Enzyme Promiscuity in Enolase Superfamily.

A Theoretical Study of *o*-Succinylbenzoate Synthase Using QM/MM Methods

María Sánchez-Tarín[‡], Katarzyna Swiderek[‡†], Maite Roca[‡]*, Iñaki Tuñón[‡]*

[‡]Departament de Química Física, Universitat de València, 46100 Burjassot, (Spain)
[†] Institute of Applied Radiation Chemistry, Lodz University of Technology (Poland)

*to whom correspondence should be addressed

ignacio.tunon@uv.es
M.Teresa.Roca@uv.es

Abstract

The promiscuous activity of the enzyme o-Succinylbenzoate Synthase (OSBS) from the actinobacteria *Amycolatopsis* is investigated by means of QM/MM methods, using both Density Functional Theory and Semiempirical Hamiltonians. This enzyme catalyzes not only the dehydration of 2-succinyl-*6R*-hydroxy-2,4-cyclohexadiene-*1R*-carboxylate but also catalyzes racemization of different acylaminoacids, being N-succinyl-R-phenylglycine the best substrate. We investigated the molecular mechanisms for both reactions exploring the Potential Energy Surface. Then, Molecular Dynamics simulations were performed to obtain the free energy profiles and the averaged interaction energies of enzymatic residues with the reacting system. Our results confirm the plausibility of the reaction mechanisms proposed in the literature, with a good agreement between theoretical and experimentally-derived activation free energies. Our simulations unravel the role played by the different residues in each of the two possible reactions. The presence of flexible loops in the active site and the selection of structural modifications in the substrate seem to be key elements to promote the promiscuity of this enzyme.

1. Introduction

Many efforts have been devoted during the last years to improve the understanding of the mechanism of action of natural enzymes[1-5] and apply this knowledge to the development of new biocatalysts.[6,7] Scientists began to use evolutionary strategies (Directed Evolution)[8] to tailor the properties of individual molecules. Random mutations or recombination can, in many cases, be done efficiently, leading in this way to molecular evolution in the laboratory. On the other side, rational design can be used to guide the direct mutations of residues at specific positions of the enzyme.[9] Molecular simulations appear to be nowadays a promising option for computer-based design. Nevertheless, the most active computationally designed enzymes usually perform modestly in comparison with the natural enzymes.[10-13]

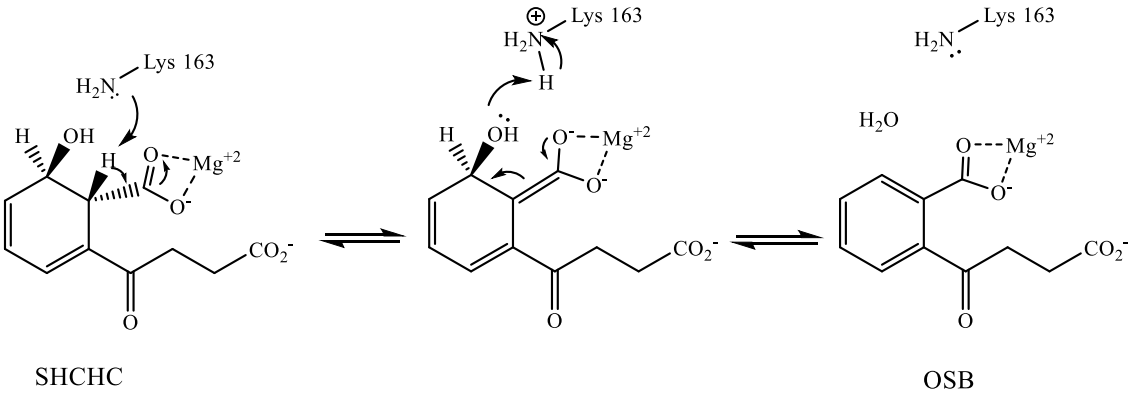
The development of new specific functions in an enzyme is something that Nature has already ~~made during~~ done through evolution, although the ~~specific paths~~ evolutionary routes followed to reach this goal remain~~, in general, raveled~~ to be revealed. While closely related proteins can have different functions, some distantly related enzymes can have the same or similar function. Evolution could have taken advantage of enzymes presenting promiscuous activities. This promiscuity can be classified in three categories: substrate promiscuity (the enzyme accepts different substrates catalyzing the same reaction), catalytic promiscuity (the enzyme accepts different substrates catalyzing different reactions) and product promiscuity (the enzyme accepts a single substrate and catalyzes the formation of different products).[14-18] Promiscuity provides a raw starting point for the evolution of enzymes, as a new duplicated gene presenting low activity would be the germ for adaptive evolution.[15] This ability endows organisms with a selective advantage and thereby enables their survival and further evolution. Studying protein evolution from promiscuous enzymes offers the possibility to obtain valuable experiences for the design of new biocatalysts.

In this paper we focus on a promiscuous enzyme member of the enolase superfamily, a large group of enzymes that catalyzes different reactions distributed in 7 subgroups.[19] All enzymes belonging to this superfamily use a common mechanistic first step consisting in the abstraction of an alpha proton next to a carboxylate group, leading to the formation of an enediolate anion intermediate stabilized by the presence of a divalent cation in the active site (usually Mg^{2+}). The active site ~~nestled on~~ positioned in

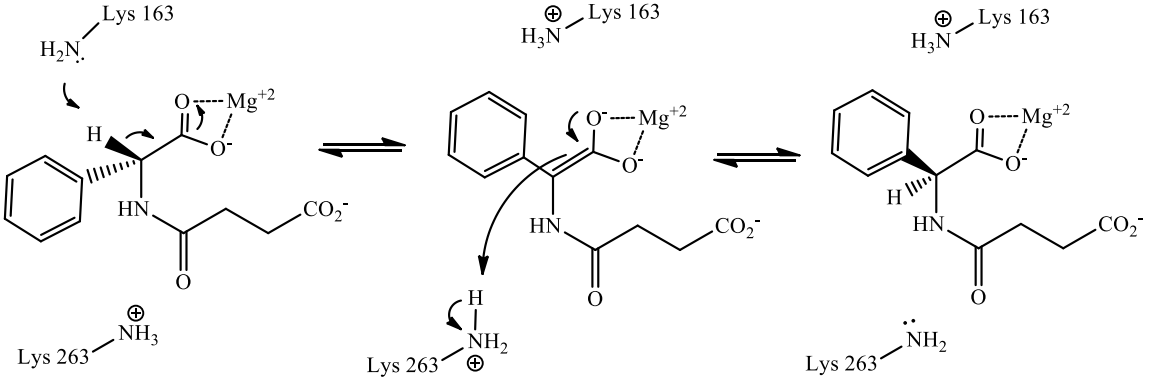
a cavity formed by the C-terminal ends of the β-strands of the barrel domain and the substrate lies sandwiched between the catalytic acid/base residues.[20] The Muconate lactonizing subgroup of the family is characterized by the presence of a conserved Lys-X-Lys motif at the end of the second β-strand and a lysine at the end of the sixth β-strand. These lysine residues are located at opposite faces of the active site, allowing proton transfer reactions to/from both faces of the substrate. The *o*-Succinylbenzoate Synthase (OSBS) enzyme from *Amycolatopsis* belongs to this subgroup and was first assigned as a racemase and designated as N-acylamino acid racemase (NAAAR). However it was later discovered that this enzyme is a better catalyst for a specific *syn* dehydration reaction that happens during *o*-succinylbenzoate synthesis and was then designated as *o*-Succinylbenzoate Synthase (OSBS).[21] OSBS is thus an example ~~of catalytic promiscuity~~ of an enzyme with catalytic promiscuity. Other members of the enolase superfamily also present the ability to catalyze dehydration and racemization reactions either for the same (product promiscuity) or different substrates (catalytic promiscuity).[22,23]

OSBS from *Amycolatopsis* catalyzes the two different reactions mentioned before (see Scheme I): i) the dehydration of 2-succinyl-*6R*-hydroxy-2,4-cyclohexadiene-*1R*-carboxylate (SHCHC) to *o*-Succinylbenzoate (OSB) (OSBS reaction, see Scheme 1 a)) and ii) the N-acyloamino acids racemization (NAAAR reaction). The NAAAR reaction can be enhanced using substrates that closely resembles ~~to~~ SHCHC. One of the best substrates is N-succinyl-(*R*)-phenylglycine (NSRPG) that is converted into N-succinyl-(*S*)-phenylglycine (NSSPG) (see Scheme I b)). The first step in both reactions requires the abstraction of a proton next to a carboxylate group of the substrate carried out by the deprotonated Lys 163. It has been proposed that the second step in the OSBS reaction consists in the dehydration of the substrate assisted by the conjugate acid of Lys 163, while the NAAAR reaction utilizes Lys 263 in the 1,1-proton transfer reaction to yield the substrate enantiomer.[24]

a) OSBS Reaction



b) NAAAR Reaction



Scheme I. Proposed reaction mechanisms for the OSBS (a) and NAAAR (b) reactions catalyzed by OSBS of *Amycolatopsis*.

In this work we present a theoretical analysis of the reaction mechanisms for the OSBS/NAAAR enzyme from *Amycolatopsis*, using SHCHC and NSRPG as substrates. The Potential Energy Surfaces of both reactions have been explored using Quantum Mechanics/Molecular Mechanics (QM/MM) hybrid methods in order to determine the reliability of the proposed mechanisms. Molecular Dynamics (MD) simulations have been carried out in order to obtain the corresponding Potentials of Mean Force (PMFs) or free energy profiles and to investigate the role of the different protein-substrate interactions during the binding of the substrate and in the stabilization of the Transition State (TS). These simulations can highlight into the origin of enzyme promiscuity and

thus serve as a useful guide for the computational design of new functions into existing proteins.

