S-HpETE (after abstracting one hydrogen atom at C$_{13}$ of AA), 3 % of 12S-HpETE (after abstracting one hydrogen atom at C$_{10}$ of AA). No signal of 9S-HpETE (after abstracting one hydrogen atom at C$_7$ of AA) was detected.[64]

In the present paper we will study the hydrogen transfer reactions from C$_{13}$ and C$_7$ of AA catalyzed by the enzyme rabbit 15-lipoxygenase-1. The case of the hydrogen abstractions from C$_7$ was chosen because their energy barriers are very high. To our purpose, no matter that hydrogen abstraction from C$_7$ is actually too slow to be competitive.

4.1. Methodological details for the enzyme rabbit 15-lipoxygenase-1.

The monomer A and the ligand bound at the active site of monomer B in the x-ray structure of rabbit 15-LOX-1 dimer [63] were removed. The monomer B was protonated with the H++ web-server.[65,66] A pH = 6.5 for the titratable residues was employed. The protonation state for the iron coordination sphere was corrected by hand to ensure a correct description.

The program GOLD5.8.0[67] was employed to carry out docking calculations for AA within the pocket of the monomer B. The conformational space was explored using the genetic algorithm. The binding site was defined as a 20 Å radius sphere centered on the iron atom. Binding free energies were estimated by the ChemScore fitness function.

The best pose was chosen to run an MD simulation. The ff14SB[43] Amber force field was used for the protein atoms, whereas the specific parameters for AA were taken from a previous work.[68] The GAFF2 [69,70] library was employed as the source for these parameters. Employing the MCPB.py procedure[71] within the bonded model and the Seminario method for the force constants calculations,[72] specific MM parameters were developed for the iron atom and its first coordination sphere (His361, His366, His541, His545, Ile663 and the OH$^-$ group).[73] This first coordination sphere was optimized

employing the B3LYP/6-31G(d) level of theory and their atomic charges were assigned using the Merz-Kollman RESP procedure.[74]

The tLeap program was used to combine enzyme and substrate files into the protein-substrate complex, and to solvate it with an orthorhombic box of pre-equilibrated TIP3P[41] waters as well as to neutralize its total charge by addition of sodium cations. The final system contains nearly 92150 atoms of which about 10600 belong to the protein. The rest of atoms correspond to water molecules and salt ions. All molecular dynamics simulations were run using the AMBER 14 GPU (CUDA) version of the PMEMD package[47,48] as mentioned above. To begin with, the Michaelis complex was submitted to 22000 energy minimization steps combining the steepest descent and conjugate gradient methods to remove bad contacts. In the first 6000 steps, harmonic restraints were applied to the protein and substrate atoms with a force constant of 5.0 kcal mol^{-1} Å^{-2}, so that only the solvent and ions were relaxed. In the following 6000 steps, harmonic restraints were applied to the protein backbone and the substrate heavy atoms with the same force constant as before. Nonetheless, the whole complex was kept free during the last 10000 steps. After that, MD simulations using periodic boundary conditions were carried out. The system was gradually heated from 0 K to 300 K for a period of 200 ps. The following step consisted of running a MD simulation of 1 ns within the NTP ensemble (300 K, 1 bar) to adjust the volume of the orthorhombic box so that a density of around 1 g cm^{-3} was achieved. During the heating and the compressing, harmonic restraints were applied to the protein backbone and the substrate heavy atoms with a force constant of 5.0 kcal mol^{-1} Å^{-2}, whereas the rest of the system was kept without restraints. The temperature was controlled by Langevin dynamics,[42] while the pressure was adjusted by the Berendsen barostat.[75] Next, an equilibration stage of 10 ns at constant temperature (300 K) and volume was performed. Finally, a production period with a length of 100 ns was run within the NVT ensemble. A time step of 2 fs was used along the whole MD trajectory. All bonds and bends containing hydrogen atoms were also constrained by the SHAKE algorithm.[45] The non-bonding interactions have been calculated with a cutoff of 9 Å.

The QMM/MM potential energy profiles were calculated at the QM(DFT)/AMBER level following the procedure described in section 3.1. To carry out the initial QM/MM optimizations for each selected snapshot all water molecules outside a 17 Å radius volume centered on AA were removed. The active region includes all residues and water molecules inside a 15 Å radius sphere centered on either C_{13} of AA when the $H_{13\text{-proS}}$ hydrogen abstraction is considered, or C_7 when the $H_{7\text{-proS}}$ hydrogen abstraction is studied.

This region moves freely ($\approx$ 2100 atoms) while the rest of atoms were kept frozen. Roughly 12000 atoms were included into the QM/MM calculations.