2. Methodology

Preparation of the system

Theoretical studies were performed starting with the crystallized structure of the OSBS enzyme that appears in the *Protein Data Bank* with code 1SJB.[25] This structure includes the OSB substrate and the magnesium ion. According to experimental data, OSBS from *Amycolatopsis* assembles as an octamer.[25] Active sites are found inside each subunit, with residues of the closer subunit placed at 15 Å from the active site chosen to model the reaction (chain B). Our model consisted of two subunits, chains A and B and the distance between the active sites of these two subunits is about 30 Å. The crystal structure contains the product of the OSBS reaction (OSB, see Scheme I). A schematic representation of the main interactions established by OSB in the 1SJB structure is shown in Figure S1 as Supporting Information. The OSB molecule was transformed into the reactant (SHCHC) adding a hydrogen atom to the C^{α} atom and a hydroxyl group to the C^{β} atom. After the chemical transformation the internal geometry of the substrate was fully relaxed. The protonation state of titratable residues at pH 7 was determined using the program PropKa 3.1.[26-29] All the His residues were protonated because they are located at the surface of our model and are accessible to the solvent. The pK_a values predicted by PropKa for Lys 163 and 263 of chain B (the selected one to model the reaction in its active site) were 7.15 and 9.14 respectively. Theses pK_a values depend on protein conformation and on the nature of the substrate. Results were similar for PDB structure 1SJC[25] that contains N-succinylmethionine in the active site; 6.47 and 9.00, respectively. However, the computed values were smaller (4.55 and 6.52 for Lys 163 and Lys 263, respectively) for structure 1SJA[25] (chain B) that contains a different substrate: N-succinylacetate. In any case, these values are consistent with a protonation state adequate for the forward reactions as depicted in Scheme I: Lys 163, which acts as a base in both reactions, is considered to be unprotonated at the reactant state, whereas Lys 263, that a acts as acid in the NAAAR reaction, is protonated. We did not ~~use~~ consider the 1SJD[25] structure because it lacks the magnesium ion and the substrate (NSRPG) is found in a non-productive orientation with respect the catalytic lysines.

Hydrogen atoms were added using the fDynamo program.[30,31] The system was solvated with a pre-equilibrated water box of 100 x 80 x 80 Å of size centered on the substrate. Water molecules placed at less than 2.8 Å from any other non-hydrogen atom of the system were deleted. To neutralize the system, 5 sodium atoms were added. In order to reduce the computational time, a sphere of 25 Å centered on the substrate was defined and the atoms outside that sphere were kept fixed. A representation of the simulated system is provided as Supporting Information (Figure S2). To study the reaction a hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Hamiltonian was employed. The quantum region (QM) contains the substrate and the side-chains of those residues directly involved in the reactions: Lys 163 and Lys 263. In addition, in order to model correctly the possible charge transfer between the negatively charged substrate and the metal center,[32] we also included in the QM subsystem the magnesium ion and the side-chains of the residues that are part of its coordination shell (Asp189, Glu 214, Asp 239 and a water molecule) (see Figure 1). Figure 1 displays the QM regions of the OSBS reaction on the left (a) and the QM region of the NAAAR reaction on the right (b). The QM subsystem was described at AM1[33] and M06-2X/6-31+G(d,p)[34,35] levels. The M06 level is clearly superior to the semiempirical AM1 treatment but, because of the computational cost, the former was used only for the exploration of the Potential Energy Surface, while the latter was employed also for QM/MM Molecular Dynamics simulations. In any case our results show (see below) that the geometrical description obtained at the semiempirical level is in good agreement with the higher QM level and that single-point energy corrections can be applied to obtain a reasonable quantitative picture of the reaction energetics. For the classical region (MM), the OPLS-AA[36,37] and TIP3P[38] force fields, as implemented in fDynamo,[30,31] were used to describe the protein and the water molecules, respectively. The link atom approach[39] was used to saturate the valence of the QM/MM frontier. These link atoms were placed between C^{α} and C^{β} atoms of each residue in the QM subsystem (see Figure 1).

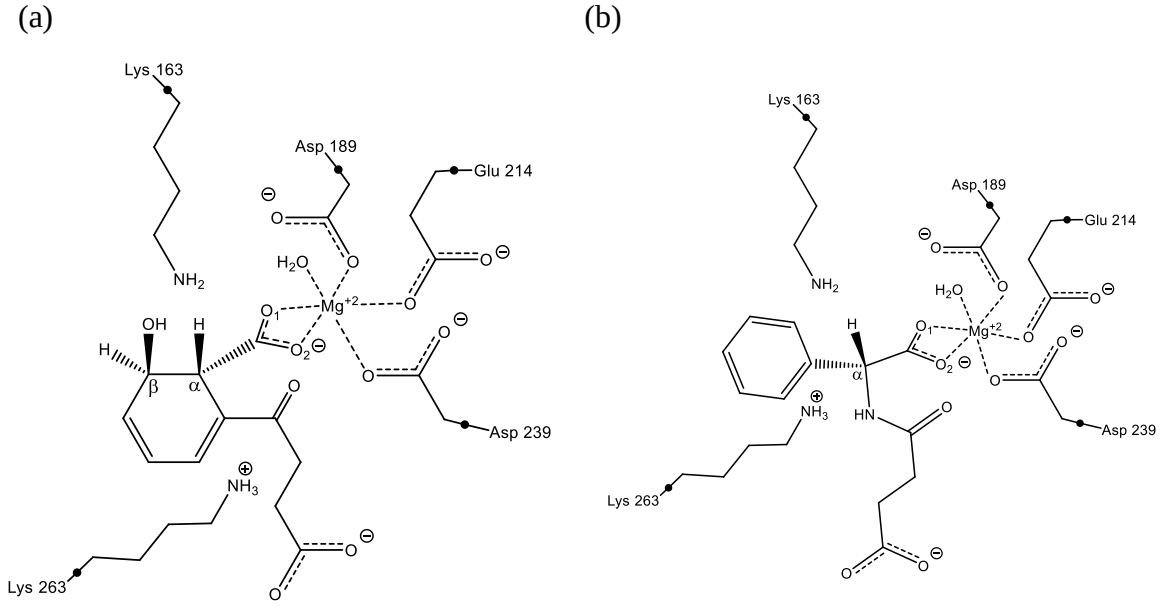


Figure 1. QM regions used for the analysis of the OSBS (a) and NAAAR (b) reactions. Link atoms are depicted as •.

In order to prepare the system for the study of the NAAAR reaction, the SHCHC molecule was replaced by NSRPG overlapping the optimized substrate structures to get an adequate initial pose of the new substrate into the active site. The final systems consisted of a dimeric enzyme complexed with the substrate (SHCHC for the OSBS reaction and NSRPG for the NAAAR reaction) and a box of water molecules and counterions. The OSBS system contains 60868 atoms of which 10565 were considered as mobile atoms including 83 atoms in the QM region; whereas the NAAAR system contains 60870 atoms in total of which 10567 were mobile and 85 QM (see Figure S2).

Simulations, performed using the fDynamo library, employed a switched cutoff from 12 Å to 14 Å for all nonbonded interactions, including those established between atoms in the rigid and flexible regions. The system was first optimized and equilibrated by means of a combination of steepest-descent and conjugate gradient optimizations and molecular dynamics (MD) simulations. First, solvent molecules were minimized and equilibrated while the rest of the system was kept frozen. Second, solvent molecules and

side chains were relaxed and finally the whole system was minimized and gradually heated from 0 K to 298 K. The system was equilibrated with a series of QM/MM MD simulations covering a total of 550 ps in the NVT ensemble using a Langevin-Verlet integrator[40] with a time a step of 0.5 fs. Snapshots of SHCHC and NSRPG equilibrated in the active site of the enzyme are provided in Figures 2a and 2b. The overlap of these configurations is shown in Figure 2c.

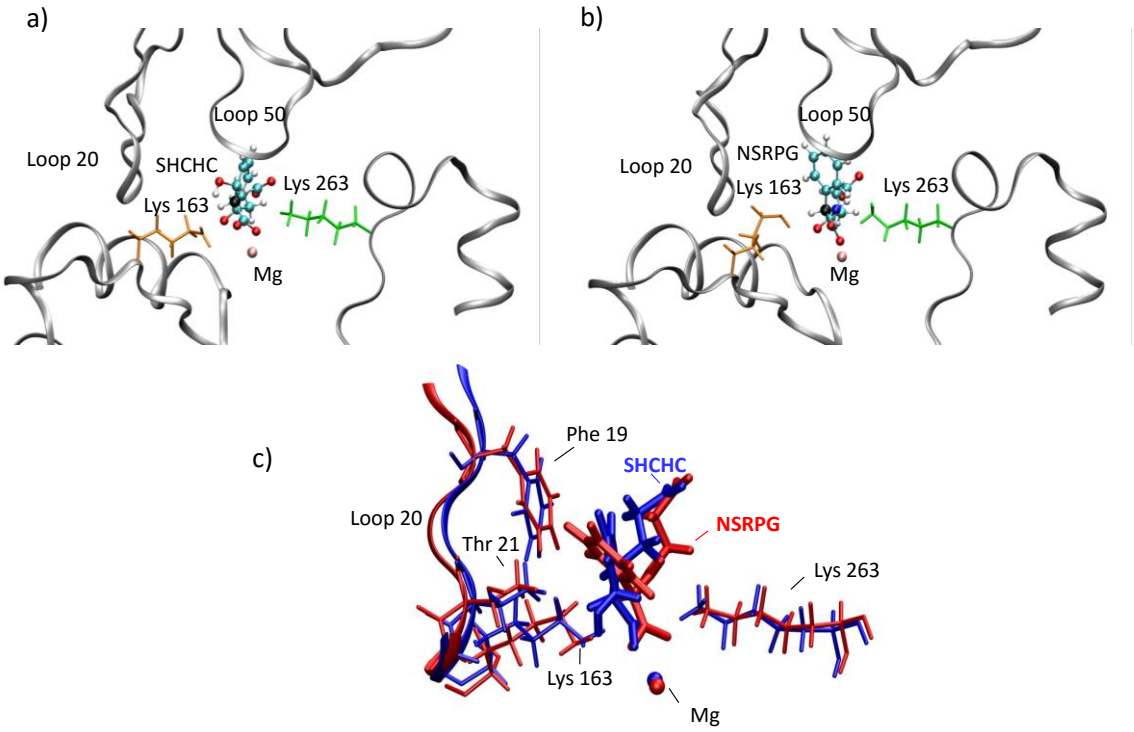


Figure 2. Snapshots of SHCHC (a) and NSRPG (b) substrates in the active site of the *Amycolatopsis* enzyme. Substrate has been drawn using ball & stick model (the C^{α} atoms are drawn in black), while the catalytic lysines are represented in licorice representation. C) Overlap of the SHCHC (blue) and NSRPG (red) structures at the active site, showing the position of the catalytic lysines and some residues of loop 20.