The B3LYP hybrid functional[76] was used to describe the QM region. The 6-31G(d) People basis set[77] was employed for the C, H, O and N atoms, while the LANL2DZ basis set[78] was used for the Fe atom. The QM region (Figure 5) was defined by all atoms of AA which are found between C_5 and C_{15}, 11 atoms for each His residue in the iron coordination sphere (His361, His366, His541 and His545), 3 atoms of the Ile terminal residue (Ile 663) in the iron coordination sphere and the Fe^{III} - OH^- cofactor. A multiplicity corresponding to a sextet was used.[79] The net charge of the QM region is 1 a.u. Seven link atoms were used, five between the bonds Cα-QM atoms of the five residues in the iron coordination sphere and two bonded to the aliphatic carbon atoms of the lipid substrate (placed between C_4-C_5 and C_{15}-C_{16}).
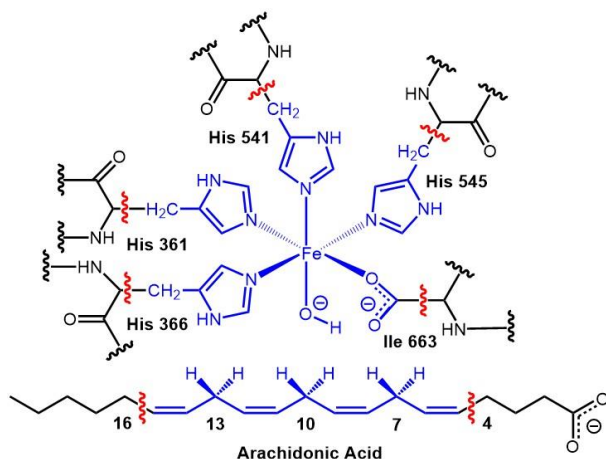

Figure 5: QM/MM partition for the hydrogen abstraction reactions from arachidonic acid catalyzed by the enzyme rabbit 15-lipoxygenase-1. QM atoms are depicted in blue. The boundary between QM and MM regions is indicated by red wavy lines.

4.2. Results for the hydrogen transfer reactions from AA

As explained in section 4.1., a production period with a length of 100 ns was run within the NVT ensemble using the force field describe there. We selected N = 10000 snapshots uniformly distributed at equal intervals along that MD simulation. Then we looked for the precatalytic structures. We searched precatalytic snapshots where the initial distance between the hydrogen atom that will be abstracted, $H_{13\text{-} proS}$ or $H_{7\text{-} proS}$, and the oxygen atom of the hydrogen acceptor, that is, the oxygen of the Fe(III)-OH^- cofactor, is smaller than 3.0 Å or 4.0 Å, respectively, for the reaction to take place with a reasonable potential

energy barrier. Furthermore, the distance from the corresponding carbon atom, C_{13} or C_7, to that oxygen atom had to be greater than the distance from $H_{13\text{-proS}}$ or $H_{7\text{-proS}}$, respectively, so ensuring that the corresponding C−H bond is properly oriented for the hydrogen abstraction (see Figure 5). Applying this filter, only $n = 3383$ or $n = 33$ precatalytic snapshots were identified for the $H_{13\text{- proS}}$ or $H_{7\text{-proS}}$ abstractions, respectively. 25 structures were randomly selected as representative of each set of precatalytic snapshots. The energy of each one was minimized at the QM(B3LYP)/AMBER level to obtain the associated reactant minimum. From these structures the corresponding QM(B3LYP)/AMBER potential energy paths were calculated using as reaction coordinate the difference between the distance of the breaking bond (C_{13}−$H_{13\text{-proS}}$ or C_7−$H_{7\text{-proS}}$, for the $H_{13\text{-proS}}$ or $H_{7\text{-proS}}$ abstractions, respectively) and the forming bond ($H_{13\text{-proS}}$−O or $H_{7\text{-proS}}$−O, respectively). The corresponding potential energy profiles are displayed in Figures 6 and 7. Again, a wide dispersion of potential energy barriers was obtained. They range from 12.8 kcal/mol up to 26.5 kcal/mol in Figure 6, and from 39.4 kcal/mol up to 66 kcal/mol in Figure 7. The arithmetic mean is 18.9 kcal/mol with a standard deviation of 4.2 kcal/mol for the $H_{13\text{-proS}}$ hydrogen abstraction, whereas the arithmetic mean is 49.6 kcal/mol with a standard deviation of 6.3 kcal/mol for the $H_{7\text{-proS}}$ hydrogen abstraction. As expected, the potential energy barriers corresponding to the $H_{13\text{-proS}}$ abstraction are quite smaller than the ones corresponding to the $H_{7\text{-proS}}$ abstraction.
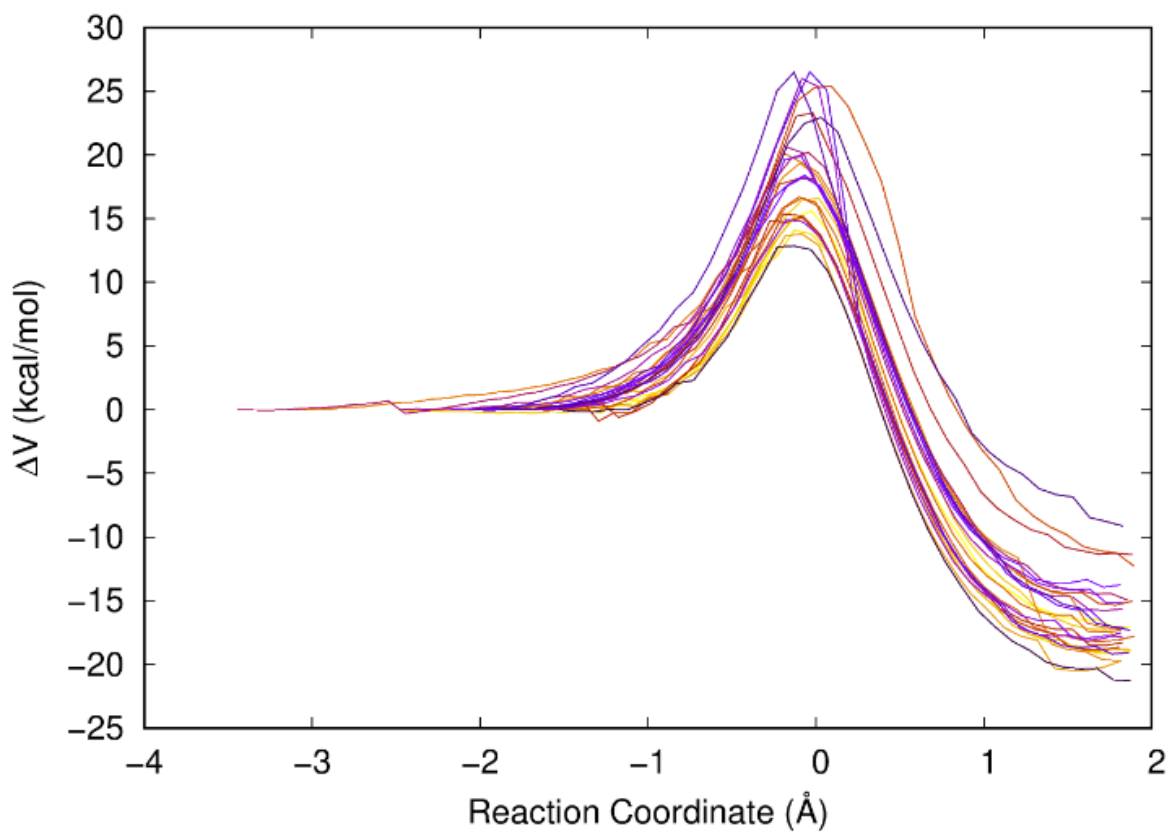
Figure 6. Potential energy paths calculated starting from each one of the 25 randomly selected precatalytic snapshots for the $H_{13\text{-pro}S}$ abstraction reaction catalyzed by the enzyme rabbit 15-lipoxygenase-1.
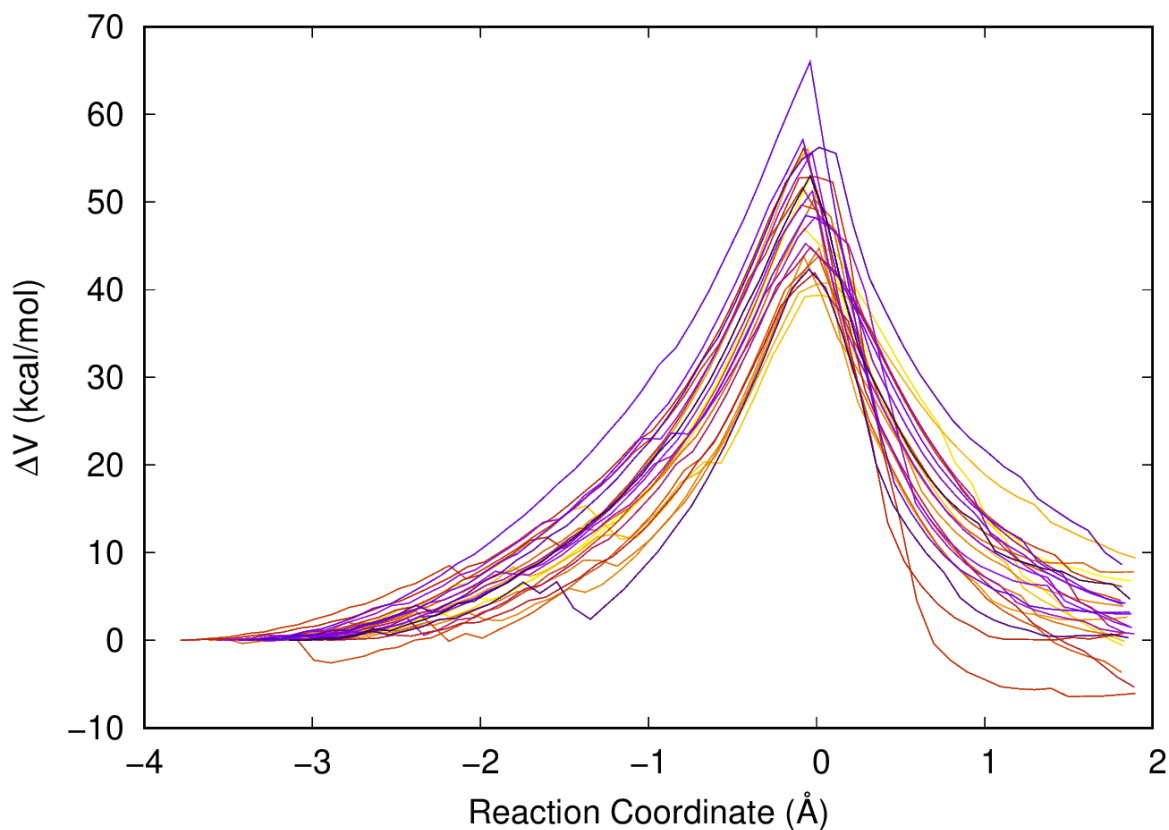
Figure 7. Potential energy paths calculated starting from each one of the 25 randomly selected precatalytic snapshots for the $H_{7\text{-proS}}$ abstraction reaction catalyzed by the enzyme rabbit 15-lipoxygenase-1.