Potential Energy Surface Exploration

The potential energy surfaces (PES) for both the OSBS and NAAAR reactions were obtained starting from optimized structures of reactants and reaching the transition state structures through scans of ~~distinguished~~ reaction coordinates. Putative TS structures were located and their nature determined by diagonalization of the Hessian matrix corresponding to the subset of QM atoms. TS structures have a single imaginary

frequency and thus they are true saddle points on the PES (there is just one direction in which the energy decreases while the structure is a minimum in all the rest). Intrinsic reaction coordinate (IRC)[41] paths were traced down to reactant/intermediate/product valleys from the saddle points. The final structures were optimized and their nature confirmed from analysis of the Hessian. All these calculations were performed using the micro/macroiteration optimization algorithm.[42-44] The PESs were explored using the antisymmetric combination of the distances associated to the proton transfer from C^{α} of the substrate to Lys 163 (d(C^{α}-H^{α}) and d(H^{α}-N^{ε} Lys 163)) for the first step of the mechanisms for both substrates. For the second mechanistic step of the NAAAR reaction we obtained the PES using as reaction coordinate the antisymmetric combination of the distances that control the proton transfer from Lys 263 to the C^{α} of the substrate (d(H^{α}-N^{ε} Lys 263) and d(C^{α}-H^{α})). For the second step of the OSBS reaction we used the antisymmetric combination of the distances associated to the proton transfer from Lys 163 to the hydroxyl group in β position and the distance of the oxygen atom of this group to C^{β} (d(H^{α}-N^{ε} Lys 163), d(C^{α}-H^{α})) and d(C^{β}-O^{β})). The later PES exploration was two-dimensional while the rest was unidimensional. The PES exploration was carried first out at the AM1/MM level and the obtained structures were used as initial guess for M06-2X/6-31+G(d,p)/MM calculations using a combination of fDynamo and GAUSSIAN09 programs.[45]

Free Energy Calculations

Potentials of Mean Force (PMFs) were obtained for the OSBS and NAAAR reactions at the AM1/MM level. These PMFs provide the free energy change along selected reaction coordinates. Antisymmetric combinations of the distances describing the breaking and forming bonds were employed as distinguished reaction coordinates (RCs). The selected distances depend on the reaction and the step studied (see below). The umbrella sampling approach[46] was used to constraint the system close to a particular value of the RC by means of the addition of a harmonic potential with a force constant of 2500 $kJ \cdot mol^{-1} \cdot Å^{-2}$, which allowed good overlap between consecutive simulation windows. The probability distributions obtained from MD simulations within each individual window were combined by means of the weighted histogram analysis method (WHAM)[47] to obtain the whole PMF. Each simulation window consisted in 10 ps of relaxation and 20 ps of production, with a time step of 0.5 fs. Simulations were carried

out in the NVT ensemble with a reference temperature of 298 K and the Langevin–Verlet integrator.[40]

In the OSBS reaction, at the first step, the PMF was performed exploring the reaction coordinate with a window width of 0.03 Å, and at the second step (a bidimensional PMF) with a window width of 0.05 Å (the total number of windows were 100 and 1600, respectively). At the first step of the NAAAR reaction, the reaction coordinate was then explored with a window width of 0.05 Å, and at the second step the reaction coordinate was explored with a window width of 0.03 Å. The total number of windows were 70 and 150, respectively.

The free energy barriers were computed at AM1/MM level and corrected at M06-2X/MM level for the two steps of both OSBS and NAAAR reactions. In particular, we obtained corrected free energy barriers by applying the following equation:

$$\Delta G^{\ddagger}_{\text{corr}} = \Delta G^{\ddagger}_{\text{AM1/MM}} + \left(\left\langle E_{TS} - E_R \right\rangle_{HL} - \left\langle E_{TS} - E_R \right\rangle_{AM1} \right) \quad (1)$$

The averaged energy differences appearing in the above expression have been obtained by means of single-point calculations of a set of ¿10? different reactant and TS structures localized using a micro-macro iteration approach[48-50] at the AM1/MM level. These structures are optimized using a Hessian-based search for the QM atoms while the MM atoms are fully optimized at each step of this search using only gradients.[48-50] In equation 1, HL stands for the high level chosen to obtain the corrected estimation of the activation free energy. In our case M06-2X/6-31+G(d,p)[34,35] has been used. HL calculations are carried out under the effect of the field created by the MM environment, where the electrostatic coupling between the QM and MM subsystems is calculated using point charges on the QM atoms.[48-50]

Interaction Energies

In order to elucidate the role of the different residues in each of the two reactions catalyzed by this enzyme, we computed the averaged interaction energies of these residues with the reacting subsystem (the substrate and lysines 163 and 263, see Figure 1). The average was performed over 100 ps of MD simulations constrained at values of the reaction coordinate corresponding to reactant and transition states, as determined

from the corresponding PMFs. The averaged interaction energy of a given residue j was obtained as:

$$\langle \Delta E_{int,j} \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle \Psi_i | \widehat{H}_{QM/MM,j} | \Psi_i \rangle \quad (2)$$

where N is the number of configurations analyzed (1000 in our case), the wave function (Ψ_i) corresponds to that of the polarized reacting subsystem and the interaction Hamiltonian with residue j is defined as:

$$\widehat{H}_{QM/MM,j} = -\sum_{i} \sum_{M \epsilon j} \frac{q_M}{r_{iM}} + \sum_{\alpha} \sum_{M \epsilon j} \frac{Z_a q_M}{R_{aM}} + \sum_{\alpha} \sum_{M \epsilon j} \left\{ \frac{A_{aM}}{R_{aM}^{12}} - \frac{B_{aM}}{R_{aM}^{6}} \right\} \quad (3)$$

including the electrostatic and van der Waals terms. These calculations were carried out at the AM1/MM level.

We have also compared the averaged interaction energies obtained for each residue at the transition states of both reactions with those obtained for their corresponding reactants. The contribution of the interaction energy of each residue to the barrier can be calculated as the difference of the interaction established in the transition state and reactant state:

$$\Delta\Delta E^{\ddagger}_{int,j} = \langle \Delta E_{int,j} \rangle_{TS} - \langle \Delta E_{int,j} \rangle_{R} \quad (4)$$

where $\langle \ldots \rangle_{TS}$ and $\langle \ldots \rangle_{R}$ denote averages over the transition and reactant states of the reaction, respectively, and the subscripts (j) refers to the residue number.

3. Results and Dicussion

3.1. Exploring the Reaction Mechanisms

Location and characterization of the stationary structures corresponding to a given reaction mechanism (reactants, products, transition structures and possible intermediates) allow determining its feasibility from the inspection of the energy barriers and the total energy change. Tracing the reaction paths along these structures provides detailed information about the molecular characteristics of the mechanism: the step-wise or concerted nature of the process and the degree of synchronicity between bond breaking and forming events. The use of QM/MM methods has facilitated the application of accurate QM methods to the exploration of reaction mechanisms in large

systems, such as in enzymatic processes. We investigated the mechanistic proposal of Gerlt and coworkers (see Scheme I)[24] at the M06-2X/MM level. This method has been successfully used to determine the energetics associated to chemical reactions.[34,35] ~~The PESs were explored using the distances associated to the proton transfer from C^{α} of the substrate to Lys 163 (d(C^{α}-H^{α}) and d(H^{α}-N^{ε} Lys 163)) for the first step of the mechanisms for both substrates. For the second mechanistic step of the NAAAR reaction we obtained the PES using as distinguished reaction coordinates the distances that control the proton transfer from Lys 263 to the C^{α} of the substrate (d(H^{α}-N^{ε} Lys 263) and d(C^{α}-H^{α})). For the second step of the OSBS reaction we used the distances associated to the proton transfer from Lys 163 to the hydroxyl group in β position and the distance of the oxygen atom of this group to C^{β} (d(H^{α}-N^{ε} Lys 163), d(C^{α}-H^{α})) and d(C^{β}-O^{β})).~~ As explained in the Methodology section, the PESs were explored by means of QM/MM methodology using the AM1/MM level. These PESs were used to select good initial guesses for the optimization of all the relevant structures (reactant, products, TSs and intermediates). Some key geometrical parameters of the optimized structures, together with the values of the relative energies and the imaginary frequencies are provided in Table 1. Figure 3 shows the corresponding Potential Energy Profiles together with a representation of the chemical subsystem at the TSs. All the stationary structures are shown in Figures S4 and S5 in the Supporting Information. We also explored the reaction mechanisms using the much less computationally demanding (and usually less accurate) AM1/MM level. Results are presented as Supporting Information (Table S2 and Figures S3-S5). In general, the AM1 calculations provide the same qualitative picture for the reaction mechanisms. Geometrically, the resulting stationary structures are very similar to those obtained at the M06-2X level. This agreement is important because gives support to the use of the much less computationally demanding AM1/MM method during QM/MM MD simulations. During these simulations (see below) the number of generated structures is much larger than during the exploration of the PES and thus these calculations can be only carried out at the semiempricial level. The larger difference observed between both QM levels is found in the energy of the intermediate relative to the reactants, that seems to be overestimated at the AM1/MM level. For the OSBS reaction the difference between AM1/MM and M06-2X/MM energies for the intermediate is 13.2 kcal·mol^{-1} and 10.2 kcal·mol^{-1} for the NAAAR reaction, being higher for the AM1/MM level.

The first step, the proton abstraction from C^{α} by the unprotonated Lys 163, is very similar for both substrates. The proton is being transferred from C^{α} to the N^{ε} atom of Lys 163 and at the transition structures (TS1) the distances of the proton to the donor and acceptor atoms (H^{α}-C^{α} and H^{α}-N^{ε}, see Table 1) are similar for both substrates. The proton transfer is somewhat earlier for SHCHC where the proton-donor distance (H^{α}-C^{α}) is 1.34 Å, while for NSRPG the distance at TS1 is 1.38 Å. In both the OSBS and the NAAAR reactions this proton transfer is assisted by the presence of the Mg^{2+} ion, as reflected in the reduction of the distances between this ion and the carboxylate oxygen atoms of the substrate observed at the enediolate intermediate (see O1-Mg and O2-Mg distances for structures in Table S3 of Supporting Information). The remaining distances defining the coordination shell of the ion are almost unaltered in both reactions (see Table S3 of Supporting Information). At the M06-2X level the energy barrier for this first step is more favorable for the SHCHC substrate (9.1 kcal·mol^{-1}) than for NSRPG (12.5 kcal·mol^{-1}). However, the differences are small considering that these values can fluctuate depending on the particular conformation analyzed (see discussion below). Our calculations show that, in the active site, the enediolate intermediates are almost isoenergetic with NSRPG (the energy difference being 0.5 kcal· mol^{-1}) or even slightly more stable than SHCHC (-2.5 kcal· mol^{-1}).

The OSBS and NAAAR reactions differ in the second step, the transformation followed by the enediolate intermediate to form the final product (OSB or NSSPG, see Scheme I). In the case of the OSBS reaction, Lys 163 transfers the proton to the oxygen atom (O^{β}) of the hydroxyl group placed in position β with respect to the carboxylate group, forming a water molecule. Dehydration of the intermediate leads to the formation of the final product, OSB, an aromatic compound, resulting in an exothermic process (see Figure 3). At the corresponding transition structure (TS2) both the proton transfer and the O^{β}-C^{β} bond cleavage are advanced. The O^{β}-C^{β} distance has been increased from 1.49 to 1.83 Å, while the proton transferred from Lys 163 is found at 1.37 Å of the hydroxyl oxygen atom (see O^{β}-H^{ε} in Table 1). The energy barrier for the dehydration of the enediolate intermediate is 15.7 kcal·mol^{-1} at the M06-2X/MM level.

For the NSRPG substrate the second step proceeds by means of a proton transfer from protonated Lys 263 back to the C^{α} atom to produce the substrate enantiomer NSSPG. The geometry of the transition structure (TS2-NAAAR) corresponds to an earlier

structure with the proton-donor distance (H^{ε}-N^{ε}) shorter than the proton-acceptor one (H^{ε}-C^{α}); 1.32 and 1.43 Å, respectively. The energy barrier determined from the reaction intermediate is 13.4 kcal·mol^{-1} at the M06-2X level. This reaction is only slightly exothermic at the M06-2X level.

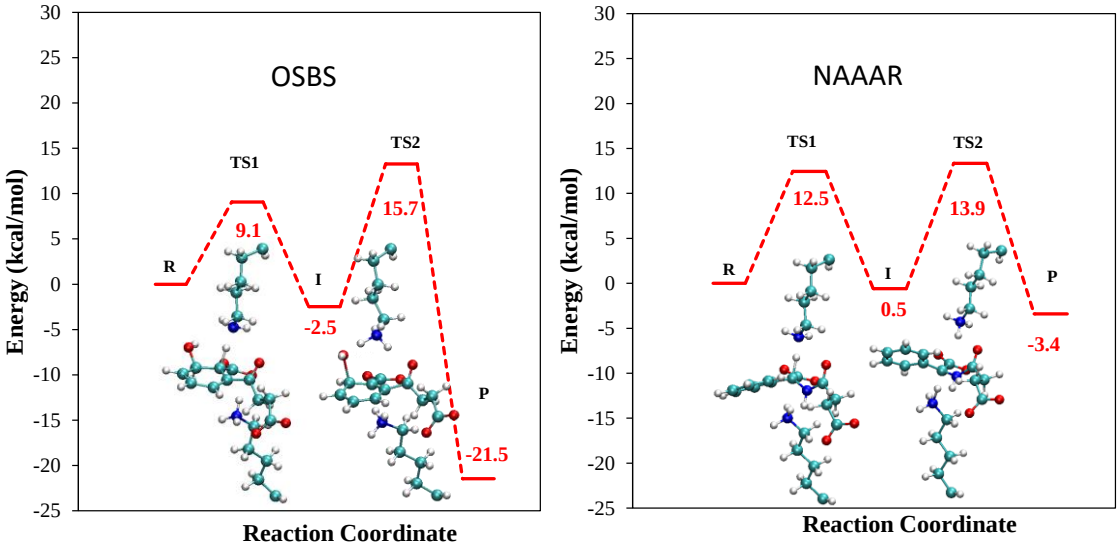


Figure 3. Potential Energy Profiles obtained for the OSBS (left) and NAAAR (right) reactions determined at the M06-2X/MM level. The geometries of the reacting subsystem at the transition states are also represented.

It is interesting to note that the distance from C^{α} to the H^{ε} atoms of Lys 263 at the intermediate is larger for the SHCHC substrate than for NSRPG (2.88 and 2.37 Å at the M06-2X level, respectively, see Table S3 of Supporting Information). This larger distance could prevent the racemization reaction in the case of SHCHC. The energy barrier for a hypothetic proton transfer from Lys 263 would increase above the value obtained for the second step of the OSBS process. There are several factors that contribute to the different reaccommodation of the substrate in the active site. First, C^{α} atom of SHCHC is part of a ring, rendering it less flexible. Second, the hydroxyl substituent on C^{β} establishes a hydrogen bond with Lys 163, favoring the displacement of the substrate in the active site away from Lys 263. Finally, in the case of the NSRPG substrate the NH group of the amide substituent makes a hydrogen bond with the carbonyl oxygen atom of Gly 291 (2.37 Å at the M06-2X level in the intermediate state), favoring the approach of this substrate to Lys 263. In any case, because of the

large exothermicity of the OSBS reaction, a hypothetical production of the SHCHC enantiomer would probably be transient and this product would not accumulate.

Table 1. Structures optimized at the M06-2X level for the OSBS (up) and NAAAR (down) reactions.

OSBS reaction					
Distances (Å)	R	TS1	I	TS2	P
H^{α}-C^{α}	1.11	1.34	2.51	2.68	3.12
H^{α}-N^{ε}(Lys 163)	2.00	1.41	1.04	1.03	1.02
H^{ε} (Lys 163)-N^{ε} (Lys 163)	1.02	1.02	1.04	1.17	1.75
O^{β}- H^{ε} (Lys 163)	3.41	2.77	2.02	1.37	1.00
O^{β}-C^{β}	1.42	1.43	1.49	1.83	3.05
Imaginary frequency (cm^{-1})	-	1153	-	533	-
Energy ($kcal{\cdot}mol^{-1}$)	0.0	9.1	-2.5	13.2	-21.5
$\Delta E^{\ddagger}$ ($kcal{\cdot}mol^{-1}$)	-	9.1	-	15.7	-

NAAAR reaction					
Distances (Å)	R	TS1	I	TS2	P
H^{α}-C^{α}	1.11	1.38	2.34	2.69	2.98
H^{α}-N^{ε} (Lys 163)	2.12	1.39	1.04	1.03	1.03
C^{α}-H^{ε} (Lys 263)	3.38	2.69	2.37	1.43	1.11
H^{ε} (Lys 263)-N^{ε} (Lys263)	1.03	1.02	1.03	1.32	2.00
Imaginary frequency (cm^{-1})	-	1236	-	1019	-
Energy ($kcal{\cdot}mol^{-1}$)	0.0	12.5	0.5	13.9	-3.4
$\Delta E^{\ddagger}$ ($kcal{\cdot}mol^{-1}$)	-	12.5	-	13.4	-

Proofs of a reaction mechanism can be obtained from the determination of Kinetic Isotope Effects (KIEs), this is the ratio between the rate constants obtained upon isotopic substitution into a particular position of the chemical system. Up to our knowledge, only the primary deuterium isotope effect was measured for the kinetic constants for the OSBS enzyme from *E. coli*. The KIE value obtained from the catalytic rate constants was 6.1 ± 0.4, indicating a mechanism dominated by the proton transfer.[51] We determined the M06-2X values of semiclassical KIEs corresponding to isotopic substitutions of non-hydrogen atoms during the rate-limiting step of the OSBS and NAAAR reactions. More details about KIEs calculations are given in the Computational Methodology section of Supplementary Information and the results and discussion are reported in the Kinetic Isotope Effects Results section of Supplementary Information.