There is a quite considerable dispersion of values of relevant geometrical parameters along the 100 ns molecular dynamics simulation. This can be seen both following the evolution on time of the distances $C_{13}-OH^-$ and C_7-OH^- (see Figure 8) and watching the histograms of those distances (see Figure 9). Indeed, the filtering leading to the precatalytic snapshots largely reduce this geometrical dispersion, although a wide dispersion of potential energy barriers holds and no correlation with the distances $C_{13}-OH^-$ or C_7-OH^- can be found (Figure 10).
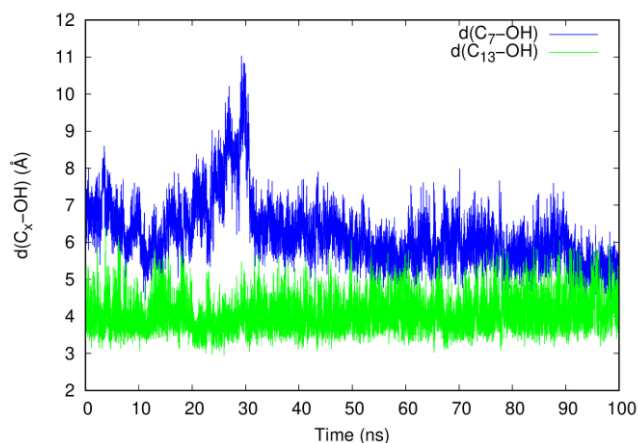
Figure 8. Distances C_{13}−OH (in Å, green line) and C_7−OH (in Å, blue line) along the 100 ns MM(AMBER) molecular dynamics simulations for the 15-LOX-1:AA Michaelis complex. OH stands for the OH$^-$ group of the Fe(III)−OH− cofactor of 15-LOX-1.
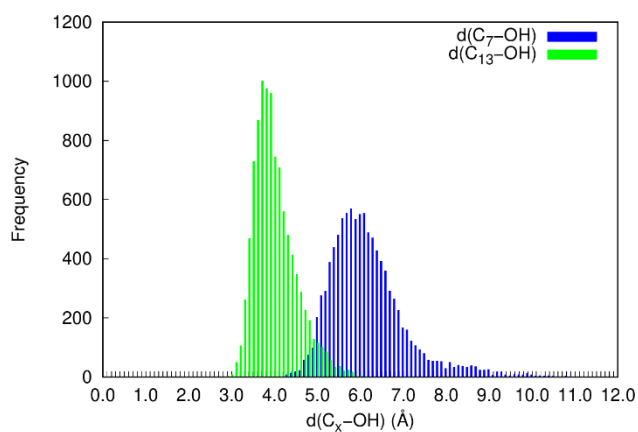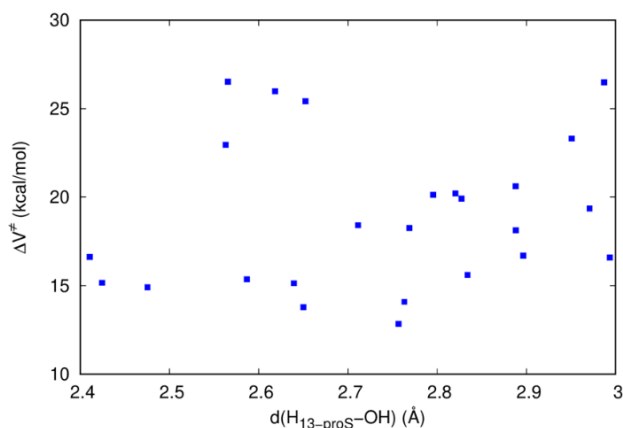