3.2. ~~Potentials of Mean Force~~ Free Energy Profiles

For reactions involving a ~~reduced~~ small number of atoms, the characterization of a few stationary structures is usually enough to characterize the process and to calculate free energies that can be compared to experimental equilibrium and rate constants. These free energies are usually obtained by means of the application of the harmonic approximation for the displacements from the stationary structures. However, this approach does not hold for reactions involving many degrees of freedom (such as enzymatic reactions) where the PES is usually rough and many different configurations must be considered in order to obtain averaged values to be compared to experimental properties. The energy profiles presented above must be considered with caution because they represent the energy paths for a particular conformation of the system but, in principle, many reaction paths are possible in an enzymatic environment as a consequence of the large number of nearly-degenerate conformations accessible for the system on the complex PES.[5,52,53] Thus, we carried out QM/MM MD simulations to determine the Free Energy Profiles or PMFs corresponding to the mechanisms described above. During these simulations many configurations are generated and the free energy changes associated to the reaction were obtained from the probability distribution of distinguished reaction coordinates selected from the previous exploration of the PES. Because of their computational cost these simulations were carried out only at the AM1/MM level. According to the results presented in the Supporting Information, the AM1 method gives a reasonable geometric and energetic description of the reaction mechanism, except for an overestimation of the energies of the enediolate intermediate and TS2 with respect to the reactants. We thus used the AM1/MM hybrid potential to obtain the free energy profiles along the distinguished coordinates and we corrected *a posteriori* the free energy values by means of single-point energy calculations at the higher M06-2X level (as explained in the Methodology section).

For the first step in both OSBS and NAAAR reactions we used the antisymmetric combination of the distances of the transferred proton to the N^{ε} atom of Lys 163 and to the C^{α} atom of the substrate (d(C^{α}-H^{α})-d(H^{α}-N^{ε} Lys 163)) as the distinguished reaction coordinate. The total number of simulation windows used to obtain the PMF was 100 in each case. A similar procedure was used for the second step of the NAAAR reaction

where the coordinate was the antisymmetric combination of the distances of the transferred proton to the N^{ε} atom of Lys 263 and to the C^{α} atom of the substrate (d(H^{α}-N^{ε} Lys263)-d(C^{α}-H^{α})). For the OSBS reaction, two distinguished coordinates were employed to obtain a smooth two dimensional PMF (2D-PMF) corresponding to the second reaction step. These coordinates were the antisymmetric combination of the distances of the transferred proton to the N^{ε} atom of Lys 163 and to the O^{β} atom of the substrate (d(H^{α}-N^{ε} Lys 163)-d(C^{α}-H^{α})) and the C^{β}-O^{β} distance. A total of 1600 simulation windows were employed to calculate this two-dimensional PMF. The PMFs obtained for the two steps of both reactions are shown in Figure S6 of the Supporting Information, while the relative AM1/MM free energies and averaged distances are given in Table 3. Free energy differences corrected by means of single-point calculations at the M06-2X/MM level are also presented in Table 3.

Table 3. AM1/MM Average distances (with their standard deviations, in Å) and relative free energies (in kcal·mol^{-1}) obtained for the stationary states of the OSBS and NAAAR reactions. Corrected free energies (ΔG_{corr}) M06-2X values were obtained adding single-point M06-2X corrections to the AM1/MM results.

OSBS reaction					
	R	**TS1**	**I**	**TS2**	**P**
H^{α}-C^{α}	1.14±0.03	1.56±0.03	2.92±0.37	2.65±0.19	2.86±0.24
H^{α}-N^{ε}(Lys 163)	2.46±0.04	1.20±0.03	1.03±0.03	1.01±0.02	1.00±0.03
H^{ε}(Lys 163)-N^{ε}(Lys 163)	0.99±0.02	1.01±0.02	1.03±0.03	1.02±0.03	1.01±0.02
O^{β}- H^{ε}(Lys 163)	3.84±0.31	2.73±0.24	1.97±0.04	1.22±0.03	0.99±0.02
O^{β}-C^{β}	1.42±0.03	1.43±0.02	1.44±0.02	1.54±0.03	2.98±0.03
ΔG	0.0	13.7	10.5	28.4	-10.3
ΔG_{corr}	0.0	12.9	2.3	13.1	-17.5

NAAAR reaction					
	R	**TS1**	**I**	**TS2**	**P**
H^{α}-C^{α}	1.14±0.03	1.60±0.04	2.21±0.04	2.68±0.14	2.87±0.22
H^{α}-N^{ε}(Lys 163)	2.59±0.09	1.17±0.03	1.03±0.03	1.04±0.03	1.03±0.02
C^{α}-H^{ε}(Lys 263)	3.08±0.16	2.72±0.17	2.61±0.20	1.45±0.03	1.15±0.03
H^{ε}(Lys 263)-N^{ε}(Lys263)	1.03±0.03	1.04±0.02	1.04±0.03	1.26±0.04	2.86±0.05
ΔG	0.0	16.9	14.9	22.4	11.3
ΔG_{corr}	0.0	16.6	8.5	17.3	0.3

The free energy profiles obtained from these simulations roughly match the potential energy profiles presented in Figure 3, validating the mechanistic description given

above. This is also confirmed by an analysis of the averaged distances presented in Table 3, which are in reasonable agreement with the optimized values presented in Table 1 and Table S2 of Supporting Information. In addition our best estimation of the activation free energies obtained from the highest energy TSs (the single-point corrected free energies, 13.1 and 17.3 kcal·mol^{-1} corresponding to TS2 for the OSBS and NAAAR processes, respectively) are compatible with the values expected for enzymatic processes. In particular, the free energy barriers obtained from the application of Transition State Theory (TST) to the experimental rate constants[24] determined at 25ºC are 14.3 and 14.6 kcal·mol^{-1} for NAAAR and OSBS reactions, respectively. These values must be considered with caution because the experimental assays used Mn^{2+} because it provides a 3-fold enhancement in the rate of the NAAAR reaction relative to Mg^{2+}.[24] However, this difference in the rate constants translates into a small change in the activation free energy (less than 1 kcal·mol^{-1}). Our free energy profiles also agree with the experimental observation that the values of the rate constant for the proton exchange exceed those for racemization[24] because our energy profiles show that TS2 of the NAAAR reaction presents a free energy slightly higher than for TS1. It should be taken into account that we have not considered quantum corrections to the classical activation free energy (zero point energies and tunneling) that typically contribute to decrease the calculated activation free energies in H-transfer processes.[54,55]

With respect to the reaction energies, our results show that the OSBS process is clearly exergonic (the reaction free energy being -17.5 kcal·mol^{-1}, see Table 3) in agreement with the experimental observation.[20,51] Instead, the NAAAR reaction is almost thermoneutral (0.3 kcal·mol^{-1}) This implies that the activation free energies and the catalytic rate constant will be very similar for the reverse reaction, the transformation of NSSPG into NSRPG, in agreement with the observation that this reaction occurs in both directions at nearly equal rates.[24]

3.3. Analysis of interactions relevant for catalysis

Once we have presented the results of the QM/MM calculations and checked that these results reasonably agree with experimental observations, we can use our simulations to provide a detailed molecular picture of the processes and the basis for the catalytic

promiscuity of the OSBS/NAAAR enzyme from *Amycolatopsis*. As explained in the Methodology section, the average interaction energy of the different residues with the reacting subsystem (the substrate and lysines 163 and 263) was computed and the results of these interaction energies for the most relevant residues and their standard deviations are provided as Supporting Information (Tables S2 and S3).

We first analyzed the role of each residue in the formation of the corresponding Michaelis complexes for the OSBS and NAAAR reactions. Figure 4 represents the interaction energies for each residue at the OSBS and NAAAR reactant states. A negative value means that a particular residue stabilizes the reactant state, favoring the binding of the substrate and/or the positioning of the reacting fragments (including the catalytic lysines); while a positive value indicates a destabilizing contribution. As expected, in both cases the largest contributions in absolute values come from the charged residues. The substrates (SHCHC and NSRPG) are both negatively charged and thus positively charged residues contribute to binding, while negatively charged residues destabilize the reactant complex. These charge-charge interactions are expected to make a similar global contribution in different residues with the same charge and different enzymes with similar active sites (as the OSBS enzyme from *E. coli*). The largest contribution comes from the Mg^{2+} ion, -211.1 and -222.7 kcal·mol^{-1} for SHCHC and NSRPG, respectively. However, this value is largely reduced when considering together the contributions from the ion and the residues of its coordination shell (Asp 239, Asp 189, Glu 214 and a water molecule) as deduced from the comparison of the green and black lines in Figure 4. The interaction energies of the ion plus its coordination shell are -79.3 and -89.7 kcal·mol^{-1} for SHCHC and NSRPG, respectively. Other residues with a large stabilizing contribution are Lys 161, Arg 266, Arg 299 and Arg 322, all of them positively charged. Destabilizing contributions, apart from the coordination shell of the Mg^{2+} ion, come from Glu 294 and Asp 316. The contacts made by these and other residues with the chemical system are presented in Scheme II.