Figure 9. Histogram of the distance C_{13}−OH (in Å, green lines) and C_7−OH (in Å, blue lines) corresponding to the 100 ns MM(AMBER) molecular dynamics simulations for the 15-LOX-1:AA Michaelis complex. OH stands for the OH$^-$ group of the Fe(III)−OH− cofactor of 15-LOX-1.
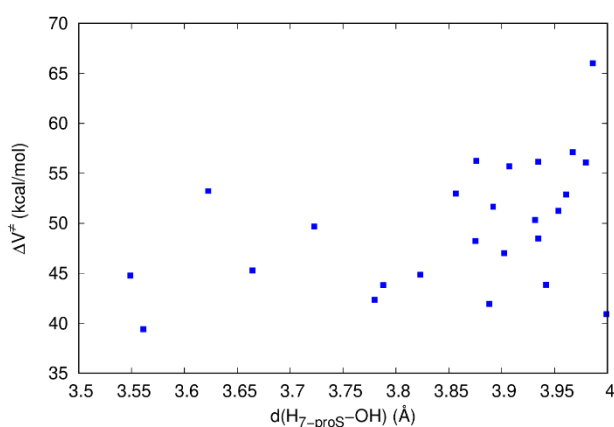
a)

b)

Figure 10. Potential energy barriers as a function of the distances $H_{13\text{-proS}}$ −OH (a) or the distances $H_{7\text{-proS}}$ −OH (b) in the randomly selected precatalytic snapshots in the case of the enzyme rabbit 15-lipoxygenase-1. OH stands for the OH⁻ group of the Fe(III)−OH⁻ cofactor of 15-LOX-1.

Using now equation 5, for the $H_{13\text{-proS}}$ abstraction $n = 3383$ and $N = 10000$. So, the contribution of the non-precatalytic structures (first term of the second member of the equation) to the Gibbs free energy barrier is 0.6 kcal/mol. The contribution of the precatalytic snapshots can be estimated from the exponential average (second term of the second member of the equation) including the representative 25 randomly selected structures, thus leading to 14.6 kcal/mol. In all, the Gibbs free energy barrier is 15.2 kcal/mol. As for the $H_{7\text{-proS}}$ abstraction, $n = 33$ and $N = 10000$. This gives 3.4 kcal/mol for the first term, 41.3 kcal/mol for the second, and 44.7 kcal/mol in all. The contribution of the Gibbs free energy corresponding to the first term is the cost that must be paid to

reduce the wide set of structures corresponding to the Michaelis complex to the narrow set of precatalytic structures, and it cannot be neglected. An analysis of the convergence of these results is presented in the Electronic Supplementary Information.

A final consideration is worth making. Let us suppose that we carry out umbrella sampling simulations starting from the selected precatalytic structure that originates the smallest potential energy barrier for each hydrogen atom abstraction. That is, the one corresponding to a barrier of 12.8 kcal/mol for the $H_{13-proS}$ abstraction reaction, and the one corresponding to 39.4 kcal/mol for the H_{7-proS} abstraction reaction. A reasonable window width for QM(DFT)/MM umbrella sampling simulations in enzyme reactions is 3 ps (50 ps if a semi-empirical method is used to describe the QM zone). The point here is up to what extent the structures generated along this short sampling (but forced by the long computer time involved in the QM/MM molecular dynamics simulations) are representative of the nuclear configurations of the whole reaction valley. In Figure 11 we have pictured the evolution of the distances C_{13}−OH and C_7−OH along 3 ps MM(AMBER) molecular dynamics simulation for the 15-LOX-1:AA Michaelis complex, and the corresponding histograms (Figure 12). These two Figures must be compared with the quite longer 100 ns MM(AMBER) molecular dynamics simulations shown above in Figures 8 and 9. It is clear that the short simulations used in umbrella sampling simulations just cover a narrow neighborhood of the starting structure (only a sub-valley), in such a way that the contribution to the Gibbs free energy of a huge number of nuclear configurations distributed in the whole reaction valley is absolutely lost. This way, the choice of only a particular structure to calculate the Gibbs free energy barrier for an enzyme reaction using either umbrella sampling simulations or the free energy perturbation method neglects very considerable contributions of the reaction valley of the enzyme-substrate Michaelis complex and can lead to a significant error. We think that the present paper introduces a meaningful warning in this sense that deserves consideration. Two final points are worth mentioning. Firstly, the generation of snapshots along molecular dynamics simulations to evaluate the catalytic effect of an enzyme has already been done previously. [80, 81] On the other hand, the lengths of the molecular dynamics simulations have to be chosen carefully, because the convergence of the free energy perturbation calculations can require a considerable time, even for gas phase reactions.[82]
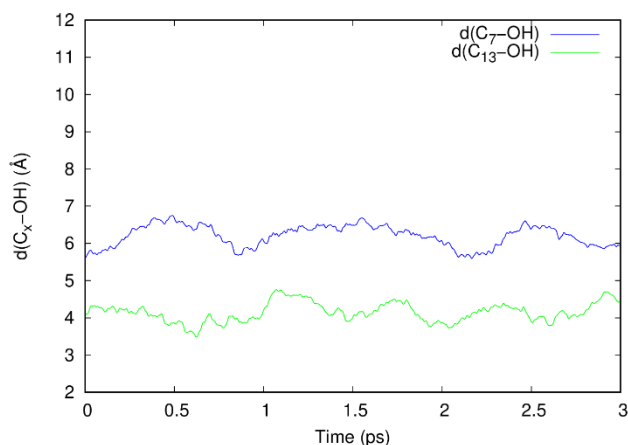
27

Figure 11. Distances $C_{13}-OH$ (in Å, green line) and C_7-OH (in Å, blue line) along the 3 ps MM(AMBER) molecular dynamics simulations for the 15-LOX-1:AA Michaelis complex starting from the selected precatalytic structure that originates the smallest potential energy barrier for each hydrogen atom abstraction. OH stands for the OH^- group of the $Fe(III)-OH^-$ cofactor of 15-LOX-1.

Figure 12. Histogram of the distance $C_{13}-OH$ (in Å, green lines) and C_7-OH (in Å, blue lines) corresponding to the 3 ps MM(AMBER) molecular dynamics simulations for the 15-LOX-1:AA Michaelis complex starting from the selected precatalytic structure that originates the smallest potential energy barrier for each hydrogen atom abstraction. OH stands for the OH^- group of the $Fe(III)-OH^-$ cofactor of 15-LOX-1.
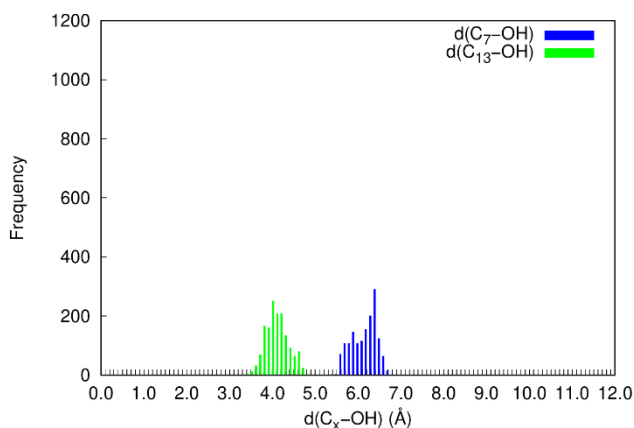
28

5. Conclusions

Instantaneous disorder consists of instantaneous dynamic fluctuations in the enzyme-substrate Michaelis complex with timescales much smaller than k_{cat}. This leads to different energy barriers that cannot be experimentally detected. They are the consequence of a huge number of nuclear configurations in local equilibrium with each other within a reaction valley. Quantum mechanics/molecular mechanics methods can build multiple minimum-energy paths connecting a potential energy minimum inside the reactant valley with a potential energy minimum in the associated product valley, so providing a wide dispersion of potential energy barriers. An exponential average has been the most used method to estimate the Gibbs free energy barrier of the enzyme reaction. However, the way how this exponential average has been justified and applied in practice is questionable. In this paper we propose a protocol based on the free energy perturbation theory that more properly justify the suitable use of the exponential average and that provides a practical way to determine the Gibbs free energy barrier of the reaction. This value includes the contribution of both the non-precatalytic and the precatalytic structures as a result of a long sampling over the reaction valley. This procedure does not require the arbitrary choice of a particular structure in whose quite narrow neighborhood a potential of mean force is calculated.

We have used the hydrolysis of a glycosidic bond catalyzed by the enzyme *Thermus thermophilus* β-glycosidase, and the mutant Y284P Ttb-gly, and the hydrogen abstraction reactions from C_{13} and C_7 of arachidonic acid catalyzed by the enzyme rabbit 15-lipoxygenase-1 to exemplify our procedure.

Electronic Supplementary Information

Test of convergence of the Gibbs free energy barrier and its contributions appearing in equation 5; Results using a slightly more restrictive filter to identify the precatalytic snapshots.

References

1   R. Wolfenden and M. J. Snider, *Acc. Chem. Res.*, 2001, **34**, 938–945.

2   B. G. Miller and R. Wolfenden, *Annu. Rev. Biochem.*, 2002, **71**, 847–885.

3   J. P. Richard, *Biochemistry*, 2013, **52**, 2009–2011.

4   A. Warshel and R. P. Bora, *J. Chem. Phys.,* 2016, **144**, 180901-1 to 180901-17.

5   A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder and R. D. Young, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 5000–5004.