Small favorable interactions are also observed between the substrate and residues found in two loops that close over the active site, known as loop 20 and loop 50 because of the position of these loops in the sequences of many members of the superfamily (see Figure 2c). These interaction energies are shown as an insert in Figure 4. Although the values are significantly smaller than for the charged residues, these interactions could be

important to differentiate between residues with equal charges. It has been proposed that the different orientation of the loop 20 in *Amycolatopsis* OSBS/NAAAR and *E. coli* OSBS could explain the absence of the NAAAR activity in this last enzyme.[56] This loop seems to play different roles in different members of the enolase superfamily, being essential for both the synthase and racemization activities in *Amycolatopsis* OSBS/NAAAR.[57] Our interaction energies show that Phe 19, Thr 21 and Tyr 55 make favorable contacts at the reactant states of both the OSBS and NAAAR reactions. The first two residues contact and orient the catalytic lysine (Lys 163) in agreement with the recent suggestion that the main role of loop 20 would be to hold the catalytic lysine close to the active site and to orient it properly.[56,58] Phe 19 and Tyr 55 make hydrophobic contacts with the ring of the substrates. It has been recently observed that mutating Phe 19 in *Amycolatopsis* OSBS/NAAAR decreased substrate binding in both activities.[57] Our simulations show that the role of this loop in *Amycolatopsis* OSBS/NAAAR is similar in the reactant state of both reactions, indicating that the loop 20 could have the flexibility needed to allow the promiscuous activity of this enzyme. The position of the loop could be easily adapted to accommodate slightly different orientations of the substrate ring in the active site, as observed in Figure 2c. The positioning of loop 20 on the active site seems to be slightly different at the reactant states of the OSBS and NAAAR processes, optimizing the interactions with the different substrates (SHCHC and NSRPG, respectively). Instead, it has been suggested that the *E. coli* OSBS could be evolutionarily optimized to increase substrate binding of the SHCHC substrate, rendering this enzyme less likely for a promiscuous function.[56] Mutagenesis studies of the 20s loops of the *E. coli* OSBS and *Amycolatopsis* OSBS/NAAAR indicate that their loops are more important to keep the structural integrity of the active site than for determining the enzyme specificity. In fact, deletion of the loop 20 of both enzymes results in a larger catalytic efficiency reduction for the *Amycolatopsis* OSBS/NAAAR enzyme than in the *E. coli* enzyme.[57] This reduction is due to loop 20 of *E. coli* enzyme is partially open when the substrate is bound, thus the residues of this loop do not strongly interact with the substrate. In contrast, the loop 20 of *Amycolatopsis* OSBS/NAAAR enzyme completely closes the active site and their residues form an important part of the substrate binding pocket.[57] Thus, differences in the position and sequences of loop 20 between promiscuous *Amycolatopsis* OSBS/NAAAR enzyme and nonpromiscuous *E. coli* OSBS enzyme could be an important determinant for NAAAR activity. However, it must be pointed out that

mutations on the loop 20 of *Amycolatopsis* OSBS/NAAAR enzyme reduced both the synthase and the racemase activities, suggesting that evolution of this loop were not the only changes needed to evolve the latter activity.[57] As discussed below other residues contribute to accelerate the racemase activity in *Amycolatopsis* OSBS/NAAAR.

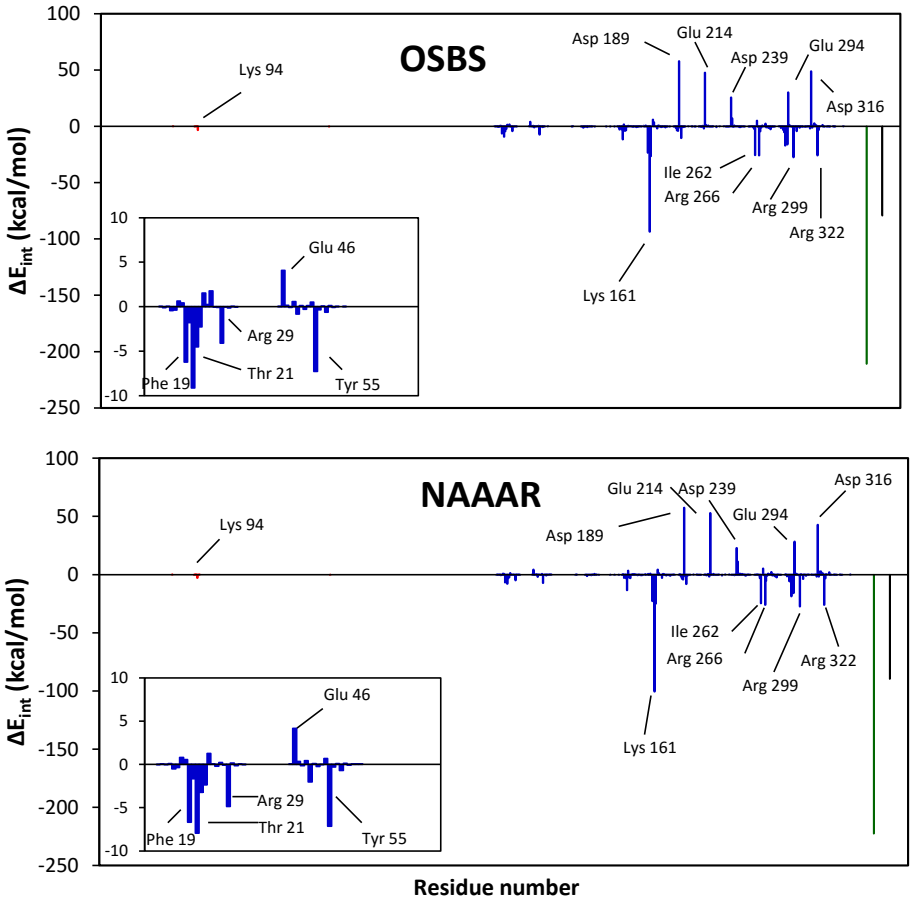
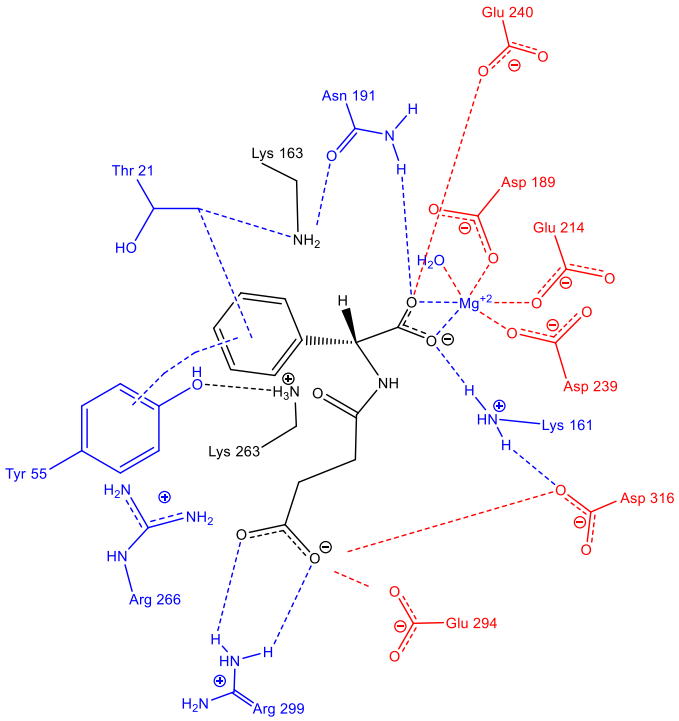


Figure 4. Average interaction energies of each enzymatic residue with the reacting subsystem (substrate + Lys 163 + Lys 263) for the reactant state of the OSBS and NAAAR reactions. The red lines corresponds to subunit-A residues, the blue one to subunit-B, green color is for the magnesium ion and the black one for the sum of this ion and its coordination shell. The insert shows the interaction energies of the residues of loop 20 and 50.



Scheme II. Schematic representations of the interactions observed at the active site of OSBS/NAAAR. Blue color is used for residues displaying stabilizing interactions with the reacting system (substrate + Lys 163 + Lys 263) and red color for destabilizing interactions. The reacting system is shown in black.

~~We have also compared the averaged interaction energies obtained for each residue at the transition states of both reactions with those obtained for their corresponding reactants. The contribution of the interaction energy of each residue to the barrier can be calculated as the difference of the interaction established in the transition state and reactant state:~~

$$\Delta\Delta E^{\ddagger}_{int,j} = \langle \Delta E_{int,j} \rangle_{TS} - \langle \Delta E_{int,j} \rangle_{R}$$

~~where $\langle \cdots \rangle_{TS}$ and $\langle \cdots \rangle_{R}$ denote averages over the transition and reactant states of the reaction, respectively, and the subscripts (j) refers to the residue number.~~

We can analyze the contribution of the interaction energy of each residue to the barrier (calculated using Eq. 4). A negative value means that a particular residue stabilizes the TS more than the reactant state and this contributes to increase k_{cat}. A positive value of the difference means that that residue stabilizes the reactant state more than the TS and this contributes to decrease k_{cat}. Figure 5 presents the results obtained for TS2 of both reactions. A similar representation for TS1 is given as Supporting Information (Figure S7). Here we only analyze the results for TS2 because this transition state corresponds to the step that differs between the OSBS and the NAAAR reactions and, in addition, this step seems to be the rate-limiting step in both cases. When analyzing these differential interaction energies it is important to take into account the different nature of the second step in the OSBS and NAAAR reactions. In the first case there is a net charge transfer of +1 from Lys 163 to the substrate while in the second step of the NAAAR reaction there is a net charge transfer of +1 from Lys 263 to the substrate. The effect of active site residues depends on the relative position of these with respect to the direction of the charge transfer. This explains the different role played by the Mg^{2+} ion and its coordination shell in both reactions. For the OSBS process, the ion (together with its coordination shell) stabilizes TS2 with respect to the reactant state (-24.0 kcal·mol^{-1}) and thus contributes to catalysis. The ion also plays a catalytic role during the first step of the two reactions (proton abstraction by Lys 163) as observed in Figure S7 in the Supporting Information. Instead, in the second step of the NAAAR reaction the contribution of the ion (with its coordination shell) to stabilize TS2 relative to the reactant state is negligible (-0.1 kcal·mol^{-1}). In this case, Arg ~~266 play a more important role~~ has a larger catalytic contribution in catalyzing the second step of the NAAAR reaction (-14.8 kcal·mol^{-1}) than in the OSBS reaction (-5.5 kcal·mol^{-1}). This residue is close to Lys 263 and ~~is decisive to~~ helps lower its pK_a, facilitating the proton transfer from Lys 263 to the substrate. This residue also plays an important structural role by means of a strong hydrogen bond interaction with Asp 239, which makes part of the coordination shell of the Mg^{2+} ion. It is important to realize that a similar role is played by Arg 217 in *E. coli* OSBS, an enzyme that does not show NAAAR activity. Thus the low pK_a of the proton donor needed for this reaction (Lys 263/235 in *Amycolatopsis/E.coli*) is more likely to be a consequence of the characteristics of an active site designed to bind a negatively charged substrate and not a requirement to catalyze the second step of the NAAAR reaction. In fact, it has been suggested that *E. coli* OSBS could also exhibit racemase activity against other, yet unidentified,

compounds.[16] Another important stabilizing interaction at TS2 relative to the reactant state is provided by Asn 191. This is most probably due to the change of the protonation state of Lys 163 (neutral at the reactant state and protonated at TS2) and thus contributes to catalyze the two reactions, although the effect is ~~quantitatively~~ larger for the NAAAR reaction. Loop 20 residues do not seem to contribute to stabilize TS2 relatively to the reactant. While Lys 161 stabilizes both reactant states and contributes to increase the energy barrier in both reactions. Finally, the total contribution of the interaction energies of all residues to the barrier of the rate-limiting step is negative for the OSBS reaction (-12.4 kcal/mol con aguas y contraiones, -21.5 kcal/mol sin aguas y sin contraiones) and positive for the NAAAR reaction (9.1 kcal/mol con aguas y contraiones, 3.6 kcal/mol sin aguas y sin contraiones). Thus the OSBS reaction is more efficiently catalyzed by the enzyme.