6   S. Ferrer, I. Tuñón, S. Martí, V. Moliner, M. Garcia-Viloca, À. González-Lafont and J. M. Lluch, *J. Am. Chem. Soc.*, 2006, **128**, 16851–16863.

7   W. Min, B. P. English, G. Luo, B. J. Cherayil, S. C. Kou and X. S. Xie, *Acc. Chem. Res.*, 2005, **38**, 923–931.

8   N. M. Antikainen, R. D. Smiley, S. J. Benkovic and G. G. Hammes, *Biochemistry*, 2005, **44**, 16835–16843.

9   I. F. Thorpe and C. L. Brooks, *J. Am. Chem. Soc.*, 2005, **127**, 12997–13006.

10   Q. Xue and E. S. Yeung, *Nature*, 1995, **373**, 681–683.

11   W. Tant and E. S. Yeung, *Anal. Chem.*, 1997, **69**, 4242–4248.

12   S. C. Kou, B. J. Cherayil, W. Min, B. P. English and X. S. Xie, *J. Phys. Chem. B*, 2005, **109**, 19068–19081.

13   R. D. Smiley and G. G. Hammes, *Chem. Rev.*, 2006, **106**, 3080–3094.

14   H. P. Lu, Xun, L. Y., Xie, X. S, *Science.*, 1998, **282**, 1877–1882.

15   H. Yang, G. Luo, P. Karnchanaphanurach, T. M. Louie, I. Rech, S. Cova, L. Xun and X. S. Xie, *Science.*, 2003, **302**, 262–266.

16   A. Gupta, A. Milias-Argeitis and M. Khammash, *J. R. Soc. Interface*, 2017, **14**,

17   A. J. M. Ribeiro, D. Santos-Martins, N. Russo, M. J. Ramos and P. A. Fernandes, *ACS Catal.*, 2015, **5**, 5617–5626.

18   M. Sanchez-Martinez, E. Marcos, R. Tauler, M. Field and R. Crehuet, *J. Phys. Chem. B*, 2013, **117**, 14261–14272.

19   R. Lonsdale, K. T. Houghton, J. Zurek, C. M. Bathelt, N. Foloppe, M. J. De Groot, J. N. Harvey and A. J. Mulholland, *J. Am. Chem. Soc.*, 2013, **135**, 8001–8015.

20   P. Saura, R. Suardíaz, L. Masgrau, J. M. Lluch and À. González-Lafont, *ACS Catal.*, 2014, **4**, 4351–4363.

21   J. Soler, P. Saura, D. Garciá-López, L. Masgrau, J. M. Lluch and À. González-Lafont, *J. Phys. Chem. B*, 2016, **120**, 1950–1960.

22   S. Adel, F. Karst, À. González-Lafont, M. Pekárová, P. Saura, L. Masgrau, J. M. Lluch, S. Stehling, T. Horn, H. Kuhn and D. Heydeck, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, E4266–E4275.

23   P. Saura, I. Kaganer, D. Heydeck, J. M. Lluch, H. Kühn and À. González-Lafont, *Chem. - A Eur. J.*, 2018, **24**, 962–973.

24   B. Von Der Esch, J. C. B. Dietschreit, L. D. M. Peters and C. Ochsenfeld, *J. Chem. Theory Comput.*, 2019, **15**, 6660–6667.

25   Y. Zhang, J. Kua and J. A. McCammon, *J. Phys. Chem. B*, 2003, **107**, 4459–4463.

26   P. Hu and and Y. Zhang, *J. Am. Chem. Soc.*, 2006, **128**, 1272-1278

27   A. Lodola, M. Mor, J. Zurek, G. Tarzia, D. Piomelli, J. N. Harvey and A. J. Mulholland, *Biophys. J.*, 2007, **92**, L20.

28   M. Roca, B. Messer, D. Hilvert and A. Warshel, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 13877–13882.

29   S. J. Benkovic, G. G. Hammes and S. Hammes-Schiffer, *Biochemistry*, 2008, **47**, 3317–3321.

30   A. Lodola, J. Sirirak, N. Fey, S. Rivara, M. Mor and A. J. Mulholland, *J. Chem. Theory Comput.*, 2010, **6**, 2948–2960.

31   R. Lonsdale, J. N. Harvey and A. J. Mulholland, *J. Phys. Chem. B*, 2010, **114**, 1156–1162.

32   A. J. M. Ribeiro, M. J. Ramos and P. A. Fernandes, *J. Am. Chem. Soc.*, 2012, **134**, 13436–13447.

33   A. M. Cooper and J. Kästner, *ChemPhysChem*, 2014, **15**, 3264–3269.

34   I. Logunov and K. Schulten, *J. Am. Chem. Soc.*, 1996, **118**, 9727–9735.

35   R. Lonsdale, S. Hoyle, D. T. Grey, L. Ridder and A. J. Mulholland, *Biochemistry*, 2012, **51**, 1774–1786.

36   L. Masgrau, À. González-Lafont and J.M. Lluch, *Computational Enzymology: Modelling Biological Catalysts*, in *Computational Tools for Chemical Biology*, Royal Society of Chemistry books, Chemical Biology No. 3, Chapter 5, p. 108 - 144, ed. by S. Marín-Santamaría, Cambridge, UK, 2018.