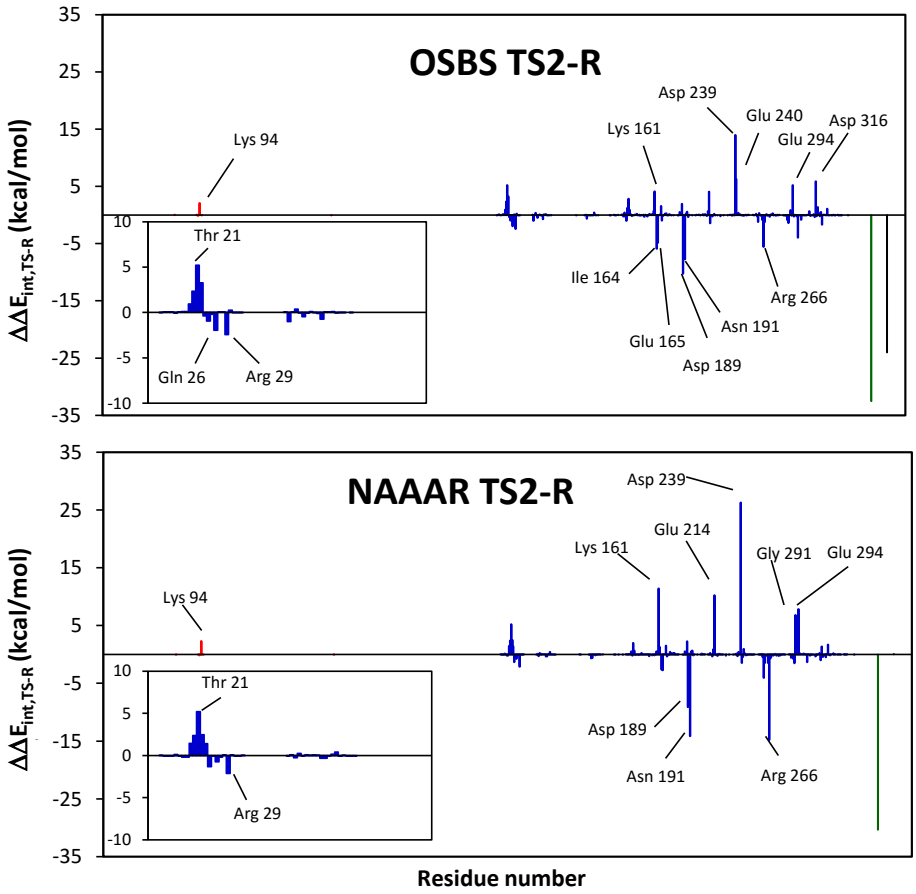


Figure 5. Differences in the interaction energies of each residue with the reacting subsystem (substrate + Lys 163 + Lys 263) averaged at TS2 and reactants state, for the

OSBS and NAAAR reactions. The red lines corresponds to subunit-A residues, the blue lines to subunit-B, green color is for the magnesium ion and the black one for the sum of this ion and its coordination shell. A negative value means that a particular residue stabilizes the TS more than the reactant state. The insert shows the differences in the average interaction energies of the residues found in loops 20 and 50.

4. Conclusions

As a summary of our work, QM/MM simulation methods offer a detailed molecular picture of enzymatic reactions and constitute, together to kinetic, structural and mutagenesis studies, a fundamental tool to understand how enzyme specificity evolves. In the particular case of the *Amycolatopsis* OSBS/NAAAR enzyme, the evolution of the racemase activity seems to be the consequence of several factors. As a general conclusion, an active site designed to bind a negatively charged residue and to lower the pK_a of the base can also be adequate to catalyze a different reaction (the racemization process) provided that the new substrate binds to the active site in the appropriate conformation. In the example analyzed in this work the presence of loops in the active site (and in particular the flexible loop 20) seems to be a key factor to promote promiscuity. This loop is able to provide similar interactions to different substrates, even if they bind into the active site in slightly different orientations. However, other residues, apart from this loop, must also contribute to evolve the racemase activity, lowering the barrier for the proton transfer back to the C^{α} atom. Obviously, another important element is the selection of the substrate that must contain those structural elements needed to compensate the interactions lost by the suppression of the hydroxyl group and to reaccommodate the substrate in the adequate binding pose. The present analysis shows that computational studies of enzymatic promiscuity may provide us guiding principles for the design of new functions into existing enzymes or for the *de novo* design of new activities in protein scaffolds.[6]

5. Acknowledgements

This work was supported by the Spanish *Ministerio de Economía y Competitividad* project CTQ2012-36253-C03-03 and FEDER funds. K.S. thanks the Polish National Science Center (NCN) for grant 2011/02/A/ST4/00246. The authors acknowledge computational facilities of the Servei d'Informàtica de la Universitat de València in the 'Tirant' supercomputer, which is part of the Spanish Supercomputing Network.

Supporting Information

Figure S1 shows the interactions of OSB substrate in the 1SJB structure. A representation of the simulated system is provided as Supporting Information in Figure S2. In Figure S2 is depicted the overlap of two snapshots of SHCHC and NSRPG equilibrated in the active site of the enzyme. Figure S3 shows the energy profiles obtained for the OSBS and NAAAR reactions at the AM1/Mm level. The stationary structures obtained from the potential energy exploration are shown in Figures S4 and S5. Figure S6 shows the corresponding Potentials of Mean Force obtained for the two steps of both reactions. Finally, Figure S7 shows the difference in the averaged interaction energies of each protein residue with the reacting subsystem at TS1 and reactants states, for both OSBS and NAAAR reactions. Table S1 shows the semiclassical KIEs determined at the M06-2X/MM level at 298 K of the rate-limiting step for the OSBS and NAAAR reactions. Table S2 and S3 provides key coordinates of the structures optimized at the AM1/level and M06-2X/MM, respectively, for the OSBS and NAAAR reactions. Table S4 and S5 give the interaction energies of some relevant residues along the stationary states of OSBS and NAAAR reactions, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., and Olsson, M. H. M. Electrostatic basis for enzyme catalysis, *Chem. Rev.* **2006** *106*, 3210-3235.

(2) Garcia-Viloca, M., Gao, J., Karplus, M., and Truhlar, D. G. How enzymes work: Analysis by modern rate theory and computer simulations, *Science* **2004** *303*, 186-195.

(3) Wolfenden, R. Benchmark Reaction Rates, the Stability of Biological Molecules in Water, and the Evolution of Catalytic Power in Enzymes, *Annu. Rev. Biochem.* **2011** *80*, 645-667.

(4) Benkovic, S. J., and Hammes-Schiffer, S. A perspective on enzyme catalysis, *Science* **2003** *301*, 1196-1202.

(5) Marti, S., Roca, M., Andres, J., Moliner, V., Silla, E., Tunon, I., and Bertran, J. Theoretical insights in enzyme catalysis, *Chem. Soc. Rev.* **2004** *33*, 98-107.

(6) Kiss, G., Celebi-Oelcuem, N., Moretti, R., Baker, D., and Houk, K. N. Computational Enzyme Design, *Angew. Chem., Int. Ed.* **2013** *52*, 5700-5725.

(7) Kries, H., Blomberg, R., and Hilvert, D. De novo enzymes by computational design, *Curr. Biol.* **2013** *17*, 221-228.

(8) Arnold, F. H. Design by directed evolution, *Acc. Chem. Res.* **1998** *31*, 125-131.

(9) Morley, K. L., and Kazlauskas, R. J. Improving enzyme properties: when are closer mutations better?, *Trends Biotechnol.* **2005** *23*, 231-237.

(10) Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. Kemp elimination catalysts by computational enzyme design, *Nature* **2008** *453*, 190-194.

(11) Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., Clair, J. L. S., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction, *Science* **2010** *329*, 309-313.

(12) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., III, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. De novo computational design of retro-aldol enzymes, *Science* **2008** *319*, 1387-1391.

(13) Zhang, X. Y., and Houk, K. N. Why enzymes are proficient catalysts: Beyond the Pauling paradigm, *Acc. Chem. Res.* **2005** *38*, 379-385.

(14) Toscano, M. D., Woycechowsky, K. J., and Hilvert, D. Minimalist active-site redesign: Teaching old enzymes new tricks, *Angew. Chem., Int. Ed.* **2007** *46*, 3212-3236.

(15) O'Brien, P. J., and Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities, *Chem. Biol.* **1999** *6*, R91-R105.

(16) Copley, S. D. Enzymes with extra talents: moonlighting functions and catalytic promiscuity, *Curr. Biol.* **2003** *7*, 265-272.

(17) Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. The 'evolvability' of promiscuous protein functions, *Nat. Genet.* **2005** *37*, 73-76.

(18) Khersonsky, O., Roodveldt, C., and Tawfik, D. S. Enzyme promiscuity: evolutionary and mechanistic aspects, *Curr. Biol.* **2006** *10*, 498-508.