37   G. K. Schenter, B. C. Garrett and D. G. Truhlar, *J. Chem. Phys.*, 2003, **119**, 5828–5833.

38   *Series in Chemical Physics: Free Energy Calculations. Theory and Applications in Chemistry and Biology*, ed. Ch. Chipot and A. Pohorille, Springer, New York, 2007, vol. 86, pp. 1–517.

39   S. Romero-Téllez, J. M. Lluch, À. González-Lafont and L. Masgrau, *Front. Chem.*, 2019, **7**, 200.

40   The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

41   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

42   A. R. Leach, *Molecular modelling : Principles and Applications*, Eddison Wesley Longman Limited, Essex, England, 1996.

43   J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

44   K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622–655.

45   J.-P. Ryckaert, G. Ciccotti and H. J. . Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

46   U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

47   A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542–1555.

48   R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.

49   P. Sherwood, A. H. De Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer and C. Lennartz, *J. Mol. Struct. THEOCHEM*, 2003, **632**, 1–28.

50   S. Metz, J. Kästner, A. A. Sokol, T. W. Keal and P. Sherwood, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2014, **4**, 101–110.

51   R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.

52   M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk and T. Frauenheim, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1998, **58**, 7260–7268.

53   W. Smith and T. R. Forester, *J. Mol. Graph.*, 1996, **14**, 136–141.

54   D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580–10594.

55   D. C. Liu and J. Nocedal, *Math. Program.*, 1989, **45**, 503–528.

56   S. R. Billeter, A. J. Turner and W. Thiel, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2177–2186.

57   A. Catalano and A. Procopio, *Histol. Histopathol.*, 2005, 20, 969–975.

58   A. D. Dobrian, D. C. Lieb, B. K. Cole, D. A. Taylor-Fishwick, S. K. Chakrabarti and J. L. Nadler, *Prog. Lipid Res.*, 2011, 50, 115–131.

59   J. Z. Haeggström and C. D. Funk, *Chem. Rev.*, 2011, 111, 5866–5896.

60   Y. C. Joo and D. K. Oh, *Biotechnol. Adv.*, 2012, 30, 1524–1532.

61   M. E. Newcomer and A. R. Brash, *Protein Sci.*, 2015, 24, 298–309.

62   *Fundamentals of Inflammation*, Serhan, C. N.; Ward, P. A.; Gilroy, D. W., Eds.; Cambridge University Press; Cambridge, 2010.

63   J. Choi, J. K. Chon, S. Kim and W. Shin, *Proteins Struct. Funct. Bioinform.*, 2008, **70**, 1023–1032.

64   S. Borngräber, M. Browner, S. Gillmor, C. Gerth, M. Anton, R. Fletterick and H. Kühn, *J. Biol. Chem.*, 1999, **274**, 37345–37350.

65   J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res.*, 2005, **33**, W368–W371.

66   R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537–W541.

67   G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, 1997, **267**, 727–748.

68   P. Tosco, *J. Am. Chem. Soc.*, 2013, **135**, 10404–10410.

69   D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGra, D. M. York and P. A. Kollman, AMBER 2018. University of California: San Francisco 2018.

70   J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

71   P. Li and K. M. Merz, *J. Chem. Inf. Model.*, 2016, **56**, 599–604.

72   J. M. Seminario, *Int. J. Quantum Chem.*, 1996, **60**, 1271–1277.

73   A. Cruz, A. Di Venere, G. Mei, A. Zhuravlev, A. Golovanov, S. Stehling, D. Heydeck, J. M. Lluch, À. González-Lafont, H. Kuhn and I. Ivanov, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, 2020, **1865**, 158680.

74   C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

75   H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

76   A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

77   P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, 1973, **28**, 213–222.

78   P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270–283.

79   R. Suardíaz, P. G. Jambrina, L. Masgrau, À. González-Lafont, E. Rosta and J. M. Lluch, *J. Chem. Theory Comput.*, 2016, **12**, 2079–2090.

80.   G. Oanca, M. Purg, J. Mavri, J. C. Shihdef and J. Stare, *Phys. Chem. Chem. Phys.* 2016, **18**, 13346−13356.

81.   A. Prah, E. Franciskovic, J. Mavri and J. Stare, *ACS Catalysis*, 2019, **9**, 1231-1240.

82.   J. Stare, *RSC Advances*, 2017, **7**, 8740–8754.

Table of Contents Entry

A protocol based on the free energy perturbation theory justifies the suitable use of the exponential average and provides a practical way to determine the Gibbs free energy barrier of an enzyme reaction.