(19) Babbitt Lab, S. T. Structure-Function Linkage Database, University of California, San Francisco, 2004-2014.

(20) Gerlt, J. A., and Babbitt, P. C. Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies, *Annu. Rev. Biochem.* **2001** *70*, 209-246.

(21) Palmer, D. R. J., Garrett, J. B., Sharma, V., Meganathan, R., Babbitt, P. C., and Gerlt, J. A. Unexpected divergence of enzyme function and sequence: "N-acylamino acid racemase" is o-succinylbenzoate synthase, *Biochemistry* **1999** *38*, 4252-4258.

(22) Palmer, D. R. J., and Gerit, J. A. Evolution of enzymatic activities: Multiple pathways for generating and partitioning a common enolic intermediate by glucarate dehydratase from Pseudomonas putida, *J. Am. Chem. Soc.* **1996** *118*, 10323-10324.
(23) Yew, W. S., Fedorov, A. A., Fedorov, E. V., Almo, S. C., and Gerlt, J. A. Evolution of enzymatic activities in the enolase superfamily: L-talarate/galactarate dehydratase from Salmonella typhimurium LT2, *Biochemistry* **2007** *46*, 9564-9577.
(24) Taylor Ringia, E. A., Garrett, J. B., Thoden, J. B., Holden, H. M., Rayment, I., and Gerlt, J. A. Evolution of enzymatic activity in the enolase superfamily: Functional studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis, *Biochemistry* **2004** *43*, 224-229.
(25) Thoden, J. B., Taylor Ringia, E. A., Garrett, J. B., Gerlt, J. A., Holden, H. M., and Rayment, I. Evolution of enzymatic activity in the enolase superfamily: Structural studies of the promiscuous o-succinylbenzoate synthase from Amycolatopsis, *Biochemistry* **2004** *43*, 5716-5727.
(26) Li, H., Robertson, A. D., and Jensen, J. H. Very fast empirical prediction and rationalization of protein pK(a) values, *Proteins* **2005** *61*, 704-721.
(27) Bas, D., Rogers, D., and Jensen, J. Very fast prediction and rationalization of pKa values for protein-ligand complexes, *Proteins* **2008** *73*, 765-783.
(28) Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions, *J. Chem. Theory Comp.* **2011** *7*, 525-537.
(29) Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values, *J. Chem. Theory Comp.* **2011** *7*, 2284-2295.
(30) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, Cambridge University Press, Cambrige, U.K., 1999.
(31) Field, M. J., Albe, M., Bret, C., Proust-De Martin, F., and Thomas, A. The Dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials, *J. Comp. Chem.* **2000** *21*, 1088-1100.
(32) Lopez-Canut, V., Marti, S., Bertran, J., Moliner, V., and Tunon, I. Theoretical Modeling of the Reaction Mechanism of Phosphate Monoester Hydrolysis in Alkaline Phosphatase, *J. Phys. Chem. B* **2009** *113*, 7816-7824.
(33) Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. The Development and Use of Quantum-Mechanical Molecular-Models .76. AM1 - A New General-Purpose Quantum-Mechanical Molecular-Model, *J. Am. Chem. Soc.* **1985** *107*, 3902-3909.
(34) Zhao, Y., and Truhlar, D. G. Density functionals with broad applicability in chemistry, *Acc. Chem. Res.* **2008** *41*, 157-167.
(35) Zhao, Y., and Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theor. Chem. Acc.* **2008** *120*, 215-241.
(36) Pranata, J., Wierschke, S. G., and Jorgensen, W. L. OPLS Potential Functions for Nucleotide Bases - Relative Association Constants of Hydrogen-Bonded Base-Pairs in Chloroform, *J. Am. Chem. Soc.* **1991** *113*, 2810-2819.
(37) Jorgensen, W. L., and Tiradorives, J. The OPLS Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin, *J. Am. Chem. Soc.* **1988** *110*, 1657-1666.
(38) Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids .10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers - Application to Liquid Water, *J. Am. Chem. Soc.* **1981** *103*, 335-340.
(39) Singh, U. C., and Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular

systems: Applications to the CH3Cl + Cl− exchange reaction and gas phase protonation of polyethers, *J. Comp. Chem.* **1986** *7*, 718-730.

(40) Paterlini, M. G., and Ferguson, D. M. Constant temperature simulations using the Langevin equation with velocity Verlet integration, *Chem. Phys.* **1998** *236*, 243-252.

(41) Fukui, K. The path of chemical reactions - the IRC approach, *Acc. Chem. Res.* **1981** *14*, 363-368.

(42) Moliner, V., Turner, A. J., and Williams, I. H. Transition-state structural refinement with GRACE and CHARMM: Realistic modelling of lactate dehydrogenase using a combined quantum/classical method, *Chem. Comm.* **1997**, 1271-1272.

(43) Turner, A. J., Moliner, V., and Williams, I. H. Transition-state structural refinement with GRACE and CHARMM: Flexible QM/MM modelling for lactate dehydrogenase, *Phys. Chem. Chem. Phys.* **1999** *1*, 1323-1331.

(44) Marti, S., Moliner, V., and Tunon, I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems, *J. Chem. Theory Comp.* **2005** *1*, 1008-1016.

(45) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. Gaussian 09, Revision D.01 ed., Gaussian Inc., Wallingford CT, 2009.

(46) Torrie, G. M., and Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comp. Phys.* **1977** *23*, 187-199.

(47) Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. The Weighted Histogram Analysis Method For Free-Energy Calculations On Biomolecules .1. The method, *J. Comp. Chem.* **1992** *13*, 1011-1021.

(48) Moliner, V., Turner, A. J., and Williams, I. H. Transition-state structural refinement with GRACE and CHARMM: Realistic modelling of lactate dehydrogenase using a combined quantum/classical method, *Chem. Commun. (Cambridge, U. K.)* **1997**, 1271-1272.

(49) Turner, A. J., Moliner, V., and Williams, I. H. Transition-state structural refinement with GRACE and CHARMM: Flexible QM/MM modelling for lactate dehydrogenase, *Phys. Chem. Chem. Phys.* **1999** *1*, 1323-1331.

(50) Marti, S., Moliner, V., and Tunon, I. Improving the QM/MM description of chemical processes: A dual level strategy to explore the potential energy surface in very large systems, *J. Chem. Theory Comput.* **2005** *1*, 1008-1016.

(51) Taylor Ringia, E. A., Palmer, D. R. J., and Gerlt, J. A. The lesser "burden borne" by o-succinylbenzoate synthase: An "easy" reaction involving a carboxylate carbon acid, *J. Am. Chem. Soc.* **2001** *123*, 5824-5825.

(52) Zhang, Y. K., Kua, J., and McCammon, J. A. Influence of structural fluctuation on enzyme reaction energy barriers in combined quantum mechanical/molecular mechanical studies, *J. Phys. Chem. B* **2003** *107*, 4459-4463.

(53) Lodola, A., Sirirak, J., Fey, N., Rivara, S., Mor, M., and Mulholland, A. J. Structural Fluctuations in Enzyme-Catalyzed Reactions: Determinants of Reactivity in Fatty Acid Amide Hydrolase from Multivariate Statistical Analysis of Quantum Mechanics/Molecular Mechanics Paths, *J. Chem. Theory Comp.* **2010** *6*, 2948-2960.

(54) Garcia-Viloca, M., Truhlar, D. G., and Gao, J. L. Reaction-path energetics and kinetics of the hydride transfer reaction catalyzed by dihydrofolate reductase, *Biochemistry* **2003** *42*, 13558-13575.

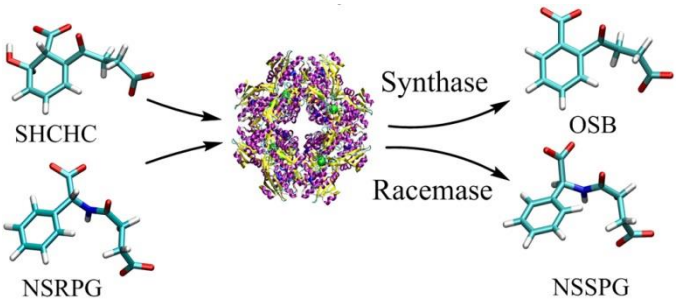
(55) Luk, L. Y. P., Javier Ruiz-Pernia, J., Dawson, W. M., Roca, M., Joel Loveridge, E., Glowacki, D. R., Harvey, J. N., Mulholland, A. J., Tunon, I., Moliner, V., and Allemann, R. K. Unraveling the role of protein dynamics in dihydrofolate reductase catalysis, *Proc. Natl. Acad. Sci. U.S.A.* **2013** *110*, 16344-16349.

(56) Glasner, M. E., Fayazmanesh, N., Chiang, R. A., Sakai, A., Jacobson, M. P., Gerlt, J. A., and Babbitt, P. C. Evolution of structure and function in the o-succinylbenzoate synthase/N-acylamino acid racemase family of the enolase superfamily, *J. Mol. Biol.* **2006** *360*, 228-250.

(57) McMillan, A. W., Lopez, M. S., Zhu, M., Morse, B. C., Yeo, I.-C., Amos, J., Hull, K., Romo, D., and Glasner, M. E. Role of an Active Site Loop in the Promiscuous Activities of Amycolatopsis sp T-1-60 NSAR/OSBS, *Biochemistry* **2014** *53*, 4434-4444.

(58) Zhu, W. W., Wang, C., Jipp, J., Ferguson, L., Lucas, S. N., Hicks, M. A., and Glasner, M. E. Residues Required for Activity in Escherichia coli o-Succinylbenzoate Synthase (OSBS) Are Not Conserved in All OSBS Enzymes, *Biochemistry* **2012** *51*, 6171-6181.

Graphical Abstract for the table of contents


SHCHC
NSRPG
Synthase
Racemase
OSB
NSSPG